Auditory Perception, Plasticity, and Cognition in Biological and Artificial Neural Networks

Submitted to the

Faculty of Engineering Friedrich-Alexander University Erlangen-Nürnberg

in fulfillment of the requirements for the

Habilitation

from

Dr. Achim Schilling

Mentors: Prof. Dr. Tobias Reichenbach, Prof. Dr. Andreas Maier, Prof. Dr. Max Happel

Erlangen 2024

Ich versichere, dass ich die Arbeit ohne fremde Hilfe und ohne Benutzung anderer als der angegebenen Quellen angefertigt habe und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat und von dieser als Teil einer Prüfungsleistung angenommen wurde. Alle Ausführungen, die wörtlich oder sinngemäß übernommen wurden, sind als solche gekennzeichnet.

Erlangen, den 11.12.2024 _____

Dr. Achim Schilling

Übersicht

Heutzutage verschwimmen die Grenzen zwischen verschiedenen Disziplinen zunehmend, was erst kürzlich durch die Vergabe des Physik- sowie des Chemie-Nobelpreises für die Entwicklung und Anwendung von künstlichen neuronalen Netzen unterstrichen wurde. Die Forschung, auf der die vorliegende Habilitationsschrift basiert, macht deutlich, dass interdisziplinäre Forschungsansätze aus den Fachgebieten, künstliche Intelligenz (KI), Neurowissenschaft und biomedizinische Technik zu Durchbrüchen in allen der genannten Disziplinen führen können. In der Arbeit, wurde das Hörsystem von Säugetieren als Inspirationsquelle für die KI-Forschung genutzt, sowie die KI wiederum benutzt wurde, um das Hörsystem zu erforschen und zu modellieren. Ein spezieller Fokus lag auf dem Verständnis der neuronalen Folgeerscheinungen eines Schadens im peripheren Hörsystem (Ohr), nämlich Tinnitus. Wir konnten zeigen, dass Tinnitusmechanismen auf Basis des 'Stochastischen Resonanzprinzips' aus der Physik, sowie auf Basis der Regelungstechnik aus den Ingenieurswissenschaften und Baysscher Statistik aus der Mathematik erklärt werden können. So führt ein Hörschaden zu einer Verringerung der Amplitude des Signals, das vom Ohr ins Gehirn weitergeleitet wird, was wiederum zu einem Rückgang der übertragenen Information entlang der auditorischen Bahn führt. Das Gehirn nutzt den stochastischen Resonanzmechanismus aus, indem es zusätzliches neuronales Rauschen zum Restsignal addiert. Dieses neuronale Rauschen ändert aber die Charakteristik des Signals im Gehirn auf eine Art, die dazu führt, dass das Gehirn das Signal als echten Ton missinterpretiert, was zur Wahrnehmung eines akuten Tinnitus führt. Zusätzlich verändert das dauerhafte Rauschen die Standardvorhersage (Prior) des Gehirns, so dass das neuronale Rauschen und der damit assoziierte virtuelle Ton als Standardvorhersage manifestiert werden, was einer Chronifizierung des Tinnitus entspricht. Basierend auf diesen Erkenntnissen, wurde ein Therapieansatz entwickelt, der sich in einer Pilotstudie als vielversprechend herausgestellt hat. Dieser Therapieansatz zeigt wie biomedizinische Technik von einem soliden experimentellen und theoretischen neurowissenschaftlichen Ansatz profitieren kann. Basierend auf der Idee einer bewussten Tinnituswahrnehmung, können auch allgemeinere Prinzipien über das Gehirn als Vorhersagemaschine abgeleitet werden. Wir konnten zeigen, dass das Gehirn ohne äußere Reize durch den Raum von möglichen zukünftigen Ereignissen navigiert, was als das Durchspielen von möglichen zukünftigen Szenarien interpretiert werden kann. All diese Erkenntnisse basieren auf der Auswertung von neuronalen Daten mit fortschrittlichen KI-Methoden. Zusammenfassend kann man sagen, dass künstliche neuronale Netze als Modell und als Werkzeug verwendet wurden, um Informationsverarbeitung im Gehirn im gesunden und kranken Fall zu untersuchen, was bereits zu einem medizinischen Therapieansatz geführt hat, der nun in einer großen klinischen Studie validiert wird. Zusätzlich wurde das Gehirn als

Inspirationsquelle zur Entwicklung effizienter KI-System genutzt. Ich hoffe, dass diese Habilitation zukünftige Forschende dazu inspiriert, Interdisziplinarität zu leben und KI im Bereich biomedizinische Technik zu nutzen, um das Leben von vielen Betroffenen zu verbessern.

Abstract

Nowadays the boundaries between scientific discipline are blurring, as recently demonstrated by the awarding of the Nobel prize in Physics and Chemistry for the development and application of artificial neural networks. The research that serves as basis for the following habilitation thesis, illustrates that interdisciplinary research approaches from the fields of artificial intelligence, neuroscience, and biomedical engineering can lead to breakthroughs in all of the fields. In this thesis, the mammal auditory system was used as a source of inspiration for Artificial Intelligence (AI) research on the one hand, and AI was used to investigate and model the auditory system on the other hand. A special focus was placed on understanding the neuronal consequences of a damage of the auditory system, namely hearing loss and tinnitus. We could show that tinnitus mechanisms could be understood using the stochastic resonance principles from physics, principles from control technology from engineering, and Bayesian statistics from mathematics. Thus, a hearing damage causes decreased neural output from the ear to the brain and thus a drop of the transmitted information along the auditory pathway. The brain adds neuronal noise to re-enhance the amount of transmitted information by means of stochastic resonance. However, this added neuronal noise changes the characteristics of the neuronal signal in the brain in a sense that the brain misinterprets the neuronal signal as real tone and thus an acute tinnitus is perceived. The neuronal noise furthermore changes the prediction (prior distribution) of the brain, which sets the noise as default value and thus the tinnitus is chronically manifested. Based on this idea, a therapy approach has been developed, which has proven to be promising in a pilot-study. This approach illustrates how biomedical engineering can be boosted by a solid experimental and theoretical neuroscience background. Furthermore, based on the ideas on conscious tinnitus perception, conclusions about the brain as prediction machine in general could be drawn. Thus, it was shown that the brain moves through the space of possible stimulus evoked activity patterns during spontaneous activity, which might be interpreted as the brain playing through, respectively preparing for future events. All of these insights are based on the evaluation of neuronal data with advanced machine learning techniques. In summary, AI was used as a model and a tool to understand healthy and impaired information processing in the brain and has already led to an individualized medical treatment strategy, which will be validated in a large clinical study. Furthermore, the brain was used as a source of inspiration for the development of efficient AI systems. Thus, I hope that the thesis provides some inspiration for future scientist to live interdisciplinarity and to use AI in biomedical engineering and thus to improve the lives of many people.

Contents

1 Introduction **1**

1.1 Motivation 1

1.2 The human auditory system 4

1.2.1 General facts on the efficiency of the auditory system 4

1.2.2 Anatomy and physiology of hearing 5

1.2.3 The auditory pathway 9

1.3 Beyond the auditory pathway: Speech and language processing 13

1.4 Neuroimaging 14

1.5 Combining AI and Neuroscience 15

2 Main **17**

2.1 Biologically-inspired neuron models and training algorithms 17

2.1.1 Spiking neural networks: The surrogate gradient approach 17

2.1.2 Spiking Long-Short-Term-Memory networks 20

2.1.3 Biologically-inspired restrictions and the influence on robustness 24

2.1.4 Summary on biologically-inspired neuron models and training algorithms 26

2.2 Insights from lesion studies in the auditory system: Auditory phantom perception 26

2.2.1 Relevance of lesion studies in the auditory system for neuroscience and medicine 26

2.2.2 Zwicker tone: the 'little brother' of tinnitus 29

2.2.3 A mechanistic model for Zwicker tone and acute tinnitus 34

2.2.4 From AI and Computational Neuroscience to Biomedical-Engineering . . 41

2.2.5 Conscious tinnitus perception, the Bayesian brain, and internal world models 42

2.3 Higher cognitive functions of the brain: the prediction machine 45

3 Conclusion **49**

4 Acknowledgments **55**

Abbreviations **57**

List of Figures **61**

Chapter 1

Introduction

1.1 Motivation

The human brain is actually the most complex and fascination entity in our world [Krauss, 2024, Sporns, 2011]. This organ consists of approximately 86 billion nerve cells, the so called neurons [Lent et al., 2012]. Neurons are the building elements of the human brain and can be compared to the transistors in a computer [Mulaosmanovic et al., 2018, Li et al., 2022]. Each neuron itself is not intelligent at all, it can just sum up signals from other neurons and if a certain threshold is reached a so-called spike is released, which means the neuron produces its' own signal, which is transmitted to further neurons [Platkiewicz and Brette, 2010]. Due to the fact that each neuron shows a relatively simple behavior (the molecular processes are not simple at all) the human brain is able to perform complex tasks and generate complex feelings and abilities such as reading, remembering, loving and finally consciousness [Damasio and Damasio, 2023, Friederici et al., 2017]. Despite the fact that the brain is able to achieve this incredible performance and therefore might be the optimal template to build an Artificial General Intelligence (AGI), we are just at the very beginning of understanding how the brain works and at the moment it is nearly impossible to build robots with the intelligence of a bug [Sharkey, 2006, Pfeifer and Iida, 2004]. How is it possible that we are not able to understand the brain, although a lot of effort is undertaken and a lot of money is spent to unravel the underlying processing mechanisms? One famous example is the human brain project, where the European Union (EU) has spent approximately one billion dollars to simulate and investigate the brain [Markram, 2012, Frégnac and Laurent, 2014]. Unfortunately, in contrast to the human genome project, a project that unraveled the whole human genome [Collins et al., 2003], the human brain project, was highly criticized as the brain remains still a mystery [Frégnac and Laurent, 2014]. The reason why the

human brain is so difficult to understand, is the fact that the complex interplay of the simple building blocks (neurons) leads to an emergent behavior [Thiebaut de Schotten and Forkel, 2022], that produces complex feelings and intelligence. As we already know from other examples of emergent behavior such as Conway's game of life, [Bays, 2010], or the dynamics of animal swarms, it is easy to produce complex behavior by implementing the simple underlying rules, but it is nearly impossible to infer the rules, when just the complex behavior is observed [Hamann and Schmickl, 2012, Eriksson et al., 2010, Gerum et al., 2013, Helbing, 2012, Szabo and Birdsey, 2017]. The swarm dynamics of e.g., birds is a good metaphor for the human brain, as birds like neurons have no overview over the complete system (swarm) and the related complex patterns such a swarm can form [Hamann and Schmickl, 2012, Eriksson et al., 2010]. However, if every bird of the swarm follows easy rules, the complex swarm dynamics emerges from these underlying simple rules (see e.g. [Eriksson et al., 2010, Gerum et al., 2013]). Nowadays we are able to understand how a bird swarm behaves, as a lot of effort was made and many computer simulations were written with the goal to understand the dynamics of bird swarms [Eriksson et al., 2010]. However, the brain consists of 86 billion elements, which is orders of magnitudes higher than the size of a bird swarm. Thus, we did not yet find the underlying rules of the complex mechanisms in the brain. How complex the brain actually is, can be understood by a deeper look at the actual scales of the brain. Indeed, the average brain weight is about 1200 to more than 1500 g [Miller and Corsellis, 1977, Herculano-Houzel, 2009] and the energy consumption of the brain is with 20 W nearly negligible [Furber, 2012, Yu et al., 2018]. Thus, the brain is the ideal template to build green AI systems [Verdecchia et al., 2023, Mehonic and Kenyon, 2022]. However, the fact that the brain consists of approximately 10^{11} neurons means that the connectivity matrix (contains all pairwise connections between neurons) has $(10^{11})^2 = 10^{22}$ entries. The values in the matrix can be any real number, as the connections of a neuron to another neuron called synapse are not quantized to a first approximation. But if we consider that there are just ten different values for each entry, there are $10^{(10^{22})}$ different connectivity matrices, which means that there are more possible brains in the universe than protons (approximately 10^{80} [Rees, 1983, Barrow, 1979]). These numbers give a first idea on how complex the brain indeed is. However, the options opened up by a deep understanding of the brain are overwhelming. Thus, brain like computers would perform complex cognitive tasks without consuming much energy. Furthermore, understanding and simulating the brain could have a huge impact on medicine, speaking of illnesses such as dementia, depression, or Parkinson's disease [Panksepp, 2010]. In summary, the implications of understanding the brain are incredible for computer science as well as biology and medicine. However, as described above also the hurdles are high. Therefore, a highly interdisciplinary approach is needed to move

forward in both fields, AI-research as well as neuroscience [Zhuang et al., 2020, Wang et al., 2022, Kriegeskorte and Douglas, 2018, Hassabis et al., 2017]. The problem of isolated disciplines with a lack of scientific exchange is illustrated by the metaphor of the five blind scientists and the elephant, an often-used metaphor in cognitive science. Thus, the blind scientists all investigate different parts of the elephant and all scientists have a completely different picture in mind. The only way to build a congruent image of the elephant is to connect, to share information, and to cooperate [Buckle et al., 2023]. The idea of this habilitation thesis is to connect neuroscience with computer science in two different ways. First of all, neuroscience should be used as source of inspiration for computer science to build neuroscience-inspired AI-architectures as described by among others the new Nobel Prize in chemistry laureate Demis Hassabis (Neuroscience-Inspired AI, [Hassabis et al., 2017]). On the other hand, artificial neural networks should serve as a model to understand the brain called Cognitive Computational Neuroscience (CCN) as proposed by Nikolaus Kriegeskorte and Pamela Douglas [Kriegeskorte and Douglas, 2018]. This approach should be further combined with another method from classic neuro-psychology, the so-called lesion studies [Vaidya et al., 2019]. Thus, by investigating an impaired system e.g., a lesioned brain, conclusion can be drawn about the way it works [Vaidya et al., 2019]. This methodological approach has led to major breakthroughs such as the localization of two of the most important brain regions related to language processing (Broca's and Wernicke's area, [Dronkers et al., 2004]). Furthermore, the most famous example of a lesion study is the example of Phineas Gage, a former railroad construction foreman, who was hit by an iron road leading to severe damages in the brain and to changes of his personality, illustrating that personality is indeed determined by brain architecture [Macmillan, 2000, Gazzaniga et al., 2014]. It is not necessary that there is always a real damage in the brain, but also some mis-processing of stimuli such as optical illusions can be a good possibility to draw conclusions on the processing principles of the brain. Thus, in a recent study it was shown that an artificial neural network suffers from the same optical illusion as the human brain, indicating that some processing principles are similar [Zhang and Yoshida, 2024, Cheng et al., 2023]. Up to now, cognitive neuroscience as well as AI research have focused a lot on the visual system and image processing. However, in this thesis the focus lies on the auditory system and the higher brain areas connected to the auditory system, as these systems are the key to unravel speech and language processing in the brain [Shamma, 1985]. To put it in a nutshell, the main idea of this habilitation thesis as well as my whole scientific career is to use the human auditory system (and language related areas of the brain) as a source of inspiration to develop novel AI-algorithms suited to process speech on the one hand, and to use these algorithms as model for the brain to build new hypotheses, which I can test in further experiments. The

general approach -also a key component of my research- of using natural sciences (e.g. physics) as a model for the brain and to develop artificial neural networks that help to understand the brain but also boost AI-research, was honored with the Nobel prize in physics for John Hopfield and Geoffrey Hinton [Fattaruso, 2024]. The usage of AI to understand nature (e.g. protein structures) was appropriately honored with the Nobel prize in chemistry in the same year (2024) for Demis Hassabis, John Jumper, and David Baker [Abriata, 2024]. In this thesis, I want to illustrate how these principles also apply to the auditory system and language processing in the human brain. Thus, to get an in-depth understanding on how the human auditory system works and what we can learn from it, the auditory system is briefly explained in the next section.

1.2 The human auditory system

1.2.1 General facts on the efficiency of the auditory system

The ear is considered the most sensitive sensory organ of the human organism [Schmidt et al., 2010]. The sensory cells in the ear and the connected neurons translate a mechanical signal -sound- into an electro-chemical signal -the activity of nerve cells. The ear and the auditory system in general are an incredibly evolutionary fine-tuned system, which is able to process a huge range of different sound intensities (amplitude of sound waves) as well as sound frequencies (frequency of sound waves). The huge dynamic range of the auditory system is caused by the fact that the ear works logarithmically, which means that the perceived loudness increases logarithmically with the amplitude of pressure fluctuations of sound waves [Schmidt et al., 2010, Hudspeth et al., 2013]. Therefore, the actual loudness (L) of a tone is in most cases given in decibels [Schmidt et al., 2010].

$$L = 20 \cdot \log_{10}\left(\frac{P_x}{P_0}\right) \tag{1.1}$$

The loudness L in (decibel sound pressure level (dB[SPL])) is given by the tens logarithm of the ratio between the actual sound pressure P_x (in Pa) and a reference sound pressure $P_0 = 2 \cdot 10^{-5}$ Pa. The physiological range of hearing lies in between -4 and 120 dB[SPL], which means that the human auditory system can process pressure fluctuations from approximately 10^{-5} Pa to 20 Pa or i.e. pressure fluctuations varying by 6 orders of magnitudes [Schmidt et al., 2010, Shatnawi et al., 2009]. The sensitivity and the subjectively perceived loudness vary with the frequency of the signal. Thus, the auditory system is most sensitive in the range between 2 and 5 kHz. Nevertheless, humans can here tones from approximately 20 Hz to 20 kHz [Hudspeth et al., 2013, Schmidt et al.,

2010]. This means that we can hear tones covering ca. ten octaves, which means ten doublings of the frequency [ten Donkelaar et al., 2020, Kunchur, 2023]. The impressive abilities of our hearing system are a result from evolutionary fine-tuning over millions of years [Fay and Popper, 2000]. In the following sections, the underlying principles and mechanisms are explained.

1.2.2 Anatomy and physiology of hearing

The process of transducing and transforming a sound wave into an electro-chemical signal that could be consciously perceived is very complex and many different anatomical and physiological structures are involved. All structures, associated with hearing are summarized under the term 'auditory system'. The auditory system can be divided into the ear (periphery) and the auditory pathway. The ear consists of structures optimized to efficiently collect sound waves and to transmit them without significant losses to the actual sensory cells in the inner ear called inner hair cells (IHCs) [Schmidt et al., 2010, Hudspeth et al., 2013]. There the signal is transduced to a chemical (transduction) and following this to an electrical signal (transformation), which means that neurons are activated [Hudspeth et al., 2013]. All structures of the nervous system related to processing auditory stimuli are summarized under the term 'auditory pathway' [Starr et al., 2001].

The ear

The ear consists of 3 main parts called: outer ear, middle ear, and inner ear. Each of these parts fulfills a different role in the hearing process.

Outer ear: The outer ear consists of the so-called pinna and the ear canal. The pinna is shaped like a shell and acts as a funnel. It has the purpose of collecting as much sound as possible on the one hand, but is also important for directional hearing, especially in the vertical axis [Hudspeth et al., 2013]. The shape of the pinna is such that different distortions are produced depending on the location of the sound source. These distortions can be used by the brain to reconstruct the origin of the sound, which means the location of the sound source [Hudspeth et al., 2013]. However, this is only one of the mechanisms involved in directional hearing. Further parts of the ear and the auditory pathway play a significant role for directional hearing. In addition to the pinna, the ear canal also has an interesting property. Thus, the ear canal serves as an acoustic cavity, amplifying incoming sound waves especially at a frequency of approximately 3 kHz (resonance frequency) contributing to the fact that this is the frequency range of best hearing [Hudspeth et al., 2013, Dempster and Mackenrie, 1990, Keefe et al., 1994, Henoch and Chesky, 1999].

Middle Ear: After passing through the ear canal, the sound hits the tympanic membrane causing it to oscillate. The tympanic membrane is connected to three bones called ossicles (malleus, incus, and stapes), which are flexibly connected to each other [Schmidt et al., 2010, Gazzaniga et al., 2014, Tucker et al., 2016]. These three ossicles move when the tympanic membrane oscillates. The main purpose of this apparatus is to achieve what is called impedance adaptation, which means that the ossicles are necessary to transmit the sound wave from air to a liquid without major reflections at the boundary surface. This is achieved through three principles: First, the area of the tympanic membrane is larger than the area of the oval window being a part of the cochlea, where the actual sensory cells are located. Second, the connections between the three ossicles are optimized for good leverage, and third, the velocity of the stapes is reduced to achieve optimal force transmission. This design reduces the reflected sound intensity at the air/liquid interface by a factor of ca. 30 [Schmidt et al., 2010, Hudspeth et al., 2013].

Inner ear: The task of the outer ear and middle ear is as described above, exclusively a proper collection and efficient transmission of sound waves. The inner ear is, however, the crucial structure for hearing, as there the mechanical signal is converted to an electrical signal. For hearing the most important part of the inner ear is the so-called cochlea. The cochlea is a bone structure, which looks like a snail and has three cavities - the scala tympani, scala media, and scala vestibuli- filled with fluid ([Schmidt et al., 2010], for a cross-section of the cochlea see Fig. 1.1, [Vlajkovic and Thorne, 2022]). The scala vestibuli and the scala tympani are filled with sodium ion rich perilymph, whereas the scala media is filled with potassium ion rich endolymph. As described in the previous paragraph, the last ossicle called stapes pushes rhythmically against the oval window in front of the scala vestibuli causing pressure fluctuations in the perilymph. This pressure fluctuations cause the the Reissner membrane between scala vestibuli and scala media to move, causing pressure fluctuations in the scala media. This leads to the movement of the basilar membrane separating scala media and scala tympani [Schmidt et al., 2010, Hudspeth et al., 2013]. On top of the oscillating basilar membrane are located one row of so-called IHCs and three rows of outer hair cells (OHCs), which are connected to a galert structure called tectorial membrane (TM) [Vlajkovic and Thorne, 2022]. Despite the fact that there are more OHCs than IHCs, the critical structure for hearing is the inner hair cell row, that is not connected to the tectorial membrane at all. Thus, endolymph flow due to the pressure fluctuations cause bending of small structures on top of the IHCs called stereocilia. These stereocilia are connected via so called tip links on top of them [Ahmed et al., 2006]. These tip links pull on ion channels on the top of the stereocilia leading to a potassium influx from the endolymph of the scala media into the intra-cellular space of the IHCs. The intra-cellular potential -normally at -70 mV- is increased due to the

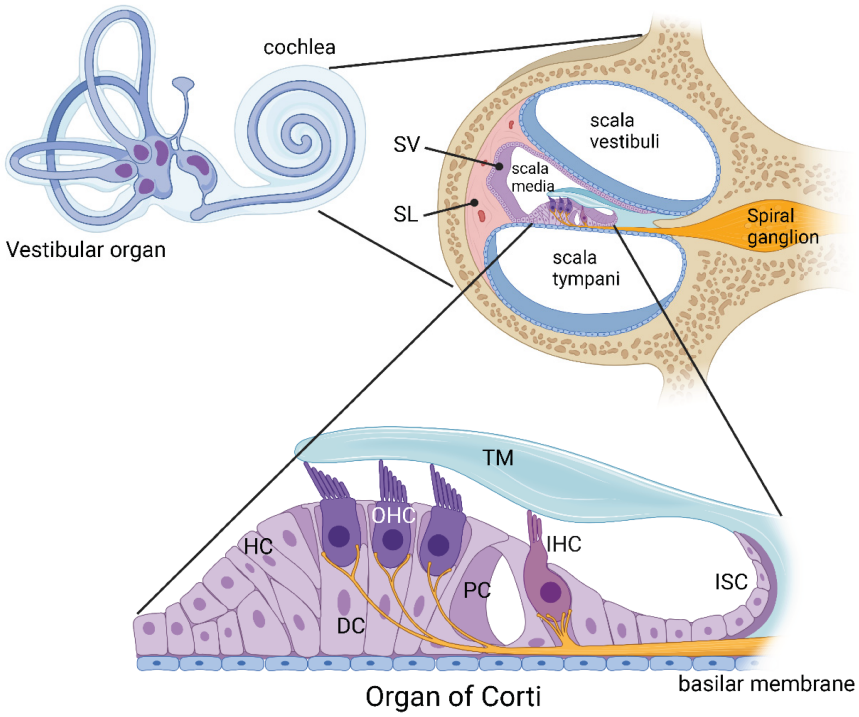

Figure 1.1: **The inner ear**
Left: The figure shows inner ear structures: the vestibular system and the cochlea. The cochlea is the crucial organ for hearing. Right: Section through the cochlea showing the three liquid filled cavities called: scala vestibuli, scala media, and scala tympani. In the scala media, the so-called organ of corti (bottom) is located. The organ of corti is the actual sensory epithelium consisting of a galert mass (tectorial membrane (TM)), the effector cells (outer hair cells (OHCs)) and the sensory cells (IHCs). When the basilar membrane moves, the TM moves, causing a liquid flow. The liquid flow shears the stereocilia of the IHCs causing a K^+-ion influx into the IHCs called depolarization. The OHCs amplify the vibration, through rhythmic contraction. (Figure taken from [Vlajkovic and Thorne, 2022])

influx of positively charged potassium ions [Schmidt et al., 2010]. Thus, the IHCs are depolarized and the mechanical signal has been transduced to a change in electrical potential, which is proportional to the sound intensity. This process is called transduction and as a mechanical signal has been transduced into an electrical signal, it is specified as mechanotransduction [Gillespie and Müller, 2009]. IHCs are no neurons and are therefore called secondary sensory cells [Caicci et al., 2007, Burighel et al., 2011]. In the next step, the depolarized IHCs have to stimulate neurons, that transmit the signal to the brain. Thus, the IHCs release glutamate -an excitatory neurotransmitter- that diffuses through the synaptic cleft (special synapse called ribbon synapse [Nouvian et al., 2006]) to the dendrites of the so called spiral ganglion, which is a conglomerate of neurons in the middle of the cochlea ([Schmidt et al., 2010], see Fig. 1.1)). The spiral ganglion neurons can,

when a threshold is reached, produce action potentials also called spikes, which are short electric voltage changes, with always the same height and duration (some ms [Platkiewicz and Brette, 2010]).

Tonotopy: The mechanisms described above do not explain how different frequencies of tones are encoded. Indeed, the frequency of a sound is encoded in the location of the IHCs in the cochlea. If the cochlea was unrolled into a straight line, the location of the oscillating IHCs on the line would determine the frequency (also called pitch) of the tone [Tramo et al., 2002, Tramo et al., 2005]. Thus, the basilar membrane, which can be considered a skin stripe is less flexible at the oval window (at the beginning of the cochlea) and becomes broader and more flexible at the end of the cochlea [Inselberg and Von Foerster, 1970, Zweig, 1976]. Thus, depending on the pitch of the tone, a different part of the basilar membrane is oscillating depending on the resonance frequency of the according part. Higher tone pitches lead to oscillations of the basal parts of the basilar membrane near the oval window, whereas lower frequency tones cause vibrations far away from the oval window at the apex. The location dependent oscillations of the basilar membrane are called traveling wave [von Békésy, 1970]. The fact that the location of the IHCs encodes the sound frequency is called tonotopy [Kandler et al., 2009]. The fact that the location encodes the stimulus frequency is not only true for the cochlea but for the whole auditory pathway. Thus, all neuron conglomerates (called nuclei) show this kind of tonotopic organization. In all brain nuclei the neurons are somehow ordered like a piano keyboard, which means that neurons located nearby each other encode for similar tone pitches [Kandler et al., 2009, Saenz and Langers, 2014]. The fact that location plays a crucial role in encoding stimuli is also true for other modalities such as vision (retinotopy) and somato-sensation (somatotopy) [Duncan et al., 2007, Hluštík et al., 2001]. The OHCs are also important for pitch processing in the cochlea. OHCs are no receptor cells but mainly effector cells. Thus, these cells amplify the oscillation of the tectorial membrane above the IHCs and thus the whole signal is amplified. Indeed, OHCs can contract similar to a muscle (note that the mechanisms are different) using a protein called prestin [Schmidt et al., 2010, Dallos, 2008].

Hearing Loss: The cochlea is a fascinating sensory organ, but damages of the cochlea can lead to decreased hearing abilities called hearing loss (HL). HL can be divided into conductive hearing loss, which is caused for example by the occlusion of the outer ear or damages of the middle ear, and the sensorineural hearing loss, caused by damages of the cochlea and/or neuronal structures of the auditory pathway [Payne and Wong, 2022]. In this part, we exclusively focus on sensorineural hearing loss and especially damages of the cochlea. There are various reasons for

cochlear damage such as noise traumas, ototoxicity, infections, vascular diseases, and age related effects [Nadol Jr, 1993, Payne and Wong, 2022]. Thus, even moderate noise traumas can induce a so called synaptopathy, which means that the ribbon synapses connecting the IHCs to the spiral ganglion disappear [Tziridis et al., 2021, Liberman and Kujawa, 2017, Kujawa and Liberman, 2015]. The cochlea and especially the IHCs are sensitive to other factors such as ototoxic drugs, vascular diseases such as an occlusion of vessels, auto-immune diseases, and inflammations [Payne and Wong, 2022, Nadol Jr, 1993]. These, factors can also have a genetic component and the number of damaged hair cells and synapses increases with age. The age-dependent hearing loss is called presbycusis and starts always at higher sound frequencies, which means the parts of the cochlea near the oval window are more and earlier affected [Nadol Jr, 1993, Wiley et al., 1998]. A further interesting but severe form of hearing loss is the sudden sensorineural hearing loss, which means that the hearing loss appears surprisingly and instantaneously [Hughes et al., 1996, Schreiber et al., 2010]. Unfortunately, the reasons are not well understood yet, but could also be related to vascular diseases [Hughes et al., 1996, Schreiber et al., 2010]. In this context, an underestimated phenomenon is of general interest -the so called transient auditory dysfunction (TAD). The TAD is also a form of sudden sensorineural hearing loss, with a duration of a few seconds to minutes often accompanied with a phantom sound the so-called tinnitus [Almond et al., 2013]. As described in the motivation section, in cognitive neuroscience lesions are a widely used method to unravel the actual processing principles of the brain [Fellows et al., 2005]. A hearing loss is nothing else than a lesion in the periphery of the nervous system, causing reduced auditory input. Therefore, this thesis uses tinnitus and HL as a tool to understand the nature of human auditory perception and cognition. HL has an influence on the neural structures of the auditory pathway (see Fig. 1.2) and thus in the following sections the basics of the auditory pathway are briefly explained.

1.2.3 The auditory pathway

The brainstem

Cochlear nuclei: The axons of the spiral ganglion (called auditory nerve) leave the cochlea and project to the so-called medulla oblongata in the brain or more precisely in the brainstem. In the medulla oblongata are located two conglomerates of cell bodies the dorsal cochlear nucleus (DCN) and the ventral cochlear nucleus (VCN). These two nuclei have different tasks for hearing but in both nuclei input from the auditory nerve is transmitted to the next-higher neuron via one synapse [Hudspeth et al., 2013]. The VCN is connected in such a way that the temporal fine

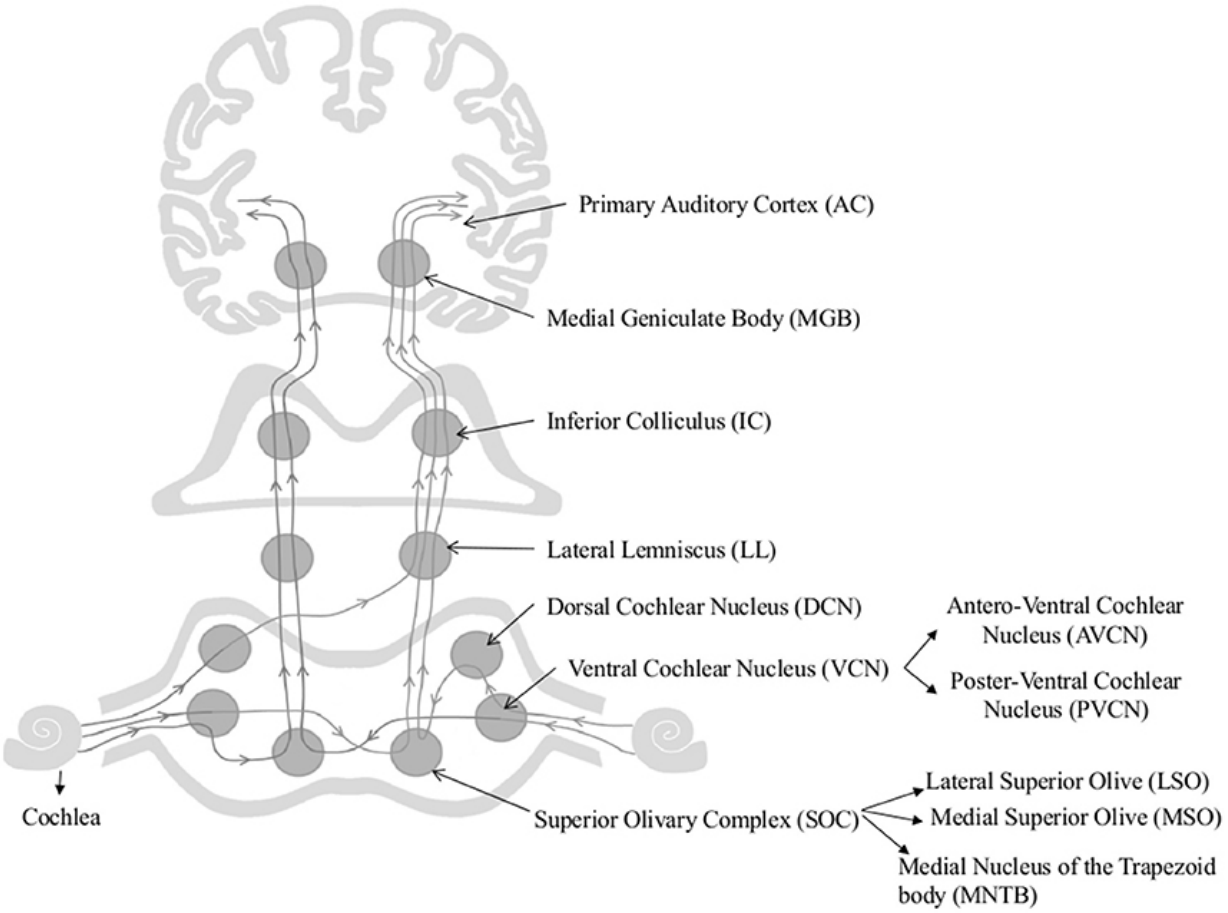

Figure 1.2: **The auditory pathway**
The sound is transformed into an electrical signal in the cochlea including the spiral ganglion. Via the auditory nerve the signal is transmitted to the brainstem. In the brainstem the signal is processed in four different nuclei (cochlear nucleus, superior olivary complex, lateral lemniscus, inferior colliculus). The signal is filtered in the thalamus, more precisely in the medial geniculate body (MGB). The last stage is the cortex, where conscious perception takes place. The information flow is not completely bottom-up, but there are also always back-projections form higher brain areas to lower ones. Thus, there are so called circuits or recurrences. (Figure taken from [Jayakody et al., 2018])

structure can be extracted from the input signal from the cochlea [Hudspeth et al., 2013]. The DCN in contrast has different tasks and a very interesting structure, which is very relevant for the research presented in this thesis. Therefore, the DCN is explained in more detail. The DCN does not only receive input from the auditory system but also from the somato-sensory system (system of touch, e.g. trigeminal nerve) [Shore and Zhou, 2006, Koehler et al., 2011, Young et al., 1995]. It is assumed that the brain might use this information for directional hearing, but

the exact function of the somato-sensory input in the DCN, which is actually specialized on hearing, remains elusive [Hudspeth et al., 2013]. The DCN has a layered structure and consists of three layers. The layer in the middle (layer 2) contains the cell bodies of the so-called fusiform cells (special form of neurons also called pyramidal cells). The fusiform cells have dendrites, which reach the first and the third layer of the DCN [Young et al., 1995]. The so-called basal dendrites of the fusiform cells lie in the third layer, where they get input from the auditory nerve, the VCN and DCN-inhibitory inter-neurons (neurons that inhibit other neurons via an inhibitory neurotransmitter [Studer and Barkat, 2022]). The apical dendrites of the fusiform cells lie in the first layer and receive input from e.g., inhibitory stellate and cartwheel cells [Young et al., 1995] (for a detailed wiring scheme of the DCN see [Oertel and Young, 2004]).

The DCN has a very interesting structure, as it somehow looks like another important brain structure, the so-called cerebellum, which contains approximately 80 % of all neurons in the brain [Herculano-Houzel, 2010]. The cerebellum is not related to hearing, but to motor learning [Ito, 2000]. Despite the fact that the cerebellum is a completely different brain part with disjoint tasks, the structure is similar to the DCN [Oertel and Young, 2004]. The cerebellum has so called delay lines. These structures are well suited to calculate some error signal or to calculate cross-correlations resp. auto-correlations of signals [MacKay and Murphy, 1976, Ivry and Keele, 1989]. Indeed, in the auditory system also auto-correlation functions are calculated, probably via these delay lines [Licklider, 1951]. As the DCN has a similar structure as the cerebellum, the DCN might be the structure in the brainstem, where auto-correlation functions are calculated [Schilling et al., 2023c]. The main principle behind the calculation of an auto-correlation via a delay line is relatively simple. The signal is split into two signal streams. The delay of the signal can be generated via two different mechanisms. The first option is, that the signal can be connected via another neuron, since a chemical synapse leads to a temporal delay of 0.5 ms to several milliseconds [Miyashita and Hikosaka, 1996, Katz and Miledi, 1965]. The second option is to send the signal along one long axon. The conduction velocity in neuronal axons is somewhere between 10-100 m/s [Waxman, 1980, Møller et al., 1989] and varies a lot between different studies. Thus, for example in the auditory nerve a velocity of 20 m/s was measured [Møller and Jannetta, 1983]. Indeed, the conduction velocity depends on the diameter of the axon as well as the thickness of the isolation of the axon called myelin sheath [Salami et al., 2003, Waxman, 1980]. Thus, the velocity linearly increases with the thickness of the myelin sheath and therefore variations of the thickness can be used to tune for the required time delay. Then the delayed stream and the non-delayed stream can be fed to a coincidence detector neuron that "multiplies" the signal [Kapfer et al., 2002, Licklider, 1951]. The multiplied signal has to be summed up for

several timesteps via e.g., integrator neurons or circuits [König et al., 1996]. The auto-correlation can be a valid approximation for the amount of information of a certain signal [Krauss et al., 2017]. The fact that the DCN has the structural prerequisites to estimate the information content of the signal will be important for the research presented in this thesis. Another interesting and universal processing principle along the whole auditory pathway is the so-called lateral inhibition. As described above all parts of the auditory pathway are organized tonotopically [Cheung et al., 2012, Saenz and Langers, 2014, Mann and Kelley, 2011]. Lateral inhibition means that a neuron inhibits its' neighboring neurons, to sharpen the transfer function and to enhance the contrast of the signal [Suga, 1995, Gerken, 1996]. Thus, when a pure-tone (only one frequency, Fourier transform δ-peak) is presented, the traveling wave is no sharp δ-peak at all. However, when the neuron representing the stimulation frequency (center frequency) inhibits the neurons on both sides of the tonotopic map, the transfer function can be re-sharpened again and the blurred transfer function can be balanced out by the brain [Johnstone et al., 1986, Suga, 1995, Gerken, 1996]. This is a universal mechanism along the whole auditory pathway and can be found in other modalities, too [Balboa and Grzywacz, 2000, Urban, 2002].

Brainstem nuclei above cochlear nuclei: After the signal has passed the cochlear nuclei, it is transmitted to three further brainstem nuclei (the superior olivary complex, the lateral lemniscus, and the inferior colliculus (IC)), where the signal is always processed and passed on to the next neuron via one synapse [Hudspeth et al., 2013]. As these regions are less important to understand the content of the thesis, they are only explained very briefly. The superior olivary complex gets input from both ears and also uses the effect of time delays to calculate the direction of the sound resp. the location of the sound source [van der Heijden et al., 2013, Gazzaniga et al., 2014]. The fact that the mechanisms in the superior olivary complex also rely on delays [Smith et al., 1993, Grothe, 2000], illustrates that some mechanisms in the brain are universal and a deep understanding could help to make also progress in AI research or general engineering. Also, the lateral lemniscus plays a role for directional hearing. The IC is assumed to suppress reflections of sound waves on surfaces. The IC is the highest brainstem nucleus the signal has to pass and to reach the thalamus [Hudspeth et al., 2013].

The thalamus: The gate to consciousness: The IC is connected to a nucleus called medial geniculate body (MGB) of the so called thalamus, often called 'gate to consciousness' [Ward, 2013, Berger and García, 2016, Velíšek, 2018]. The thalamus is connected to the brainstem as well as the cortex and is also highly inter-connected with the limbic system (system of emotions, some thalamus parts are even considered to be part of the limbic system, [Vertes et al., 2015]). The

main purpose of the thalamus is to filter out irrelevant information and to transmit only relevant information to the cortex, which is assumed to be the part of the brain, where conscious perception takes place [Jääskeläinen et al., 2004].

The auditory cortex: The cortex is an approx. 2.5 mm thick [Fischl and Dale, 2000] layered structure of neuron cell bodies covering the whole brain. When you look at a brain from a bird's eye view, you only see this thin sheet of neurons. The cortex is assumed to contain all higher cognitive functions such as language understanding, math, or consciousness [Amalric and Dehaene, 2019]. However, the primary auditory cortex (A1) nearly exclusively responds to auditory stimuli and is the first cortical station for the auditory signal coming from the thalamus [Wang et al., 2024, Hudspeth et al., 2013]. The primary auditory (A1) cortex is also organized tonotopically, which means that different stimulus frequencies are still represented by the piano keyboard structure of the cochlea. Neighboring frequencies are still represented by neighboring parts of the auditory cortex [Humphries et al., 2010]. Tonotopic organization is a basic principle of the whole auditory pathway [Kandler et al., 2009]. One further basic principle is that the neocortex (nearly whole cortex, evolutionary newer) looks very similar throughout the whole brain and consists of exactly 6 layers, which can be easily observed in histological sections [Super and Uylings, 2001, Karten, 2015, Rakic, 2009]. It is even more fascinating that there is some kind of canonical wiring scheme in the cortex, which means the layers are wired in a certain way, which indeed applies to all neocortex regions [Budd and Kisvárday, 2012, Shipp, 2007, Nelson, 2002]. Thus, the cortex is no fully connected network. The self-similarity of the cortex suggests a general principle behind it. Potentially, general intelligence does not need an incredibly complicated wiring scheme, but emerges from wiring many similar modules in neuroscience often called micro columns [Jones, 2000, Köster et al., 2014]. Therefore, using the brain as template to build intelligent machines might be a fruitful approach [Russin et al., 2020, Hassabis et al., 2017].

1.3 Beyond the auditory pathway: Speech and language processing

As described above, the cortex shows a canonical wiring scheme. This fact allows for defining some kind of hierarchy of cortical areas [Wessinger et al., 2001, Hilgetag and Goulas, 2020]. After the signal is processed in the primary auditory cortex, the sound must be classified and assigned a meaning. Thus, for example to understand, if a sound is related to e.g., speech or music, certain features have to extracted from the sound. Indeed, there are areas of the cortex

connected to the primary auditory cortex that exclusively respond to speech related sounds such as phonemes, words, and sentences [DeWitt and Rauschecker, 2012]. The superior temporal sulcus and the superior temporal gyrus are specialized on distinguishing between speech and non-speech sounds [Gazzaniga et al., 2014, DeWitt and Rauschecker, 2012]. Lesions to these regions can cause e.g. pure word deafness [Gutschalk et al., 2015]. These brain regions are the last cortex regions related to mainly auditory processing [Gazzaniga et al., 2014]. Higher cortex regions integrate signal from different modalities and do more universal processing. Thus, for example the networks related to language processing have to deal with signals e.g., from the visual system -as language could also be perceived by reading- and the auditory system. Furthermore, also memory and experiences play a crucial role for language processing [Schwering and MacDonald, 2020, Gazzaniga et al., 2014]. Examples for important cortex regions related to language processing are the already mentioned Wernicke's and Broca's areas [Fridriksson et al., 2015]. These two brain regions are connected via a nerve fiber bundle called arcuate fasciculus, which might play a crucial role in processing grammar and semantics [Ardila, 2021, Ivanova et al., 2021].

1.4 Neuroimaging

Due to the enormous progress in neuroimaging techniques and especially the increasing spatial resolution (sub mm) of modern functional magnetic resonance imaging (fMRI) devices [Saha et al., 2021], nowadays we have a good idea where things happen in the brain [Koelbl et al., 2023]. However, the bad temporal resolution of fMRI and more precisely of the blood oxygenation level dependent (BOLD) response, limits the attempt to unravel the processing mechanisms. Furthermore, fMRI is based on a metabolic signal and not on the actual neuronal activity [Saha et al., 2021]. Thus, in this thesis I rely on electrophysiological methods such as intracranial recordings, magnetoencephalography (MEG), and electroencephalography (EEG). In intracranial recordings electrodes are implanted directly in the brain. In humans this technique is called stereotactic EEG (sEEG) or intracranial EEG (iEEG) and is for ethical reasons exclusively applied for diagnostic and therapeutic purposes [Parvizi and Kastner, 2018]. In MEG measurements magnetic fields produced by currents in the brain are recorded using supra-conductive coils [Takeda et al., 2009], whereas far cheaper EEG-devices rely on electrodes placed on the skull of the subject to be measured [Saha et al., 2021]. Unfortunately, EEG and MEG have a bad spatial resolution and the sum activity of millions of neurons is measured with one sensor unit (e.g. electrode) [Murugavel et al., 2016]. However, the good temporal resolution and especially the

combination with intra-cranial recordings offer a promising approach to unravel the mechanisms and mysteries of the human brain [Saha et al., 2021].

1.5 Combining AI and Neuroscience

The aim of my research and this thesis is to use the brain as a source of inspiration to develop efficient AI systems on the one hand (neuroscience-inspired AI [Hassabis et al., 2017]. Thus, the auditory system and the language processing system of the brain are highly efficient and many universal principles such as delay lines, lateral inhibition, tonotopy, recurrences, spiking, briefly described above, could make machine learning (ML)-algorithms more efficient especially in terms of energy consumption. However, this is only one side of the coin. On the other hand, AI can also be used as a tool and a model to understand the brain. Thus, we rely on neuroimaging techniques, that record activity patterns in the brain. However, to translate this data into mechanisms is very challenging and many pitfalls lurk on the way. Thus, without a good hypothesis the data can for example be over-interpreted. In fact, it is not easy to develop a good hypothesis. Therefore, artificial neural networks trained on complex cognitive tasks -which already exist- can be analyzed and potentially broken down into sub-systems ore effects, which could be understood. In a next step, these sub-systems can be searched for in the neuroimaging data. This research philosophy is called CCN [Kriegeskorte and Douglas, 2018]. My complete research approach can be seen as circuit, where AI serves as model for the brain and boosts neuroscience and neuroscience serves as source of inspiration to develop brain-inspired artificial neural networks. In this thesis, I will furthermore illustrate that this approach has indeed the potential to trigger progress in biomedical engineering and therefore is well suited to boost the development of novel therapy approaches. In the main part of this thesis (Chapter 2 Main) I will summarize the most important findings of my research and already discuss the meaning of these findings in the light of existing literature. In the conclusion, I briefly summarize this research, put a special focus on the big picture, and draw conclusion on the meaning of my research for the scientific field in general.

The thematic structure of the cumulative thesis is as follows: Based on my own publications I will first explain how spiking and sparse connections implemented in the brain can be used to develop AI-algorithms. In the second part, I will show based on phantom perception of a sound that has no physical sound source, called tinnitus, how HL can lead to the perception of a tone that is not there and based on that draw conclusions how the auditory system works. Furthermore, I will show how this knowledge can be used to develop a novel therapy approach for that pathological condition. In the final part, I will show that the brain is always predicting future

events and how this property of the brain relates to tinnitus, language processing, and conscious perception. I will show in this context how AI can be used to interpret neural data and how the results might affect the development of novel AI-algorithms.

Chapter 2

Main

2.1 Biologically-inspired neuron models and training algorithms

In the first section of the main part of the thesis, I will describe how we tried to translate the basic principles of information processing in the brain -spiking activity of neurons, sparse coding and sparsity of the connectome [Beyeler et al., 2019, Valdés-Sosa et al., 2005]- to artificial neural networks to build efficient bio-inspired AI systems on the one hand, and to investigate biological mechanisms in computer-models, on the other hand.

2.1.1 Spiking neural networks: The surrogate gradient approach

As described above the brain is incredibly energy efficient with an energy consumption of approx. 20 W [Furber, 2012, Yu et al., 2018]. The brain can fulfill a variety of complex cognitive tasks without consuming much energy, because it uses spikes (action potentials) to process information. Spikes are short ms-voltage-pulses of always the same amplitude [Bean, 2007, Schmidt et al., 2010]. Thus, spikes can be seen as some kind of digital signal. However, the so-called inter-spike-intervals (time-interval between two spikes) can be any real number. Thus, the actual information of a signal processed in the brain is encoded in the inter-spike-intervals [Reich et al., 2000]. Normally, stimuli are encoded by only a few spikes of a few neurons. This principle is called sparse-coding and is one further major reason for the energy efficiency of the brain [Beyeler et al., 2019]. However, standard ML algorithms work totally different. They work with real numbers instead of using digital spikes [Taherkhani et al., 2020]. Thus, standard ML-algorithms can be interpreted as neural networks, that exclusively use resp. compute the frequency of occurrence of

spikes, called spike rate. However, therefore, these networks are less energy efficient and the fast and efficient temporal dynamics of spikes cannot be exploited by them [Gerstner, 2000, Brette, 2015]. Why don't we just develop algorithms that work similar to the brain to achieve comparable energy efficiency and thus to make a step towards Green-AI? The main problem is that it is difficult to train spiking neural networks, for two reasons. First, spiking neuron models such as the Leaky-Integrate-and-Fire (LIF)-Neuron have a self-connection (recurrence). However, recurrent neural networks are difficult and inefficient to train using back-propagation [Li et al., 2016]. The second problem is, that the δ-function of the LIF-Neuron has a gradient that is, except at one single point, always zero. Therefore, back-propagation does not work very well [Lee et al., 2016]. This is the reason why spiking neural networks cannot compete with standard ML-algorithms. To solve this issue, the so called 'surrogate gradient' approach was developed [Zenke and Vogels, 2021]. Thus, the gradient of LIF-Neuron is replaced by an artificial gradient [Zenke and Vogels, 2021]. The choice of the gradient-function is arbitrary; however, a clever choice is necessary to get valid results [Cramer et al., 2022].

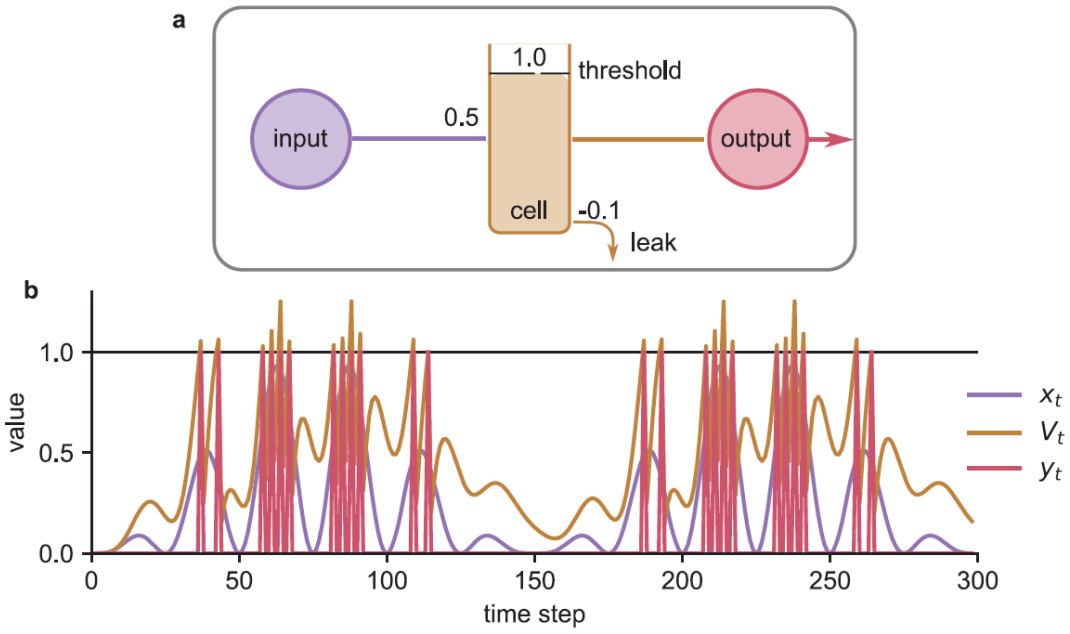

Figure 2.1: **Leaky-Integrate-and-Fire-Neuron**
a: The plot illustrates the mechanism in a single LIF-neuron. The input x_t from e.g., other neurons is multiplied with a weight value and summed up over time. When a certain threshold of the inner state (potential, V_t) is reached, the neuron releases a spike (output, y_t). The summation over time is from an algorithmic point of view a self-connection of the neuron (recurrence). (Figure taken from [Gerum and Schilling, 2021])

In our publication "Integration of Leaky-Integrate-and-Fire Neurons in Standard Machine Learning Architectures to Generate Hybrid Networks: A surrogate gradient approach" (**Publication 1**, [Gerum and Schilling, 2021]), we have shown that a simple surrogate gradient is well suited to train LIF-Networks supervisedly. The LIF-neuron is a simplified form of the Hodgkin-Huxley and Fitzhugh-Nagumo neurons consisting of several coupled differential equations simulating the membrane currents of real neurons [Izhikevich and FitzHugh, 2006, Hodgkin and Huxley, 1952]. The LIF-Neuron can be described with one differential equation [Koch and Segev, 1998]:

$$I(t) = \frac{V_\mathrm{m}(t)}{R_\mathrm{m}} = C_\mathrm{m} \cdot \dot{V}_\mathrm{m}(t) \tag{2.1}$$

($I(t)$: input signal coming e.g., from other neurons, R_m resistance of membrane, C_m: capacity of membrane, $\dot{V}_\mathrm{m}(t)$: derivative (by time) of the membrane potential). This differential equation can be converted into a recursive notation (for detailed steps see [Gerum and Schilling, 2021]):

$$V_{\mathrm{t_n}} = w_\mathrm{input} \cdot x_{\mathrm{t_n}} + (1 - w_\mathrm{leak} \Delta t) \cdot V_{\mathrm{t_{n-1}}} \cdot \Theta_2(V_\mathrm{thresh} - V_{\mathrm{t_{n-1}}}) \tag{2.2}$$

$$y_{\mathrm{t_n}} = \Theta_1(V_{\mathrm{t_n}} - V_\mathrm{thresh}) \tag{2.3}$$

($\Theta_{1,2}$: Heaviside step function. Index of Θ is only used for later explanations. It is always the same step function. w_input: weight factor of input signal $x_{\mathrm{t_n}}$, V_thresh: membrane potential threshold for spike release. $y_{\mathrm{t_n}}$: output value, spike 0 or 1). This formula can be used to simulate the dynamics of the LIF-Neuron in time (see Fig. 2.1). To train a neural network, which consists of these LIF-neurons, the gradient of the output $y_{\mathrm{t_n}}$ has to be calculated (see [Gerum and Schilling, 2021]). If we want to achieve that the gradient does not vanish, we can choose a surrogate gradient. We found that the simplest solution is $\Theta_1'(x) = 1$ and $\Theta_2'(x) = 0$ (for complete derivation see [Gerum and Schilling, 2021]). In, addition to that surrogate gradient, we found a trick to improve the training accuracy and the performance of the LIF-network. As the network, was often stuck in solutions, where the membrane potential became too negative and the threshold potential (V_thresh) was never reached, we forced the potential to be always positive (note that we normalized the potential and it has not the same values as in a real neuron). We just used a ReLu-function to enforce the membrane potential to be positive [Gerum and Schilling, 2021].

$$V_{\mathrm{t_n}} = \mathrm{ReLu}[w_\mathrm{input} \cdot x_{\mathrm{t_n}} + (1 - w_\mathrm{leak} \Delta t) \cdot V_{\mathrm{t_{n-1}}} \cdot \Theta_2(V_\mathrm{thresh} - V_{\mathrm{t_{n-1}}})] \tag{2.4}$$

Using this surrogate gradient in combination with our formulation of the membrane potential, we trained a neural networks supervisedly on the MNIST image data set [LeCun et al., 1995].

Thus, we translated the images into spikes using 'Poisson-rate' coding and trained the LIF network on image classification [Cramer et al., 2020]. We could show that the surrogate gradient is sufficient to train the network and that more complicated gradients, used in other studies, are not necessary at all (cf. [Neftci et al., 2019]). However, the spiking networks are up to now far away to achieve classification accuracies comparable to state-of-the-art networks [Tavanaei et al., 2019]. Nevertheless, this is a first important step towards a spiking-neuron-based machine intelligence. Additionally, these networks are an interesting model for brain activity and thus could be used to perform in-silico neuroscientific studies. As the spiking neural networks are difficult to train, up to now often dynamics of randomly connected spiking neural networks is investigated [Metzner et al., 2024]. Thus, our approach has indeed some relevance for computational neuroscience.

2.1.2 Spiking Long-Short-Term-Memory networks

In the previous section, we tried to integrate a concept from computational neuroscience in ML-algorithms to generate biologically plausible and efficient artificial neural networks. However, it is also possible to approach this problem from the other side, which means to start from AI-research. There already exist neuron models with an internal memory comparable to the LIF-neurons, the Long-Short-Term-Memory (LSTM)-neurons. Thus, in the next study we did not use a concept from computational neuroscience- the LIF-neurons and tried to train them supervisedly with backpropagation, but we used the in AI-research already established LSTM-neuron model introduced by Schmidhuber and Hochreiter [Hochreiter, 1997] and tried to make this neuron model behave like a biological inspired LIF-neuron [Seenivasan et al., 2024]. The previous study could be regarded as an 'Neuroscience-Inspired AI' approach as suggested by Hassabis and coworkers [Hassabis et al., 2017]. Based on our award-winning publication, I here illustrate the 'CCN' approach suggested by Kriegeskorte and Douglas [Kriegeskorte and Douglas, 2018].

In our study "Leaky-Integrate-And-Fire Neuron-Like Long-Short-Term-Memory Units as a Model System in Computational Biology" (**Publication 2**, [Gerum et al., 2023]), we used so called peephole-LSTM units (see Fig. 2.2 a) to investigate under which conditions these units behave similar to biologically plausible LIF-neurons (see Fig. 2.2 b, [Gers and Schmidhuber, 2000, Gers and Schmidhuber, 2001, Yang et al., 2018]). Peephole-LSTM units have more intrinsic connections than standard LSTM-units [Bo et al., 2019, Rahman and Siddiqui, 2019]. Thus, the input, the output, the bias, and the internal state can influence all gates [Gerum et al., 2023] of the neuron model (see Fig. 2.2 a, [Gers and Schmidhuber, 2000, Gers and Schmidhuber, 2001, Yang et al., 2018]). The crucial connection for our study is the connection of the internal

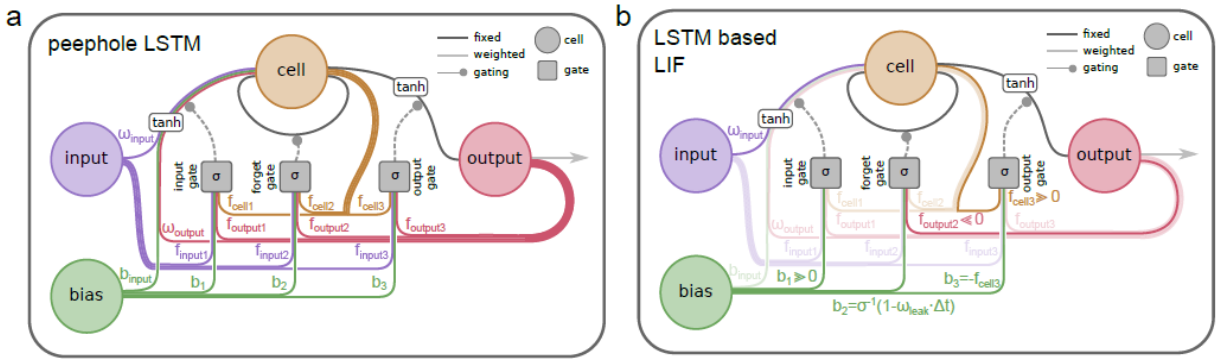

Figure 2.2: **Long-Short-Term-Memory-Neuron** a: Graphical illustration of a peephole-LSTM neuron. The fact that all inner states of the neuron are connected to all gates, allows for implementing a spike-threshold in analogy to the LIF-neurons. b: Fine-tuned peephole-LSTM with LIF-like spiking behavior.
(Figure adapted from [Gerum et al., 2023], reduced figure)

state of the neuron to the output gate that decides if an activation is transmitted to the next neuron. This connection can be used to implement an output-threshold (or spike threshold) typical for a biological neuron and also biologically inspired neuron model such as the LIF neuron [Gerum and Schilling, 2021, Gers and Schmidhuber, 2000]. The peephole-LSTM-unit is defined by the following equations (from [Gers and Schmidhuber, 2000, Gers and Schmidhuber, 2001, Yang et al., 2018]):

$$\begin{aligned} V_{t_n} &= \tanh(b_{\text{input}} + w_{\text{input}} \cdot x_{t_n} + w_{\text{output}} \cdot y_{t_{n-1}}) \\ &\cdot\sigma(b_1 + f_{\text{input1}} \cdot x_{t_n} + f_{\text{cell1}} \cdot V_{t_{n-1}} + f_{\text{output1}} \cdot y_{t_{n-1}}) \\ &+V_{t_{n-1}} \cdot \sigma(b_2 + f_{\text{input2}} \cdot x_{t_n} + f_{\text{cell2}} \cdot V_{t_{n-1}} + f_{\text{output2}} \cdot y_{t_{n-1}}) \end{aligned} \tag{2.5}$$

$$\begin{aligned} y_{t_n} = \tanh(V_{t_n-1}) \cdot \sigma(b_3 + f_{\text{input3}} \cdot x_{t_n} + f_{\text{cell3}} \cdot V_{t_{n-1}} \\ +f_{\text{output3}} \cdot y_{t_{n-1}}) \end{aligned} \tag{2.6}$$

(same notation as for the previous study [Gerum and Schilling, 2021], V_t: internal state, x_t: input, $y_t \epsilon\ ]-1, 1[$: output, $\sigma(x)$: sigmoid activation function, $\tanh(x)$: hyperbolic tangent, $w_{\text{input,output}}$: input resp. output weights, $b_{1,2,3}$, $f_{\text{input1,2,3}}$, $f_{\text{output1,2,3}}$: gate weights, for graphical scheme see Fig. 2.2 a). The equations 2.5, 2.6 look relatively similar to the equation 2.2, 2.3 of the LIF-neuron. We could show that a clever choice of the parameters can make the peephole-LSTM units behave like LIF-neurons (for complete mathematical derivation see [Gerum et al.,

2023]). Furthermore, we have illustrated that the inter-connection of two standard LSTM-units could show similar behavior to one spiking peephole-LSTM-unit (see [Gerum et al., 2023]). This additional finding, is important, as often LSTM units without peepholes are used in ML research [Bo et al., 2019, Rahman and Siddiqui, 2019]. We tested the validity of our approach by training two neural networks on image classification of images of four different data sets (MNIST, FashionMNIST, EMNIST Letter, CIFAR10, [LeCun et al., 1995, Xiao et al., 2017, Cohen et al., 2017, Krizhevsky, 2009]) and Poisson-rate-coding analogously to the previous study [Gerum and Schilling, 2021, Cramer et al., 2020]. One network consisted of LIF-neurons and was trained using the surrogate gradient approach described above [Gerum and Schilling, 2021]. The other network consisted of our spiking peephole-LSTM neurons and was trained with normal gradient descent [Gerum et al., 2023]. After training the weight matrices were kept constant but the LIF neurons were replaced by spiking LSTM neurons and the other way round. We found no significant difference in classification accuracy as well as in the spatio-temporal patterns of the neurons in the hidden layers after switching neuron models (See Fig. 2.3). These results indicate that the spiking characteristics of tuned peephole-LSTM units is similar to biologically inspired LIF neurons [Gerum et al., 2023].

Why should the fact that a peephole-LSTM unit or two normal LSTM units could be tuned to behave like a LIF neuron be interesting or relevant for the ML community or the computational neuroscience community? Indeed, these insights open up new doors for both disciplines. On the one hand, LSTM-units are a well-established neuron model, which can be efficiently trained with backpropagation [Hochreiter, 1998]. Thus, it is perhaps possible to use spiking LSTM-units to generate AI-models that are able to perform complex cognitive tasks exploiting the great advantages of spiking neurons as described in the previous section (see also [Gerstner, 2000, Brette, 2015]). These networks could be trained with backpropagation and the spiking-characteristics of the peephole-LSTM units can be varied by means of various adjustment screws [Gerum et al., 2023]. The paper shows in detail how different parameter combinations affect the spiking characteristics of the peephole-LSTM units. However, the study is also of relevance for the computational neuroscience community. Nowadays, we already have neural networks based on LSTM-units with already fascination capabilities. One great example is ELMO [Peters et al., 2018, Liu et al., 2020, Zhang et al., 2020], a neural network based on LSTM-units, used to do efficient language processing [Ronran et al., 2020, Zhang et al., 2020] (Note that language processing in artificial and biological neural networks will be further discussed in later parts of the thesis.). The fact that we already have LSTM-networks that are able to solve cognitive tasks, opens up further doors. These networks can be regarded as artificial brains and thus can

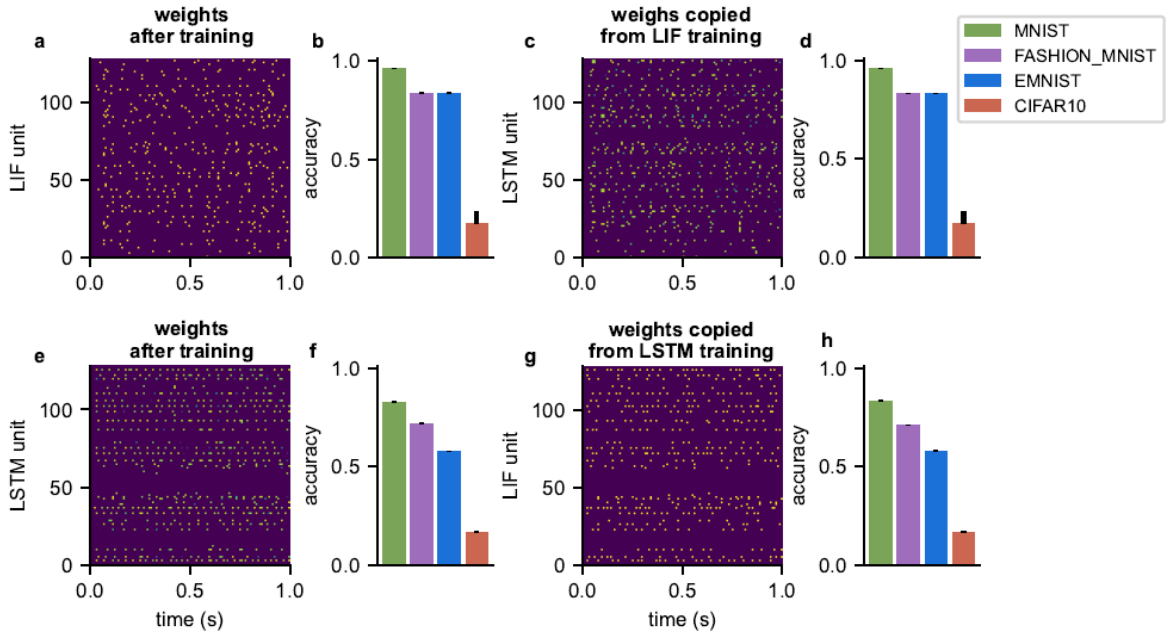

Figure 2.3: **Comparison of spiking LSTM and LIF neurons in supervisedly trained neural networks**
We trained one LIF network and one spiking LSTM network on image classification. a: Activation patterns of the hidden LIF neurons and classification accuracies of the trained network for 4 different data sets (b). Exemplary activation patterns of hidden peephole-LSTM units (e) and classification accuracies of LSTM network (f) are shown. We replaced the LIF neurons with peephole LSTM neurons and the other way round and analyzed the neuron activation patterns (c, g) as well as the classification accuracies again (d, h). Replacement of LIF neurons as well as peephole-LSTM neurons had no significant effect on activation patterns as well as the classification accuracies. (Figure taken from [Gerum et al., 2023])

be investigated with methods from neuroscience (see e.g. [Jonas and Kording, 2017]). The fact that we have shown that the LSTM-units are indeed to some extend biologically plausible, makes these networks interesting for computational neuroscience as well as CCN research [Kriegeskorte and Douglas, 2018].

The relevance of the work for various communities is emphasized by the fact that the paper was awarded best paper at the International Joint Conference on Neural Networks 2023 (IJCNN 2023), the largest conference on neural networks. The paper was chosen out of more than 1000 accepted and more than 1800 submitted scientific publications [IJCNN, 2023].

2.1.3 Biologically-inspired restrictions and the influence on robustness

There are more abstract features and facts on the brain, which might help to improve existing ML algorithms. As described in the introduction section, the brain of humans consists of approximately 86 billion neurons [Herculano-Houzel, 2009], which means that there are approximately 10^{22} possible connections between neurons in the brain. However, the brain has only 10^{15} synapses, which means that only 1 out of 10 million possible connections between neurons does actually exist [Gerum et al., 2020, Hagmann et al., 2008, Sporns et al., 2005]. Up to now, I exclusively presented facts on human brains, however, it might be a good idea to look at other species to identify general principles of cognition. Birds are an interesting target, as birds are able to perform incredibly difficult cognitive tasks, but the brains of birds contain less neurons and are far lighter than the brains of mammals. Indeed, the overall cognitive performance of corvids is similar to the performance of apes [Emery and Clayton, 2004]. However, corvids' palliums contain approximately 6 time less neurons than apes' pallium [Güntürkün et al., 2017]. Indeed, there exist very special tasks, where pigeons perform better than humans [Herbranson and Schroeder, 2010]. How, is this possible? The actual reason is not fully understood. However, there is a good theory. Weight plays a critical role for birds, as they need to fly. Thus, the additional restriction of low-weight has potentially led to the evolution of smaller, lighter, and more efficiently connected brains [Emery, 2006, Güntürkün et al., 2017]. Especially, in the context of developments in modern AI, where the number of parameters and calculation power is always increasing, more efficient but less energy consuming networks might be an interesting step towards a 'Green AI' [Verdecchia et al., 2023].

Our aim was to adapt, the principle of sparse connectivity and the introduction of additional restriction, to create small neural networks, which are able to navigate efficiently through a maze (see Fig. 2.4)). In our study "Sparsity through Evolutionary Pruning Prevents Neural Networks from Overfitting" (**Publication 3**, [Gerum et al., 2020]) we used small (16 neurons) fully connected neural networks (for method description see [Gerum et al., 2020]) and trained them on running as fast as possible through a maze (shown in Fig. 2.4 c). The network (agent) gets input information on the distance to walls in all spatial directions and about its own orientation. Furthermore, the networks are trained using an evolutionary algorithm on fulfilling the maze task. The portion of the individuals that perform better, are more likely to reproduce [Gerum et al., 2020]. Thus, the evolutionary selection pressure is the performance of the networks in the maze. After a new generation of individuals was created depending on the performance of the parent generation, the networks are 'mutated'. Thus, the weights as well as the mutation rate are slightly changed by chance. The most important mutation for this project is the so called

'connection mutation'. Thus, also some neuron-connections are removed by chance (for exact numbers see [Gerum et al., 2020]).

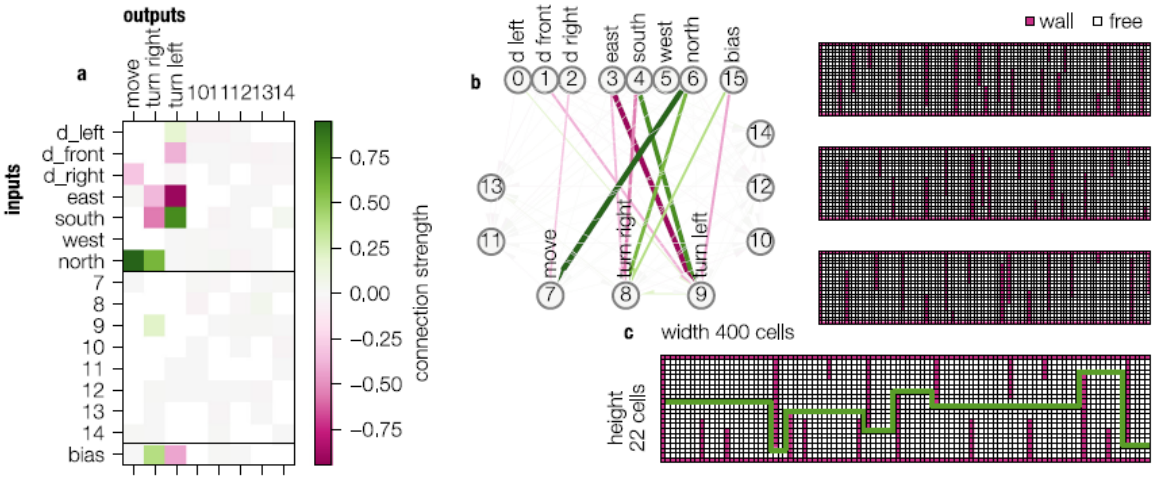

Figure 2.4: **Maze task for evolutionary pruning study**
a: The connectivity (weight) matrix of the network, which is able to navigate through the maze; b: Connectivity matrix shows as graph; c: Exemplary mazes used for training and testing and the ideal trajectory through one exemplary maze. (Figure taken from [Gerum et al., 2020])

The 'connection mutation' is the critical method of this study. Thus, networks which only rely on a small number of neurons to run through the maze, are less likely to lose a critical connection. Thus, there is a selection pressure towards using small networks. Indeed, small feed-forward networks emerged from the fully connected networks. These networks showed a very good performance in the maze task. The fact that this biologically, inspired learning algorithm has led to the emergence of very small but efficient neural networks, that are able to reach a goal by walking around obstacles, gives us an idea on how the nematode (worm) c. elegans is able to fulfill these tasks. C. elegans is a widely used model in neuroscience, as it only possesses 302 neurons and thus the full connectome is already completely known [Jarrell et al., 2012, White et al., 1986]. Indeed, this worm is able to perform thermo- and electro-taxis, which means it can follow a thermal and electrical gradient [Gabel et al., 2007] using only 302 neurons. The 'compass' neurons in our model might relate to the sensory cells of that worm. Thus, our simulation provides some insights in biology on the one hand and suggests some improvements for AI research on the other hand. Potentially, it is a good idea to add some restrictions in the training process to enforce neural networks to find solutions, which are more efficient in terms of connectivity and size.

2.1.4 Summary on biologically-inspired neuron models and training algorithms

In summary, I have presented three studies that used principles from neuroscience to better understand and potentially to improve efficiency of deep neural networks. Thus, we developed a simple surrogate gradient to train biologically-inspired LIF-neurons through backpropagation [Gerum and Schilling, 2021]. In addition to that, we showed that peephole-LSTM units can also behave like spiking LIF-units [Gerum et al., 2023]. Therefore, these tuned LSTM-units might be one further option to train spiking neural networks supervisedly. Potentially, these spiking LSTM units have already emerged automatically in pre-trained LSTM-networks, which now can serve as a model for biology as they are already capable of performing complex cognitive tasks on the one hand and exploit the dynamic processing of spiking neural networks on the other hand [Gerstner, 2000, Brette, 2015]. We furthermore showed that sparsity is a property, which is not only advantageous in biological neural networks but might also be useful in artificial neural networks. We found that using an evolutionary algorithm including a pruning step leads to the emergence of efficient feed-forward networks that are able to navigate through mazes [Gerum et al., 2020]. Therefore, we provide some ideas how navigation in e.g., the worm c. elegans might work. In the next part of the thesis, the concept of using artificial neural networks as a tool and as a model to understand the brain is extended to the so-called lesion studies. As already, described in the Introduction, in lesion studies the consequences of a damaged neural networks are analyzed to reverse-engineer the actual function of certain brain circuits [Gazzaniga et al., 2014]. Thus, in the next part, I will explain the consequences of damages along the auditory pathway and especially in the cochlea. Cochlear damage is often the origin of hearing loss (HL) and can induce a cascade of secondary effects in the brain, which I have investigated throughout approximately the last decade.

2.2 Insights from lesion studies in the auditory system: Auditory phantom perception

2.2.1 Relevance of lesion studies in the auditory system for neuroscience and medicine

As described above, a common approach to understand the brain are so called lesion studies. Lesions in the brain or the peripheral nervous system can lead to effects that can be used to reverse-

engineer mechanisms of certain brain regions [Gazzaniga et al., 2014]. This thesis mainly deals with the auditory system. The most common lesion in the auditory system is HL. Thus, damages in the cochlea can lead to reduced input from the ear to the central nervous system (CNS) [Schaette and McAlpine, 2011]. A damage of the cochlea is sometimes not easy to detect, as normal so called pure-tone audiograms used in clinical practice are often not fine-grained enough to detect the hearing loss caused by the cochlear damage [Schaette and McAlpine, 2011, Tziridis et al., 2021]. Thus, this damage is then called hidden HL [Schaette and McAlpine, 2011]. Histologic studies can help to identify hidden damages at the IHCs- in most cases a loss of ribbon synapses connecting IHCs with the spiral ganglion neurons ([Tziridis et al., 2021], for an exact explanation see Introduction). A very common consequence of HL is an auditory phantom percept, widely known as tinnitus. Nowadays, there is a large consensus that cochlear damage and thus (hidden) HL is the inducer of tinnitus [Krauss et al., 2019, Knipper et al., 2013, Eggermont and Roberts, 2004]. However, in this context the word tinnitus has to be specified. There are two forms of tinnitus: the objective tinnitus and the subjective tinnitus [Prengel et al., 2023]. The objective tinnitus is the perception of a sound produced in the inner ear, that can be measured using a really sensitive microphone [Prengel et al., 2023]. Objective tinnitus is rare and has no neuronal origin and is thus no real tinnitus at all and not interesting for lesion studies [Prengel et al., 2023]. Subjective tinnitus in contrast refers to the perception of a sound without any physical sound source (intrinsic or extrinsic, [Prengel et al., 2023]). From now on, the word tinnitus refers to subjective tinnitus. Tinnitus can be divided into different sub-forms depending on how the phantom percept sounds and how long it is present. Thus, approximately 50 % of the affected hear a pure-tone like sound (one frequency) [Prengel et al., 2023]. Approximately 25 % hear some broad-band noise and ca. 25 % hear some mixture often called complex tinnitus [Prengel et al., 2023]. In clinical practice, a tinnitus is called acute, when the affected person suffers less than 3 months from the tinnitus and chronic when the person suffers more than 3 months. Unfortunately, there are no internationally valid standards yet [Prengel et al., 2023]. Why is it so important to use tinnitus as a model to understand auditory processing on the one hand but also to understand the pathological mechanism on the other hand? In Europe 8.7 %-28.3 % of the population suffer from tinnitus, with a prevalence of 12 % in Germany [Biswas et al., 2022]. Alone in Germany, tinnitus leads to socio-economic costs of 21.9 billion euros, which is within the same order of magnitude as the costs of diabetes mellitus [Tziridis et al., 2022b]. Thus, a cure for tinnitus would have a huge impact on the quality of life of the affected persons but also on the economy of a country. For these reasons, much effort is spent on understanding and treating tinnitus. Thus, e.g., the number of scientific publications has significantly increased over the last decades form

approximately 20 in 1980 to approximately 600 in 2020 [Yaz et al., 2023]. Despite the fact that a lot of money is spent on developing treatment strategies, up to now the neural mechanisms of tinnitus are not fully understood and there is no real mechanistic treatment yet [Prengel et al., 2023]. Tinnitus has certain properties that could help to understand the underlying mechanisms.

- As described above, tinnitus is presumably always related to a cochlear damage [Krauss et al., 2016, Knipper et al., 2013, Eggermont and Roberts, 2004]. Thus, tinnitus itself is no illness but just a symptom of the impaired auditory processing [Baguley et al., 2013].
- However, not everyone with HL perceives tinnitus [Prengel et al., 2023, König et al., 2006].
- It was found that tinnitus is related to increased spontaneous firing of the neurons along the auditory pathway [Kaltenbach and Afman, 2000, Kaltenbach, 2007].
- 30 %- 80 % of tinnitus patients suffer from an over-sensitivity against mild sounds called hyperacusis [Pienkowski, 2019, Bigras et al., 2022], whereas 86 % of hyperacusis patients suffer from tinnitus [Baguley, 2003]. Indeed, there exist patients with and without hyperacusis with and without tinnitus.
- Somatosensory stimuli can change the loudness of the perceived tinnitus called somatic tinnitus [Marks et al., 2018, Shore et al., 2007].
- More recent studies have shown, that the hearing thresholds of tinnitus patients out of a huge cohort of people with hearing loss are lower compared to the people without tinnitus [Krauss et al., 2016, Gollnast et al., 2017, König et al., 2006]. Thus, tinnitus is somehow related to better hearing.
- In a very recent study, it has been reported that tinnitus patients suffer under certain circumstances less from cognitive decline in terms of short-term memory than an age and HL-matched control group [Hamza and Zeng, 2021]. This is an surprising finding, as tinnitus is often related to concentration issues or even dementia [Andersson and McKenna, 2006, Bauer, 2018, Yang et al., 2024].

At this point, we have collected a lot of knowledge on tinnitus symptoms and certain side effects, but despite all efforts, the neural mechanisms behind tinnitus are not fully understood yet. How is this possible? The first problem is that tinnitus can be considered on many different levels. Thus, the tinnitus related neuronal hyperactivity in the brainstem could be seen as tinnitus [Krauss et al., 2016]. However, tinnitus could also be interpreted as the conscious perception of a sound

in the cortex that disturbs affected people. The conscious perception of a tone that is not there rises the question: Why is the neuronal signal not filtered out by the thalamus assumed to be the gate to consciousness? Different theories try to explain that ([De Ridder et al., 2011, De Ridder et al., 2015, Koops and Eggermont, 2021], for a detailed discussion see [Schilling et al., 2023c]). Other theories assume tinnitus as a mis-prediction resp. mis-interpretation of the brain and explain it by means of the Bayesian brain or the predictive coding framework [Friston, 2012, Sedley et al., 2016, Schilling and Krauss, 2024, Yasoda-Mohan et al., 2024]. However, tinnitus can be also regarded as a precept that leads to suffering of the affected. Thus, the way how the phantom perception is interpreted by the affected persons plays a crucial role, too [Andersson and McKenna, 2006, Mazurek et al., 2012, Mazurek et al., 2015]. Tinnitus has to be and is investigated on all different levels. Another problem for tinnitus research is that we lack good model systems, which we could use for deeper investigations of the neural mechanisms. Thus, patients go to the ENT-hospital when they already suffer from tinnitus. However, for neuroimaging approaches, it would be helpful to contrast the findings in the impaired system with the healthy condition. However, this is not possible as the 'before-tinnitus'-measurements do not exist. Thus, only a population comparison can be done [Dauman et al., 2015]. However, inter-individual differences are often too big to get valid results. Thus, a model system, which could be switched off and on at will, would be a great step forward.

2.2.2 Zwicker tone: the 'little brother' of tinnitus

Indeed, there exists one auditory phantom percept that may be caused by mechanisms similar to tinnitus, the so called "Zwicker tone" [Zwicker, 1964, Franosch et al., 2003]. In 1964 Eberhard Zwicker observed that after the presentation of a so-called notched noise (NN) stimulus, study participants perceived a pure-tone like sound. In analogy to the visual system, Eberhard Zwicker called this sound an auditory "negative after image" [Zwicker, 1964]. NN is white noise (WN), where certain frequencies are notched out, i.e., WN filtered with a band-stop filter [Fastl and Stoll, 1979]. The huge advantage of Zwicker tone is that this phantom percept is perceived for just a few seconds [Wiegrebe et al., 1996]. The exact duration of a Zwicker tone depends on the duration of the preceding NN stimulus [Schilling et al., 2023a]. Furthermore, the frequency of the perceived Zwicker tone lies within the spectral gap (notch) of the NN [Zwicker, 1964]. These properties would make Zwicker tone the ideal model for tinnitus [Norena et al., 2000, Mohan et al., 2020], as the 'tinnitus' frequency could be tuned by changing the notch frequency and the presentation of NN does not lead to any permanent damage in the auditory system. Thus, the presentation of a NN is by some researchers interpreted as a transient resp. simulated hearing loss [Hullfish

et al., 2019, Krauss and Tziridis, 2021, Schilling et al., 2021c], as the auditory system is fooled to believe that there is a cochlear damage at the notch frequency. One further finding that points to the direction that Zwicker tone and tinnitus are based on similar neural mechanisms is that tinnitus patients are more likely to perceive a Zwicker tone after NN presentation than people without tinnitus [Parra and Pearlmutter, 2007]. The possible applications of the Zwicker tone as model for tinnitus are versatile. Thus, experiments in humans are ethically safe and it is possible to investigate the time course of such a phantom percept. Intra-individual comparisons between different conditions can be analyzed and it is potentially even possible to test certain tinnitus treatment approaches in humans [Dauman et al., 2015]. However, to use Zwicker tone as a model for tinnitus it is necessary to find out, whether the underlying mechanisms of tinnitus and Zwicker tone are really similar. The problem is that as described above, tinnitus is caused by HL, which is an irreversible damage of the auditory system. Thus, the neural mechanisms of Zwicker tone and tinnitus cannot be easily compared in human subjects. Therefore, it is a sophisticated approach to compare tinnitus and Zwicker tone in animals. The ultimate goal of these experiments is to establish Zwicker tone as model for tinnitus and thus to reduce animal burden and the number of animal experiments on the long run [Krauss and Tziridis, 2021, Loss et al., 2021].

However, to do so another issue has to be solved. It is not clear, if animals actually do perceive tinnitus and Zwicker tone at all. Thus, behavioral tests were developed, to find out if animals show tinnitus related behavioral signs. Some tinnitus tests are based on operant conditioning, i.e., on learning experiments [Hayes et al., 2023, Zuo et al., 2017]. However, also learning induces plasticity in the brain, that might interfere with the plasticity caused by or related to tinnitus. Thus, Turner and coworkers developed a method based on animal reflexes [Turner et al., 2006]. The idea is to present the animal a narrow band noise with a small gap of silence. The gap of silence is followed by a loud startle stimulus leading to twitching (startling) of the animal. This twitching (startle amplitude) is recorded using a force or acceleration sensor [Turner et al., 2006, Gerum et al., 2019]. The condition with gap of silence is compared to a control condition without any gap. The critical measure is the so called pre-pulse-inhibition (PPI) ($PPI = 1 - \frac{A_{\text{gap}}}{A_{\text{nogap}}}$, A_{gap}: startle amplitude for stimulus with gap of silence before startle stimulus, A_{nogap}: startle amplitude without gap of silence, [Gerum et al., 2019, Schilling et al., 2017]). The gap of silence serves as some kind of 'warning' for the animal causing reduced startle amplitudes. Thus, an animal the perceives the gap of silence has lower startle amplitudes and thus a PPI-value near 1. However, if the animal perceives a tinnitus, the tinnitus could mask the gap of silence, which leads to a decreased PPI-value. Thus, a decreased PPI-value is interpreted as a behavioral sign of tinnitus [Turner et al., 2006]. However, this method should only work, if the tinnitus sounds

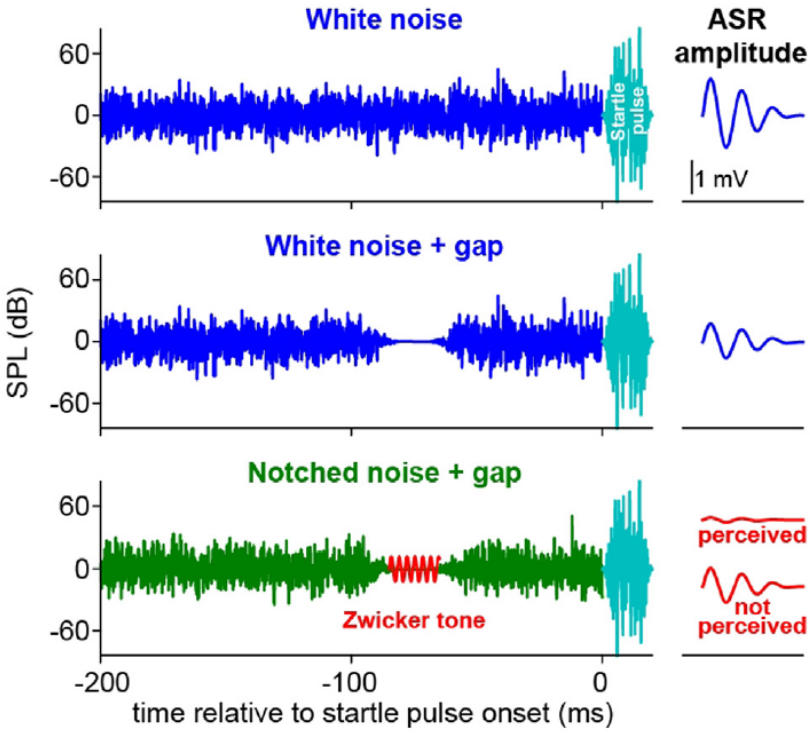

Figure 2.5: **GPIAS paradigm for Zwicker tone testing**
The figure shows the novel GPIAS paradigm for testing for the presence of Zwicker tone in animals. Reference condition: White noise stimulus (dark blue) with startle pulse. Second Control: White noise stimulus (dark blue) with gap of silence. This leads to a normal pre-pulse inhibition. Test stimulus: Notched noise (green) stimulus mutually causing a Zwicker tone percept that fills the gap of silence (red). Hypothesis: Notched noise stimulus leads to altered pre-pulse inhibition compared to white noise. (Figure taken from [Schilling et al., 2023d])

similar to the narrow band noise, which means the tinnitus frequency must be similar to the center frequency of the narrow band noise [Wilson et al., 2020]. This tinnitus test is called gap pre-pulse inhibition of the acoustic startle response (GPIAS) paradigm and is often criticized, as it is not validated enough to be sure that the decreased PPI is really a sign of the perception of tinnitus [Galazyuk and Hébert, 2015].

In our study "Behavioral Assessment of Zwicker Tone Percepts in Gerbils" (**Publication 4**, [Schilling et al., 2023d]) we developed an altered GPIAS paradigm to test the validity of the paradigm for tinnitus testing from Turner and coworkers on the one hand and to find out [Turner et al., 2006], if animals also perceive a Zwicker tone. This information is needed to better understand, if the neural mechanisms of Zwicker tone and tinnitus are similar and thus to establish Zwicker tone as transient tinnitus model as described above. To do so, we used our self-developed

open-source startle setup [Gerum et al., 2019]. The animal to be measured is kept in a restrainer placed on a movable construction including a 3D acceleration sensor (for detailed description of the setup see [Gerum et al., 2019]). Two loudspeakers are used to present the noise stimuli (WN, NN) and the noise burst used as loud startle stimulus. Twitching of the animal is recorded with the acceleration sensor. Three stimulus conditions are used. First, a reference stimulus is presented consisting of WN followed by a startle pulse. The second stimulus is the same noise but a gap of silence is included to induce a PPI. The third stimulus is used to test for Zwicker tone. The stimulus looks like the second stimulus except the fact that the WN is replaced by the potentially Zwicker tone inducing NN (see Fig. 2.5). We also had a control condition for NN, where no gap of silence was added. This stimulus is not really necessary, as the WN control stimulus sounds similar to the NN control stimulus. We calculated the PPI values for the NN condition and the WN condition. We found a significant higher PPI in the NN condition, which indicates that the Zwicker tone does not fill the gap at all, but serves as pre-stimulus itself. Thus, the PPI is increased by the Zwicker tone percept as the contrast of Zwicker tone in silence and the surrounding NN is higher than the contrast of silence and the surrounding NN. The significant difference indicates that indeed a Zwicker tone is perceived by the animals. Additionally, this finding indicates that the original Turner-paradigm might be indeed suited to estimate the frequency of the perceived tinnitus [Turner et al., 2006, Schilling et al., 2023d], as our measurement shows that the PPI is only decreased when the stimulus in the gap has a similar frequency to the surrounding noise. Therefore, we have validated a paradigm used for tinnitus research over the last nearly two decades [Turner et al., 2006]. Additionally, we found evidence that our animal model might be useful for Zwicker tone research. In addition to the GPIAS paradigm, we developed a learning paradigm that also indicated that animals perceive Zwicker tone [Schilling et al., 2023d]. Thus, the animals were trained in a so-called shuttle box on dividing between a WN stimulus followed by a pure tone (used to simulate the Zwicker tone) and WN followed by silence. After the training period, the animals have to discriminate between WN followed by silence and NN followed by silence in a test session. The NN is assumed to induce a pure-tone like Zwicker tone similar to the pure-tone presented during training. If the animals perceive the Zwicker tone, they should be able to distinguish between NN and WN. In tendency, the animals show behavioral signs of Zwicker tone [Schilling et al., 2023d]. In summary, we have collected evidence for the idea that Zwicker tone can be investigated in animal models. On basis of this finding, electrophysiological measurements were performed to find Zwicker tone related neural activity in the brain.

In the study, which I conducted during my 11 months stay at the Aix-Marseille-University, Marseille, France, together with colleagues, we analyzed the propagation of neural activity through

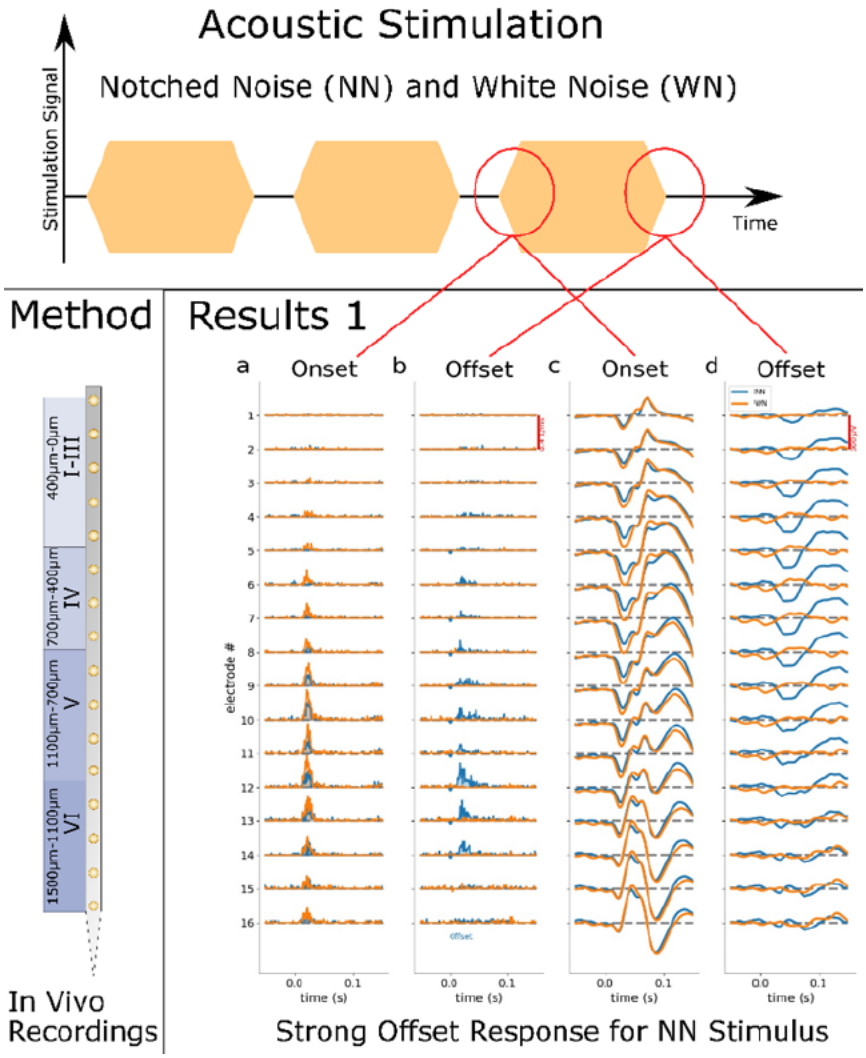

Figure 2.6: **Measurement of onset and offset responses in the cortex of guinea pigs** Acoustic stimulation: The stimuli consisted of several 3 s resp. 1 s WN resp. NN stimuli interrupted by 1 s-pauses. The onset response is the neural activity caused by the onset (beginning) of the noise stimulus and the offset response is the neural activity after the offset of the noise (red circles). Method: We implanted electrodes in wake and anesthetized guinea pigs and measured the activity in the 6 layers of the cortex. Results 1: a, b) Spiking activity at the different cortical layers for NN (blue) and WN (orange) onsets resp. offsets. c, d) Local field potentials (LFPs) for NN and WN onsets and offsets. Local field potentials are the low frequency signal from neurons surrounding the electrode. (Figure adapted form graphical abstract of [Schilling et al., 2023a])

the layers of the primary auditory cortex after the offset of WN and NN stimuli [Schilling et al., 2023a]. Thus, supported by the behavioral evidence described above, we expect the perception of a Zwicker tone after the offset of a NN stimulus. To analyze the exact activity patterns in the different layers of the primary auditory cortex, we used linear shank electrodes consisting of 1

shank with 16 contact points that can record signal from all layers of the auditory cortex in parallel (see Fig. 2.6 Method). As stimulus we used a NN stimulus (3 s resp. 1 s) followed by 1 s of silence (see Fig. 2.6 Acoustic Stimulation). The silence after noise offset is the most interesting period of the measurement, as we expect to see Zwicker tone related neural activity there. Interestingly, we found significantly higher offset responses after the offset of the NN stimulus than after the offset of the WN stimulus (see Fig. 2.6 Results 1). However, the onset responses are stronger for the WN stimulus. Indeed, the strongest NN-offset responses occur in the middle layers of the cortex and especially in layer 4 [Schilling et al., 2023a]. Layer 4 is the input layer of the cortex receiving input from the sub-cortical parts of the brain (MGB in thalamus, [Parameshwarappa and Norena, 2024, Smith et al., 2012]). A so-called current source density (CSD) analysis has been performed, to further analyze the time course of information transmission in the cortex. The CSD is defined as minus the second spatial derivative in z-direction (z-direction: perpendicular to the cortex surface, $-\mathrm{CSD}(z) = \frac{\partial^2 \Phi(z)}{\partial z}$) [Mitzdorf, 1985, Happel et al., 2010, Jeschke et al., 2021]. The CSD analysis shows a clear sink (location of activity) in layer 4 for the onset of WN and NN (see 2.7 CSD WN, NN Onset). This sink disappears for the WN offset (see Fig. 2.7 CSD WN Offset). However, the offset responses in layer 4 after NN offset look very similar to the onset responses (see Fig. 2.7 CSD NN Offset). This finding might point to the direction that after NN offset, a new stimulus onset is perceived. This new onset might be the beginning of Zwicker tone perception (for in-depth discussion of results see [Schilling et al., 2023a]). If we assume that the NN offset is indeed the onset of the Zwicker tone and we see a clear and early sink in layer 4, Zwicker tone might be processed in the cortex similar to a real tone. This means, that the neural correlate of Zwicker tone is already produced in brain areas below the cortex such as the brainstem (see Introduction).

To put it in a nutshell, we know a lot of different characteristics and properties of tinnitus and Zwicker tone and have already some ideas on the neural correlate of Zwicker tone. Thus, the next step is to find a mechanistic model that provides a universal explanation for Zwicker tone and tinnitus and can be falsified in further experiments. Furthermore, the model could be translated into a computer simulation to perform in-silico experiments.

2.2.3 A mechanistic model for Zwicker tone and acute tinnitus

In 2016 the so-called stochastic resonance (SR) model of tinnitus development was developed in our laboratory [Krauss et al., 2016]. The idea is based on the fact that noise (in this context refers to a random signal) is not always bad for the sensory system. Thus, it was shown that the addition of acoustic WN to very mild sounds made it easier to hear those sounds. This means the

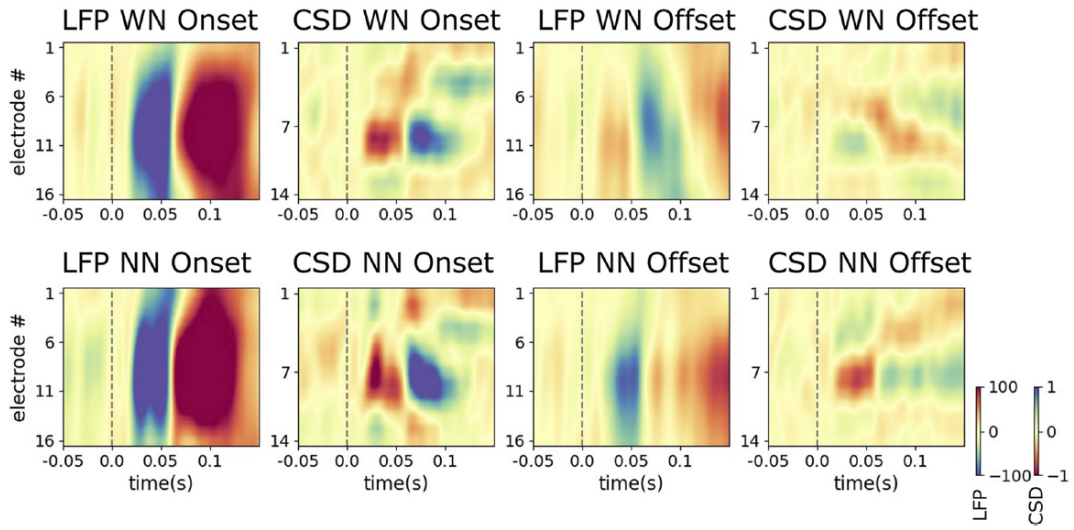

Figure 2.7: **Current source density analysis**
Local field potentials (LFPs) for noise onset and offsets. From this signal the second spatial derivative can be formed to get the CSD. The CSD looks similar for the onset responses of NN and WN and the NN-offset response, but not for the WN offset response. This finding might indicate that after NN offset a Zwicker tone onset can be measured, which is not present after WN offset. (Figure taken from [Schilling et al., 2023a])

addition of white noise decreased hearing thresholds of the probands of the experiment [Zeng et al., 2000]. However, the effect was not very big (<5 dB) [Zeng et al., 2000]. However, potentially the auditory system is capable of adding the amount of noise it needs to optimize hearing thresholds intrinsically. This could indeed help to improve hearing ability in a flexible and fast way. Thus, it was proposed that in the DCN a control circuit is implemented, that always adds the exact amount of noise needed to decrease hearing thresholds. In a healthy auditory system without any damage of the cochlea, the addition of noise is not needed as the signal from the cochlea is above the detection threshold and thus a signal cascade along the auditory pathway is induced by an auditory stimulus. However, a damage in the cochlea leads to a decreased signal amplitude and shifts the signal below the detection threshold [Schaette and McAlpine, 2011]. Thus, no signal cascade is induced. The auditory system can react by adding neural noise to the remaining signal and shifting it stochastically above the detection threshold. Indeed, the SR mechanism is very robust and works for variable thresholds and many different signals [Krauss et al., 2017]. However, there are two problems about the implementation of this mechanism in the brain that need to be regarded. First, where does the noise come from and how is the correct noise amplitude added to the remaining signal? Second, the noise itself is a neuronal signal that might lead to the perception of a sound in the brain. The SR model of tinnitus development proposes that this

neural noise is indeed the neural correlate of tinnitus [Krauss et al., 2016, Krauss et al., 2019]. It is very likely that the neural noise or i.e., the spontaneous neural activity is generated in the somatosensory system and transmitted to the auditory system [Krauss et al., 2016]. This would explain why the DCN receives input signals from somatosensory nerves as already described in the Introduction (see [Shore and Zhou, 2006, Young et al., 1995]). The hypothesis that the neural noise is generated in the somatosensory system is supported by the fact that tinnitus can be modulated by somatosensory stimulation (somatic tinnitus [Shore et al., 2007]). Thus, for example jaw movement or electric stimulation modulate the tinnitus loudness [Shore et al., 2007]. The idea is that the somatosensory stimuli increase the noise amplitude in the somatosensory system and thus this increased spontaneous activity is perceived as louder tinnitus [Krauss et al., 2016]. Potentially we hear what we 'feel'.

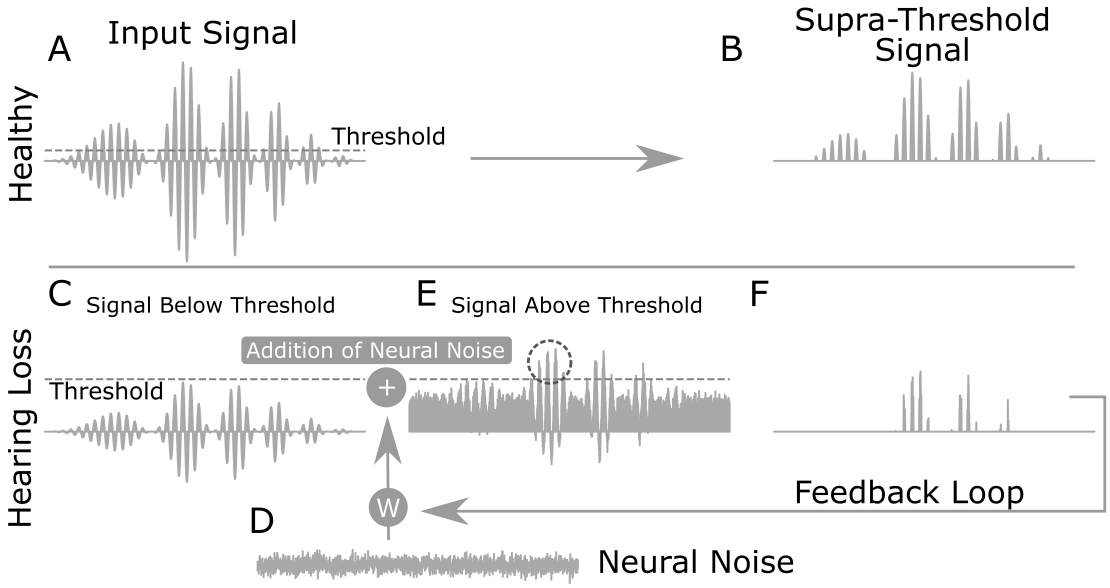

Figure 2.8: **The stochastic resonance control circuit**
A: In the healthy auditory system the signal from the cochlea induces a signal cascade along the auditory pathway (neural pathway related to hearing). The signal is above the detection threshold (supra-threshold signal) (B). A damage of the cochlea (e.g. synaptopathy [Tziridis et al., 2021, Schaette and McAlpine, 2011]) leads to a decrease of the amplitude of the cochlear output (C). No signal cascade is induced. The brain adds neural noise to lift the sub-threshold signal above the detection threshold (E). The noise amplitude has to be tuned so that the information of the signal is optimized (F and D). The loop potentially works like a simple control circuit. (Figure taken from [Schilling et al., 2023c])

However, one major problem remains. How does the auditory system resp. brain know what the ideal noise amplitude is? The model is an example for a great interdisciplinary approach.

Thus, the SR mechanism itself comes from physics and the control circuit is an idea inspired by engineering science. We propose that within the DCN a simple control circuit is implemented that measures the information of the signal and optimizes the noise amplitude until the information reaches a maximum value (see Fig. 2.8). Basically, it is a simple dynamic optimization problem. However, measuring the information is not trivial, as it has to be done using a biological neural network. It could be shown that the information of the signal could be approximated with the auto-correlation function [Krauss et al., 2017]. The auto-correlation can be easily calculated in the DCN. As already shown in the Introduction, the DCN surprisingly looks like the cerebellum and contains so called delay lines [Oertel and Young, 2004]. These delay lines can be used to generate time shifts, which then could be used to calculate the auto-correlation of the signal [MacKay and Murphy, 1976, Ivry and Keele, 1989]. The idea that the DCN is used to estimate the information of the signal by calculating auto-correlations is the first idea that could explain the anatomy of the DCN [Krauss et al., 2017].

To calculate the auto-correlation, in addition to the delay lines so called co-incidence detector neurons are needed, that release a spike when they get input from at least two input streams co-incidentally. In our study "Coincidence Detection and Integration Behavior in Artificial Neural Networks" (**Publication 5**, [Stoll et al., 2023]), we investigated the properties of these co-incidence detector neurons in spiking neural networks. The LIF networks were trained with the surrogate gradient approach developed in an earlier study and the properties of the neurons in the trained networks were investigated [Gerum and Schilling, 2021, Stoll et al., 2023]. We found that the degree of integration resp. coincidence detection characteristics of the neurons can be tuned by just adapting the leak term (resp. inverse leak term: decay times). Thus, co-incidence detector neurons differ from integrator neurons in a sense that the decay times are shorter [Stoll et al., 2023]. Thus, the neuron has not much time to sum up (or integrate) signals and therefore co-incident input spikes are needed to lift the membrane potential above the spike threshold. Note that this study is in line with further studies on spiking characteristics in spiking neural network (SNN), demonstrating that different operating modes of the neurons influence the performance of the networks [Perez-Nieves et al., 2021].

The tinnitus model proposes that in the auditory system [Krauss et al., 2016, Schilling et al., 2021c] the auto-correlation calculated through interplay of delay lines and co-incidence detector neurons, is used to estimate the information content of the sum signal (noise + sub-threshold remaining signal). The DCN uses this value to add exactly the amount of neural noise to the sub-threshold signal to maximize the information transmission [Krauss et al., 2016]. This model on tinnitus development has the major advantage that it operates on small time scales. The noise

amplitude can be changed within seconds. This fits to the fact that tinnitus can start directly after a hearing loss such as a TAD [Almond et al., 2013]. Other models based on (homeostatic) plasticity [Noreña, 2011] resp. 'learning' cannot explain, why tinnitus can occur immediately after a hearing loss as synaptic plasticity is not fast enough [Zenke et al., 2017, O'Donnell, 2023].

In our review paper, "The stochastic resonance model of auditory perception: A unified explanation of tinnitus development, Zwicker tone illusion, and residual inhibition" (**Publication 6**, [Schilling et al., 2021c]) we illustrated that the SR models described above can also explain Zwicker tone and residual inhibition. Thus, as I have mentioned before, the big advantage of the SR model compared to plasticity based models such as the central gain model (see e.g. [Noreña, 2011]) is that the control circuit works fast, which fits to the observation that tinnitus arises directly after a noise trauma and that Zwicker tone occurs directly after the offset of the NN and disappears after a few hundred milliseconds or a few seconds [Zwicker, 1964, Norena and Eggermont, 2003]. The fact that Zwicker tone is induced by NN, which is assumed to be a transient artificial hearing loss [Hullfish et al., 2019, Krauss and Tziridis, 2021] points to the direction that the neural mechanisms of Zwicker tone and tinnitus are similar. This idea is supported by the finding that Zwicker tone as well as tinnitus can be modulated by somatosensory stimulation [Shore et al., 2007, Ueberfuhr et al., 2017] due to the connection between somatosensory system and the DCN [Krauss et al., 2016, Shore and Zhou, 2006]. However, up to this point it is not yet clear if the addition of neural noise has indeed an effect on hearing. The hypothesis is that tinnitus related neural noise (SR effect) should increase the hearing ability of patients. This is indeed the case. The comparison of hearing thresholds of patients of the Ear Nose and Throat (ENT) hospital showed significant better hearing thresholds for patients with tinnitus compared to patients without tinnitus [Krauss et al., 2016, Gollnast et al., 2017]. Indeed, it was also shown that during the perception of Zwicker tone the hearing thresholds were better at the center notch frequency, which means directly at the spectral location of the simulated hearing loss [Wiegrebe et al., 1996]. There is good evidence that tinnitus and Zwicker tone are induced by the SR mechanism. There is another tinnitus related phenomenon that nicely fits to the SR model, the so-called residual inhibition [Galazyuk et al., 2019, Fournier et al., 2018]. The tinnitus percept disappears for a few seconds after the presentation of WN [King et al., 2021]. This phenomenon might also be a consequence of the SR control circuit. The increased neural noise, which is the neural correlate of tinnitus, is tuned down by the brain as it is replaced by external acoustic noise. Thus, the tinnitus percept disappears for a few seconds until the noise level is up-regulated again [Schilling et al., 2021c]. This effect is the basis of a therapy approach that will be discussed later in the thesis. In summary, we could argument that the SR model is the only model that fits to the anatomical

structures of the DCN and the experimental observations of many tinnitus studies. However, one question remains. Is it really possible that during evolution, a mechanism evolved that leads to a decrease of hearing thresholds by approx. 4 dB on the one hand, but causes severe side effects such as depression and concentration issues on the other hand [Gollnast et al., 2017, Krauss et al., 2016, Mazurek et al., 2012, Mazurek et al., 2015]. Our hypothesis was that the effect of SR could be stronger for speech comprehension, which is essential for humans. To test this idea, we developed a hybrid neural network and combined methods from computational neuroscience and AI [Schilling et al., 2022].

From tinnitus to speech comprehension to intelligent speech interfaces

We combined methods from computational neuroscience and AI to build a valid model of the auditory pathway, which we could use as in-silico model for our experiments [Schilling et al., 2022]. Thus, the cochlea is approximated with 30 bandpass filters covering the whole human hearing range, in order to simulate the tonotopy caused by the traveling wave of the basilar membrane (see Introduction). The cochlea performs some kind of Fourier transform but keeps also information on the phase of the signal (see Fig. 2.9 B). The DCN is simulated using 30 LIF units, which represent a whole conglomerate of neurons each. Thus, the signal split according to the signal frequency is transformed into a spiking signal [Gerum and Schilling, 2021]. Note that there is a tonotopic organization along the whole auditory pathway, which means that certain frequencies are processed by certain parts of the DCN. Thus, the frequency channels of the auditory pathway are ordered somehow like a piano keyboard. The rest of the auditory pathway (remaining brainstem nuclei, see Introduction), the thalamus, and the cortex are modeled using a convolutional neural network (CNN). This CNN receives spiking input from the DCN module.

The hybrid neural network (see Fig. 2.9) is trained on classifying the 207 most common German words based on input audio files, from a custom-made data set recorded in our laboratory. The data of 10 speakers was used as training data set and the data of two further speakers was used as test data set. After the training procedure, the weight values were kept constant. This trained network is considered as healthy auditory system. We used this system and applied a hearing loss by reducing the output amplitude of the cochlea, which leads to a worse classification accuracy, and thus can be interpreted as worsened speech understanding. To test if SR has indeed an effect on speech understanding, we added white noise to the signal before it was fed to the DCN. For an optimal noise amplitude, the classification accuracy was increased by a factor of more than two. Thus, the SR could indeed have a significant effect on speech understanding, far beyond of just 4 dB hearing threshold improvement. Our model suggests that tinnitus patients understand

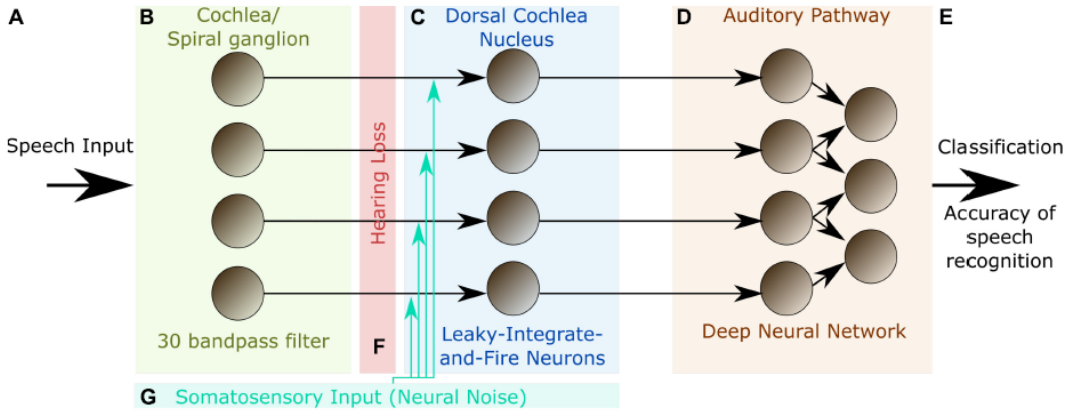

Figure 2.9: **Hybrid neural network as model of the auditory pathway**
A: The input signal consists of sound files of the 207 most common German words spoken by 12 human speakers. B: The cochlea is simulated as row of 30 bandpass filters simulating the tonotopic organization of the basilar membrane. C: The DCN consists of 30 neuron-'conglomerates' implemented as single LIF neurons. D: As it is not totally clear how to simulate the rest of the auditory pathway up to the cortex, we used a convolutional neural network. The network is trained on classifying the words based on the audio files. F: After training, a simulated hearing loss is applied. G: The neural noise from the somatosensory system is added to restore classification accuracy. (Figure taken from [Schilling et al., 2022])

speech potentially better than non-tinnitus patients. However, in contrast to that, tinnitus patients are often found to have worse speech understanding than control groups [Ivansic et al., 2017]. Hamza and Zeng argued that this is the case, because the control groups are not well matched and that the main cause for the problems is the tinnitus inducing hearing loss and not the tinnitus mechanism itself [Hamza and Zeng, 2021]. They found that under certain conditions, tinnitus patients suffer from less cognitive decline than an age and HL matched control group. We discuss in one further opinion paper that this observation may be a consequence of the SR mechanism improving speech comprehension [Schilling and Krauss, 2022].

Our study shows that artificial neural networks are indeed a useful tool to generate hypothesis on mechanisms in the human brain as proposed by the CCN field [Kriegeskorte and Douglas, 2018]. However, the brain is also a good inspiration for AI research as proposed by the neuroscience-inspired AI field [Hassabis et al., 2017]. Thus, the SR mechanism exploited in several sensory modalities of humans [Krauss et al., 2018, Moss et al., 2004, Levin and Miller, 1996], could be an interesting target to build sensitive sensors or improve existing intelligent speech interfaces [Rousseau et al., 2003, Verma and Yadava, 2016, Repperger et al., 2005].

2.2.4 From AI and Computational Neuroscience to Biomedical-Engineering

Up to this point, we have gained a deep understanding on the neural mechanisms of tinnitus, Zwicker tone and residual inhibition. However, the knowledge on the mechanisms must be at some point translated into an individualized therapy approach. Our idea was to replace the internal neural noise (the neural correlate of tinnitus) with external acoustic noise [Krauss et al., 2016, Schilling et al., 2021c]. In collaboration with a hearing aid company (Sivantos), we have performed a pilot study to find out if external acoustic noise could help to decrease the subjectively perceived tinnitus loudness. In our study "Reduktion der Tinnituslautstärke: Pilotstudie zur Abschwächung von tonalem Tinnitus mit schwellennahem, individuelle spektral optimiertem Rauschen" (**Publication 7**, [Schilling et al., 2021a]), we measured 22 subjects perceiving a tonal tinnitus and presented an individualized acoustic noise stimulus. Note that we used near-threshold narrow band noise, which has nothing to do with louder broad band noise used for tinnitus masker, which are used to mask the tinnitus and shift the attention away from the tinnitus [Cuesta et al., 2022, McNeill et al., 2012, Kikidis et al., 2021]. Thus, in our study, we first determined hearing thresholds and tinnitus frequency and then created a noise stimulus with a loudness close to the hearing threshold, which is adapted to the hearing loss resp. the tinnitus frequency. 16 out of 22 tinnitus patients, which means more than 70 %, reported a decrease of the subjectively perceived tinnitus loudness [Schilling et al., 2021a]. Furthermore, we found that the therapy works best for narrow band noise centered around the tinnitus frequency and for patients with only mild or no significant hearing loss [Schilling et al., 2021a]. Note that the tinnitus often occurs at the frequency with the highest hearing loss [Schecklmann et al., 2012]. Thus, our approach can be interpreted as fooling the brain into believing that there is no cochlear damage at the tinnitus frequency, as the narrow band noise always stimulates the auditory pathway exactly at the tinnitus frequency. This pilot study has led to a further study, which was conducted to fine-tune the noise stimuli [Tziridis et al., 2022a]. After this second study, a clinical study has been started. Thus, in cooperation with a hearing aid company, a larger cohort of patients is measured using small hearing aids with noise generators. This individualized therapy approach is a result of a mechanistic model tested in-silico [Schilling et al., 2022] and proves that biomedical engineering can profit from the interplay of experimental, computational neuroscience, and AI. This therapy approach is a good example for a structured interdisciplinary cooperation that has the potential to help many people suffering from tinnitus.

Nevertheless, it is still unclear why not everyone with hearing loss develops a conscious tinnitus percept (tinnitus heterogeneity, [Cederroth et al., 2019]) and why restoring the hearing ability with a hearing aid does not cure tinnitus in all patients [Trotter and Donaldson, 2008].

2.2.5 Conscious tinnitus perception, the Bayesian brain, and internal world models

Up to this point, tinnitus was mainly regarded as a mal-adaptation of brainstem structures to a HL [Schilling et al., 2021c]. However, to cover the full phenomenon, it is important to understand tinnitus on a cognitive level. Indeed, the SR model does not explain why the neural hyperactivity produced in the brainstem is transmitted upwards to the cortex, where it becomes a conscious perception of a sound. As described in the Introduction, the thalamus is a part of the brain that filters out unwanted signal that should not enter consciousness. Thus, many different models describe why the thalamus does not fulfill this task in tinnitus [Rauschecker et al., 2015, Llinás et al., 1999, De Ridder et al., 2015]. In our review paper "Predictive coding and stochastic resonance as fundamental principles of auditory phantom perception" (**Publication 8**, [Schilling et al., 2023c]), we discuss these models regarding explanatory power and concreteness. Thus, we apply the "tri-level"-framework of David Marr, to structure the different tinnitus models [Huskey et al., 2020, Kitcher, 1988, Marr, 2010, Schilling et al., 2023c]. The "tri-level"-hypothesis says that any information processing procedure can be explained on three different levels: computational, algorithmic, and implementational. The computational level refers to the exact task that is fulfilled by the system [Huskey et al., 2020]. In the case of tinnitus development, the exact formulation might be something like: 'Consciousness perception of a pure-tone like sound without the presence of any physical sound source'. This level provides no information on how things are processed. The algorithmic level refers to the algorithm itself, which means that it refers to which calculations are performed but not to the hardware on which the algorithm is run [Huskey et al., 2020, Schilling et al., 2023d]. Thus, if the algorithmic level was known, a computer program could be written producing a tinnitus percept. The implementational level refers to the exact implementation of the algorithm on a certain hardware. Thus, in our case, to understand the implementation of tinnitus means to understand every molecular process in the brain related to tinnitus perception. This level is so fine-grained that it is not realistic to understand the concrete implementation of tinnitus. In our review paper, we argue that tinnitus must be understood on an algorithmic level [Schilling et al., 2023c], which is in line with the idea of the famous physicist and Nobel prize winner Richard Feynman: "What I cannot create, I do not understand" [Feynman, 1988, Samantsidis et al., 2020]. In the publication, we provide a detailed discussion on existing tinnitus models [Schilling et al., 2023c]. From our point of few, there is only one cognitive model that fulfills the requirements to explain tinnitus on an algorithmic level, the so called 'Predictive Coding' model of tinnitus development [Sedley et al., 2016]. This model is based on the idea that the brain is a prediction machine always trying to predict future events [Smith et al., 2021].

The idea can be formalized using the Bayesian brain framework [Friston, 2012, Friston, 2010]. Thus, the brain is assumed to process information according to the famous formula of Bayes on conditional probabilities [O'Reilly et al., 2012, Webb and Sidebotham, 2020, Schilling et al., 2023c]:

$$p(x|o) \propto p(o|x) \cdot p(x) \tag{2.7}$$

($p(x|o)$: posterior, $p(o|x)$: likelihood, $p(x)$: prior). In cognitive neuroscience, the posterior is seen as the actual conscious percept, which means that it refers to the actual interpretation of the signals coming from the sensory system [Sedley et al., 2016]. The likelihood is the neuronal signal originating in the ear, which is transmitted via the brainstem and the thalamus to the cortex, whereas the prior is the prior expectation i.e., the prediction of the brain [Sedley et al., 2016]. The brain continuously updates the prior so that the prediction matches the input coming from the sensory system. Thus, the brain minimizes surprise [Clark, 2018]. The brain has an internal model of the world, learned through a huge number of experiences (world model) and derives the concrete prior expectation form this model [Schilling and Krauss, 2024]. This framework can be unified with the SR model of tinnitus development described above [Krauss et al., 2016, Schilling et al., 2021c, Schilling et al., 2023c]. The unified model is described in detail in [Schilling et al., 2023c].

In the healthy system, the standard prior is silence, which means no tone is perceived, if there is no physical sound source (see Fig. 2.10). When a tone from an external sound source is recorded by the ear, the prior is overruled for a short time by a larger likelihood [Schilling et al., 2023c]. The prior is not changed due to the presence of a short tone. A HL causes a decreased input amplitude coming from the ear [Schaette and McAlpine, 2011]. Thus, the likelihood is shifted to lower values and the prior remains unchanged. Therefore, the posterior is shifted to lower values, which means no tone is perceived at all. However, the DCN increases the neural noise, in order to restore hearing thresholds by means of the SR principle [Krauss et al., 2016, Schilling et al., 2021c]. This neural noise is a continuous signal coming from the brainstem to the cortex. If the amplitude of this noise is big enough, the prior is again overruled and the brain thinks there is a real sound source. This means an acute tinnitus is perceived. Additionally, the sensory precision (inverse variance) of the likelihood is probably not changed or increased by the added neural noise [Sedley et al., 2016, Schilling et al., 2023c]. A higher sensory precision further drives the misinterpretation (according to equation 2.7, [Sedley et al., 2016]). The continuous neural noise transmitted to the cortex causes an adaptation of the prior. Thus, the brain starts

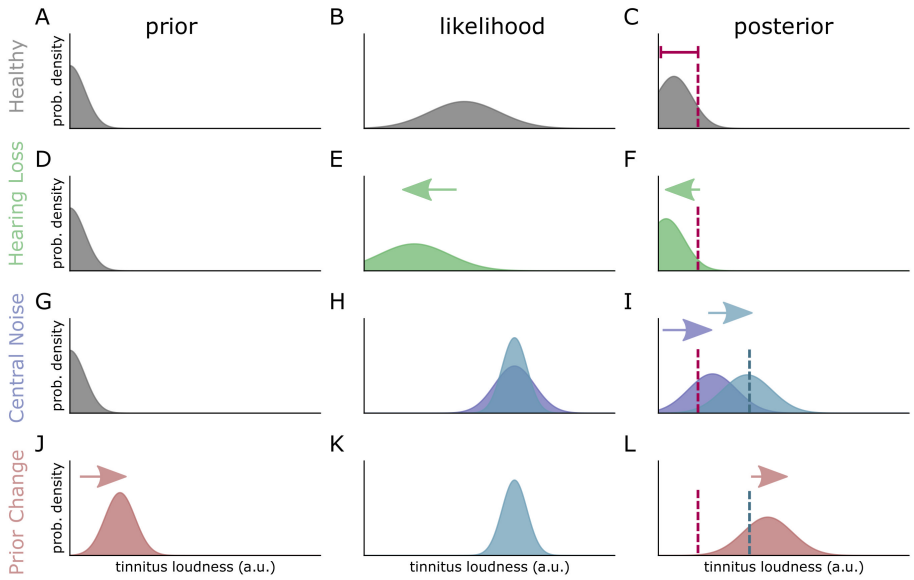

Figure 2.10: **Integrated model of conscious tinnitus perception**
The figure shows the cause of conscious perception of a phantom sound. In the healthy system (A-C) the prior distribution is centered around zero (no tone is predicted). The likelihood (B) distribution is broad and centered around low values representing the spontaneous activity of the auditory pathway. This results in a posterior distribution near zero (C), which means that no tone is perceived. A hearing loss (D-F) reduces the mean of the likelihood distribution, which represents the reduced input from the damaged cochlea. The posterior remains low and no tinnitus is perceived (F). The addition of neural noise by means of SR shifts the likelihood to higher values (H) and potentially increases the sensory precision (H). The likelihood overrules the prior and the posterior is shifted to higher values (I). Tinnitus is perceived. J: After a while, the prior is adapted to the continuous input (likelihood). The standard prediction is now tinnitus. This might be the neural correlate of chronic tinnitus, which is difficult resp. impossible to cure. (Figure taken from [Schilling et al., 2023c])

to predict the presence of a continuous sound source and therefore the standard prediction is set to a higher value, which means the standard prediction is now the presence of a continuous tone. We propose that the adaptation of the prior is the correlate of the chronic manifestation of tinnitus [Wallhäusser-Franke et al., 2017, Schilling et al., 2023c]. An altered prior distribution might explain why the restoration of hearing with hearing aids does not necessarily cure tinnitus. Thus, the brainstem probably reduces the neural noise, as it is not needed any more to drive the SR mechanism. However, the expectation of the brain remains that there is an external sound source [Schilling et al., 2023c]. The prior is highly individualized and depends on many personal experiences. Thus, this might be the key to tinnitus heterogeneity [Cederroth et al., 2019].

2.3 Higher cognitive functions of the brain: the prediction machine

As described above, the brain builds its own inner model of the world [Schilling and Krauss, 2024, Perlis, 1997, Krauss and Maier, 2020, Yasoda-Mohan et al., 2024]. This inner model might be one key component of consciousness [Krauss and Maier, 2020]. To find hidden processing principles in the brain based on electrophysiological data such as EEG and MEG, we established AI-based methods that help to extract valuable insights for the very high dimensional electrophysiological data. Thus, we established the so-called general discrimination value (GDV) to quantify how good internal representation of neural networks fit to certain labels (for details see [Schilling et al., 2021b]). Thus, the GDV compares the average distances of vectors within clusters with the average distances between clusters. These vectors can be e.g., activation patterns of hidden neurons in deep neural networks as well as activation patterns of brain neuron populations recorded using electrophysiological methods. Thus, the method can be used to find out how good different stimulus conditions such as different tone frequencies can be distinguished on the basis of neural activation patterns by analyzing the evoked potentials (see e.g. [Schilling et al., 2024]). To validate the methods, we applied them to EEG data recorded in a sleep laboratory. In our study "Analysis and visualization of sleep stages based on deep neural networks" (**Publication 9**, [Krauss et al., 2021]), we trained neural networks supervisedly on classifying different sleep stages to identify for example pathological sleep. This can be very useful, as in clinical practice sleep stage classification is done by hand and is therefore very time consuming [Patanaik et al., 2018]. It has to be mentioned that sleep is indeed crucial for learning and cognition in general [Walker, 2009, Deak and Stickgold, 2010]. Therefore, we tried to find out more on local patterns of sleep in different brain areas by comparing the sleep stage curves created on base of the data from different EEG channels by calculating cross-correlations. We found some indication for the hypothesis that sleep stages are not homogeneously distributed over the whole cortex but that different brain parts might be in different sleep stages at the same time. However, these results are preliminary and the hypothesis testing is still work in progress.

Nevertheless, we could apply the GDV and the idea of using AI models to evaluate neuronal data to data from intracranial recordings in rodents as well as in humans. Thus, in our study "Deep learning based decoding of single local field potential events" (**Publication 10**, [Schilling et al., 2024]), we used a self- supervised approach to extract hidden features from intracranial recordings (electrodes implanted directly in the brain). The idea was to observe the brain during spontaneous processing, when no external stimulus was presented. The increased signal-to-noise

(S/N) ratio of intracranial recording (such as iEEG) compared to standard methods such as EEG and MEG, makes it possible to evaluate the data on single trial basis. Note that in experimental neuroscience usually the signal has to be averaged over many measurement repetitions to increase the S/N ratio and thus to identify the underlying neuronal signal. However, averaging across several measurement trials causes vanishing of the temporal fine structure of the signal. The brain does not average at all, but the actual information processing is hidden in the continuous, serial stream of neuronal signals. It is well known that the brain is always active even when no stimulus is present or also during sleep as described above [Krauss et al., 2021]. However, it is difficult to decode what the actual purpose of this activity is. To evaluate the continuous signal stream measured via intracranial electrodes, we have set a local threshold to find so called local field potential (LFP)-events in the stream, in a first step. A variety of different LFP-event shapes have been extracted from the data stream (see Fig. 2.11 b, [Schilling et al., 2024]). An LFP-event is a compound signal of the activity of neurons within ca. 1 mm around the electrode and is thus less focused than so called multi-unit activity (spiking activity) [Kajikawa and Schroeder, 2011, Lindén et al., 2011, Buzsáki et al., 2012, Kreiman et al., 2006], which is the compound signal from action potentials of only a few neurons around the measurement electrode [Buzsáki et al., 2012]. Indeed, the low frequency components of LFP signals are caused by post-synaptic currents, whereas to the high frequency parts, the spiking activity of the neurons contributes [Buzsáki et al., 2012, Kreiman et al., 2006, Kraskov et al., 2007, Logothetis, 2003]. Often LFP events are assumed to be the input signal to the neurons, as they mainly consist of post-synaptic potentials [Kreiman et al., 2006]. EEG electrodes however measure the compound signal of millions of neurons and thus EEG data has a bad spatial resolution [Murugavel et al., 2016]. Thus, LFP signals are a good compromise between the spiking signal from very few neurons, which does not allow for general assumption and EEG signals with a very bad spatial resolution.

As described above, a variety of differently shaped LFP events have been extracted through the threshold method (see Fig. 2.11). To evaluate and compare the shapes, a method to project the LFP-events into lower dimensions is needed. Therefore, we trained an autoencoder network on these LFP events and used the encoding layer as complex representation of the different shapes (encodings, see 2.11a). Autoencoder networks are useful to project data to lower dimensions with a minimum loss of information [Kensert et al., 2021, Meng et al., 2017, Gondara, 2016, Bourlard and Kabil, 2022]. We compared the projections to principal component analysis (PCA) projections [de Cheveigné et al., 2007] and found that the autoencoder provided more valid results in e.g. terms of better GDV values for different stimulus conditions in experiments on auditory evoked activity [Schilling et al., 2024, Krauss et al., 2021]. We used the same autoencoder trained on spontaneous

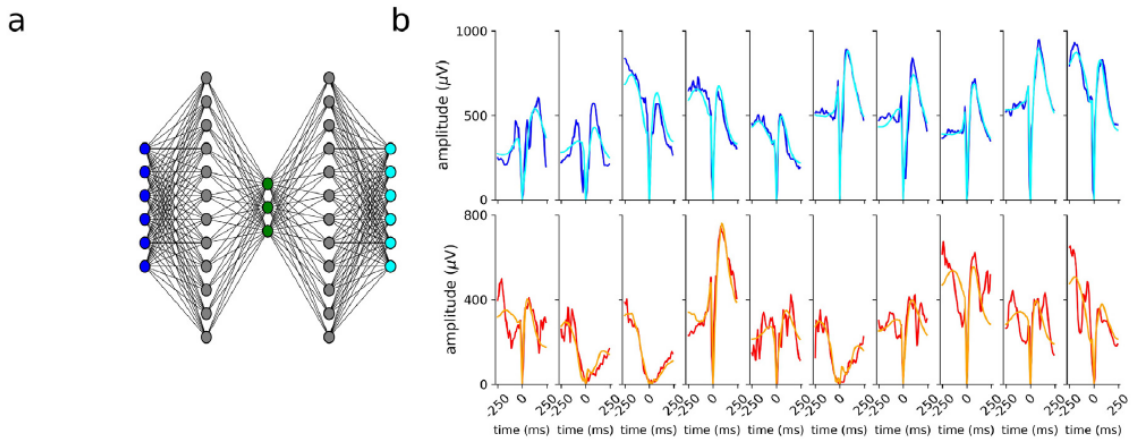

Figure 2.11: **Autoencoder and LFP-events with reconstructions**
a: Scheme of the autoencoder used to encode the LFP events identified through the thresholding-algorithm. An autoencoder with dimensionality expansion layer was used (e.g. proposed in [Bourlard and Kabil, 2022]). b) Examples of LFP event shapes for training data set (dark blue) and test data set (red) and the decoded reconstructions after autoencoding (cyan, orange). (Figure taken from [Schilling et al., 2024])

activity to project these stimulus-evoked LFP-events to lower dimensions. We observed that the shapes of the LFP events during spontaneous activity were similar compared to the LFP shapes during auditory stimulation (see Fig. 2.12). Due to this finding, we conclude that we could show that the brain spontaneously samples from brain states evoked by all possible stimulus conditions during the default mode with no external stimulation [Schilling et al., 2024]. A similar effect was observed in the spiking activity recorded with several implanted electrodes [Luczak et al., 2009]. We could reproduce this observation through the application of autoencoders on LFP events using exclusively data from single electrodes. These findings can be interpreted as follows. During default mode, the brain runs through different possible states or i.e., the brain plays out several future scenarios [Schilling et al., 2024]. This idea fits nicely to the Bayesian brain hypothesis postulating that the brain is a prediction machine, always trying to predict future events. In a next step, these results need to be reproduced in humans. However, intracranial recordings in humans are due to ethical reasons rare, as they can exclusively be performed in combination with diagnostic or therapeutic purposes (e.g. epilepsy, [Grova et al., 2016, Bartolomei et al., 2017]). Nevertheless, we have already shown in one patient that our method can be also applied to human iEEG data [Schilling et al., 2024]. The implication of these experiments could be far-reaching, as the sampling from possible stimulus-evoked brain states could be something like a continuous exploration of the inner world model of the brain. Thus, these techniques could be useful to get

some idea about the nature of these world models and as a final consequence about the basis of consciousness [Schilling and Krauss, 2024]. However, we are up to this point, at the very beginning of the journey. Thus, many more experiments and especially control experiments are needed to explore consciousness with methods from experimental neuroscience [Dehaene and Changeux, 2011] and these experiments have to be combined with computer simulations (see e.g. [Immertreu et al., 2024] for preliminary approach).

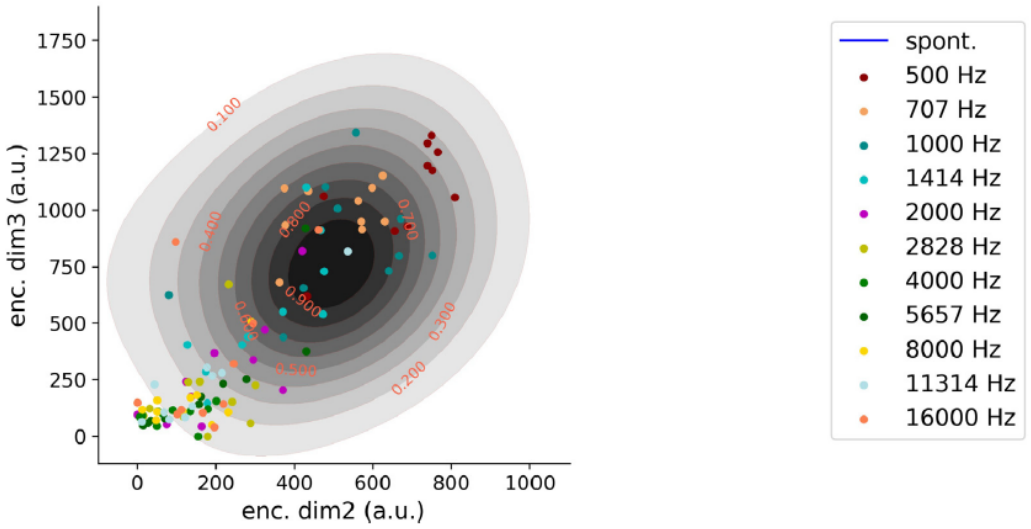

Figure 2.12: **Brain samples from possible stimulus driven sates**
The figure shows two dimensions of encoding layer representations of stimulus driven (colored dots) and spontaneous LFP-events (black contour lines). We observed, that the spontaneously occurring LFP events look similar to the stimulus evoked activity. (Figure adapted from [Schilling et al., 2024])

At this point, the circle is completed, because as described at the very beginning of the thesis, it is very useful to combine experimental methods with computer simulations and AI algorithms. The incredible progress in the development of large language models (LLMs) opens up novel pathways for neuroscience. Modern LLMs are trained on predicting the most probable next word [Gruver et al., 2024]. Thus, also transformers (basis of LLMs) are prediction machines potentially similar to the brain, which are often combined with the biology-inspired reinforcement learning [Wu et al., 2023, Neftci and Averbeck, 2019], and are able to solve complex cognitive tasks. Therefore, according to the philosophy of CCN, they are an interesting target to understand the brain [Kriegeskorte and Douglas, 2018]. Thus, we already started to investigate the activation patterns of hidden units of LLMs with the purpose to compare them with neuronal data on the long run (see preliminary results in [Krauss et al., 2024]).

Chapter 3

Conclusion

In the present thesis, I have demonstrated how different disciplines can interact fruitfully to make progress in all of them. The main purpose of the research, was to use artificial neural networks as a tool and a model to understand the healthy and the lesioned brain, with a special focus on the auditory system, on the one hand [Kriegeskorte and Douglas, 2018]. On the other hand, the unraveled principles of information processing in the brain were used to develop brain-inspired AI-algorithms [Hassabis et al., 2017]. First, basic processing principles of the brain such as neuron communication via action potentials (spikes, [Bean, 2007, Schmidt et al., 2010]), sparse coding [Beyeler et al., 2019], and sparse connectivity [Hagmann et al., 2008, Sporns et al., 2005] were used and adapted to, resp. integrated in artificial neural networks. However, spiking neural networks are difficult to train, as the gradient of δ- peak-shaped spikes is everywhere zero and thus it is not possible to train the networks supervisedly with standard algorithms. Therefore, we developed a simple surrogate gradient, which could be used to train SNN supervisedly with a minimum number of parameters ([Gerum and Schilling, 2021], **Publication 1**). As the information processing mechanisms of these neural networks are biologically plausible, they might be an interesting target also for computational neuroscience. Furthermore, this study triggered the question, if we potentially already have biologically plausible models, which are able to perform complex cognitive tasks. Thus, an in-depth analysis of so called peephole-LSTM units [Hochreiter, 1997, Bo et al., 2019, Rahman and Siddiqui, 2019] showed that these units can be run in a similar spiking mode as LIF neurons [Seenivasan et al., 2024]. Therefore, in our study, we demonstrated how to tune these peephole LSTM units, so that they behave like LIF neurons and analyzed the spiking characteristics as a function of the parameter combinations ([Gerum et al., 2023], **Publication 2**). This finding opens up two pathways. Thus, novel LSTM based SNN can be developed and be trained without any surrogate gradient as needed for LIF

neurons [Gerum and Schilling, 2021]. Furthermore, our research points out the possibility that the spiking characteristics has already emerged in LSTM networks, trained on complicated cognitive tasks (e.g. ELMO [Peters et al., 2018, Liu et al., 2020, Zhang et al., 2020]) and thus these networks can potentially be used as a model to further understand the principles of spike-based information processing in the brain. The brain and biology in general underlie further principles, which might be very interesting for AI research. Thus, the brain is very energy efficient and only consumes approximately 20 W [Furber, 2012, Yu et al., 2018]. Additionally, it has be shown that the size of the brain does not necessarily correlate with intelligence, in a sense that corvids with significantly smaller brains have similar cognitive capabilities as apes [Emery and Clayton, 2004, Güntürkün et al., 2017]. This difference comes from additional evolutionary pressure rep. restrictions from nature, as birds fly and thus need lightweight brains [Emery, 2006, Güntürkün et al., 2017]. Inspired by these natural principles, we trained neural networks with an evolutionary algorithm on orienting in a maze and added an additional pruning parameter. Thus, in each generation, by chance, some neural connections were removed ([Gerum et al., 2020], **Publication 3**). This pruning step has led to the effect that only very efficient and small networks survived and fulfilled the task with a minimum number of connections and neurons. This biologically inspired algorithm shows that efficiency and sparse connectivity can be achieved by adding additional restrictions to the training algorithm. Up to this point, mainly basic processing principles of the brain and artificial neural networks were investigated. The methods and ideas described above are now applied and extended to the complex information processing in the mammalian auditory system, with a special focus on the impaired auditory system. Thus, so called lesion studies are an important method, to reverse-engineer information processing in the brain by analyzing the damaged brain [Gazzaniga et al., 2014]. We focused mainly on effects caused by damages in the ear, which cause decreased signal amplitudes of the signal transmitted from the cochlea to the CNS [Schaette and McAlpine, 2011]. The Zwicker tone [Zwicker, 1964] is an auditory phantom percept, caused by a simulated transient hearing loss, which is simulated through the presentation of a so called NN (noise with spectral gap). The spectral gap simulates that at the center frequency of the gap is a hearing loss [Hullfish et al., 2019, Krauss and Tziridis, 2021]. This NN can induce a so called Zwicker tone percept, which is an auditory phantom percept potentially similar to acute tinnitus [Zwicker, 1964, Franosch et al., 2003, Norena et al., 2000]. We established a behavioral paradigm, which could be used to identify a Zwicker tone perception in animals ([Schilling et al., 2023d], **Publication 4**). This method is necessary to scientifically investigate Zwicker tone in animals and thus to make further progress in tinnitus research. In 2016, a possible explanation for tinnitus perception was developed in our group, the so-called SR model of tinnitus

development [Krauss et al., 2016]. The idea is that the brain adds neural noise to the decreased signal coming from the cochlea [Schaette and McAlpine, 2011] (caused by a hearing damage), to lift that signal above the detection threshold [Krauss et al., 2016]. The SR model of auditory phantom perception proposes a control circuit in the DCN in the brainstem and is thus highly inspired by engineering science [Schilling et al., 2021c, Krauss et al., 2016]. This control circuit tries to maximize the information transmission in the auditory system, by approximating the information of the compound signal of neural nose and auditory signal with the auto-correlation function [Krauss et al., 2016, Krauss et al., 2017]. The anatomy of DCN is well suited to calculate the auto-correlation function [Oertel and Young, 2004, MacKay and Murphy, 1976, Ivry and Keele, 1989]. Thus, so called delay lines and coincidence detector neurons are needed [Licklider, 1951]. We used this ideas and implemented a SNN, where we found that integration resp. coincidence detection neurons can be easily created by just varying one term of the LIF neuron parameters ([Stoll et al., 2023]), **Publication 5**), which might be also useful for LIF neuron based AI systems (see e.g. [Perez-Nieves et al., 2021]) and is a good example for neuroscience-inspired AI [Hassabis et al., 2017]. Based on numerous experimental findings on Zwicker tone and tinnitus, we were able to gain deeper insights in the mechanisms of tinnitus, Zwicker tone perception, and suppression of tinnitus through the presentation of acoustic noise (residual inhibition, [Galazyuk et al., 2019, Fournier et al., 2018]). In a next step, we discussed the SR model in the light of various experimental results and other existing tinnitus models in a review paper ([Schilling et al., 2021c], **Publication 6**). Thus, we proposed the SR model to mainly explain acute tinnitus and showed how this jigsaw piece fits to the numerous other models of acute and especially chronic tinnitus [Schilling et al., 2021c]. The SR model proposes better hearing thresholds due to the addition of neural noise by means of SR, which has indeed been found in a large clinical data base [Krauss et al., 2016, Gollnast et al., 2017]. However, it is still unknown, if this mechanism has any real advantages for speech understanding in humans. To answer this question, we developed a hybrid neural network based on spiking neurons and a CNN [Schilling et al., 2022]. This network was trained on word classification with spoken words as input signal. We used this network as a model for the auditory system and interpreted the classification accuracy as a measure of speech comprehension. After training the network on speech classification, we applied a virtual HL. This HL trivially has led to a decreased classification accuracy and thus impaired speech comprehension. The addition of neural noise, however, has re-increased the accuracy by a factor of more than 2. Thus, we conclude that hearing impaired persons can significantly profit from this mechanism [Schilling et al., 2022] and argue that the SR model might be an interesting target for the development of sensitive and intelligent sensors and speech interfaces (as also proposed

by [Rousseau et al., 2003, Verma and Yadava, 2016, Repperger et al., 2005]). Furthermore, the model suggests that tinnitus patients might have better speech comprehension abilities than an age-matched and especially HL-matched control group. Indeed, Hamza and Zeng found that tinnitus patients suffer from less cognitive decline than an HL- and age-matched control group, which is an unexpected result, as tinnitus is often associated with concentration issues [Hamza and Zeng, 2021, Mazurek et al., 2012, Mazurek et al., 2015]. We discuss these findings in the light of the SR model and argue that improved speech comprehension of tinnitus patients might contribute to the decreased cognitive decline [Schilling and Krauss, 2022]. Our results gained through a tight connection between AI, neuroscience, and medicine were used to develop an individualized therapy approach for tinnitus. Thus, in a pilot study, we could show that a near-threshold narrow band noise can be used to suppress tinnitus ([Schilling et al., 2021a], **Publication 7**). More than 70 % of the patients reported a positive effect of the mild noise stimulation [Schilling et al., 2021a]. In the next step, the stimuli were fine-tuned to optimize the effect [Tziridis et al., 2022a]. These results now serve as basis for a large clinical trial conducted together with a hearing aid company. This therapeutic approach shows impressively that AI, experimental neuroscience, as well as computational neuroscience can significantly boost Biomedical Engineering. We have furthermore boosted interdisciplinary tinnitus research by initiating a special issue looking at tinnitus mechanisms from many different viewpoints of different scientific fields [Schilling et al., 2023b]. Up to this point, tinnitus was discussed exclusively in terms of neural hyperactivity in the brainstem resp. DCN [Krauss et al., 2016, Schilling et al., 2021c]. However, tinnitus has also a conscious dimension. Thus, we consciously perceive a tone that is not filtered out by our thalamus [Rauschecker et al., 2015, Rauschecker, 2024]. We were able to unify the brainstem mechanisms of the SR model with the Bayesian brain framework [Friston, 2010, Friston, 2012] based model of tinnitus development [Sedley et al., 2016]. In our review paper, we discussed the majority of relevant tinnitus models in the light of David Marr's 'tri-level'-hypothesis [Marr, 2010] and showed that the Bayesian brain model [Sedley et al., 2016] and the SR model [Krauss et al., 2016, Schilling et al., 2021c] of tinnitus development can be unified and that the SR is a crucial part of the integrated tinnitus model that can explain tinnitus on all temporal scales ([Schilling et al., 2023c], **Publication 8**). Thus, the perception of tinnitus depends on previous experiences that form the internal model of the external world and the neuronal signal coming from the cochlea [Schilling et al., 2023c, Schilling and Krauss, 2024, Yasoda-Mohan et al., 2024]. Chronic tinnitus might be a result of the brain to believe too strongly in its world model, ignoring that potential brainstem mechanism resp. the possibility that the internal neural noise was already switched off or tuned down due to e.g. hearing aid supply [Trotter and Donaldson, 2008, Yasoda-

Mohan et al., 2024, Schilling and Krauss, 2024]. To further unravel the mechanisms of information processing in the brain with a special focus on conscious perception, we developed AI- based methods (see e.g. [Schilling et al., 2021b]) to evaluate electrophysiological data and to extract meaningful information from e.g. sleep EEG data ([Krauss et al., 2021], **Publication 9**). We used this method to further investigate the internal world model of the brain [Perlis, 1997, Krauss and Maier, 2020]. Thus, we analyzed LFP events recorded with intracranial electrodes during spontaneous activity and found that the brain samples events from the space of all possible stimulus-evoked activations ([Schilling et al., 2024], **Publication 10**). Thus, our study has confirmed and supported similar findings from a study that was based on a different measurement and evaluation approach, namely the analysis of spatio-temporal spiking activity patterns [Luczak et al., 2009]. The interpretation of these results is not trivial at all. However, these findings point to the direction that the brain plays through possible future events in default mode [Schilling et al., 2024]. On could go even one step further and say, potentially the brain takes a walk through its own world model. However, this is only a first approach to understand these complex findings and the interpretation must be reconsidered again and again. This is the ideal starting point to again switch to the AI side and try to reproduce these findings in artificial agents (preliminary results in [Immertreu et al., 2024]). Furthermore, as described above, the brain always tries to predict future events, a mechanism that is termed 'predictive coding' [Friston, 2012, Sedley et al., 2016, Schilling et al., 2023c, Kocagoncu et al., 2021]. Indeed, in the last few years the LLMs, which are also based on predicting the next word resp. token, have become extremely popular and attracted great attention [Gruver et al., 2024, Wu et al., 2023, Neftci and Averbeck, 2019, Sejnowski, 2023]. Thus, these networks might be a great model to further investigate the principles of predictive coding. The present thesis proves that AI can profit from inspirations from the natural sciences [Hassabis et al., 2017], on the one hand, and that neuroscience and also medicine can profit from AI [Kriegeskorte and Douglas, 2018, Lesica et al., 2021], on the other hand. Thus, this thesis is written exactly in the year, in which scientists, who have chosen these approaches, were honored for the first time with the Nobel prize for chemistry and physics. Thus, the physics Nobel prize was awarded for the realization of principles of the natural sciences (in this case physics) in artificial neural networks to develop efficient AI-systems (Nobel prize for John Hopfield and Geoffrey Hinton [Fattaruso, 2024]), whereas the Nobel prize in chemistry was awarded for using AI to better understand molecular processes in nature (Nobel prize for Demis Hassabis, John Jumper, David Baker [Abriata, 2024]). Thus, I hope that this interdisciplinary approach will lead to major breakthroughs in all of the fields and that it will boost the field of

Biomedical Engineering to give hope to the people suffering from certain pathological conditions such as tinnitus.

Chapter 4

Acknowledgments

Was ich im Laufe der letzten Dekade meiner wissenschaftlichen Laufbahn gelernt habe, ist, dass manche Glaubenssätze nicht immer auf die Wissenschaft zutreffen. Einer davon ist, dass es reicht an sich selbst zu glauben, um weit zu kommen. Das ist definitiv nicht der Fall, denn man braucht das Glück, andere Menschen zu finden, die an einen glauben und einen unterstützen. Dieses Glück hatte ich während meiner Habilitation. Ich danke Prof. Dr. Reichenbach, dass er sich ohne Zögern dazu bereit erklärt hat, dem Fachmentorat vorzustehen und mich auf meinem Weg zu begleiten. Er hatte und hat immer ein offenes Ohr für mich, einen guten Rat, oder eine Idee. Vielen Dank Tobias! Ich habe viel von dir gelernt.

Eine weitere für mich unglaublich wichtige Person, die mir immer einen Vertrauensvorschub gegeben hat, ohne irgendeine Bedingung zu stellen, ist Prof. Dr. Andreas Maier. Lieber Andreas, ohne dich wäre das Habilitationsvorhaben gescheitert, bevor es angefangen hat! Ich bin dir unglaublich dankbar dafür, dass du dich vor mich gestellt hast, wenn es mal Schwierigkeiten gab, dass du sofort eingewilligt hast, meine Stelle zu sichern und dass du immer ein offenes Ohr hast. Man kann unfassbar viel von dir als Wissenschaftler, als Führungspersönlichkeit und als Mensch lernen. Das ist alles so selten in der Wissenschaftswelt und wird viel zu wenig gewürdigt! Deswegen hier ein von Herzen kommendes Danke, dafür, dass du das alles für mich getan hast.

Komplettiert wird das Fachmentorat von Prof. Dr. Max Happel aus Berlin. Lieber Max, ich danke dir von Herzen für deine bedingungslose Unterstützung, deine beruhigende und lockere Art, die dazu beigetragen hat, dass ich aus jedem Fachmentoratsmeeting immer besser gelaunt herausgekommen bin, als ich hineingegangen bin. Das hat mir sehr geholfen, wenn ich mal das Gefühl hatte, dass der Weg etwas steiniger wird. Für die Tatsache, dass ich dieses super Fachmentorat gefunden habe, ist in erster Linie mein Glück verantwortlich. Ich hoffe aber, dass ich es schaffe, mir euch als Vorbild zu nehmen und zu helfen, dass sich auch die nächste Generation

Forschende während ihrer akademischen Laufbahn wohlfühlt und dass ich den Mut habe, mich auch gegen Widerstände vor andere zu stellen, wenn der Wind von vorne weht.

Ich möchte wie immer meinen Kollegen, Freunden, meiner Familie und meiner Freundin danken, die mal wieder diesen Weg mit mir gegangen sind. Nein, ihr seid in keiner Zeitschleife gefangen und ja, das war die letzte Abschlussarbeit meines Lebens! Vielen Dank dafür, dass ihr erneut und zum letzten Mal meinen Stress und manchmal auch schlechte Laune weggesteckt habt. Das ist keine Selbstverständlichkeit! Ich habe wirklich sehr viel Glück mit euch allen!

Abbreviations

AGI Artificial General Intelligence

EU European Union

AI Artificial Intelligence

CCN Cognitive Computational Neuroscience

dB[SPL] decibel sound pressure level

IHCs inner hair cells

OHCs outer hair cells

TAD transient auditory dysfunction

HL hearing loss

TM tectorial membrane

DCN dorsal cochlear nucleus

VCN ventral cochlear nucleus

IC inferior colliculus

MGB medial geniculate body

A1 primary auditory cortex

STS superior temporal sulcus

STG superior temporal gyrus

fMRI functional magnetic resonance imaging

BOLD blood oxygenation level dependent

MEG magnetoencephalography

EEG electroencephalography

sEEG stereotactic EEG

iEEG intracranial EEG

ML machine learning

LIF Leaky-Integrate-and-Fire

LSTM Long-Short-Term-Memory

CNS central nervous system

NN notched noise

WN white noise

PPI pre-pulse-inhibition

GPIAS gap pre-pulse inhibition of the acoustic startle response

CSD current source density

LFP local field potential

SR stochastic resonance

ENT Ear Nose and Throat

CNN convolutional neural network

SNN spiking neural network

GDV general discrimination value

DNN deep neural network

S/N signal-to-noise

PCA principal component analysis

LLMs large language models

List of Figures

1.1 The inner ear . . . 7
1.2 The auditory pathway . . . 10

2.1 Leaky-Integrate-and-Fire-Neuron . . . 18
2.2 Long-Short-Term-Memory-Neuron . . . 21
2.3 Comparison of spiking LSTM and LIF neurons in supervisedly trained neural networks . . . 23
2.4 Maze task for evolutionary pruning study . . . 25
2.5 GPIAS paradigm for Zwicker tone testing . . . 31
2.6 Measurement of onset and offset responses in the cortex of guinea pigs . . . 33
2.7 Current source density analysis . . . 35
2.8 The stochastic resonance control circuit . . . 36
2.9 Hybrid neural network as model of the auditory pathway . . . 40
2.10 Integrated model of conscious tinnitus perception . . . 44
2.11 Autoencoder and LFP events with reconstructions . . . 47
2.12 Brain samples from possible stimulus driven states . . . 48

Bibliography

[Abriata, 2024] Abriata, L. A. (2024). The nobel prize in chemistry: past, present, and future of ai in biology. *Communications Biology*, 7(1):1409.

[Ahmed et al., 2006] Ahmed, Z. M., Goodyear, R., Riazuddin, S., Lagziel, A., Legan, P. K., Behra, M., Burgess, S. M., Lilley, K. S., Wilcox, E. R., Riazuddin, S., et al. (2006). The tip-link antigen, a protein associated with the transduction complex of sensory hair cells, is protocadherin-15. *Journal of Neuroscience*, 26(26):7022–7034.

[Almond et al., 2013] Almond, L. M., Patel, K., and Rejali, D. (2013). Transient auditory dysfunction: A description and study of prevalence. *Ear, Nose & Throat Journal*, 92(8):352–356.

[Amalric and Dehaene, 2019] Amalric, M. and Dehaene, S. (2019). A distinct cortical network for mathematical knowledge in the human brain. *NeuroImage*, 189:19–31.

[Andersson and McKenna, 2006] Andersson, G. and McKenna, L. (2006). The role of cognition in tinnitus. *Acta Oto-Laryngologica*, 126(sup556):39–43.

[Ardila, 2021] Ardila, A. (2021). Grammar in the brain: Two grammar subsystems and two agrammatic types of aphasia. *Journal of Neurolinguistics*, 58:100960.

[Baguley et al., 2013] Baguley, D., McFerran, D., and Hall, D. (2013). Tinnitus. *The Lancet*, 382(9904):1600–1607.

[Baguley, 2003] Baguley, D. M. (2003). Hyperacusis. *Journal of the Royal Society of Medicine*, 96(12):582–585.

[Balboa and Grzywacz, 2000] Balboa, R. M. and Grzywacz, N. M. (2000). The role of early retinal lateral inhibition: more than maximizing luminance information. *Visual Neuroscience*, 17(1):77–89.

[Barrow, 1979] Barrow, J. D. (1979). The proton half life and the dirac hypothesis. *Nature*, 282(5740):698–699.

[Bartolomei et al., 2017] Bartolomei, F., Lagarde, S., Wendling, F., McGonigal, A., Jirsa, V., Guye, M., and Bénar, C. (2017). Defining epileptogenic networks: contribution of seeg and signal analysis. *Epilepsia*, 58(7):1131–1147.

[Bauer, 2018] Bauer, C. A. (2018). Tinnitus. *New England Journal of Medicine*, 378(13):1224–1231.

[Bays, 2010] Bays, C. (2010). Introduction to cellular automata and conway's game of life. In *Game of Life Cellular Automata*, pages 1–7. Springer.

[Bean, 2007] Bean, B. P. (2007). The action potential in mammalian central neurons. *Nature Reviews Neuroscience*, 8(6):451–465.

[Berger and García, 2016] Berger, M. and García, P. S. (2016). Anesthetic suppression of thalamic high-frequency oscillations: evidence that the thalamus is more than just a gateway to consciousness?

[Beyeler et al., 2019] Beyeler, M., Rounds, E. L., Carlson, K. D., Dutt, N., and Krichmar, J. L. (2019). Neural correlates of sparse coding and dimensionality reduction. *PLoS computational biology*, 15(6):e1006908.

[Bigras et al., 2022] Bigras, C., Villatte, B., Duda, V., and Hébert, S. (2022). The electrophysiological markers of hyperacusis: a scoping review. *International Journal of Audiology*, 62(6):489–499.

[Biswas et al., 2022] Biswas, R., Lugo, A., Akeroyd, M. A., Schlee, W., Gallus, S., and Hall, D. (2022). Tinnitus prevalence in europe: a multi-country cross-sectional population study. *The Lancet Regional Health–Europe*, 12.

[Bo et al., 2019] Bo, Y., Tang, J., Yu, M., and Wei, W. (2019). Ultra-short-term pv power forecasting based on lstm with peepholes connections. In *2019 IEEE Sustainable Power and Energy Conference (iSPEC)*, pages 1222–1226. IEEE.

[Bourlard and Kabil, 2022] Bourlard, H. and Kabil, S. H. (2022). Autoencoders reloaded. *Biological cybernetics*, 116(4):389–406.

[Brette, 2015] Brette, R. (2015). Philosophy of the spike: rate-based vs. spike-based theories of the brain. *Frontiers in systems neuroscience*, 9:140675.

[Buckle et al., 2023] Buckle, K. L., Poliakoff, E., and Gowen, E. (2023). The blind men and the elephant: The case for a transdiagnostic approach to initiation. *Frontiers in Psychology*, 13:1113579.

[Budd and Kisvárday, 2012] Budd, J. M. and Kisvárday, Z. F. (2012). Communication and wiring in the cortical connectome. *Frontiers in neuroanatomy*, 6:42.

[Burighel et al., 2011] Burighel, P., Caicci, F., and Manni, L. (2011). Hair cells in non-vertebrate models: lower chordates and molluscs. *Hearing research*, 273(1-2):14–24.

[Buzsáki et al., 2012] Buzsáki, G., Anastassiou, C. A., and Koch, C. (2012). The origin of extracellular fields and currents—eeg, ecog, lfp and spikes. *Nature reviews neuroscience*, 13(6):407–420.

[Caicci et al., 2007] Caicci, F., Burighel, P., and Manni, L. (2007). Hair cells in an ascidian (tunicata) and their evolution in chordates. *Hearing Research*, 231(1-2):63–72.

[Cederroth et al., 2019] Cederroth, C. R., Gallus, S., Hall, D. A., Kleinjung, T., Langguth, B., Maruotti, A., Meyer, M., Norena, A., Probst, T., Pryss, R., et al. (2019). Towards an understanding of tinnitus heterogeneity. *Frontiers in aging neuroscience*, 11:446253.

[Cheng et al., 2023] Cheng, F. L., Horikawa, T., Majima, K., Tanaka, M., Abdelhack, M., Aoki, S. C., Hirano, J., and Kamitani, Y. (2023). Reconstructing visual illusory experiences from human brain activity. *Science Advances*, 9(46):eadj3906.

[Cheung et al., 2012] Cheung, M. M., Lau, C., Zhou, I. Y., Chan, K. C., Cheng, J. S., Zhang, J. W., Ho, L. C., and Wu, E. X. (2012). Bold fmri investigation of the rat auditory pathway and tonotopic organization. *Neuroimage*, 60(2):1205–1211.

[Clark, 2018] Clark, A. (2018). A nice surprise? predictive processing and the active pursuit of novelty. *Phenomenology and the Cognitive Sciences*, 17(3):521–534.

[Cohen et al., 2017] Cohen, G., Afshar, S., Tapson, J., and Van Schaik, A. (2017). Emnist: Extending mnist to handwritten letters. In *2017 international joint conference on neural networks (IJCNN)*, pages 2921–2926. IEEE.

[Collins et al., 2003] Collins, F. S., Morgan, M., and Patrinos, A. (2003). The human genome project: lessons from large-scale biology. *Science*, 300(5617):286–290.

[Cramer et al., 2022] Cramer, B., Billaudelle, S., Kanya, S., Leibfried, A., Grübl, A., Karasenko, V., Pehle, C., Schreiber, K., Stradmann, Y., Weis, J., et al. (2022). Surrogate gradients for analog neuromorphic computing. *Proceedings of the National Academy of Sciences*, 119(4):e2109194119.

[Cramer et al., 2020] Cramer, B., Stradmann, Y., Schemmel, J., and Zenke, F. (2020). The heidelberg spiking data sets for the systematic evaluation of spiking neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 33(7):2744–2757.

[Cuesta et al., 2022] Cuesta, M., Garzón, C., and Cobo, P. (2022). Efficacy of sound therapy for tinnitus using an enriched acoustic environment with hearing-loss matched broadband noise. *Brain Sciences*, 12(1):82.

[Dallos, 2008] Dallos, P. (2008). Cochlear amplification, outer hair cells and prestin. *Current opinion in neurobiology*, 18(4):370–376.

[Damasio and Damasio, 2023] Damasio, A. and Damasio, H. (2023). Feelings are the source of consciousness. *Neural Computation*, 35(3):277–286.

[Dauman et al., 2015] Dauman, N., Erlandsson, S., Lundlin, L., and Dauman, R. (2015). Intra-individual variability in tinnitus patients. *HNO*, 63(4):302–306.

[de Cheveigné et al., 2007] de Cheveigné, A., Le Roux, J., and Simon, J. Z. (2007). Meg signal denoising based on time-shift pca. In *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*, volume 1, pages I–317. IEEE.

[De Ridder et al., 2011] De Ridder, D., van der Loo, E., Vanneste, S., Gais, S., Plazier, M., Kovacs, S., Sunaert, S., Menovsky, T., and Van de Heyning, P. (2011). Theta-gamma dysrhythmia and auditory phantom perception: case report. *Journal of neurosurgery*, 114(4):912–921.

[De Ridder et al., 2015] De Ridder, D., Vanneste, S., Langguth, B., and Llinas, R. (2015). Thalamocortical dysrhythmia: a theoretical update in tinnitus. *Frontiers in neurology*, 6:124.

[Deak and Stickgold, 2010] Deak, M. C. and Stickgold, R. (2010). Sleep and cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(4):491–500.

[Dehaene and Changeux, 2011] Dehaene, S. and Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, 70(2):200–227.

[Dempster and Mackenrie, 1990] Dempster, J. and Mackenrie, K. (1990). The resonance frequency of the external auditory canal in children. *Ear and hearing*, 11(4):296–298.

[DeWitt and Rauschecker, 2012] DeWitt, I. and Rauschecker, J. P. (2012). Phoneme and word recognition in the auditory ventral stream. *Proceedings of the National Academy of Sciences*, 109(8):E505–E514.

[Dronkers et al., 2004] Dronkers, N. F., Wilkins, D. P., Van Valin Jr, R. D., Redfern, B. B., and Jaeger, J. J. (2004). Lesion analysis of the brain areas involved in language comprehension. *Cognition*, 92(1-2):145–177.

[Duncan et al., 2007] Duncan, R. O., Sample, P. A., Weinreb, R. N., Bowd, C., and Zangwill, L. M. (2007). Retinotopic organization of primary visual cortex in glaucoma: Comparing fmri measurements of cortical function with visual field loss. *Progress in retinal and eye research*, 26(1):38–56.

[Eggermont and Roberts, 2004] Eggermont, J. J. and Roberts, L. E. (2004). The neuroscience of tinnitus. *Trends in neurosciences*, 27(11):676–682.

[Emery, 2006] Emery, N. J. (2006). Cognitive ornithology: the evolution of avian intelligence. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 361(1465):23–43.

[Emery and Clayton, 2004] Emery, N. J. and Clayton, N. S. (2004). The mentality of crows: convergent evolution of intelligence in corvids and apes. *science*, 306(5703):1903–1907.

[Eriksson et al., 2010] Eriksson, A., Nilsson Jacobi, M., Nyström, J., and Tunstrøm, K. (2010). Determining interaction rules in animal swarms. *Behavioral Ecology*, 21(5):1106–1111.

[Fastl and Stoll, 1979] Fastl, H. and Stoll, G. (1979). Scaling of pitch strength. *Hearing Research*, 1(4):293–301.

[Fattaruso, 2024] Fattaruso, L. (2024). Leaders in artificial neural network development share 2024 nobel prize in physics.

[Fay and Popper, 2000] Fay, R. R. and Popper, A. N. (2000). Evolution of hearing in vertebrates: the inner ears and processing. *Hearing research*, 149(1-2):1–10.

[Fellows et al., 2005] Fellows, L. K., Heberlein, A. S., Morales, D. A., Shivde, G., Waller, S., and Wu, D. H. (2005). Method matters: an empirical study of impact in cognitive neuroscience. *Journal of cognitive neuroscience*, 17(6):850–858.

[Feynman, 1988] Feynman, R. (1988). What i cannot create i do not understand.

[Fischl and Dale, 2000] Fischl, B. and Dale, A. M. (2000). Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences*, 97(20):11050–11055.

[Fournier et al., 2018] Fournier, P., Cuvillier, A.-F., Gallego, S., Paolino, F., Paolino, M., Quemar, A., Londero, A., and Norena, A. (2018). A new method for assessing masking and residual inhibition of tinnitus. *Trends in hearing*, 22:2331216518769996.

[Franosch et al., 2003] Franosch, J.-M. P., Kempter, R., Fastl, H., and van Hemmen, J. L. (2003). Zwicker tone illusion and noise reduction in the auditory system. *Physical review letters*, 90(17):178103.

[Frégnac and Laurent, 2014] Frégnac, Y. and Laurent, G. (2014). Neuroscience: Where is the brain in the human brain project? *Nature*, 513(7516):27–29.

[Fridriksson et al., 2015] Fridriksson, J., Fillmore, P., Guo, D., and Rorden, C. (2015). Chronic broca's aphasia is caused by damage to broca's and wernicke's areas. *Cerebral Cortex*, 25(12):4689–4696.

[Friederici et al., 2017] Friederici, A. D., Chomsky, N., Berwick, R. C., Moro, A., and Bolhuis, J. J. (2017). Language, mind and brain. *Nature human behaviour*, 1(10):713–722.

[Friston, 2010] Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature reviews neuroscience*, 11(2):127–138.

[Friston, 2012] Friston, K. (2012). The history of the future of the bayesian brain. *NeuroImage*, 62(2):1230–1233.

[Furber, 2012] Furber, S. (2012). To build a brain. *IEEE spectrum*, 49(8):44–49.

[Gabel et al., 2007] Gabel, C. V., Gabel, H., Pavlichin, D., Kao, A., Clark, D. A., and Samuel, A. D. (2007). Neural circuits mediate electrosensory behavior in caenorhabditis elegans. *Journal of Neuroscience*, 27(28):7586–7596.

[Galazyuk and Hébert, 2015] Galazyuk, A. and Hébert, S. (2015). Gap-prepulse inhibition of the acoustic startle reflex (gpias) for tinnitus assessment: current status and future directions. *Frontiers in neurology*, 6:88.

[Galazyuk et al., 2019] Galazyuk, A., Longenecker, R., Voytenko, S., Kristaponyte, I., and Nelson, G. (2019). Residual inhibition: From the putative mechanisms to potential tinnitus treatment. *Hearing research*, 375:1–13.

[Gazzaniga et al., 2014] Gazzaniga, M. S., Ivry, R. B., and Mangun, G. (2014). Cognitive neuroscience. the biology of the mind,(2014).

[Gerken, 1996] Gerken, G. M. (1996). Central tinnitus and lateral inhibition: an auditory brainstem model. *Hearing research*, 97(1-2):75–83.

[Gers and Schmidhuber, 2001] Gers, F. A. and Schmidhuber, E. (2001). Lstm recurrent networks learn simple context-free and context-sensitive languages. *IEEE transactions on neural networks*, 12(6):1333–1340.

[Gers and Schmidhuber, 2000] Gers, F. A. and Schmidhuber, J. (2000). Recurrent nets that time and count. In *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium*, volume 3, pages 189–194. IEEE.

[Gerstner, 2000] Gerstner, W. (2000). Population dynamics of spiking neurons: fast transients, asynchronous states, and locking. *Neural computation*, 12(1):43–89.

[Gerum et al., 2023] Gerum, R., Erpenbeck, A., Krauss, P., and Schilling, A. (2023). Leaky-integrate-and-fire neuron-like long-short-term-memory units as model system in computational biology. In *2023 international joint conference on neural networks (IJCNN)*, pages 1–9. IEEE.

[Gerum et al., 2013] Gerum, R., Fabry, B., Metzner, C., Beaulieu, M., Ancel, A., and Zitterbart, D. (2013). The origin of traveling waves in an emperor penguin huddle. *New Journal of Physics*, 15(12):125022.

[Gerum et al., 2019] Gerum, R., Rahlfs, H., Streb, M., Krauss, P., Grimm, J., Metzner, C., Tziridis, K., Günther, M., Schulze, H., Kellermann, W., et al. (2019). Open (g) pias: An open-source solution for the construction of a high-precision acoustic startle response setup for tinnitus screening and threshold estimation in rodents. *Frontiers in behavioral neuroscience*, 13:140.

[Gerum et al., 2020] Gerum, R. C., Erpenbeck, A., Krauss, P., and Schilling, A. (2020). Sparsity through evolutionary pruning prevents neuronal networks from overfitting. *Neural Networks*, 128:305–312.

[Gerum and Schilling, 2021] Gerum, R. C. and Schilling, A. (2021). Integration of leaky-integrate-and-fire neurons in standard machine learning architectures to generate hybrid networks: A surrogate gradient approach. *Neural Computation*, 33(10):2827–2852.

[Gillespie and Müller, 2009] Gillespie, P. G. and Müller, U. (2009). Mechanotransduction by hair cells: models, molecules, and mechanisms. *Cell*, 139(1):33–44.

[Gollnast et al., 2017] Gollnast, D., Tziridis, K., Krauss, P., Schilling, A., Hoppe, U., and Schulze, H. (2017). Analysis of audiometric differences of patients with and without tinnitus in a large clinical database. *Frontiers in neurology*, 8:31.

[Gondara, 2016] Gondara, L. (2016). Medical image denoising using convolutional denoising autoencoders. In *2016 IEEE 16th international conference on data mining workshops (ICDMW)*, pages 241–246. IEEE.

[Grothe, 2000] Grothe, B. (2000). The evolution of temporal processing in the medial superior olive, an auditory brainstem structure. *Progress in neurobiology*, 61(6):581–610.

[Grova et al., 2016] Grova, C., Aiguabella, M., Zelmann, R., Lina, J.-M., Hall, J. A., and Kobayashi, E. (2016). Intracranial eeg potentials estimated from meg sources: A new approach to correlate meg and ieeg data in epilepsy. *Human brain mapping*, 37(5):1661–1683.

[Gruver et al., 2024] Gruver, N., Finzi, M., Qiu, S., and Wilson, A. G. (2024). Large language models are zero-shot time series forecasters. *Advances in Neural Information Processing Systems*, 36.

[Güntürkün et al., 2017] Güntürkün, O., Ströckens, F., Scarf, D., and Colombo, M. (2017). Apes, feathered apes, and pigeons: differences and similarities. *Current Opinion in Behavioral Sciences*, 16:35–40.

[Gutschalk et al., 2015] Gutschalk, A., Uppenkamp, S., Riedel, B., Bartsch, A., Brandt, T., and Vogt-Schaden, M. (2015). Pure word deafness with auditory object agnosia after bilateral lesion of the superior temporal sulcus. *Cortex*, 73:24–35.

[Hagmann et al., 2008] Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Wedeen, V. J., and Sporns, O. (2008). Mapping the structural core of human cerebral cortex. *PLoS biology*, 6(7):e159.

[Hamann and Schmickl, 2012] Hamann, H. and Schmickl, T. (2012). Modelling the swarm: Analysing biological and engineered swarm systems. *Mathematical and Computer Modelling of Dynamical Systems*, 18(1):1–12.

[Hamza and Zeng, 2021] Hamza, Y. and Zeng, F.-G. (2021). Tinnitus is associated with improved cognitive performance in non-hispanic elderly with hearing loss. *Frontiers in Neuroscience*, 15:735950.

[Happel et al., 2010] Happel, M. F., Jeschke, M., and Ohl, F. W. (2010). Spectral integration in primary auditory cortex attributable to temporally precise convergence of thalamocortical and intracortical input. *Journal of Neuroscience*, 30(33):11114–11127.

[Hassabis et al., 2017] Hassabis, D., Kumaran, D., Summerfield, C., and Botvinick, M. (2017). Neuroscience-inspired artificial intelligence. *Neuron*, 95(2):245–258.

[Hayes et al., 2023] Hayes, S. H., Beh, K., Typlt, M., Schormans, A. L., Stolzberg, D., and Allman, B. L. (2023). Using an appetitive operant conditioning paradigm to screen rats for tinnitus induced by intense sound exposure: Experimental considerations and interpretation. *Frontiers in Neuroscience*, 17:1001619.

[Helbing, 2012] Helbing, D. (2012). *Social self-organization: Agent-based simulations and experiments to study emergent social behavior*. Springer.

[Henoch and Chesky, 1999] Henoch, M. A. and Chesky, K. (1999). Ear canal resonance as a risk factor in music-induced hearing loss. *Medical Problems of Performing Artists*, 14(3):103–106.

[Herbranson and Schroeder, 2010] Herbranson, W. T. and Schroeder, J. (2010). Are birds smarter than mathematicians? pigeons (columba livia) perform optimally on a version of the monty hall dilemma. *Journal of Comparative Psychology*, 124(1):1.

[Herculano-Houzel, 2009] Herculano-Houzel, S. (2009). The human brain in numbers: a linearly scaled-up primate brain. *Frontiers in human neuroscience*, 3:857.

[Herculano-Houzel, 2010] Herculano-Houzel, S. (2010). Coordinated scaling of cortical and cerebellar numbers of neurons. *Frontiers in neuroanatomy*, 4:952.

[Hilgetag and Goulas, 2020] Hilgetag, C. C. and Goulas, A. (2020). 'hierarchy'in the organization of brain networks. *Philosophical Transactions of the Royal Society B*, 375(1796):20190319.

[Hluštík et al., 2001] Hluštík, P., Solodkin, A., Gullapalli, R. P., Noll, D. C., and Small, S. L. (2001). Somatotopy in human primary motor and somatosensory hand representations revisited. *Cerebral Cortex*, 11(4):312–321.

[Hochreiter, 1998] Hochreiter, S. (1998). Recurrent neural net learning and vanishing gradient. *International Journal Of Uncertainity, Fuzziness and Knowledge-Based Systems*, 6(2):107–116.

[Hochreiter, 1997] Hochreiter, S, S. J. (1997). Long short-term memory. *Neural Computation MIT-Press*.

[Hodgkin and Huxley, 1952] Hodgkin, A. L. and Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology*, 117(4):500.

[Hudspeth et al., 2013] Hudspeth, A. J., Jessell, T. M., Kandel, E. R., Schwartz, J. H., and Siegelbaum, S. A. (2013). *Principles of neural science, 5th Edition*. McGraw-Hill, Health Professions Division.

[Hughes et al., 1996] Hughes, G. B., Freedman, M. A., Haberkamp, T. J., and Guay, M. E. (1996). Sudden sensorineural hearing loss. *Otolaryngologic Clinics of North America*, 29(3):393–405.

[Hullfish et al., 2019] Hullfish, J., Sedley, W., and Vanneste, S. (2019). Prediction and perception: Insights for (and from) tinnitus. *Neuroscience & Biobehavioral Reviews*, 102:1–12.

[Humphries et al., 2010] Humphries, C., Liebenthal, E., and Binder, J. R. (2010). Tonotopic organization of human auditory cortex. *Neuroimage*, 50(3):1202–1211.

[Huskey et al., 2020] Huskey, R., Bue, A. C., Eden, A., Grall, C., Meshi, D., Prena, K., Schmälzle, R., Scholz, C., Turner, B. O., and Wilcox, S. (2020). Marr's tri-level framework integrates biological explanation across communication subfields. *Journal of Communication*, 70(3):356–378.

[IJCNN, 2023] IJCNN (2023). Ijcnn 2023: Celebrating 80 years of neural networks. In *2023 International Joint Conference on Neural Networks (IJCNN)*, pages 1–45.

[Immertreu et al., 2024] Immertreu, M., Schilling, A., Maier, A., and Krauss, P. (2024). Probing for consciousness in machines. *arXiv preprint arXiv:2411.16262*.

[Inselberg and Von Foerster, 1970] Inselberg, A. and Von Foerster, H. (1970). A mathematical model of the basilar membrane. *Mathematical Biosciences*, 7(3-4):341–363.

[Ito, 2000] Ito, M. (2000). Mechanisms of motor learning in the cerebellum. *Brain research*, 886(1-2):237–245.

[Ivanova et al., 2021] Ivanova, M. V., Zhong, A., Turken, , Baldo, J. V., and Dronkers, N. F. (2021). Functional contributions of the arcuate fasciculus to language processing. *Frontiers in human neuroscience*, 15:672665.

[Ivansic et al., 2017] Ivansic, D., Guntinas-Lichius, O., Müller, B., Volk, G. F., Schneider, G., and Dobel, C. (2017). Impairments of speech comprehension in patients with tinnitus—a review. *Frontiers in aging neuroscience*, 9:224.

[Ivry and Keele, 1989] Ivry, R. B. and Keele, S. W. (1989). Timing functions of the cerebellum. *Journal of cognitive neuroscience*, 1(2):136–152.

[Izhikevich and FitzHugh, 2006] Izhikevich, E. M. and FitzHugh, R. (2006). Fitzhugh-nagumo model. *Scholarpedia*, 1(9):1349.

[Jääskeläinen et al., 2004] Jääskeläinen, I. P., Ahveninen, J., Bonmassar, G., Dale, A. M., Ilmoniemi, R. J., Levänen, S., Lin, F.-H., May, P., Melcher, J., Stufflebeam, S., et al. (2004). Human posterior auditory cortex gates novel sounds to consciousness. *Proceedings of the National Academy of Sciences*, 101(17):6809–6814.

[Jarrell et al., 2012] Jarrell, T. A., Wang, Y., Bloniarz, A. E., Brittin, C. A., Xu, M., Thomson, J. N., Albertson, D. G., Hall, D. H., and Emmons, S. W. (2012). The connectome of a decision-making neural network. *science*, 337(6093):437–444.

[Jayakody et al., 2018] Jayakody, D. M., Friedland, P. L., Martins, R. N., and Sohrabi, H. R. (2018). Impact of aging on the auditory system and related cognitive functions: a narrative review. *Frontiers in neuroscience*, 12:125.

[Jeschke et al., 2021] Jeschke, M., Happel, M. F., Tziridis, K., Krauss, P., Schilling, A., Schulze, H., and Ohl, F. W. (2021). Acute and long-term circuit-level effects in the auditory cortex after sound trauma. *Frontiers in neuroscience*, 14:598406.

[Johnstone et al., 1986] Johnstone, B., Patuzzi, R., and Yates, G. (1986). Basilar membrane measurements and the travelling wave. *Hearing research*, 22(1-3):147–153.

[Jonas and Kording, 2017] Jonas, E. and Kording, K. P. (2017). Could a neuroscientist understand a microprocessor? *PLoS computational biology*, 13(1):e1005268.

[Jones, 2000] Jones, E. G. (2000). Microcolumns in the cerebral cortex. *Proceedings of the National Academy of Sciences*, 97(10):5019–5021.

[Kajikawa and Schroeder, 2011] Kajikawa, Y. and Schroeder, C. E. (2011). How local is the local field potential? *Neuron*, 72(5):847–858.

[Kaltenbach, 2007] Kaltenbach, J. A. (2007). The dorsal cochlear nucleus as a contributor to tinnitus: mechanisms underlying the induction of hyperactivity. *Progress in brain research*, 166:89–106.

[Kaltenbach and Afman, 2000] Kaltenbach, J. A. and Afman, C. E. (2000). Hyperactivity in the dorsal cochlear nucleus after intense sound exposure and its resemblance to tone-evoked activity: a physiological model for tinnitus. *Hearing research*, 140(1-2):165–172.

[Kandler et al., 2009] Kandler, K., Clause, A., and Noh, J. (2009). Tonotopic reorganization of developing auditory brainstem circuits. *Nature neuroscience*, 12(6):711–717.

[Kapfer et al., 2002] Kapfer, C., Seidl, A. H., Schweizer, H., and Grothe, B. (2002). Experience-dependent refinement of inhibitory inputs to auditory coincidence-detector neurons. *Nature neuroscience*, 5(3):247–253.

[Karten, 2015] Karten, H. J. (2015). Vertebrate brains and evolutionary connectomics: on the origins of the mammalian 'neocortex'. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1684):20150060.

[Katz and Miledi, 1965] Katz, B. and Miledi, R. (1965). The effect of temperature on the synaptic delay at the neuromuscular junction. *The Journal of physiology*, 181(3):656.

[Keefe et al., 1994] Keefe, D. H., Bulen, J. C., Campbell, S. L., and Burns, E. M. (1994). Pressure transfer function and absorption cross section from the diffuse field to the human infant ear canal. *The Journal of the Acoustical Society of America*, 95(1):355–371.

[Kensert et al., 2021] Kensert, A., Collaerts, G., Efthymiadis, K., Van Broeck, P., Desmet, G., and Cabooter, D. (2021). Deep convolutional autoencoder for the simultaneous removal of baseline noise and baseline drift in chromatograms. *Journal of Chromatography A*, 1646:462093.

[Kikidis et al., 2021] Kikidis, D., Vassou, E., Markatos, N., Schlee, W., and Iliadou, E. (2021). Hearing aid fitting in tinnitus: A scoping review of methodological aspects and effect on tinnitus distress and perception. *Journal of Clinical Medicine*, 10(13):2896.

[King et al., 2021] King, R. O., Shekhawat, G. S., King, C., Chan, E., Kobayashi, K., and Searchfield, G. D. (2021). The effect of auditory residual inhibition on tinnitus and the electroencephalogram. *Ear and hearing*, 42(1):130–141.

[Kitcher, 1988] Kitcher, P. (1988). Marr's computational theory of vision. *Philosophy of Science*, 55(1):1–24.

[Knipper et al., 2013] Knipper, M., Van Dijk, P., Nunes, I., Rüttiger, L., and Zimmermann, U. (2013). Advances in the neurobiology of hearing disorders: recent developments regarding the basis of tinnitus and hyperacusis. *Progress in neurobiology*, 111:17–33.

[Kocagoncu et al., 2021] Kocagoncu, E., Klimovich-Gray, A., Hughes, L. E., and Rowe, J. B. (2021). Evidence and implications of abnormal predictive coding in dementia. *Brain*, 144(11):3311–3321.

[Koch and Segev, 1998] Koch, C. and Segev, I. (1998). *Methods in neuronal modeling: from ions to networks*. MIT press.

[Koehler et al., 2011] Koehler, S. D., Pradhan, S., Manis, P. B., and Shore, S. E. (2011). Somatosensory inputs modify auditory spike timing in dorsal cochlear nucleus principal cells. *European Journal of Neuroscience*, 33(3):409–420.

[Koelbl et al., 2023] Koelbl, N., Schilling, A., and Krauss, P. (2023). Adaptive ica for speech eeg artifact removal. In *2023 5th International Conference on Bio-engineering for Smart Technologies (BioSMART)*, pages 1–4. IEEE.

[König et al., 2006] König, O., Schaette, R., Kempter, R., and Gross, M. (2006). Course of hearing loss and occurrence of tinnitus. *Hearing research*, 221(1-2):59–64.

[König et al., 1996] König, P., Engel, A. K., and Singer, W. (1996). Integrator or coincidence detector? the role of the cortical neuron revisited. *Trends in neurosciences*, 19(4):130–137.

[Koops and Eggermont, 2021] Koops, E. A. and Eggermont, J. J. (2021). The thalamus and tinnitus: Bridging the gap between animal data and findings in humans. *Hearing Research*, 407:108280.

[Köster et al., 2014] Köster, U., Sohl-Dickstein, J., Gray, C. M., and Olshausen, B. A. (2014). Modeling higher-order correlations within cortical microcolumns. *PLoS computational biology*, 10(7):e1003684.

[Kraskov et al., 2007] Kraskov, A., Quiroga, R. Q., Reddy, L., Fried, I., and Koch, C. (2007). Local field potentials and spikes in the human medial temporal lobe are selective to image category. *Journal of cognitive neuroscience*, 19(3):479–492.

[Krauss, 2024] Krauss, P. (2024). The most complex system in the universe. In *Artificial Intelligence and Brain Research: Neural Networks, Deep Learning and the Future of Cognition*, pages 15–18. Springer.

[Krauss et al., 2024] Krauss, P., Hösch, J., Metzner, C., Maier, A., Uhrig, P., and Schilling, A. (2024). Analyzing narrative processing in large language models (llms): Using gpt4 to test bert. *arXiv preprint arXiv:2405.02024*.

[Krauss and Maier, 2020] Krauss, P. and Maier, A. (2020). Will we ever have conscious machines? *Frontiers in computational neuroscience*, 14:556544.

[Krauss et al., 2021] Krauss, P., Metzner, C., Joshi, N., Schulze, H., Traxdorf, M., Maier, A., and Schilling, A. (2021). Analysis and visualization of sleep stages based on deep neural networks. *Neurobiology of sleep and circadian rhythms*, 10:100064.

[Krauss et al., 2017] Krauss, P., Metzner, C., Schilling, A., Schütz, C., Tziridis, K., Fabry, B., and Schulze, H. (2017). Adaptive stochastic resonance for unknown and variable input signals. *Scientific reports*, 7(1):2450.

[Krauss et al., 2019] Krauss, P., Schilling, A., Tziridis, K., and Schulze, H. (2019). Models of tinnitus development: From cochlea to cortex. *HNO*, 67:172–177.

[Krauss and Tziridis, 2021] Krauss, P. and Tziridis, K. (2021). Simulated transient hearing loss improves auditory sensitivity. *Scientific reports*, 11(1):14791.

[Krauss et al., 2016] Krauss, P., Tziridis, K., Metzner, C., Schilling, A., Hoppe, U., and Schulze, H. (2016). Stochastic resonance controlled upregulation of internal noise after hearing loss as a putative cause of tinnitus-related neuronal hyperactivity. *Frontiers in neuroscience*, 10:597.

[Krauss et al., 2018] Krauss, P., Tziridis, K., Schilling, A., and Schulze, H. (2018). Cross-modal stochastic resonance as a universal principle to enhance sensory processing. *Frontiers in neuroscience*, 12:578.

[Kreiman et al., 2006] Kreiman, G., Hung, C. P., Kraskov, A., Quiroga, R. Q., Poggio, T., and DiCarlo, J. J. (2006). Object selectivity of local field potentials and spikes in the macaque inferior temporal cortex. *Neuron*, 49(3):433–445.

[Kriegeskorte and Douglas, 2018] Kriegeskorte, N. and Douglas, P. K. (2018). Cognitive computational neuroscience. *Nature neuroscience*, 21(9):1148–1160.

[Krizhevsky, 2009] Krizhevsky, A. (2009). Learning multiple layers of features from tiny images. *https://www. cs. toronto. edu/kriz/learning-features-2009-TR. pdf*.

[Kujawa and Liberman, 2015] Kujawa, S. G. and Liberman, M. C. (2015). Synaptopathy in the noise-exposed and aging cochlea: Primary neural degeneration in acquired sensorineural hearing loss. *Hearing research*, 330:191–199.

[Kunchur, 2023] Kunchur, M. N. (2023). The human auditory system and audio. *Applied Acoustics*, 211:109507.

[LeCun et al., 1995] LeCun, Y., Jackel, L. D., Bottou, L., Cortes, C., Denker, J. S., Drucker, H., Guyon, I., Muller, U. A., Sackinger, E., Simard, P., et al. (1995). Learning algorithms for classification: A comparison on handwritten digit recognition. *Neural networks: the statistical mechanics perspective*, 261(276):2.

[Lee et al., 2016] Lee, J. H., Delbruck, T., and Pfeiffer, M. (2016). Training deep spiking neural networks using backpropagation. *Frontiers in neuroscience*, 10:508.

[Lent et al., 2012] Lent, R., Azevedo, F. A., Andrade-Moraes, C. H., and Pinto, A. V. (2012). How many neurons do you have? some dogmas of quantitative neuroscience under revision. *European Journal of Neuroscience*, 35(1):1–9.

[Lesica et al., 2021] Lesica, N. A., Mehta, N., Manjaly, J. G., Deng, L., Wilson, B. S., and Zeng, F.-G. (2021). Harnessing the power of artificial intelligence to transform hearing healthcare and research. *Nature Machine Intelligence*, 3(10):840–849.

[Levin and Miller, 1996] Levin, J. E. and Miller, J. P. (1996). Broadband neural encoding in the cricket cereal sensory system enhanced by stochastic resonance. *nature*, 380(6570):165–168.

[Li et al., 2022] Li, H., Hu, J., Chen, A., Wang, C., Chen, L., Tian, F., Zhou, J., Zhao, Y., Chen, J., Tong, Y., et al. (2022). Single-transistor neuron with excitatory–inhibitory spatiotemporal dynamics applied for neuronal oscillations. *Advanced Materials*, 34(51):2207371.

[Li et al., 2016] Li, X., Qin, T., Yang, J., and Liu, T.-Y. (2016). Lightrnn: Memory and computation-efficient recurrent neural networks. *Advances in Neural Information Processing Systems*, 29.

[Liberman and Kujawa, 2017] Liberman, M. C. and Kujawa, S. G. (2017). Cochlear synaptopathy in acquired sensorineural hearing loss: Manifestations and mechanisms. *Hearing research*, 349:138–147.

[Licklider, 1951] Licklider, J. C. R. (1951). A duplex theory of pitch perception. *The Journal of the Acoustical Society of America*, 23(1_Supplement):147–147.

[Lindén et al., 2011] Lindén, H., Tetzlaff, T., Potjans, T. C., Pettersen, K. H., Grün, S., Diesmann, M., and Einevoll, G. T. (2011). Modeling the spatial reach of the lfp. *Neuron*, 72(5):859–872.

[Liu et al., 2020] Liu, W., Wen, B., Gao, S., Zheng, J., and Zheng, Y. (2020). A multi-label text classification model based on elmo and attention. In *MATEC Web of Conferences*, volume 309, page 03015. EDP Sciences.

[Llinás et al., 1999] Llinás, R. R., Ribary, U., Jeanmonod, D., Kronberg, E., and Mitra, P. P. (1999). Thalamocortical dysrhythmia: a neurological and neuropsychiatric syndrome characterized by magnetoencephalography. *Proceedings of the National Academy of Sciences*, 96(26):15222–15227.

[Logothetis, 2003] Logothetis, N. K. (2003). The underpinnings of the bold functional magnetic resonance imaging signal. *Journal of Neuroscience*, 23(10):3963–3971.

[Loss et al., 2021] Loss, C. M., Melleu, F. F., Domingues, K., Lino-de Oliveira, C., and Viola, G. G. (2021). Combining animal welfare with experimental rigor to improve reproducibility in behavioral neuroscience. *Frontiers in Behavioral Neuroscience*, 15:763428.

[Luczak et al., 2009] Luczak, A., Barthó, P., and Harris, K. D. (2009). Spontaneous events outline the realm of possible sensory responses in neocortical populations. *Neuron*, 62(3):413–425.

[MacKay and Murphy, 1976] MacKay, W. and Murphy, J. (1976). Integrative versus delay line characteristics of cerebellar cortex. *Canadian Journal of Neurological Sciences*, 3(2):85–97.

[Macmillan, 2000] Macmillan, M. (2000). Restoring phineas gage: A 150th retrospective. *Journal of the History of the Neurosciences*, 9(1):46–66.

[Mann and Kelley, 2011] Mann, Z. F. and Kelley, M. W. (2011). Development of tonotopy in the auditory periphery. *Hearing research*, 276(1-2):2–15.

[Markram, 2012] Markram, H. (2012). The human brain project. *Scientific American*, 306(6):50–55.

[Marks et al., 2018] Marks, K. L., Martel, D. T., Wu, C., Basura, G. J., Roberts, L. E., Schvartz-Leyzac, K. C., and Shore, S. E. (2018). Auditory-somatosensory bimodal stimulation desynchronizes brain circuitry to reduce tinnitus in guinea pigs and humans. *Science Translational Medicine*, 10(422):eaal3175.

[Marr, 2010] Marr, D. (2010). *Vision: A computational investigation into the human representation and processing of visual information*. MIT press.

[Mazurek et al., 2012] Mazurek, B., Haupt, H., Olze, H., and Szczepek, A. J. (2012). Stress and tinnitus—from bedside to bench and back. *Frontiers in systems neuroscience*, 6:47.

[Mazurek et al., 2015] Mazurek, B., Szczepek, A., and Hebert, S. (2015). Stress und tinnitus. *Hno*, 63:258–265.

[McNeill et al., 2012] McNeill, C., Távora-Vieira, D., Alnafjan, F., Searchfield, G. D., and Welch, D. (2012). Tinnitus pitch, masking, and the effectiveness of hearing aids for tinnitus therapy. *International journal of audiology*, 51(12):914–919.

[Mehonic and Kenyon, 2022] Mehonic, A. and Kenyon, A. J. (2022). Brain-inspired computing needs a master plan. *Nature*, 604(7905):255–260.

[Meng et al., 2017] Meng, Q., Catchpoole, D., Skillicom, D., and Kennedy, P. J. (2017). Relational autoencoder for feature extraction. In *2017 International joint conference on neural networks (IJCNN)*, pages 364–371. IEEE.

[Metzner et al., 2024] Metzner, C., Yamakou, M. E., Voelkl, D., Schilling, A., and Krauss, P. (2024). Quantifying and maximizing the information flux in recurrent neural networks. *Neural Computation*, 36(3):351–384.

[Miller and Corsellis, 1977] Miller, A. and Corsellis, J. (1977). Evidence for a secular increase in human brain weight during the past century. *Annals of human biology*, 4(3):253–257.

[Mitzdorf, 1985] Mitzdorf, U. (1985). Current source-density method and application in cat cerebral cortex: investigation of evoked potentials and eeg phenomena. *Physiological reviews*, 65(1):37–100.

[Miyashita and Hikosaka, 1996] Miyashita, N. and Hikosaka, O. (1996). Minimal synaptic delay in the saccadic output pathway of the superior colliculus studied in awake monkey. *Experimental brain research*, 112:187–196.

[Mohan et al., 2020] Mohan, A., Bhamoo, N., Riquelme, J. S., Long, S., Norena, A., and Vanneste, S. (2020). Investigating functional changes in the brain to intermittently induced auditory illusions and its relevance to chronic tinnitus. *Human Brain Mapping*, 41(7):1819–1832.

[Møller and Jannetta, 1983] Møller, A. R. and Jannetta, P. J. (1983). Auditory evoked potentials recorded from the cochlear nucleus and its vicinity in man. *Journal of neurosurgery*, 59(6):1013–1018.

[Møller et al., 1989] Møller, A. R., Sekiya, T., and Sen, C. N. (1989). Responses from dorsal column nuclei (dcn) in the monkey to stimulation of upper and lower limbs and spinal cord. *Electroencephalography and clinical Neurophysiology*, 73(4):353–361.

[Moss et al., 2004] Moss, F., Ward, L. M., and Sannita, W. G. (2004). Stochastic resonance and sensory information processing: a tutorial and review of application. *Clinical neurophysiology*, 115(2):267–281.

[Mulaosmanovic et al., 2018] Mulaosmanovic, H., Chicca, E., Bertele, M., Mikolajick, T., and Slesazeck, S. (2018). Mimicking biological neurons with a nanoscale ferroelectric transistor. *Nanoscale*, 10(46):21755–21763.

[Murugavel et al., 2016] Murugavel, M., Akshaya, D., Anitha, S., Manjureka, M., and Mohanapriya, T. (2016). Multiclass support vector machine with new kernel for eeg classification. *International Research Journal of Engineering and Technology*, 3:1192–1197.

[Nadol Jr, 1993] Nadol Jr, J. B. (1993). Hearing loss. *New England Journal of Medicine*, 329(15):1092–1102.

[Neftci and Averbeck, 2019] Neftci, E. O. and Averbeck, B. B. (2019). Reinforcement learning in artificial and biological systems. *Nature Machine Intelligence*, 1(3):133–143.

[Neftci et al., 2019] Neftci, E. O., Mostafa, H., and Zenke, F. (2019). Surrogate gradient learning in spiking neural networks: Bringing the power of gradient-based optimization to spiking neural networks. *IEEE Signal Processing Magazine*, 36(6):51–63.

[Nelson, 2002] Nelson, S. B. (2002). Cortical microcircuits: diverse or canonical? *Neuron*, 36(1):19–27.

[Norena and Eggermont, 2003] Norena, A. and Eggermont, J. (2003). Neural correlates of an auditory afterimage in primary auditory cortex. *Journal of the Association for Research in Otolaryngology*, 4:312–328.

[Norena et al., 2000] Norena, A., MICHEyL, C., and Chery-Croze, S. (2000). An auditory negative after-image as a human model of tinnitus. *Hearing research*, 149(1-2):24–32.

[Noreña, 2011] Noreña, A. J. (2011). An integrative model of tinnitus based on a central gain controlling neural sensitivity. *Neuroscience & Biobehavioral Reviews*, 35(5):1089–1109.

[Nouvian et al., 2006] Nouvian, R., Beutner, D., Parsons, T. D., and Moser, T. (2006). Structure and function of the hair cell ribbon synapse. *The Journal of membrane biology*, 209:153–165.

[O'Donnell, 2023] O'Donnell, C. (2023). Nonlinear slow-timescale mechanisms in synaptic plasticity. *Current opinion in neurobiology*, 82:102778.

[Oertel and Young, 2004] Oertel, D. and Young, E. D. (2004). What's a cerebellar circuit doing in the auditory system? *Trends in neurosciences*, 27(2):104–110.

[O'Reilly et al., 2012] O'Reilly, J. X., Jbabdi, S., and Behrens, T. E. (2012). How can a bayesian approach inform neuroscience? *European Journal of Neuroscience*, 35(7):1169–1179.

[Panksepp, 2010] Panksepp, J. (2010). Affective neuroscience of the emotional brainmind: evolutionary perspectives and implications for understanding depression. *Dialogues in clinical neuroscience*, 12(4):533–545.

[Parameshwarappa and Norena, 2024] Parameshwarappa, V. and Norena, A. J. (2024). The effects of acute and chronic noise trauma on stimulus-evoked activity across primary auditory cortex layers. *Journal of Neurophysiology*, 131(2):225–240.

[Parra and Pearlmutter, 2007] Parra, L. C. and Pearlmutter, B. A. (2007). Illusory percepts from auditory adaptation. *The Journal of the Acoustical Society of America*, 121(3):1632–1641.

[Parvizi and Kastner, 2018] Parvizi, J. and Kastner, S. (2018). Human intracranial eeg: promises and limitations. *Nature neuroscience*, 21(4):474.

[Patanaik et al., 2018] Patanaik, A., Ong, J. L., Gooley, J. J., Ancoli-Israel, S., and Chee, M. W. (2018). An end-to-end framework for real-time automatic sleep stage classification. *Sleep*, 41(5):zsy041.

[Payne and Wong, 2022] Payne, T. and Wong, G. (2022). Hearing loss: Conductive versus sensorineural. *InnovAiT*, 15(4):218–225.

[Perez-Nieves et al., 2021] Perez-Nieves, N., Leung, V. C., Dragotti, P. L., and Goodman, D. F. (2021). Neural heterogeneity promotes robust learning. *Nature communications*, 12(1):5791.

[Perlis, 1997] Perlis, D. (1997). Consciousness as self-function. *Journal of Consciousness Studies*, 4(5-6):509–525.

[Peters et al., 2018] Peters, M. E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., and Zettlemoyer, L. (2018). Deep contextualized word representations. arxiv: 180205365. *arXiv*.

[Pfeifer and Iida, 2004] Pfeifer, R. and Iida, F. (2004). Embodied artificial intelligence: Trends and challenges. *Lecture notes in computer science*, pages 1–26.

[Pienkowski, 2019] Pienkowski, M. (2019). Rationale and efficacy of sound therapies for tinnitus and hyperacusis. *Neuroscience*, 407:120–134.

[Platkiewicz and Brette, 2010] Platkiewicz, J. and Brette, R. (2010). A threshold equation for action potential initiation. *PLoS computational biology*, 6(7):e1000850.

[Prengel et al., 2023] Prengel, J., Dobel, C., and Guntinas-Lichius, O. (2023). Tinnitus. *Laryngo-Rhino-Otologie*, 102(02):132–145.

[Rahman and Siddiqui, 2019] Rahman, M. M. and Siddiqui, F. H. (2019). An optimized abstractive text summarization model using peephole convolutional lstm. *Symmetry*, 11(10):1290.

[Rakic, 2009] Rakic, P. (2009). Evolution of the neocortex: a perspective from developmental biology. *Nature Reviews Neuroscience*, 10(10):724–735.

[Rauschecker, 2024] Rauschecker, J. P. (2024). The frontostriatal gating model of tinnitus. In *Textbook of Tinnitus*, pages 221–230. Springer.

[Rauschecker et al., 2015] Rauschecker, J. P., May, E. S., Maudoux, A., and Ploner, M. (2015). Frontostriatal gating of tinnitus and chronic pain. *Trends in cognitive sciences*, 19(10):567–578.

[Rees, 1983] Rees, M. J. (1983). Large numbers and ratios in astrophysics and cosmology. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 310(1512):311–322.

[Reich et al., 2000] Reich, D. S., Mechler, F., Purpura, K. P., and Victor, J. D. (2000). Interspike intervals, receptive fields, and information encoding in primary visual cortex. *Journal of Neuroscience*, 20(5):1964–1974.

[Repperger et al., 2005] Repperger, D. W., Phillips, C. A., Berlin, J. E., Neidhard-Doll, A. T., and Haas, M. W. (2005). Human-machine haptic interface design using stochastic resonance methods. *IEEE Transactions on systems, man, and cybernetics-part A: Systems and Humans*, 35(4):574–582.

[Ronran et al., 2020] Ronran, C., Lee, S., and Jang, H. J. (2020). Delayed combination of feature embedding in bidirectional lstm crf for ner. *Applied Sciences*, 10(21):7557.

[Rousseau et al., 2003] Rousseau, D., Varela, J. R., and Chapeau-Blondeau, F. (2003). Stochastic resonance for nonlinear sensors with saturation. *Physical Review E*, 67(2):021102.

[Russin et al., 2020] Russin, J., O'Reilly, R. C., and Bengio, Y. (2020). Deep learning needs a prefrontal cortex. *Work Bridging AI Cogn Sci*, 107(603-616):1.

[Saenz and Langers, 2014] Saenz, M. and Langers, D. R. (2014). Tonotopic mapping of human auditory cortex. *Hearing research*, 307:42–52.

[Saha et al., 2021] Saha, S., Mamun, K. A., Ahmed, K., Mostafa, R., Naik, G. R., Darvishi, S., Khandoker, A. H., and Baumert, M. (2021). Progress in brain computer interface: Challenges and opportunities. *Frontiers in systems neuroscience*, 15:578875.

[Salami et al., 2003] Salami, M., Itami, C., Tsumoto, T., and Kimura, F. (2003). Change of conduction velocity by regional myelination yields constant latency irrespective of distance between thalamus and cortex. *Proceedings of the National Academy of Sciences*, 100(10):6174–6179.

[Samantsidis et al., 2020] Samantsidis, G.-R., Panteleri, R., Denecke, S., Kounadi, S., Christou, I., Nauen, R., Douris, V., and Vontas, J. (2020). 'what i cannot create, i do not understand':

functionally validated synergism of metabolic and target site insecticide resistance. *Proceedings of the Royal Society B*, 287(1927):20200838.

[Schaette and McAlpine, 2011] Schaette, R. and McAlpine, D. (2011). Tinnitus with a normal audiogram: physiological evidence for hidden hearing loss and computational model. *Journal of Neuroscience*, 31(38):13452–13457.

[Schecklmann et al., 2012] Schecklmann, M., Vielsmeier, V., Steffens, T., Landgrebe, M., Langguth, B., and Kleinjung, T. (2012). Relationship between audiometric slope and tinnitus pitch in tinnitus patients: insights into the mechanisms of tinnitus generation. *PloS one*, 7(4):e34878.

[Schilling et al., 2023a] Schilling, A., Choi, B., Parameshwarappa, V., and Norena, A. J. (2023a). Offset responses in primary auditory cortex are enhanced after notched noise stimulation. *Journal of Neurophysiology*, 129(5):1114–1126.

[Schilling et al., 2024] Schilling, A., Gerum, R., Boehm, C., Rasheed, J., Metzner, C., Maier, A., Reindl, C., Hamer, H., and Krauss, P. (2024). Deep learning based decoding of single local field potential events. *NeuroImage*, page 120696.

[Schilling et al., 2022] Schilling, A., Gerum, R., Metzner, C., Maier, A., and Krauss, P. (2022). Intrinsic noise improves speech recognition in a computational model of the auditory pathway. *Frontiers in Neuroscience*, 16:908330.

[Schilling and Krauss, 2022] Schilling, A. and Krauss, P. (2022). Tinnitus is associated with improved cognitive performance and speech perception–can stochastic resonance explain? *Frontiers in Aging Neuroscience*, 14:1073149.

[Schilling and Krauss, 2024] Schilling, A. and Krauss, P. (2024). The bayesian brain: world models and conscious dimensions of auditory phantom perception.

[Schilling et al., 2017] Schilling, A., Krauss, P., Gerum, R., Metzner, C., Tziridis, K., and Schulze, H. (2017). A new statistical approach for the evaluation of gap-prepulse inhibition of the acoustic startle reflex (gpias) for tinnitus assessment. *Frontiers in behavioral neuroscience*, 11:198.

[Schilling et al., 2021a] Schilling, A., Krauss, P., Hannemann, R., Schulze, H., and Tziridis, K. (2021a). Reduktion der tinnituslautstärke: Pilotstudie zur abschwächung von tonalem tinnitus mit schwellennahem, individuell spektral optimiertem rauschen. *Hno*, 69(11):891.

[Schilling et al., 2021b] Schilling, A., Maier, A., Gerum, R., Metzner, C., and Krauss, P. (2021b). Quantifying the separability of data classes in neural networks. *Neural Networks*, 139:278–293.

[Schilling et al., 2023b] Schilling, A., Schaette, R., Sedley, W., Gerum, R. C., Maier, A., and Krauss, P. (2023b). Auditory perception and phantom perception in brains, minds and machines.

[Schilling et al., 2023c] Schilling, A., Sedley, W., Gerum, R., Metzner, C., Tziridis, K., Maier, A., Schulze, H., Zeng, F.-G., Friston, K. J., and Krauss, P. (2023c). Predictive coding and stochastic resonance as fundamental principles of auditory phantom perception. *Brain*, 146(12):4809–4825.

[Schilling et al., 2021c] Schilling, A., Tziridis, K., Schulze, H., and Krauss, P. (2021c). The stochastic resonance model of auditory perception: A unified explanation of tinnitus development, zwicker tone illusion, and residual inhibition. *Progress in brain research*, 262:139–157.

[Schilling et al., 2023d] Schilling, A., Tziridis, K., Schulze, H., and Krauss, P. (2023d). Behavioral assessment of zwicker tone percepts in gerbils. *Neuroscience*, 520:39–45.

[Schmidt et al., 2010] Schmidt, R. F., Lang, F., and Heckmann, M. (2010). *Physiologie des menschen: mit pathophysiologie, 31. Auflage*. Springer-Verlag.

[Schreiber et al., 2010] Schreiber, B. E., Agrup, C., Haskard, D. O., and Luxon, L. M. (2010). Sudden sensorineural hearing loss. *The Lancet*, 375(9721):1203–1211.

[Schwering and MacDonald, 2020] Schwering, S. C. and MacDonald, M. C. (2020). Verbal working memory as emergent from language comprehension and production. *Frontiers in human neuroscience*, 14:68.

[Sedley et al., 2016] Sedley, W., Friston, K. J., Gander, P. E., Kumar, S., and Griffiths, T. D. (2016). An integrative tinnitus model based on sensory precision. *Trends in neurosciences*, 39(12):799–812.

[Seenivasan et al., 2024] Seenivasan, M., Parekkattil, A. V., Ahmed, R. U., and Saha, P. (2024). Biologically inspired tonic and bursting lif neuron model for spiking neural network: a cmos implementation. *Microsystem Technologies*, pages 1–15.

[Sejnowski, 2023] Sejnowski, T. J. (2023). Large language models and the reverse turing test. *Neural computation*, 35(3):309–342.

[Shamma, 1985] Shamma, S. A. (1985). Speech processing in the auditory system i: The representation of speech sounds in the responses of the auditory nerve. *The Journal of the Acoustical Society of America*, 78(5):1612–1621.

[Sharkey, 2006] Sharkey, A. J. (2006). Robots, insects and swarm intelligence. *Artificial Intelligence Review*, 26:255–268.

[Shatnawi et al., 2009] Shatnawi, S., Stroup-Gardiner, M., and Stubstad, R. (2009). California's perspective on concrete pavement preservation. In *National conference on preservation, repair, and rehabilitation of concrete pavements, St. Louis, Missouri*, pages 71–86.

[Shipp, 2007] Shipp, S. (2007). Structure and function of the cerebral cortex. *Current Biology*, 17(12):R443–R449.

[Shore et al., 2007] Shore, S., Zhou, J., and Koehler, S. (2007). Neural mechanisms underlying somatic tinnitus. *Progress in brain research*, 166:107–548.

[Shore and Zhou, 2006] Shore, S. E. and Zhou, J. (2006). Somatosensory influence on the cochlear nucleus and beyond. *Hearing research*, 216:90–99.

[Smith et al., 1993] Smith, P. H., Joris, P. X., and Yin, T. C. (1993). Projections of physiologically characterized spherical bushy cell axons from the cochlear nucleus of the cat: evidence for delay lines to the medial superior olive. *Journal of Comparative Neurology*, 331(2):245–260.

[Smith et al., 2012] Smith, P. H., Uhlrich, D. J., Manning, K. A., and Banks, M. I. (2012). Thalamocortical projections to rat auditory cortex from the ventral and dorsal divisions of the medial geniculate nucleus. *Journal of Comparative Neurology*, 520(1):34–51.

[Smith et al., 2021] Smith, R., Badcock, P., and Friston, K. J. (2021). Recent advances in the application of predictive coding and active inference models within clinical neuroscience. *Psychiatry and Clinical Neurosciences*, 75(1):3–13.

[Sporns, 2011] Sporns, O. (2011). The human connectome: a complex network. *Annals of the new York Academy of Sciences*, 1224(1):109–125.

[Sporns et al., 2005] Sporns, O., Tononi, G., and Kötter, R. (2005). The human connectome: a structural description of the human brain. *PLoS computational biology*, 1(4):e42.

[Starr et al., 2001] Starr, A., Sininger, Y., Nguyen, T., Michalewski, H., Oba, S., and Abdala, C. (2001). Cochlear receptor (microphonic and summating potentials, otoacoustic emissions) and

auditory pathway (auditory brain stem potentials) activity in auditory neuropathy. *Ear and Hearing*, 22(2):91–99.

[Stoll et al., 2023] Stoll, A., Maier, A., Krauss, P., Gerum, R., and Schilling, A. (2023). Coincidence detection and integration behavior in spiking neural networks. *Cognitive Neurodynamics*, pages 1–13.

[Studer and Barkat, 2022] Studer, F. and Barkat, T. R. (2022). Inhibition in the auditory cortex. *Neuroscience & Biobehavioral Reviews*, 132:61–75.

[Suga, 1995] Suga, N. (1995). Sharpening of frequency tuning by inhibition in the central auditory system: tribute to yasuji katsuki. *Neuroscience research*, 21(4):287–299.

[Super and Uylings, 2001] Super, H. and Uylings, H. (2001). The early differentiation of the neocortex: a hypothesis on neocortical evolution. *Cerebral Cortex*, 11(12):1101–1109.

[Szabo and Birdsey, 2017] Szabo, C. and Birdsey, L. (2017). Validating emergent behavior in complex systems. In *Advances in Modeling and Simulation: Seminal Research from 50 Years of Winter Simulation Conferences*, pages 47–62. Springer.

[Taherkhani et al., 2020] Taherkhani, A., Belatreche, A., Li, Y., Cosma, G., Maguire, L. P., and McGinnity, T. M. (2020). A review of learning in biologically plausible spiking neural networks. *Neural Networks*, 122:253–272.

[Takeda et al., 2009] Takeda, T., Okamoto, M., Atsuda, K., and Katagiri, K. (2009). Performance of a helium circulation system for a meg. *Cryogenics*, 49(3-4):144–150.

[Tavanaei et al., 2019] Tavanaei, A., Ghodrati, M., Kheradpisheh, S. R., Masquelier, T., and Maida, A. (2019). Deep learning in spiking neural networks. *Neural networks*, 111:47–63.

[ten Donkelaar et al., 2020] ten Donkelaar, H. J., ten Donkelaar, H. J., and Kaga, K. (2020). The auditory system. *Clinical neuroanatomy: brain circuitry and its disorders*, pages 373–407.

[Thiebaut de Schotten and Forkel, 2022] Thiebaut de Schotten, M. and Forkel, S. J. (2022). The emergent properties of the connected brain. *Science*, 378(6619):505–510.

[Tramo et al., 2005] Tramo, M. J., Cariani, P. A., Koh, C. K., Makris, N., and Braida, L. D. (2005). Neurophysiology and neuroanatomy of pitch perception: auditory cortex. *Annals of the New York Academy of Sciences*, 1060(1):148–174.

[Tramo et al., 2002] Tramo, M. J., Shah, G. D., and Braida, L. D. (2002). Functional role of auditory cortex in frequency processing and pitch perception. *Journal of neurophysiology*, 87(1):122–139.

[Trotter and Donaldson, 2008] Trotter, M. and Donaldson, I. (2008). Hearing aids and tinnitus therapy: a 25-year experience. *The Journal of Laryngology & Otology*, 122(10):1052–1056.

[Tucker et al., 2016] Tucker, R. P., Peterson, C. A., Hendaoui, I., Bichet, S., and Chiquet-Ehrismann, R. (2016). The expression of tenascin-c and tenascin-w in human ossicles. *Journal of anatomy*, 229(3):416–421.

[Turner et al., 2006] Turner, J. G., Brozoski, T. J., Bauer, C. A., Parrish, J. L., Myers, K., Hughes, L. F., and Caspary, D. M. (2006). Gap detection deficits in rats with tinnitus: a potential novel screening tool. *Behavioral neuroscience*, 120(1):188.

[Tziridis et al., 2022a] Tziridis, K., Brunner, S., Schilling, A., Krauss, P., and Schulze, H. (2022a). Spectrally matched near-threshold noise for subjective tinnitus loudness attenuation based on stochastic resonance. *Frontiers in Neuroscience*, 16:831581.

[Tziridis et al., 2021] Tziridis, K., Forster, J., Buchheidt-Dörfler, I., Krauss, P., Schilling, A., Wendler, O., Sterna, E., and Schulze, H. (2021). Tinnitus development is associated with synaptopathy of inner hair cells in mongolian gerbils. *European Journal of Neuroscience*, 54(3):4768–4780.

[Tziridis et al., 2022b] Tziridis, K., Friedrich, J., Brüeggemann, P., Mazurek, B., and Schulze, H. (2022b). Estimation of tinnitus-related socioeconomic costs in germany. *International Journal of Environmental Research and Public Health*, 19(16):10455.

[Ueberfuhr et al., 2017] Ueberfuhr, M. A., Braun, A., Wiegrebe, L., Grothe, B., and Drexl, M. (2017). Modulation of auditory percepts by transcutaneous electrical stimulation. *Hearing research*, 350:235–243.

[Urban, 2002] Urban, N. N. (2002). Lateral inhibition in the olfactory bulb and in olfaction. *Physiology & behavior*, 77(4-5):607–612.

[Vaidya et al., 2019] Vaidya, A. R., Pujara, M. S., Petrides, M., Murray, E. A., and Fellows, L. K. (2019). Lesion studies in contemporary neuroscience. *Trends in cognitive sciences*, 23(8):653–671.

[Valdés-Sosa et al., 2005] Valdés-Sosa, P. A., Sánchez-Bornot, J. M., Lage-Castellanos, A., Vega-Hernández, M., Bosch-Bayard, J., Melie-García, L., and Canales-Rodríguez, E. (2005). Estimating brain functional connectivity with sparse multivariate autoregression. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1457):969–981.

[van der Heijden et al., 2013] van der Heijden, M., Lorteije, J. A., Plauška, A., Roberts, M. T., Golding, N. L., and Borst, J. G. G. (2013). Directional hearing by linear summation of binaural inputs at the medial superior olive. *Neuron*, 78(5):936–948.

[Velíšek, 2018] Velíšek, L. (2018). "shake it off" versus "in your wildest dreams": Thalamus as a consciousness gate for temporal lobe seizures. *Epilepsy Currents*, 18(4):248–250.

[Verdecchia et al., 2023] Verdecchia, R., Sallou, J., and Cruz, L. (2023). A systematic review of green ai. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 13(4):e1507.

[Verma and Yadava, 2016] Verma, V. K. and Yadava, R. (2016). Stochastic resonance in mems capacitive sensors. *Sensors and Actuators B: Chemical*, 235:583–602.

[Vertes et al., 2015] Vertes, R. P., Linley, S. B., and Hoover, W. B. (2015). Limbic circuitry of the midline thalamus. *Neuroscience & Biobehavioral Reviews*, 54:89–107.

[Vlajkovic and Thorne, 2022] Vlajkovic, S. M. and Thorne, P. R. (2022). Purinergic signalling in the cochlea. *International Journal of Molecular Sciences*, 23(23):14874.

[von Békésy, 1970] von Békésy, G. (1970). Travelling waves as frequency analysers in the cochlea. *Nature*, 225(5239):1207–1209.

[Walker, 2009] Walker, M. P. (2009). The role of sleep in cognition and emotion. *Annals of the New York Academy of Sciences*, 1156(1):168–197.

[Wallhäusser-Franke et al., 2017] Wallhäusser-Franke, E., D'Amelio, R., Glauner, A., Delb, W., Servais, J. J., Hörmann, K., and Repik, I. (2017). Transition from acute to chronic tinnitus: predictors for the development of chronic distressing tinnitus. *Frontiers in neurology*, 8:605.

[Wang et al., 2024] Wang, C., Jiang, Z.-y., Chai, J.-y., Chen, H.-s., Liu, L.-x., Dang, T., and Meng, X.-m. (2024). Mouse auditory cortex sub-fields receive neuronal projections from mgb subdivisions independently. *Scientific Reports*, 14(1):7078.

[Wang et al., 2022] Wang, M., Liu, X., Lai, Y., Cao, W., Wu, Z., and Guo, X. (2022). Application of neuroscience tools in building construction–an interdisciplinary analysis. *Frontiers in Neuroscience*, 16:895666.

[Ward, 2013] Ward, L. M. (2013). The thalamus: gateway to the mind. *Wiley Interdisciplinary Reviews: Cognitive Science*, 4(6):609–622.

[Waxman, 1980] Waxman, S. G. (1980). Determinants of conduction velocity in myelinated nerve fibers. *Muscle & Nerve: Official Journal of the American Association of Electrodiagnostic Medicine*, 3(2):141–150.

[Webb and Sidebotham, 2020] Webb, M. and Sidebotham, D. (2020). Bayes' formula: a powerful but counterintuitive tool for medical decision-making. *BJA education*, 20(6):208–213.

[Wessinger et al., 2001] Wessinger, C. M., VanMeter, J., Tian, B., Van Lare, J., Pekar, J., and Rauschecker, J. P. (2001). Hierarchical organization of the human auditory cortex revealed by functional magnetic resonance imaging. *Journal of cognitive neuroscience*, 13(1):1–7.

[White et al., 1986] White, J. G., Southgate, E., Thomson, J. N., Brenner, S., et al. (1986). The structure of the nervous system of the nematode caenorhabditis elegans. *Philos Trans R Soc Lond B Biol Sci*, 314(1165):1–340.

[Wiegrebe et al., 1996] Wiegrebe, L., Kössl, M., and Schmidt, S. (1996). Auditory enhancement at the absolute threshold of hearing and its relationship to the zwicker tone. *Hearing research*, 100(1-2):171–180.

[Wiley et al., 1998] Wiley, T. L., Cruickshanks, K. J., Nondahl, D. M., Tweed, T. S., Klein, R., and Klein, B. E. (1998). Aging and high-frequency hearing sensitivity. *Journal of Speech, Language, and Hearing Research*, 41(5):1061–1072.

[Wilson et al., 2020] Wilson, C. A., Berger, J. I., De Boer, J., Sereda, M., Hall, D. A., and Wallace, M. N. (2020). Using gap-induced inhibition of the post-auricular muscle response as an objective measure of tinnitus in humans. *Acta Scientific Otolaryngology*, 2(12).

[Wu et al., 2023] Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Q.-L., and Tang, Y. (2023). A brief overview of chatgpt: The history, status quo and potential future development. *IEEE/CAA Journal of Automatica Sinica*, 10(5):1122–1136.

[Xiao et al., 2017] Xiao, H., Rasul, K., and Vollgraf, R. (2017). Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*.

[Yang et al., 2024] Yang, D., Zhang, D., Zhang, X., and Li, X. (2024). Tinnitus-associated cognitive and psychological impairments: a comprehensive review meta-analysis. *Frontiers in neuroscience*, 18:1275560.

[Yang et al., 2018] Yang, T., Wang, H., Aziz, S., Jiang, H., and Peng, J. (2018). A novel method of wind speed prediction by peephole lstm. In *2018 International conference on power system technology (POWERCON)*, pages 364–369. IEEE.

[Yasoda-Mohan et al., 2024] Yasoda-Mohan, A., Chen, F., Ó Sé, C., Allard, R., Ost, J., and Vanneste, S. (2024). Phantom perception as a bayesian inference problem-a pilot study. *Journal of Neurophysiology*.

[Yaz et al., 2023] Yaz, F., Büttner, M., Tekin, A. M., Bahşi, İ., and Topsakal, V. (2023). A bibliometric analysis of publications on tinnitus: A study based on web of science data from 1980 to 2020. *The Journal of International Advanced Otology*, 19(2):121.

[Young et al., 1995] Young, E. D., Nelken, I., and Conley, R. A. (1995). Somatosensory effects on neurons in dorsal cochlear nucleus. *Journal of Neurophysiology*, 73(2):743–765.

[Yu et al., 2018] Yu, Y., Herman, P., Rothman, D. L., Agarwal, D., and Hyder, F. (2018). Evaluating the gray and white matter energy budgets of human brain function. *Journal of Cerebral Blood Flow & Metabolism*, 38(8):1339–1353.

[Zeng et al., 2000] Zeng, F.-G., Fu, Q.-J., and Morse, R. (2000). Human hearing enhanced by noise. *Brain research*, 869(1-2):251–255.

[Zenke et al., 2017] Zenke, F., Gerstner, W., and Ganguli, S. (2017). The temporal paradox of hebbian learning and homeostatic plasticity. *Current opinion in neurobiology*, 43:166–176.

[Zenke and Vogels, 2021] Zenke, F. and Vogels, T. P. (2021). The remarkable robustness of surrogate gradient learning for instilling complex function in spiking neural networks. *Neural computation*, 33(4):899–925.

[Zhang and Yoshida, 2024] Zhang, H. and Yoshida, S. (2024). Exploring deep neural networks in simulating human vision through five optical illusions. *Applied Sciences*, 14(8):3429.

[Zhang et al., 2020] Zhang, Y., Wang, Y., and Yang, J. (2020). Lattice lstm for chinese sentence representation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28:1506–1519.

[Zhuang et al., 2020] Zhuang, Y., Cai, M., Li, X., Luo, X., Yang, Q., and Wu, F. (2020). The next breakthroughs of artificial intelligence: The interdisciplinary nature of ai. *Engineering*, 6(3):245–247.

[Zuo et al., 2017] Zuo, H., Lei, D., Sivaramakrishnan, S., Howie, B., Mulvany, J., and Bao, J. (2017). An operant-based detection method for inferring tinnitus in mice. *Journal of neuroscience methods*, 291:227–237.

[Zweig, 1976] Zweig, G. (1976). Basilar membrane motion. In *Cold Spring Harbor symposia on quantitative biology*, volume 40, pages 619–633. Cold Spring Harbor Laboratory Press.

[Zwicker, 1964] Zwicker, E. (1964). "negative afterimage" in hearing. *The Journal of the Acoustical Society of America*, 36(12):2413–2415.

Appendix

List of Papers for cumulative habilitation

(1) Gerum, R. C., & **Schilling, A.** (2021). Integration of leaky-integrate-and-fire neurons in standard machine learning architectures to generate hybrid networks: A surrogate gradient approach. Neural Computation, 33(10), 2827-2852.

(2) Gerum, R., Erpenbeck, A., Krauss, P., & **Schilling, A.** (2023). Leaky-integrate-and-fire neuron-like long-short-term-memory units as model system in computational biology. In 2023 international joint conference on neural networks (IJCNN) (pp. 1-9). IEEE.

(3) Gerum, R. C., Erpenbeck, A., Krauss, P., & **Schilling, A.** (2020). Sparsity through evolutionary pruning prevents neuronal networks from overfitting. Neural Networks, 128, 305-312.

(4) **Schilling, A.***, Tziridis, K.*, Schulze, H., & Krauss, P. (2023). Behavioral assessment of Zwicker tone percepts in gerbils. Neuroscience, 520, 39-45.

(5) Stoll, A., Maier, A., Krauss, P., Gerum, R.*, & **Schilling, A.*** (2023). Coincidence detection and integration behavior in spiking neural networks. Cognitive Neurodynamics, 18(4), 1753-1765.

(6) **Schilling, A.**, Tziridis, K., Schulze, H., & Krauss, P. (2021). The stochastic resonance model of auditory perception: A unified explanation of tinnitus development, Zwicker tone illusion, and residual inhibition. Progress in brain research, 262, 139-157.

(7) **Schilling, A.**, Krauss, P., Hannemann, R., Schulze, H., & Tziridis, K. (2021). Reduktion der Tinnituslautstärke: Pilotstudie zur Abschwächung von tonalem Tinnitus mit schwellennahem, individuell spektral optimiertem Rauschen. HNO, 69(11), 891.

(8) **Schilling, A.**, Sedley, W., Gerum, R., Metzner, C., Tziridis, K., Maier, A., Schulze, H., Zeng, FG, Friston KJ & Krauss, P. (2023). Predictive coding and stochastic resonance as fundamental principles of auditory phantom perception. Brain, 146(12), 4809-4825.

(9) Krauss, P., Metzner, C., Joshi, N., Schulze, H., Traxdorf, M., Maier, A., & **Schilling, A.** (2021). Analysis and visualization of sleep stages based on deep neural networks. Neurobiology of sleep and circadian rhythms, 10, 100064.

(10) **Schilling, A.**, Gerum, R., Boehm, C., Rasheed, J., Metzner, C., Maier, A., Reindl, C., Hamer, H. & Krauss, P. (2024). Deep learning based decoding of single local field potential events. NeuroImage, 120696.

(* contributed equally)

Attached Papers

In the following, the papers, which are part of the cumulative habilitation, are attached in the order shown above and described in the main and conclusion part of the thesis.

LETTER Communicated by Anthony Maida

Integration of Leaky-Integrate-and-Fire Neurons in Standard Machine Learning Architectures to Generate Hybrid Networks: A Surrogate Gradient Approach

Richard C. Gerum
gerum@yorku.ca
Department of Physics and Center for Vision Research, York University, Toronto, Ontario M3J 1P3 Canada

Achim Schilling
achim.schilling@fau.de
Experimental Otolaryngology, Neuroscience Lab, University Hospital Erlangen, 91054 Erlangen, Germany; Cognitive Computational Neuroscience Group at the Chair of English Philology and Linguistics, Friedrich-Alexander University Erlangen-Nürnberg, 91054 Erlangen, Germany; and Laboratoire Neuorsciences Sensorielles et Cognitives, Aix Marseille-University, 13331 Marseille, France

Up to now, modern machine learning (ML) has been based on approximating big data sets with high-dimensional functions, taking advantage of huge computational resources. We show that biologically inspired neuron models such as the leaky-integrate-and-fire (LIF) neuron provide novel and efficient ways of information processing. They can be integrated in machine learning models and are a potential target to improve ML performance. Thus, we have derived simple update rules for LIF units to numerically integrate the differential equations. We apply a surrogate gradient approach to train the LIF units via backpropagation. We demonstrate that tuning the leak term of the LIF neurons can be used to run the neurons in different operating modes, such as simple signal integrators or coincidence detectors. Furthermore, we show that the constant surrogate gradient, in combination with tuning the leak term of the LIF units, can be used to achieve the learning dynamics of more complex surrogate gradients.

To prove the validity of our method, we applied it to established image data sets (the Oxford 102 flower data set, MNIST), implemented various network architectures, used several input data encodings and demonstrated that the method is suitable to achieve state-of-the-art classification performance.

We provide our method as well as further surrogate gradient methods to train spiking neural networks via backpropagation as an open-source KERAS package to make it available to the neuroscience and machine learning community. To increase the interpretability of the underlying

Neural Computation **33, 2827–2852** (2021)
https://doi.org/10.1162/neco_a_01424

effects and thus make a small step toward opening the black box of machine learning, we provide interactive illustrations, with the possibility of systematically monitoring the effects of parameter changes on the learning characteristics.

1 Introduction

The interest in neuroscience-inspired AI has rapidly grown (Hassabis et al., 2017), and there are major reasons for this development. Although traditional machine learning algorithms have been massively improved by the collection of huge data sets (Russakovsky et al., 2015) and the development of modern hardware components (Steinkraus, Buck, & Simard, 2005; Sheng & Zhou, 2017), certain issues remain still unsolved by these algorithms. Up to now, these algorithms are, in contrast to our brain, highly specialized on a given task. We were not yet able to develop algorithms with general intelligence (Shevlin, Vold, Crosby, & Halina, 2019; Pontes-Filho & Nichele, 2019). Our nervous system has the ability to perform sensory tasks with enormous precision, such as the detection of very low stimuli in the eye (Field, Uzzell, Chichilnisky, & Rieke, 2019; Rieke & Baylor, 1998) or very small pressure differences in the ear on the one hand and is able to process and understand complex story plots on the other hand (Mar, 2004; Tenenbaum, Griffiths, & Kemp, 2006). Thus, we do not need huge hardware components but are limited to approximately 10^{11} neurons (Herculano-Houzel, 2009), which perform these tasks in a very efficient way.

Information can be processed faster and more efficiently in the brain, as spiking neural networks can encode data in different spatiotemporal patterns (Thorpe, Delorme, & Van Rullen, 2001; Perkel & Bullock, 1968; Krauss et al., 2018; Gross & Kowalski, 1999). Thus, the brain does not simply count spikes (rate codes) but also exploits the temporal dynamics of these spikes (Koopman, Van Leeuwen, & Vreeken, 2003; Gerstner, 1998; Vreeken, 2003; Brette, 2015), uses spontaneous spiking and neural noise to enhance sensory processing (Schilling et al., 2020; Krauss et al., 2017, 2016), and thus can quickly react to changing input stimuli (Gerstner, 1998). Furthermore, in biological neural networks, spiking neurons can run in different operating modes (e.g., as coincidence detectors; Roome & Kuhn, 2020), which can be changed dynamically by background neural activity (Wolfart, Debay, Le Masson, Destexhe, & Bal, 2005).

Different biologically inspired neuron models have been developed such as the Hodgkin-Huxley and the Fitzhugh-Nagumo model (Hodgkin & Huxley, 1952; Izhikevich & FitzHugh, 2006), but they are rarely integrated in ML applications due to a lack of hardware optimized to simulate spiking neural networks and techniques to train them efficiently (Zenke & Gerstner, 2014).

Nevertheless, biologically inspired leaky-integrate-and-fire (LIF) neurons (Burkitt, 2006) are an interesting target for ML models to potentially improve performance and increase interpretability on the one hand ("neuroscience-inspired AI"; Hassabis, Kumaran, Summerfield, & Botvinick, 2017) and create models for biology on the other hand ("cognitive computational neuroscience"; Kriegeskorte & Douglas, 2018).

The motivation for exploiting the properties of spiking neural networks in ML is emphasized by the fact that established machine learning tools are adapted to work with spiking neuron models. One example is the approach to start from already established LSTM units (Hochreiter & Schmidhuber, 1997) and try to run them in a spiking mode (Pozzi, Nusselder, Zambrano, & Bohté, 2018; Rezaabad & Vishwanath, 2020).

Due to the interesting dynamics, energy efficiency (Lee, Sarwar, Panda, Srinivasan, & Roy, 2020; Kim, Li, & Sejnowski, 2019), and future potential of spiking neural networks, in recent years much effort has been undertaken to train spiking neural networks using supervised algorithms. Some approaches focused on the biological plausibility of the learning procedure (Gilra & Gerstner, 2017), whereas other approaches are based on backpropagation (backpropagation through time (BPTT) for recurrent neural networks) to combine established methods from standard ML with spiking neuron models (Xin & Embrechts, 2001; Huh & Sejnowski, 2018). (For a review on methods see Taherkhani et al., 2020; Tavanaei, Ghodrati, Kheradpisheh, Masquelier, & Maida, 2019.)

Some major points are important for efficient training of spiking neural networks with backpropagation. First, the choice and, especially, the encoding of the training data are not trivial and depend on the problem to be solved. The simplest choice would be to use analog values as the input data for the spiking neural network (for a comparison of methods, see Rueckauer, Lungu, Hu, Pfeiffer, & Liu, 2017). However, as spiking neural networks have interesting temporal dynamics, it could be a good choice to convert the input data into spike trains. This could be done, for example, by so-called latency coding, where each neuron spikes exactly once, with a latency reversely correlated to the analog value the neuron should represent (Zenke & Vogels, 2020; Kheradpisheh & Masquelier, 2019; Bohte, Kok, & La Poutré, 2000; Schrauwen & Van Campenhout, 2004a, 2004b). Another interesting approach is to use Poisson rate coding, where each analog value is translated to a probability of the neuron to spike in each time step and thus results in a spike rate that is proportional to the analog value and a stochastic component (Lee, Delbruck, & Pfeiffer, 2016). In this study, analog input values and Poisson rate coding were used. Furthermore, a novel approach, where one input image dimension serves as the time axis, was implemented. It should be mentioned that there also exist more biologically inspired encoding techniques trying to imitate human vision by simulating saccades (Orchard, Jayawant, Cohen, & Thakor, 2015).

Besides the data encoding techniques, the main issue to be solved when training spiking neural networks via backpropagation is the fact that spiking neurons have a constant derivative of zero except at one point where the derivative is not defined. To account for that, the surrogate gradient method was introduced, which means that the zero-derivative of the spiking neuron is replaced by an arbitrary choice of a differentiable function (Neftci, Mostafa, & Zenke, 2019; Wu, Deng, Li, Zhu, & Shi, 2018; Lee et al., 2020). However, the choice of this function is not straightforward and not unique (Zenke & Vogels, 2020; Neftci, Mostafa, & Zenke, 2019). Among the surrogate gradients are piecewise linear functions (Bellec, Salaj, Subramoney, Legenstein, & Maass, 2018; Bohte, 2011; Esser et al., 2016), derivatives of the sigmoid function (Zenke & Ganguli, 2018), and exponential functions (Shrestha & Orchard, 2018; for a review of different surrogate gradients, see Zenke & Vogels, 2020).

Here, we introduce a very direct way to train LIF neurons using backpropagation by simplifying the surrogate gradient and reducing parameters. Thus, we manually fix the derivative of the LIF neuron to a constant value.

The study is structured as follows:

- We first illustrate and explain the function of LIF neurons and show a recursive formulation so that this neuron model can be integrated in neural networks. In a second step, we introduce our method of setting the derivative of the step function (Heaviside) to one in order to train LIF units with backpropagation through time. Thus, we show explicitly how the derivative can be propagated through time and several neural network layers.
- We show that the neurons can be trained on different operating modes, starting from a neuron simply summing up the input values (integrator), to neurons responding just to the coincidence of multiple input spikes (coincidence detector).
- We prove the validity of our approach to combine an LIF neuron layer with classical deep learning architectures (LSTM) by the application of the method in hybrid neural networks trained on an image classification task using the Oxford 102 flower data set (Nilsback & Zisserman, 2008).
- The study comes with interactive versions of most of the figures, which help readers gain a better understanding of the mechanisms within spiking neural networks (https://rgerum.github.io/paper_spiking_machine_intelligence).
- Finally, we compare our method to existing methods by building networks to classify standard image data sets. We provide a KERAS package (https://github.com/rgerum/tf_spiking) that can be used to integrate LIF unit layers in standard machine learning architectures and to train them with different surrogate gradient methods.

2 Methods

All simulations were run on a standard desktop PC equipped with a Nvidia TitanXp GPU device. The simulations were written in Python using KERAS (Chollet, 2018) and Tensorflow (Abadi et al., 2015) for ML and NumPy (Walt, Colbert, & Varoquaux, 2011) for further evaluations and interfaces. The visualization of the data was done in Javascript using the D^3-library (Bostock, Ogievetsky, & Heer, 2011) and in Python using Matplotlib (Hunter, 2007) and Pylustrator (Gerum, 2020). Thus, we provide interactive plots in an open github-repository. These interactive plots should provide a deeper understanding of the mechanisms described in the letter. A link to the repository is provided in the figure captions. Furthermore, we provide a KERAS package, which can be used to integrate LIF unit layers in standard machine learning architectures and train these networks via backpropagation using different surrogate gradient methods.

3 Results

3.1 Leaky-Integrate-and-Fire Neurons. As described above, the LIF neuron model is a simple spiking neuron model based on one single differential equation. The idea is that the neuron sums up all input currents, increases its membrane potential, and produces a spike if a certain threshold is reached (see Figure 1). The leak term causes a continuous decrease of the membrane potential and thus prevents long-range correlations.

The LIF neuron's (Koch & Segev, 1998) membrane potential V_m is described by the following differential equation:

$$I(t) - \frac{V_\mathrm{m}(t)}{R_\mathrm{m}} = C_\mathrm{m} \cdot \dot{V}_\mathrm{m}(t), \tag{3.1}$$

with the input current $I(t)$, the membrane resistance R_m, and the membrane capacity C_m.

When the membrane potential exceeds a threshold V_th, a spike in the form of a delta function $\delta(t)$ is emitted and the membrane potential is reset to 0. To simulate the response of the LIF neuron, this differential equation has to be integrated. The standard integration method is the Euler integration (Atkinson, 1989), which we used for this approach. First, the differential equation is solved for $\dot{V}_\mathrm{m}(t)$:

$$\dot{V}_\mathrm{m}(t) = \frac{I(t)}{C_\mathrm{m}} - \frac{V_\mathrm{m}(t)}{R_\mathrm{m} C_\mathrm{m}}. \tag{3.2}$$

This differential equation can be reformulated in a recursive manner (for one Euler step),

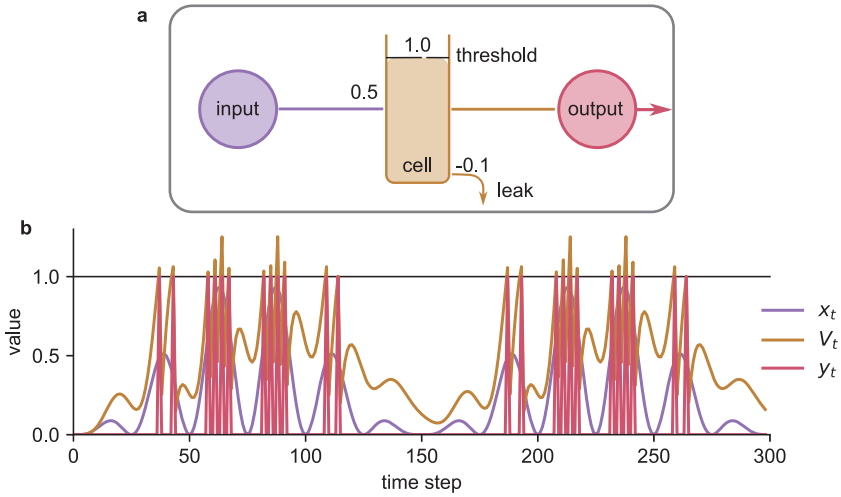

Figure 1: The response of the LIF neurons. (a) The data flow in an LIF unit. (b) The response of the LIF unit to a given input signal. The cell (V_t, orange) integrates the input (x_t, purple) until the internal state exceeds the threshold (gray line). Then it outputs a spike (y_t, red). The leaky term lets the cell state decay over time. With the parameters: $w_{\text{input}} = 0.5$, $w_{\text{leak}} = 0.1$, $V_{\text{thresh}} = 1.0$. For the interactive version, see https://rgerum.github.io/paper_spiking_machine_intelligence/#lif_unit.

$$V_{t+1} = V_t + \left(C_{\text{m}}^{-1} \cdot x_t - V_t \cdot R_{\text{m}}^{-1} C_{\text{m}}^{-1}\right) \cdot \Delta t, \tag{3.3}$$

with the time step delta Δt and the $I(t)$, now renamed x_t. We extend the update equation to include the spiking when the threshold has been reached and the resetting of V_{t+1} after the spike,

$$\tilde{V}_{t+1} = V_t + \left(C_{\text{m}}^{-1} \cdot x_t - V_t \cdot R_{\text{m}}^{-1} C_{\text{m}}^{-1}\right) \cdot \Delta t, \tag{3.4}$$

$$y_{t+1} = \Theta(\tilde{V}_{t+1} - V_{\text{thresh}}), \qquad \text{spiking} \tag{3.5}$$

$$V_{t+1} = V_{\text{m}} \cdot \Theta(-\tilde{V}_{t+1} + V_{\text{thresh}}), \qquad \text{resetting} \tag{3.6}$$

where $\Theta(x)$ is the Heaviside step function. The parameters can be renamed as follows:

$$C_{\text{m}}^{-1} \Delta t = w_{\text{input}}, \tag{3.7}$$

$$R_{\text{m}}^{-1} C_{\text{m}}^{-1} \cdot \Delta t = w_{\text{leak}}. \tag{3.8}$$

Without loss of generality, V_{thresh} can be fixed to 1, as the scaling can be absorbed in w_{input}. The update rule of the LIF unit can be summarized

as follows:

$$V_t = w_{\text{input}} \cdot x_t + (1 - w_{\text{leak}}) \cdot V_{\text{t}-1} \cdot \Theta(V_{\text{thresh}} - V_{\text{t}-1}), \tag{3.9}$$

$$y_t = \Theta(V_t - V_{\text{thresh}}). \tag{3.10}$$

These equations can be used to analytically calculate the firing rates r of the LIF neurons (for calculation, see supplementary Figure S1). The firing rates are an important property for many ML algorithms (Dominguez-Morales et al., 2016):

$$r = 1/\text{ceil}\left(\frac{\ln\left(1 - \frac{V_{\text{thresh}}}{I} \cdot \frac{w_{\text{leak}}}{w_{\text{input}}}\right)}{\ln(1 - w_{\text{leak}})} - 1\right). \tag{3.11}$$

Here, ceil(x) denotes the ceiling of a number, that is, rounding up to the closest integer.

3.2 Deep Learning with LIF Neurons. We show how the LIF neurons can be integrated in standard ML applications for image classification. For the following analysis, we use LIF layers with only one recurrence, meaning the inner state (potential) of the LIF unit, which is gated by the output (see Figure 2). This is a simple circuit, which contains, in contrast to previous studies (see Zenke & Vogels, 2020), no recurrences projecting back to the input x_t. Thus, more complex recurrences could be achieved by stacking several LIF layers.

3.2.1 Calculation of the Gradient and Backpropagation through Time. The standard method to supervisedly train neural networks with supervised training on a classification task is to minimize a loss function $L(y_{\text{out}}, y_{\text{desired}})$. It is a measure of the dissimilarity between the desired output and the output of the neural network y_{out} calculated by forward propagation. For example in a classification task with a softmax output, the loss function usually is the cross-entropy. This loss function is minimized using a gradient descent algorithm. The gradient descent works by adding the negative gradient multiplied with the learning rate γ to the weights, which have to be optimized:

$$\Delta W = -\gamma \cdot \frac{\mathrm{d}L}{\mathrm{d}W}. \tag{3.12}$$

To illustrate the calculation of the gradient, we use an example architecture, consisting of two fully connected time-distributed layers and a LIF layer in between (see Figure 3).

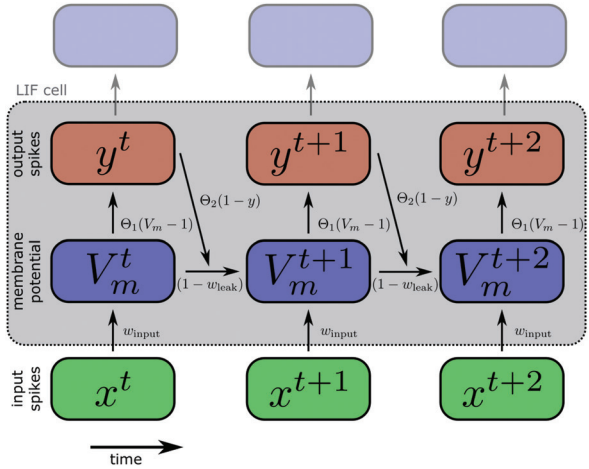

Figure 2: Scheme of forward pass through LIF neuron over time. At each time step, the input x_t multiplied by the weight ω_{input} is added to the membrane potential V_m^t. If no spike occurred in the previous time step, the membrane potential of the the previous state, multiplied by $(1-\omega_{\text{leak}})$, is added to this input. If the membrane potential exceeds the threshold, a 1 (=spike) is the output y^t; if not, a 0 (=no spike) is the output y^t.

Figure 3: Example neural network architecture used to illustrate the gradient descent algorithm.

To calculate the update of the weights W_1, we have to calculate the gradient $\frac{\mathrm{d}L}{\mathrm{d}W_1}$:

$$\frac{\mathrm{d}L}{\mathrm{d}W_1} = \frac{\partial L}{\partial y_{\text{out}}} \cdot \frac{\partial y_{\text{out}}}{\partial y} \cdot \frac{\partial y}{\partial x} \cdot \frac{\mathrm{d}x}{\mathrm{d}W_1}. \tag{3.13}$$

The term $\frac{\partial y}{\partial x}$ is the derivative of the LIF output as a function of the LIF input. If this derivative is zero, the weights of the first layer W_1 cannot be trained:

$$V_t = w_{\text{input}} \cdot x_t + (1 - w_{\text{leak}}) \cdot V_{\text{t}-1} \cdot \Theta_2(V_{\text{thresh}} - V_{\text{t}-1}), \tag{3.14}$$

$$y_t = \Theta_1(V_t - V_{\text{thresh}}). \tag{3.15}$$

However, this is exactly the case, as the LIF equations contain two Θ functions. To better reference them, we call the Θ function for the generation of the output spike Θ_1 and the Θ function for resetting the membrane potential Θ_2. As the Θ function is a completely flat function (except at zero), its gradient is zero at all points, therefore reducing all gradients to zero. Thus, no gradient can enter or pass the LIF cell. One possibility to overcome this problem would be to smooth the Θ functions. But a more elegant solution, which does not affect the forward pass, is to redefine the gradient of the LIF unit. As LIF neurons have self-connections (recurrencies), backpropagation through time has to be applied. Thus, the influence of the preceding time steps on the output error has to be taken into account. Therefore, the derivative for all previous time steps $t - i$ has to be calculated.

The derivative with respect to the inputs can be written as

$$\frac{\partial y_t}{\partial x_t} = \frac{\partial y_t}{\partial V_t}\frac{\partial V_t}{\partial x_t} = \Theta_1'(V_t - V_{\text{thresh}}) \cdot w_{\text{input}}, \tag{3.16}$$

$$\begin{aligned}\frac{\partial y_t}{\partial x_{t-1}} &= \frac{\partial y_t}{\partial V_t}\frac{\partial V_t}{\partial V_{t-1}}\frac{\partial V_{t-1}}{\partial x_{t-1}} \\ &= \Theta_1'(V_t - V_{\text{thresh}}) \cdot (1 - w_{\text{leak}}) \\ &\quad \cdot \left[\Theta_2(V_{\text{thresh}} - V_{t-1}) + V_{t-1} \cdot \Theta_2'(V_{\text{thresh}} - V_{t-1})\right] \cdot w_{\text{input}}.\end{aligned} \tag{3.17, 3.18}$$

If we define $\Theta_2'(x) = 0$, then the expression for an arbitrary derivative for a past x is

$$\frac{\partial y_t}{\partial x_{t-n}} = \Theta_1'(V_t - V_{\text{thresh}}) \cdot w_{\text{input}}(1 - w_{\text{leak}})^n \prod_{i=1}^{n} \Theta_2(V_{\text{thresh}} - V_{t-i}). \tag{3.19}$$

The crucial part here is the function Θ_1, which prevents any gradient from passing. Therefore, we redefine the gradient of Θ_1 to be 1 and keep the gradient of Θ_2 as zero. The term $\prod_{i=1}^{n} \Theta_2(V_{\text{thresh}} - V_{t-i})$ becomes zero when the number of past time steps i is bigger than the number of time steps since the last output spike. This property makes sense as the errors should not be propagated further than the most recent released spike.

In summary, the gradient enters the cell (as $\Theta_1' = 1$) and propagates to all input units that contributed to the current spike (see also Figure 4) but does not penetrate to inputs that contributed to the previous spike (as $\Theta_2' = 0$). We have shown that our gradient definition allows errors to pass the LIF neurons to preconnected layers. However, for the backpropagation procedure, the gradients with respect to the LIF parameters are also needed if they are supposed to be trainable (w_{input}, w_{leak}, for complete gradients; see the supplement).

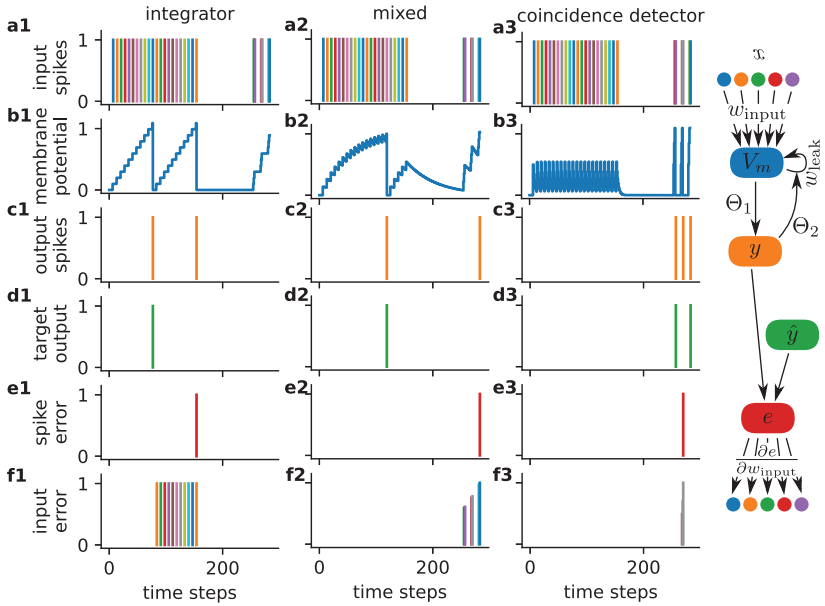

Figure 4: Changing operating modes of LIF units from integrator to coincidence detector by leak term adaptations. The figure shows that an adaptation of the leak term, w_{leak} can change the response properties of the LIF unit (column 1: $w_{\text{leak}} = 0$, integrator; column 2: intermediate operating mode; column 3: $w_{\text{leak}} = 0.99$, coincidence detector). For this analysis, 200 input units and one output unit were used as model system. a1–a3: Input spikes, which are all equally weighted (w_{input}). The first part of the input spike train consists of equidistant spikes, whereas the second part consists of three blocks of spikes with low time intervals between them (input spike bursts). b1–b3: membrane potential of the output neuron; The integrator (b1) simply sums up the input spikes. The coincidence detector (b2) only spikes when "spike bursts" are presented. c1–c3: output spikes. The integrator spikes two times during the first input spike train (c1), whereas for higher w_{leak} (c2), only one spike is released, and for higher w_{leak} (coincidence mode), no output spike is released. The input bursts lead to a spike in the coincidence mode, whereas no spike is produced in the integrator regime. d1–d3: Target output of the system and according error (e1–e3). f1–f3: Backpropagation of error through time. For the integrator, the error is constantly backpropagated in time, whereas for the coincidence, the error quickly decreases. Thus, w_{leak} can be used to tune the operating mode as well as the learning behavior.

3.2.2 The Influence of the Leak Term on the Operation Mode and Training Dynamics. The leak term w_{leak} influences the operation mode and the training dynamics of the LIF units. LIF units can simply sum up inputs until they

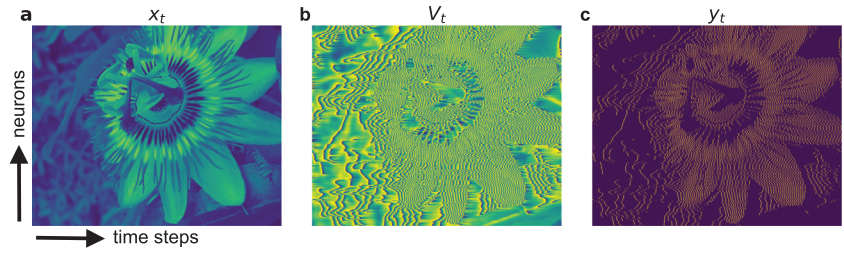

Figure 5: Image processed column-wise by a LIF layer. The figure shows the input (a), internal state (b), and output (c) of the LIF layer. For the interactive version, see https://rgerum.github.io/paper_spiking_machine_intelligence/#lif_image_processing

reach the threshold and release a spike (integrator; see Figure 4 a1, b1, and c1), or they only spike if the time difference of the inputs is small (coincidence detectors; see Figure 4 a3, b3, and c3). In contrast to an integrator, a coincidence detector spikes only if spikes with very short time intervals between them are fed to it (see Figure 4 a3). As the internal state quickly decreases, several input spikes in a short period of time are needed to reach the threshold and to cause a spike release (see Figure 4 b3). As the operation mode is exclusively changed by altering the value of w_{leak}, intermediate operating modes can easily be created (see Figure 4 a1, b1, and c1). However, not just the operating mode of the LIF units is changed by w_{leak}. The training dynamics is also influenced by w_{leak} as the error, backpropagated through time, decays with $(1 - w_{\text{leak}})^n$ (see equation 3.19), that is, it decays more quickly for higher leak term values (see Figure 4 f1–f3). For these reasons, treating w_{leak} as a trainable parameter in the learning procedure is an interesting approach to generate neural networks with a self-organized specialization of spiking neurons.

3.2.3 Analysis of Basic Computations of LIF Layers in Simple Hybrid Network Architectures.

LIF Neurons and Multidimensional Data. When the LIF neurons are applied to multidimensional data such as an image, an efficient representation is needed to optimize LIF neurons for our standard hardware and software architectures. The images fed to the LIF units have N rows and M columns ($N \times M$). However, we regard the image as a serial data set, where the time axis corresponds to the x-axis of the image. Thus, the input of the LIF neurons consists of an N-dimensional vector for each of the M time steps. Therefore, V_t and y_t are also N-dimensional vectors. Thus, when an image is fed to an LIF unit as described above (input image; see Figure 5a), it is transformed to voltage fluctuations in the LIF unit (see Figure 5b) and output spiking patterns (see Figure 5c).

The model can be extended by the use of several LIF units with different w_{input} and w_{leak} (see equation 3.10). This would make w_{input} and w_{leak} a vector instead of a scalar and the scalar product would transform into a tensor product. This is an efficient representation, which can easily be optimized for and run on GPUs.

Note that for the following analysis, we have chosen to use one image dimension as a time axis to present the data in a serial format. Using spiking neurons to analyze serial data such as speech is promising, as spiking neurons encode information in serial spike trains. Nevertheless, in section 3.4, where we compare our method with earlier studies, we use simple analog inputs and Poisson rate coding (see Lee et al., 2016).

The LIF unit implementation described above is embedded in two hybrid neural networks out of LSTM layers and a softmax layer. These hybrid neural networks are applied to an image classification task, where 10 different flower species should be identified. The first network is used to show the effects of the LIF layers when the three color channels are fed to the layer separately, although this simple architecture does not lead to state-of-the-art classification accuracies. The second network allows mixing of the color channels within the LIF layer and thus the detection of more sophisticated features. Nevertheless, both networks are exclusively used to visualize the backpropagation procedure and the validity of the surrogate gradient method. Fine-tuned spiking neural networks for image classification and a detailed comparison to other state-of-the-art approaches are provided in section 3.4.

Obviously, the LSTM units add further complex effects and trainable parameters to the model, but they are used as a simple method to integrate the spikes and transform them into a class label. The used data set is a sub–data set of the 102 category flower data set (Nilsback & Zisserman, 2008) with only 10 categories. Thus, the LIF units are supposed to preprocess and compress the images of the different blossoms.

Network Architecture 1. The classification task on different blossoms is based on the 10 most occurring flower species (see supplementary Figure S2) of the 102 category flower data set (Nilsback & Zisserman, 2008). The network consists of one LIF layer with three different LIF unit types (3×2 = trainable parameters) compressing the colored images of 500×400 pixels.

Each of the three LIF unit types gets exactly one color channel of the input images. In each time step, each LIF unit type receives one column of one color channel of the image as input and returns a spike vector still representing the same color channel. Thus, the view that the LIF layer consists of 1200 individual LIF neurons of three different sorts (in analogy to network architecture 2) with a 1D spike train output is equivalent, although for programming reasons, the tensor notation was used in the tensor flow (Abadi et al., 2015) implementation.

Thus, the 8 bit images are compressed by a factor of 8 as each 8 bit integer is replaced by a Boolean number (spike, no spike). The compressed spike

Table 1: Network Architecture 1.

Layer (type)	Output Shape	Parameters #
LIF-layer	(None, 400, 500, 3)	6
Reshape	(None, 500, 1200)	0
LSTM layer	(None, 500, 30)	147,720
Dropout	(None, 500, 30)	0
Time distributed dense	(None, 500, 30)	930
Dropout	(None, 500, 30)	0
Softmax	(None, 500, 10)	310

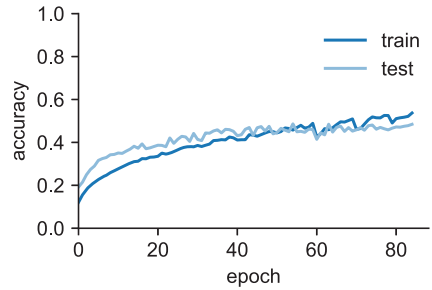

Figure 6: Accuracy of training a network with one LIF layer. The training accuracy is shown in dark blue and the test accuracy in light blue.

data are fed to an LSTM layer (30 units) connected to a fully connected output layer with softmax activation (10 units; see Table 1). As a loss function, the categorical cross-entropy is used. The parameters w_{input} and w_{leak} of the LIF units, as well as the LSTM and softmax parameters, are trained via backpropagation.

The training procedure is stopped after 30 epochs of no improvement of the test accuracy (early stopping). The classification accuracy for one image (see Figure 6) is defined as the average probability value of the correct label during the image presentation, a very conservative estimator for the accuracy. The overall test accuracy (all test images) achieves a value of over 40%, whereas the chance accuracy is 10% (10 categories). This proves that the LIF neurons can be trained via backpropagation so that they operate in a sophisticated parameter range.

Note that the data set does not allow the neural network to train on trivial features such as the color of the flower because the data set contains different color variants of the same flower species. The LIF units compress the image; nevertheless, the shape of the flowers can still be seen in the spike patterns (see Figure 7) due to the chosen network architecture. The model described here proves that spiking layers can be trained on a classification task.

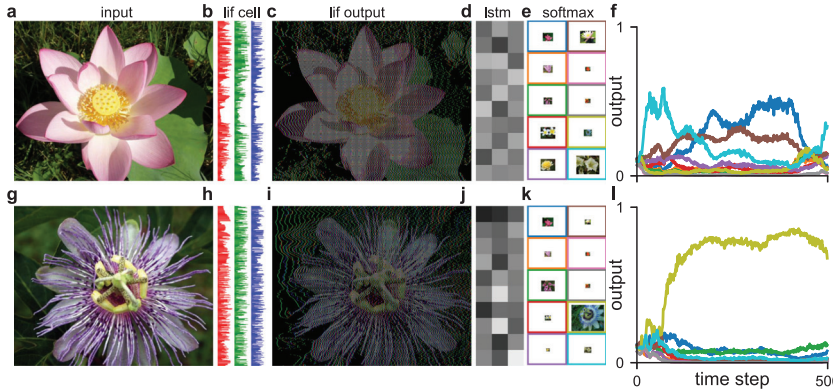

Figure 7: Spiking network processes images. The network takes images as input (a, g) and applies an LIF layer (b, h) to generate spike trains for each row and color channel (c, i). (Note that the x-dimension of the image serves as the time axis, and thus the input is a time series of the image columns. Thus, after 500 time steps, the complete image has been presented to the LIF layer.) The spike trains from the LIF layer are fed into an LSTM layer (d, j), which is followed by a softmax layer (e, k). The softmax layer predicts the category of the image (f, l). The three different color bars (lif cell) represent the internal state of the LIF units for the three different color channels. The spike data produced by this LIF layer are shown under the heading lif output. The activation of the LSTM layer is shown as a color map (lstm). The category probability calculated through the softmax layer is represented by the size of the category images (softmax). For the interactive version, see https://rgerum.github.io/paper_spiking_machine_intelligence/#lif_network1.

Network Architecture 2. We further provide evidence that a classification network can also be trained when there is a fully connected layer preconnected to the LIF layer. Here, the LIF layer consists of 1200 LIF units, generating output spike trains (1D Boolean scalar spike train). Each LIF unit receives a weighted sum of 3×400 values as input (3 color channels and 400 as the images consist of 400 rows). The x-coordinate (500 pixels width of the image) is the time axis.

In contrast to the architecture above (network 1), with only 6 trainable parameters except LSTM and softmax layer (3 LIF neuron types, 1200 LIF neurons), this network has 1,441,200 trainable parameters for the preconnected, fully connected layer and 2400 trainable parameters (w_{input}, w_{leak}) for the 1200 individual LIF units (see Table 2). The fact that the gradient can pass the LIF units can be seen when analyzing the accuracy as a function of the epochs (learning curve). The accuracy is higher for the LIF units with the activation function gradient ($\Theta_1' = 1$) set to one (blue curves). This is true for training as well as test accuracy.

Table 2: Network Architecture 2.

Layer (type)	Output Shape	Parameters #
Reshape	(None, 500, 1200)	0
Time distributed dense	(None, 500, 1200)	1,441,200
LIF-layer	(None, 500, 1200)	2400
LSTM layer	(None, 500, 30)	147,720
Dropout	(None, 500, 30)	0
Time distributed dense	(None, 500, 30)	930
Dropout	(None, 500, 30)	0
Softmax	(None, 500, 10)	310

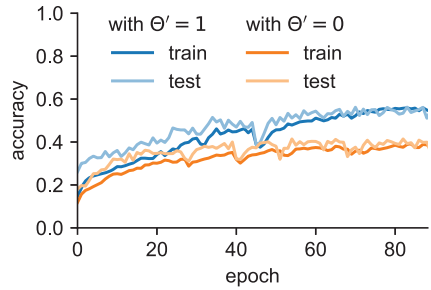

Figure 8: Accuracy of training a network with included LIF layer. The orange curves show the accuracy during training without manually setting the gradient of $\Theta'_1 = 1$ (disappearing gradient network, dark orange: training accuracy, light orange: test accuracy). The gradient cannot pass the LIF units. The blue curves in contrast show that the algorithm is able to train the LIF layer as well as the preconnected dense layer (dark blue: training accuracy, light blue: test accuracy).

The test accuracy for the network, where the gradient can pass the LIF layer, is increased by more than 10% compared to the disappearing gradient network (orange curve) and rises up to a value of approximately 55% (see Figure 8). Nevertheless, the accuracy of the disappearing gradient network is not at chance level, as the random connections lead to usable features for the higher layers (LSTM layers), an effect that was shown in biology as well as in computer science (see Dasgupta et al., 2017; Yang, Schilling, Maier, & Krauss, 2021). In the following, we show that the output of the LIF layer significantly changes over the training epochs. Thus, the time course of the output of a LIF unit for one certain image is shown in Figure 9.

The spike patterns clearly change during the training process, and the spike occurrences decrease (see Figure 9). This effect, called sparse coding (Olshausen & Field, 1997), is correlated with a higher performance of the network and shows that the gradient can pass the LIF layer. Furthermore,

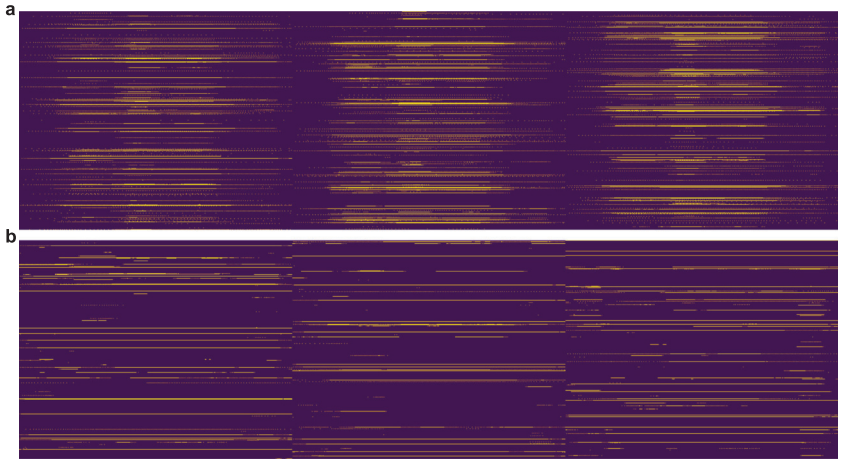

Figure 9: Training of a network with LIF units. (a) The spiking patterns of the LIF units before training (epoch 0, each line in the three blocks represents one of the 1200 neurons, the x-axis is the 500 time points, yellow represents a spike, and purple represents no spike). (b) The spiking patterns of the LIF units after training (epoch 172). The input of one LIF unit in each time step is a weighted sum of the rows of the image (each image has 400 rows and three color channels, input: weighted sum of 1200 values). It can be seen that the output spike patterns change during training (a, epoch 0; b, epoch 172) and that the spike density is reduced. Thus, the network develops a sparse coding of the input image. For the interactive version, see https://rgerum.github.io/paper_spiking_machine_intelligence/#lif_network2.

sparse coding of input stimuli is a basic principle in biological neural networks and here emerges automatically due to our training method.

3.3 Increase Learning Performance by Cutting Negative Inner States. The robustness of the learning procedure could be increased by not allowing negative inner states (potential). As LIF units are highly asymmetrical (only at a positive threshold is a spike produced), too negative potentials could decrease the LIF unit dynamics. Thus, some units could be stuck in a very negative regime and do no longer spike. This problem could be solved by simply cutting negative values, which means mathematically feeding the original potential V_t to a rectified linear function (ReLu, 3.20) (see equation 3.21).

$$\mathbf{ReLu}(x) = \begin{cases} 0 & \text{for } x \leq 0 \\ x & \text{for } x > 0, \end{cases} \tag{3.20}$$

$$V_t = \mathbf{ReLu}\left(w_{\text{input}} \cdot x_t + (1 - w_{\text{leak}}) \cdot V_{\text{t}-1} \cdot \Theta_2(V_{\text{thresh}} - V_{\text{t}-1})\right). \tag{3.21}$$

This procedure could increase the learning performance. The use of the ReLu function can be interpreted as an extra and very strict leak term forcing the neuron to get back to resting potential (in our simple case, 0). If the V_t is in the positive regime, the dynamics is the same as described above. The gradient (see equation 3.22) for backpropagation in time gets an additional term $\prod_{j=1}^{n} \Theta\left(w_{\text{input}} \cdot x_{t-j} + (1 - w_{\text{leak}}) \cdot V_{t-j} \cdot \Theta_2(V_{\text{thresh}} - V_{t-j})\right)$, which sets the derivative to zero, if the potential should become negative in any time step. The technique is very strict, but there are further possibilities to secure that there is a lower limit for the inner state (potential), thus as using a smoother function than a ReLu or shifting the ReLu to more negative values, to allow for some negative potentials:

$$\frac{\partial y_t}{\partial x_{t-n}} = \Theta_1'(V_t - V_{\text{thresh}}) \cdot w_{\text{input}}(1 - w_{\text{leak}})^n \prod_{i=1}^{n} \Theta_2(V_{\text{thresh}} - V_{t-i}) \cdot \prod_{j=0}^{n} \Theta\left(w_{\text{input}} \cdot x_{t-j} + (1 - w_{\text{leak}}) \cdot V_{t-j} \cdot \Theta_2(V_{\text{thresh}} - V_{t-j})\right). \quad (3.22)$$

In the next step, we combine the methods described above and use them to classify the MNIST (LeCun et al., 1995) data set.

3.4 Comparison of the Method with State-of-the-Art Techniques for Image Classification. The training technique works well for serial input data. In the following, we compare the method to existing methods by applying it to the standard MNIST data set. Furthermore, we use two established encoding methods of the input data: simple analog input encoding and the so-called Poisson rate coding (Lee et al., 2016).

For the simple analog coding, the analog pixel intensities are fed to the LIF layer and are presented for 100 time steps (see Figure 10 INPUT a). In contrast to that, the Poisson rate coding converts the pixel values to spike probabilities. Thus, one 784-pixel-MNIST-image results in 784 spike trains with an average spike rate proportional to the according pixel intensity (see Figure 10 INPUT c). This means that the image is converted to a serial data set. An increase in the number of calculated time steps (in this case 100), corresponds with the possibility to perform a more detailed reconstruction of the original image by simply evaluating the spike counts of the individual pixels/neurons (see Figure 10 INPUT c).

We feed these 784 spike trains to a single fully connected trainable LIF layer (see Figure 10 LIF d, e, f) and the output spikes of this hidden layer to an accumulation layer (see Figure 10 OUTPUT g). The accumulation layer transforms the spike trains to an analog signal (see Figure 10 OUTPUT h), which then can be used to calculate classification accuracies. The accumulation layer is a simple layer, which integrates the input values. The output of

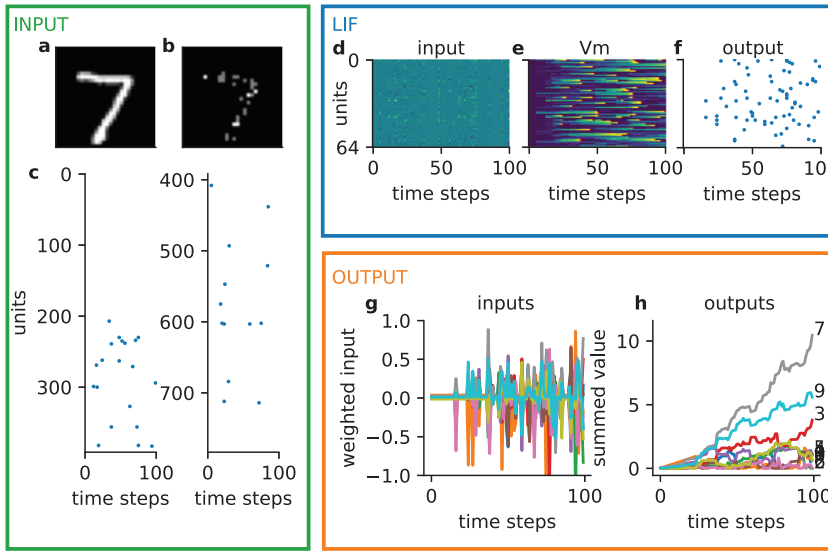

Figure 10: Classification of MNIST data set. An overview of how a spiking neural network is trained on classifying the MNIST data set, which is encoded using Poisson rate coding. The input image (a) is translated to a spike train (c), with each pixel of the image regarded as one spiking neuron. The probability of a spike is proportional to the intensity of the pixel. Thus, as more time steps are calculated, the better the image can be reconstructed by simply counting spikes (b). The spike train is fed to a trainable LIF layer (d–f). The spike train output of the LIF layer is fed to a final layer, which accumulates the inputs (an LIF without a leak and without spiking). The membrane potential of these units representing the correct number produces the highest value.

the accumulation layer at the last time step (100) is then fed to one softmax layer, which assigns the label probabilities to the input images.

The analog input coding leads to higher classification accuracies compared to the Poisson rate coding (see Figures 11a and 11b). The learning curves prove that the surrogate gradient is suited for training with backpropagation, as the validation accuracies are significantly increased in networks with an intermediate LIF layer compared to no hidden layer (see Figure 11a inlay). The classification accuracies saturate for larger hidden LIF layers; however, more LIF units lead to significantly increased accuracies. Furthermore, the Poisson rate coding leads to decreased accuracies compared to the analog input but seems to prevent the network from overfitting through the addition of noise through the Poisson process (training accuracy = validation accuracy). The decreased accuracy could be explained by the fact that the Poisson rate coding leads to an information loss, as it is

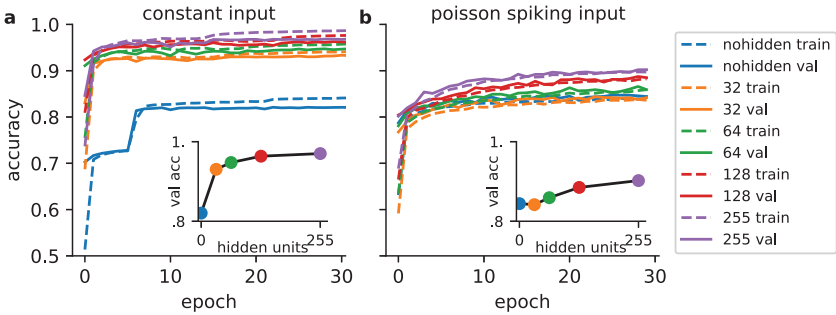

Figure 11: Learning curve for the MNIST data set. Training (dashed lines) and validation (solid lines) accuracies over 30 epochs of training on MNIST images, for (a) the images provided as a constant analog input and (b) the images provided as Poisson distributed spike trains. Final validation accuracies (insets) increase with the number of hidden LIF units (no LIF units, 32, 64, 128, 255). Training on constant inputs results in higher accuracies than training on Poisson spike trains. Overfitting is higher for analog inputs.

impossible to reconstruct all nuances of the image by only looking at 100 time steps of spiking activity (see Figure 10 INPUT b). In summary, we could show that spiking neural networks in combination with a simple surrogate gradient are sufficient for image classification and interesting targets for further classification tasks, such as speech processing.

4 Discussion

4.1 Summary. In this study, we have done an in-depth analysis of the math behind and the applicability of LIF neurons in machine learning.

We show that LIF units can be embedded in standard neural network models and can be trained with backpropagation using a simple surrogate gradient. The surrogate gradient is defined by simply fixing the derivatives of the activation functions ($\Theta'_1 = 1$, $\Theta'_2 = 0$).

This surrogate gradient contains no further arbitrary parameters, which have to be optimized to increase learning performance, in contrast to other approaches, where the surrogate gradient follows a self-customized function (Neftci et al., 2019), which vanishes in time. We show that the simple definition allows for the same training dynamics as more complex surrogate gradients.

Therefore, we show that tuning the leak parameter w_{leak} changes the spiking behavior of the neuron. Thus, a neuron that only integrates the input signal could be tuned to serve as coincidence detector by simply increasing the leak term (see Figure 4a). Additionally, the leak term influences the gradient, so that a higher leak term corresponds to a faster decrease of the

backpropagated error in time (see Figure 4f). This tunability of the learning dynamic, as well as spiking behavior of the neurons by simply changing the leak term, could be an interesting target to further establish the use of spiking neural networks for machine learning tasks. Thus, a trainable leak term could lead to neural networks consisting of different neuron types, which serve as simple signal integrators or coincidence detectors.

We increased the learning performance by cutting negative inner state values by adding a rectified linear function (ReLu) and prove the validity of the method by applying it to the Oxford 102 flower data set (Nilsback & Zisserman, 2008) as well as the MNIST data set (LeCun et al., 1995). Furthermore, we used different input value encodings such as simple analog input, serializing the analog input, or Poisson rate coding (see Lee et al., 2016). All techniques used in this letter are provided as an open-source KERAS (Chollet, 2018) add-on (https://github.com/rgerum/tf_spiking).

4.2 Limitations. The aim of our study was to establish a simple method to train neural networks consisting of LIF neurons using backpropagation, although we are aware of the fact that we cannot compete with the state-of-the-art performance of other architectures optimized for image classification (see Xia, Xu, & Nan, 2017; Feng, Wang, Zha, & Cao, 2019; Qin, Xi, & Jiang, 2019). Thus, the big advantages of spiking neural networks probably lie in the analysis of serial data such as speech. In follow-up studies, the use of complex speech or music data sets could give novel insights into how these kinds of data are efficiently processed in the brain and how these principles could be translated to machine learning applications.

However, to achieve higher performance, the recurrent neural networks have to be optimized for the used hardware components or the other way around (Bhuiyan, Pallipuram, Smith, Taha, & Jalasutram, 2010).

5 Conclusion

Despite these limitations, our methods provide a direct way to train spiking neural networks with backpropagation and a minimum set of parameters. The interactive visualizations are provided to create a deeper understanding of the computational mechanisms and are a small step towards explainable AI. To gain a deep understanding on how machine learning algorithms work, that is, to tackle the black box problem (also opacity debate), has become an important issue in AI research (Castelvecchi, 2016; De Laat, 2018).

Additionally, the results of this study have the potential to provide novel insights in the function of biological neural networks (Schemmel, Grubl, Meier, & Mueller, 2006; Jin et al., 2010). Thus, we think that the application of the analysis techniques developed for untrained or randomly connected neural networks such as stability analysis (Liapunov exponent) or motif distribution analysis (see Bertschinger & Natschläger, 2004; Krauss, Zankl,

Schilling, Schulze, & Metzner, 2019; Krauss, Prebeck, Schilling, & Metzner, 2019; Krauss, Schuster et al., 2019), can be applied to the spiking neural networks trained with backpropagation to gain new insights in brain dynamics and function.

The implementation of biological principles in machine learning, such as sparsity (Gerum, Erpenbeck, Krauss, & Schilling, 2020) or spiking properties, can help improve the performance of machine learning algorithms. Furthermore, we hypothesize that spiking neural networks unfold their full potential in tasks with serial data such as speech or music classification (Schilling et al., 2020). Thus, not only pure spiking neural networks but also hybrid networks, where the spiking layer performs efficient data encoding and an additional layer performs the signal integration, are an interesting starting point for developing novel machine learning techniques.

Acknowledgments

We thank Nvidia for the donation of two Titan Xp GPU devices. This study was supported by the Deutsche Forschungsgemeinschaft (DFG) (grant: SCHI 1482/3-1, 451810794) to A.S.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., . . . Zheng, X. (2015). *TensorFlow: Large-scale machine learning on heterogeneous systems*. Software available from tensorflow.org.

Atkinson, K. E. (1989). *An introduction to numerical analysis.* New York: Wiley

Bellec, G., Salaj, D., Subramoney, A., Legenstein, R., & Maass, W. (2018). Long short-term memory and learning-to-learn in networks of spiking neurons. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, & R. Garnett (Eds.), *Advances in neural information processing systems*, *31* (pp. 787–797). Red Hook, NY: Curran.

Bertschinger, N., & Natschläger, T. (2004). Real-time computation at the edge of chaos in recurrent neural networks. *Neural Computation*, *16*(7), 1413–1436.

Bhuiyan, M. A., Pallipuram, V. K., Smith, M. C., Taha, T., & Jalasutram, R. (2010). Acceleration of spiking neural networks in emerging multi-core and GPU architectures. In *Proceedings of the IEEE International Symposium on Parallel and Distributed Processing, Workshops & PhD Forum* (pp. 1–8). Piscataway, NJ: IEEE.

Bohte, S. M. (2011). Error-backpropagation in networks of fractionally predictive spiking neurons. In *Proceedings of the International Conference on Artificial Neural Networks* (pp. 60–68). Berlin: Springer.

Bohte, S. M., Kok, J. N., & La Poutré, J. A. (2000). Spikeprop: Backpropagation for networks of spiking neurons. In *Proceedings of the Eighth 8th European Symposium on Artificial Neural Networks* (pp. 17–37).

Bostock, M., Ogievetsky, V., & Heer, J. (2011). D^3 data-driven documents. *IEEE Transactions on Visualization and Computer Graphics*, *17*(12), 2301–2309.

Brette, R. (2015). Philosophy of the spike: rate-based vs. spike-based theories of the brain. *Frontiers in Systems Neuroscience*, *9*, 151.

Burkitt, A. N. (2006). A review of the integrate-and-fire neuron model: I. Homogeneous synaptic input. *Biological Cybernetics*, *95*(1), 1–19.

Castelvecchi, D. (2016). Can we open the black box of AI? *Nature News*, *538*(7623), 20.

Chollet, F. (2018). *Deep Learning mit Python und Keras: Das Praxis-Handbuch vom Entwickler der Keras-Bibliothek*. Bonn: MITP-Verlags.

Dasgupta, S., Stevens, C. F., & Navlakha, S. (2017). A neural algorithm for a fundamental computing problem. *Science*, *358*(6364), 793–796.

De Laat, P. B. (2018). Algorithmic decision-making based on machine learning from big data: Can transparency restore accountability? *Philosophy and Technology*, *31*(4), 525–541.

Dominguez-Morales, J. P., Jimenez-Fernandez, A., Rios-Navarro, A., Cerezuela-Escudero, E., Gutierrez-Galan, D., Dominguez-Morales, M. J., & Jimenez-Moreno, G. (2016). Multilayer spiking neural network for audio samples classification using spinnaker. In *Proceedings of the International Conference on Artificial Neural Networks* (pp. 45–53). Berlin: Springer.

Esser, S. K., Merolla, P. A., Arthur, J. V., Cassidy, A. S., Appuswamy, R., Andreopoulos, A., . . . Modha, D. S. (2016). Convolutional networks for fast, energy-efficient neuromorphic computing. In *Proceedings of the National Academy of Sciences*, *113*(41), 11441–11446.

Feng, J., Wang, Z., Zha, M., & Cao, X. (2019). Flower recognition based on transfer learning and Adam deep learning optimization algorithm. In *Proceedings of the 2019 International Conference on Robotics, Intelligent Control & Artificial Intelligence* (pp. 598–604). New York: ACM.

Field, G. D., Uzzell, V., Chichilnisky, E., & Rieke, F. (2019). Temporal resolution of single-photon responses in primate rod photoreceptors and limits imposed by cellular noise. *Journal of Neurophysiology*, *121*(1), 255–268.

Gerstner, W. (1998). *Spiking neurons* (Technical report). Cambridge, MA: MIT.

Gerum, R. (2020). pylustrator: code generation for reproducible figures for publication. *Journal of Open Source Software*, *5*(51), 1989.

Gerum, R. C., Erpenbeck, A., Krauss, P., & Schilling, A. (2020). Sparsity through evolutionary pruning prevents neuronal networks from overfitting. *Neural Networks*, *128*, 305–312.

Gilra, A., & Gerstner, W. (2017). Predicting non-linear dynamics by stable local learning in a recurrent spiking neural network. *eLife*, *6*, e28295.

Gross, G. W., & Kowalski, J. M. (1999). Origins of activity patterns in self-organizing neuronal networks in vitro. *Journal of Intelligent Material Systems and Structures*, *10*(7), 558–564.

Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. (2017). Neuroscience-inspired artificial intelligence. *Neuron*, *95*(2), 245–258.

Herculano-Houzel, S. (2009). The human brain in numbers: A linearly scaled-up primate brain. *Frontiers in Human Neuroscience*, *3*, 31.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.

Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *Journal of Physiology*, *117*(4), 500–544.

Huh, D., & Sejnowski, T. J. (2018). Gradient descent for spiking neural networks. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, & R. Garnett (Eds.), *Advances in neural information processing systems*, *31* (pp. 1433–1443). Red Hook, NY: Curran.

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science and Engineering*, *9*(3), 90–95.

Izhikevich, E. M., & FitzHugh, R. (2006). Fitzhugh-Nagumo model. *Scholarpedia*, *1*(9), 1349.

Jin, X., Lujan, M., Plana, L. A., Davies, S., Temple, S., & Furber, S. B. (2010). Modeling spiking neural networks on spinnaker. *Computing in Science and Engineering*, *12*(5), 91–97.

Kheradpisheh, S. R., & Masquelier, T. (2019). *S4nn: Temporal backpropagation for spiking neural networks with one spike per neuron*. arXiv:1910.09495.

Kim, R., Li, Y., & Sejnowski, T. J. (2019). Simple framework for constructing functional spiking recurrent neural networks. In *Proceedings of the National Academy of Sciences*, *116*(45), 22811–22820.

Koch, C., & Segev, I. (Eds.). (1998). *Methods in neuronal modeling: From ions to networks* Cambridge, MA: MIT Press.

Koopman, A., Van Leeuwen, M., & Vreeken, J. (2003). Dynamic neural networks, comparing spiking circuits and LSTM (Technical Report UU-CS-2003-007). Utrecht University.

Krauss, P., Metzner, C., Schilling, A., Schütz, C., Tziridis, K., Fabry, B., & Schulze, H. (2017). Adaptive stochastic resonance for unknown and variable input signals. *Scientific Reports*, *7*(1), 1–8.

Krauss, P., Metzner, C., Schilling, A., Tziridis, K., Traxdorf, M., Wollbrink, A., . . . Schulze, H. (2018). A statistical method for analyzing and comparing spatiotemporal cortical activation patterns. *Scientific Reports*, *8*(1), 1–9.

Krauss, P., Prebeck, K., Schilling, A., & Metzner, C. (2019). Recurrence resonance in three-neuron motifs. *Frontiers in Computational Neuroscience*, *13*.

Krauss, P., Schuster, M., Dietrich, V., Schilling, A., Schulze, H., & Metzner, C. (2019). Weight statistics controls dynamics in recurrent neural networks. *PLOS One*, *14*(4), e0214541.

Krauss, P., Tziridis, K., Metzner, C., Schilling, A., Hoppe, U., & Schulze, H. (2016). Stochastic resonance controlled upregulation of internal noise after hearing loss as a putative cause of tinnitus-related neuronal hyperactivity. *Frontiers in Neuroscience*, *10*, 597.

Krauss, P., Zankl, A., Schilling, A., Schulze, H., & Metzner, C. (2019). Analysis of structure and dynamics in three-neuron motifs. *Frontiers in Computational Neuroscience*, *13*, 5.

Kriegeskorte, N., & Douglas, P. K. (2018). Cognitive computational neuroscience. *Nature Neuroscience*, *21*(9), 1148–1160.

LeCun, Y., Jackel, L. D., Bottou, L., Cortes, C., Denker, J. S., Drucker, H., . . . Vapnik, V. (1995). Learning algorithms for classification: A comparison on handwritten digit recognition. *Neural Networks*, *261*(276), 2.

Lee, C., Sarwar, S. S., Panda, P., Srinivasan, G., & Roy, K. (2020). Enabling spike-based backpropagation for training deep neural network architectures. *Frontiers in Neuroscience*, *14*.

Lee, J. H., Delbruck, T., & Pfeiffer, M. (2016). Training deep spiking neural networks using backpropagation. *Frontiers in Neuroscience*, *10*, 508.

Mar, R. A. (2004). The neuropsychology of narrative: Story comprehension, story production and their interrelation. *Neuropsychologia*, *42*(10), 1414–1434.

Neftci, E. O., Mostafa, H., & Zenke, F. (2019). Surrogate gradient learning in spiking neural networks. *IEEE Signal Processing Magazine*, *36*, 61–63.

Nilsback, M.-E., & Zisserman, A. (2008). Automated flower classification over a large number of classes. In *Proceedings of the 2008 Sixth Indian Conference on Computer Vision, Graphics and Image Processing* (pp. 722–729). Piscataway, NJ: IEEE.

Olshausen, B. A., & Field, D. J. (1997). Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Research*, *37*(23), 3311–3325.

Orchard, G., Jayawant, A., Cohen, G. K., & Thakor, N. (2015). Converting static image datasets to spiking neuromorphic datasets using saccades. *Frontiers in Neuroscience*, *9*, 437.

Perkel, D. H., & Bullock, T. H. (1968). Neural coding. *Neurosciences Research Program Bulletin*, *6*(3), 221–348.

Pontes-Filho, S., & Nichele, S. (2019). *Towards a framework for the evolution of artificial general intelligence*. arXiv:1903.10410.

Pozzi, I., Nusselder, R., Zambrano, D., & Bohté, S. (2018). Gating sensory noise in a spiking subtractive LSTM. In *Proceedings of the International Conference on Artificial Neural Networks* (pp. 284–293). Berlin: Springer.

Qin, M., Xi, Y., & Jiang, F. (2019). A new improved convolutional neural network flower image recognition model. In *Proceedings 2019 IEEE Symposium Series on Computational Intelligence* (pp. 3110–3117). Piscataway, NJ: IEEE.

Rezaabad, A. L., & Vishwanath, S. (2020). *Long short-term memory spiking networks and their applications.* arXiv:2007.04779.

Rieke, F., & Baylor, D. A. (1998). Single-photon detection by rod cells of the retina. *Reviews of Modern Physics*, *70*(3), 1027.

Roome, C. J., & Kuhn, B. (2020). Dendritic coincidence detection in Purkinje neurons of awake mice. *eLife*, *9*, e59619.

Rueckauer, B., Lungu, I.-A., Hu, Y., Pfeiffer, M., & Liu, S.-C. (2017). Conversion of continuous-valued deep networks to efficient event-driven networks for image classification. *Frontiers in Neuroscience*, *11*, 682.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., . . . Fei-Fei, L. (2015). ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, *115*(3), 211–252.

Schemmel, J., Grubl, A., Meier, K., & Mueller, E. (2006). Implementing synaptic plasticity in a VLSI spiking neural network model. In *Proceedings of the 2006 IEEE International Joint Conference on Neural Network Proceedings* (pp. 1–6). Piscataway, NJ: IEEE.

Schilling, A., Gerum, R., Zankl, A., Schulze, H., Metzner, C., & Krauss, P. (2020). *Intrinsic noise improves speech recognition in a computational model of the auditory pathway.* bioRxiv.

Schrauwen, B., & Van Campenhout, J. (2004a). Extending SpikeProp. In *Proceedings of the 2004 IEEE International Joint Conference on Neural Networks* (pp. 471–475). Piscataway, NJ: IEEE.

Schrauwen, B., & Van Campenhout, J. (2004b). Improving SpikeProp: Enhancements to an error-backpropagation rule for spiking neural networks. In *Proceedings of the 15th ProRISC Workshop*, *11* (pp. 301–305).

Sheng, Y.-B., & Zhou, L. (2017). Distributed secure quantum machine learning. *Science Bulletin*, *62*(14), 1025–1029.

Shevlin, H., Vold, K., Crosby, M., & Halina, M. (2019). The limits of machine intelligence: Despite progress in machine intelligence, artificial general intelligence is still a major challenge. *EMBO Reports*, *20*(10), e49177.

Shrestha, S. B., & Orchard, G. (2018). Slayer: Spike layer error reassignment in time. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, & R. Garnett (Eds.), *Advances in neural information processing systems*, *31* (pp. 1412–1421). Red Hook, NY: Curran.

Steinkraus, D., Buck, I., & Simard, P. (2005). Using GPUs for machine learning algorithms. In *Proceedings of the Eighth International Conference on Document Analysis and Recognition* (pp. 1115–1120). Piscataway, NJ: IEEE.

Taherkhani, A., Belatreche, A., Li, Y., Cosma, G., Maguire, L. P., & McGinnity, T. M. (2020). A review of learning in biologically plausible spiking neural networks. *Neural Networks*, *122*, 253–272.

Tavanaei, A., Ghodrati, M., Kheradpisheh, S. R., Masquelier, T., & Maida, A. (2019). Deep learning in spiking neural networks. *Neural Networks*, *111*, 47–63.

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences*, *10*(7), 309–318.

Thorpe, S., Delorme, A., & Van Rullen, R. (2001). Spike-based strategies for rapid processing. *Neural Networks*, *14*(6-7), 715–725.

Vreeken, J. (2003). Spiking neural networks, an introduction. (Technical Report UU-CS-2003-008). Utrect University.

Walt, S. v. d., Colbert, S. C., & Varoquaux, G. (2011). The NumPy array: A structure for efficient numerical computation. *Computing in Science and Engineering*, *13*(2), 22–30.

Wolfart, J., Debay, D., Le Masson, G., Destexhe, A., & Bal, T. (2005). Synaptic background activity controls spike transfer from thalamus to cortex. *Nature Neuroscience*, *8*(12), 1760–1767.

Wu, Y., Deng, L., Li, G., Zhu, J., & Shi, L. (2018). Spatio-temporal backpropagation for training high-performance spiking neural networks. *Frontiers in Neuroscience*, 12:331.

Xia, X., Xu, C., & Nan, B. (2017). Inception-v3 for flower classification. In *Proceedings of the 2nd International Conference on Image, Vision and Computing* (pp. 783–787). Piscataway, IEEE.

Xin, J., & Embrechts, M. J. (2001). Supervised learning with spiking neural networks. In *Proceedings of the International Joint Conference on Neural Networks. Proceedings* (pp. 1772–1777). Piscataway, NJ: IEEE.

Yang, Z., Schilling, A., Maier, A., & Krauss, P. (2021). Neural networks with fixed binary random projections improve accuracy in classifying noisy data. In *Bildverarbeitung für die Medizin* (pp. 211–216). Berlin: Springer.

Zenke, F., & Ganguli, S. (2018). SuperSpike: Supervised learning in multilayer spiking neural networks. *Neural Computation*, *30*(6), 1514–1541.

Zenke, F., & Gerstner, W. (2014). Limits to high-speed simulations of spiking neural networks using general-purpose computers. *Frontiers in Neuroinformatics*, *8*, 76.

Zenke, F., & Vogels, T. P. (2020). *The remarkable robustness of surrogate gradient learning for instilling complex function in spiking neural networks.* BioRxiv.

Received September 8, 2020; accepted April 26, 2021.

Leaky-Integrate-and-Fire Neuron-Like Long-Short-Term-Memory Units as Model System in Computational Biology

Richard Gerum
Department of Physics and Astronomy
York University
Toronto, Canada
0000-0001-5893-2650

André Erpenbeck
Department of Physics
University of Michigan
Ann Arbor, United States
aerp@umich.edu

Patrick Krauss
Neuroscience Lab,
University Hospital Erlangen
Erlangen, Germany
patrick.krauss@fau.de

Achim Schilling
Neuroscience Lab,
University Hospital Erlangen
Erlangen, Germany
achim.schilling@fau.de

2023 International Joint Conference on Neural Networks (IJCNN) | 978-1-6654-8867-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/IJCNN54540.2023.10191268

***Abstract*—Biological neural networks encode information very efficiently, and dynamically react to sensory input on very small time scales. In contrast to most contemporary machine learning approaches which rely on rate neurons with continuous output, biological neural networks are based on spiking neurons with quasi-binary discrete output.**

In Artificial Intelligence (AI) Research, time series data are efficiently encoded in Long-Short-Term-Memory (LSTM) networks. Despite their strength in encoding time series data and making predictions, LSTM units are assumed to be biologically implausible. Nevertheless, recent studies show that LSTM unit networks indeed behave similar to biological neural networks.

In this study, we show that a particular choice of parameters for the weights and gates in peephole LSTM units causes these units to show similar dynamic behaviour as biologically plausible leaky integrate-and-fire (LIF) neurons, which represent a simple biologically inspired spiking neuron model. We analyzed the spiking characteristics of the restricted peephole LSTM units and characterize the parameter space, in which these units show certain spiking characteristics.

We conclude that tackling complex cognitive tasks with biologically plausible and explainable artificial neural networks is an important step to make progress in both fields, neuroscience and artificial intelligence.

I. Introduction

In recent years, the interest in neuroscience-inspired Artificial Intelligence has massively increased [1], [2]. Although standard machine learning approaches benefit from novel hardware resources (see e.g. [3]–[5]) and thus their performance in certain tasks increases, we are still far from creating an artificial neural network with human-level general intelligence. Biological brains are very energy efficient due to their spiking neurons and therefore spiking neuron models are a promising way to increase the efficiency of artificial neural networks used in machine learning [6]–[11].

Several biologically inspired spiking neuron models were developed such as the Hodgkin-Huxley neuron [12] or the Fitzhugh-Nagumo neuron [13], which depend on several coupled differential equations. A simpler neuron model is the leaky-integrate-and-fire (LIF) neuron, which temporally integrates its input currents and releases a spike when a certain threshold is crossed from below [14]. Long-time dependencies are prevented by an exponentially decreasing membrane potential (leakage). In very recent studies, the LIF neuron became a target to integrate spiking neurons in machine learning architectures to solve complicated tasks, in order to increase biological plausibility, energy efficiency and performance [15], [16].

However, in most cases the performance of spiking neural networks is still not comparable to standard machine learning architectures [9]. Nevertheless, a more established neuron model for machine learning—the so called long-short-term-memory (LSTM) neuron [17]—can also be trained to show spiking behavior [18]. This is possible as LSTM units have an internal state, which can be compared to the membrane potential of LIF units. Salaj and coworkers even describe the LSTM unit as "functional equivalent" to spiking neurons enabling for a working memory [19]. Thus, the interest in the interplay of LSTM units, spiking neurons and biological systems has recently grown [19]–[21]. Especially so called peephole-LSTM units [22]–[24] are suited to be driven in a spiking manner. The peephole connections — the connections of the internal state to the three gates of the LSTM units — allow the neuron to produce a spike, when a certain threshold is reached. This property is similar to LIF units that also spike when a certain threshold is exceeded. One major disadvantage of LSTM units is the enormous number of parameters in combination with their lack of biological plausibility [15], [25].

In this study, we show that LIF and LSTM units are tightly connected, when the parameters of the LSTM units are tuned and especially reduced in a certain way. Therefore LIF units can be implemented as adapted LSTM units in already existing machine learning frameworks such as Keras [26]. We show that, with a sophisticated choice of the LSTM parameters, the update rule of the LSTM neuron is similar to the update rule of the LIF neuron. Additionally, we numerically demonstrate

DFG (German Research Foundation, Deutsche Forschungsgemeinschaft)

the validity of our considerations. Therefore, we trained a restricted peephole LSTM network as well as a LIF network on image classification using four common image data sets (MNIST [27], fashion-MNIST [28], EMNIST [29] , CIFAR 10 [30]), which we transformed into spike trains using a Poisson rate encoding [31], [32]. We provide evidence that spiking characteristics as well as learning performance is similar in restricted LSTM and LIF neuron networks and that spiking LSTM units can be trained with simple backpropagation, whereas LIF unit networks need a surrogate gradient approach (cf. [33]).

II. Methods

A. Software Resources

All simulations and visualizations were created using a standard desktop PC. The simulation was implemented in python 3.7 using NumPy [34]. Visualizations were created using Javascript and especially the D^3-library [35] and python using Matplotlib [36] and Pylustrator [37]. An interactive version of Fig. 1 is available in a github repository (https://rgerum.github.io/paper_lstm_lif). The feed-forward neural network consisting of LIF units and LSTM units was implemented in Keras [26] with Tensorflow backend [38].

B. LIF units

In the following, we show that the equations describing LIF neurons (cf. (1)–(2)) represent a limiting case of the LSTM equations (cf. (8)–(9)). This insight is then exploited in order to employ LSTMs in such a way that they effectively behave like LIF neurons.

In a previous study [33], we could show that LIF neurons can be mathematically described by the following equations:

$$V_{t_n} = w_{\text{input}} \cdot \Delta t \cdot x_{t_n} + (1 - w_{\text{leak}} \cdot \Delta t) \cdot V_{t_{n-1}} \cdot \Theta(V_{\text{thresh}} - V_{t_{n-1}}) \tag{1}$$

$$y_{t_n} = \Theta(V_{t_n} - V_{\text{thresh}})/\Delta t \tag{2}$$

$$t_n = t_{n-1} + \Delta t \tag{3}$$

$V_{t_n}, V_{\text{thresh}}, x_{t_n}, w_{\text{input}}, t_n \in \mathbb{R}, w_{\text{leak}}, \Delta t, y_{t_n} \in \mathbb{R}^+, n \in \mathbb{N}$.

The evolution of the neuron is simulated at discrete time points t_n with a temporal resolution of Δt. The membrane potential V_{t_n}, which is the internal state of the LIF neuron at time point t_n, is the sum of the input at this time, x_{t_n}, and the previous state $V_{t_{n-1}}$, weighted with a leakage term w_{leak} preventing long-range correlations. The Heaviside function (Θ) in (1) only keeps the membrane potential of the previous time point t_{n-1} if no spike occurred, therefore acting as a reset for the membrane potential after a spike. The output y_{t_n} is 0 if no spike occurs in the time point t_n and $1/\Delta t$ otherwise. The spike is a discretized version of the Dirac delta distribution $\delta(t)$, i.e. a rectangle with width Δt and height $1/\Delta t$, ensuring the integral property of the one-dimensional Dirac delta distribution $\int_{\mathbb{R}} \delta(t)dt = 1$. This spike is released, when V_{t_n} reaches a threshold value of V_{thresh}. V_{thresh} can be set to 1 without loss of generality [33].

When staking multiple LIF layers, it is convenient to normalize the input and output values y_{t_n}, and x_{t_n}, in order to suspend the Δt scaling of these quantities. Or in other words, replacing y_{t_n} with $\hat{y}_{t_n} = y_{t_n} \cdot \Delta t$ and x_{t_n} with $\hat{x}_{t_n} = x_{t_n} \cdot \Delta t$ yields an output $\hat{y}_{t_n} \in \{0, 1\}$ that is a binary variable independently of the time step Δt. The modified LIF equations now are

$$V_{t_n} = w_{\text{input}} \cdot \hat{x}_{t_n} + (1 - w_{\text{leak}} \cdot \Delta t) \cdot V_{t_{n-1}} \cdot \Theta(V_{\text{thresh}} - V_{t_{n-1}}) \tag{4}$$

$$\hat{y}_{t_n} = \Theta(V_{t_n} - V_{\text{thresh}}). \tag{5}$$

Note that while this requires to scale the input to the first LIF layer with Δt for analog input or for spiking input to use a value of 1 to indicate a spike.

To show how V_{t_n} is linked with $\hat{y}_{t_{n-1}}$ the equations can also be written as

$$V_{t_n} = w_{\text{input}} \cdot \hat{x}_{t_n} + (1 - w_{\text{leak}} \cdot \Delta t) \cdot V_{t_{n-1}} \cdot (1 - \hat{y}_{t_n}) \tag{6}$$

$$\hat{y}_{t_n} = \Theta(V_{t_n} - V_{\text{thresh}}). \tag{7}$$

C. Tuning LSTM Units to Create LIF-Like Behavior

In the following, we show that a proper tuning of an LSTM unit leads to similar properties.

The long-short-term-memory (LSTM) unit comprises a memory cell with a recurrent connection of weight 1 to permanently store information (without a vanishing gradient, see Fig. 1a). The input, output, and memory are controlled by so called gates, which control if information can flow into the cell (input gate), if the memory is reset (forget gate, forget when gate is 0), or if the cell state is fed to the output (see Fig. 1a). These gates are connected to the input (including the bias term) and the output. Optionally, the cell's state is also connected to the gates (peephole connections). These gates allow for a controlled flow of information and allows the LSTM to release a spike when the internal state reaches a certain threshold [22]–[24].

The peephole-LSTM unit can be described by the following equations [22]–[24]:

$$\begin{aligned} V_{t_n} = {} & \tanh(b_{\text{input}} + w_{\text{input}} \cdot x_{t_n} + w_{\text{output}} \cdot y_{t_{n-1}}) \\ & \cdot \sigma(b_1 + f_{\text{input1}} \cdot x_{t_n} + f_{\text{cell1}} \cdot V_{t_{n-1}} + f_{\text{output1}} \cdot y_{t_{n-1}}) \\ & + V_{t_{n-1}} \cdot \sigma(b_2 + f_{\text{input2}} \cdot x_{t_n} + f_{\text{cell2}} \cdot V_{t_{n-1}} + f_{\text{output2}} \cdot y_{t_{n-1}}) \end{aligned} \tag{8}$$

$$\begin{aligned} y_{t_n} = \tanh(V_{t_n-1}) \cdot \sigma(b_3 + f_{\text{input3}} \cdot x_{t_n} + f_{\text{cell3}} \cdot V_{t_{n-1}} \\ + f_{\text{output3}} \cdot y_{t_{n-1}}) \end{aligned} \tag{9}$$

In this representation, $y_{t_n} \in]-1, 1[$, all other variables are $\in \mathbb{R}$. Here, the input is x_{t_n}, the cell state is V_{t_n}, the cells previous state is $V_{t_{n-1}}$, and the output is y_{t_n}. The sigmoid activation function is denoted by $\sigma(x) = 1/(1 + \exp(-x))$, with the input weight ω_{input}, the output weight ω_{output}, and the gate weights $b_{1,2,3}$, $f_{\text{input}1,2,3}$, $f_{\text{output}1,2,3}$ and the peephole weights $f_{\text{cell}1,2,3}$. Similar variable names in (1)–(2)

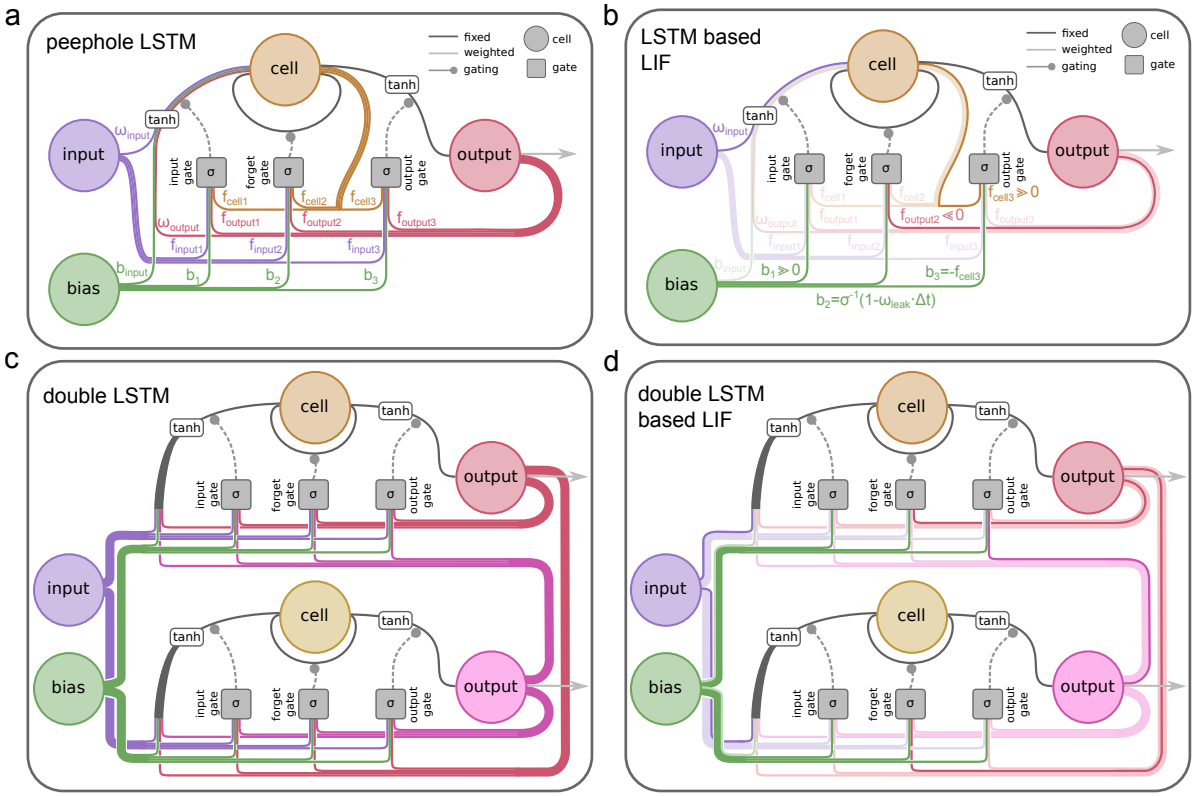

Fig. 1. **Architecture of a peephole LSTM unit and the LIF-like LSTM.**
a, Peephole LSTM unit: The input (x_{t_n}, purple) from pre-connected layers is summed up with the bias (green) and the output (red) and scaled with a tanh function (see (8)). This scaled value is summed up with the preceding internal state $V_{t_{n-1}}$ and defines the present internal state (cell: V_{t_n}). Additionally, in peephole LSTM units the input, the bias, the output, and the internal states are connected to all gates (gates: input, forget, output). The output y_{t_n} (red) is the tanh of the internal state and controlled by the output gate (see (9)). The connections of internal state to the gates (especially the output gate) is very important to achieve LIF properties, as this gate enables to define a threshold for the internal state when a spike is released (in analogy to the Θ-function in (2)). **b**, The restricted LSTM to operate in LIF mode; Transparent links are the weights that are set to zero. The input gate is forced to be always open, the forget gate imitates the leak term (connection with the bias) and the reset function after a spike (connection with the output). The output gate is controlled by the peephole connection from the cell to trigger the output when the state is above threshold. https://rgerum.github.io/paper_lstm_lif#lstm **c**, Two coupled "normal" LSTM units; As peephole LSTM units are not very common in AI applications, we show that also two "normal" LSTM units can create spiking behavior. **d** Two connected LSTM units tuned so that they show LIF like spiking behavior; The connection of the inner state (cell) to the output gate is replaced by a connection of the output of a second LSTM unit to the output gate of the spiking LSTM. The second LSTM unit's output is set to be always open to expose the cell state. Thus the second LSTM units imitates the inner state of the first LSTM.

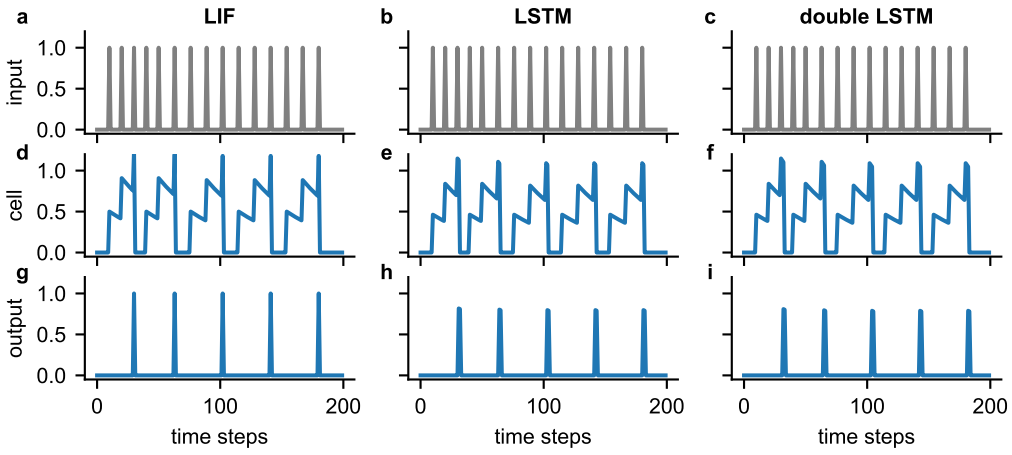

Fig. 2. **Spiking behaviour of LIF, peephole LSTM and double LSTM.**
a,b,c The input signal is the same for all three spiking units. **d,e,f** Inner state of LIF unit (d), spiking peephole LSTM unit (e), and spiking double LSTM (f, Fig. 1 d). **g,h,i** Output spikes of all three spiking units.

and (8)–(9) indicate that the variables serve the same purpose in the description of the neuron and that they share the same range of values.

In order to establish the existence of a limiting case in which an LSTM behaves like a LIF, we identify a parameter regime for $f_{...}$, $b_{...}$, and $w_{...}$, where (8)–(9) are structurally similar to (1)–(2). In the first step, we employ mathematical considerations in order to map the LSTM equations onto the LIF equations. The practicability and the biological implications together with a discussion of potential errors is provided later. In order to establish a connection between the LSTM equations and the LIF equations, we start by removing unneeded connections by setting them to 0, i.e. $b_{\text{input}} = 0$, $f_{\text{input1,2,3}} = 0$, $f_{\text{cell1,2}} = 0$, $f_{\text{output1,3}} = 0$. We also set the recurrent connection $w_{\text{output}} = 0$, note that other LIF models use a recurrent connection [39]. This yields the first set of *restricted* LSTM equations,

$$V_{t_n} = \tanh(w_{\text{input}} \cdot x_{t_n}) \cdot \sigma(b_1) + V_{t_{n-1}} \cdot \sigma(b_2 + f_{\text{output2}} \cdot y_{t_{n-1}}), \tag{10}$$

$$y_{t_n} = \tanh(V_{t_{n-1}}) \cdot \sigma(f_{\text{cell3}} \cdot V_{t_{n-1}} + b_3), \tag{11}$$

which are intermediate equations that are now subdued to a limit analysis. In particular, we consider the limits

1) $w_{\text{input}} \cdot x_{t_n} \approx 0$, which implies $\tanh(w_{\text{input}} \cdot x_t) = w_{\text{input}} \cdot x_{t_n} - \mathcal{O}((w_{\text{input}} \cdot x_{t_n})^3)$
2) $b_1 \to \infty$, such that $\sigma(b_1) \to 1$, which corresponds to keeping the **input gate** always open
3) we identify $b_2 = \sigma^{-1}(1 - w_{\text{leak}} \cdot \Delta t)$, i.e. open the forget gate so that it acts as a **leak term**,
4) $f_{\text{output2}} \to -\infty$, the necessity of this limit will be discussed subsequently, where it will become apparent that it **resets** the state after a spike
5) $f_{\text{cell3}} \to \infty$ and $b_3 = -f_{\text{cell3}}$, which will be used in the following to mimic a theta function to open the **output** when the state exceeds the threshold of $V_{\text{threshold}} = 1$

In addition to these limits, a following limit on the behavior of the sigmoid function is needed,

$$\sigma(\alpha \cdot x) = \begin{cases} 1 & \text{for } \alpha \cdot x \to +\infty \\ 0 & \text{for } \alpha \cdot x \to -\infty \end{cases} \tag{12}$$

such that given an adequate scaling factor α, the sigmoid function behaves like a Heaviside step function,

$$\sigma(\alpha \cdot x) \to \Theta(x) \quad \text{for } \alpha \to \infty. \tag{13}$$

At this point, we can use the above limits to map one of the restricted LSTM equations onto its LIF counterpart. In particular, using limit number 5 and the Heaviside identity in (13), (11) assumes the form

$$y_{t_n} = \tanh(V_{t_{n-1}}) \cdot \Theta(V_{t_{n-1}} - 1). \tag{14}$$

Notice that this equation is equivalent to the LIF equation (2) up to the pre-factor $\tanh(V_{t_{n-1}})$ as well as the index shift from V_{t_n} to $V_{t_{n-1}}$. This results in the spike being release one time step later than in the LIF model.

Now, we consider the second restricted LSTM (10). Knowing that y_{t_n} is either 0 or $\tanh(V_{t_{n-1}})$ and employing limit 4, $f_{\text{output2}} \to -\infty$, the sigmoid function in (10) allows for the following distinct cases,

$$\sigma(\sigma^{-1}(1 - w_{\text{leak}} \cdot \Delta t) + f_{\text{output2}} \cdot y_{t_{n-1}}) = \begin{cases} (1 - w_{\text{leak}} \cdot \Delta t) & \text{for } y_{t_{n-1}} = 0 \\ 0 & \text{otherwise .} \end{cases} \tag{15}$$

This allows for the identification

$$\sigma(\sigma^{-1}(1 - w_{\text{leak}} \cdot \Delta t) + f_{\text{output2}} \cdot y_{t_{n-1}}) = (1 - w_{\text{leak}} \cdot \Delta t) \cdot \Theta(1 - y_{t_{n-1}}), \tag{16}$$

which transfers the restricted LSTM equation (10) into

$$V_{t_n} = w_{\text{input}} \cdot x_{t_n} + V_{t_{n-1}} \cdot (1 - w_{\text{leak}} \cdot \Delta t) \cdot \Theta(1 - y_{t_{n-1}}) \tag{17}$$

Notice again the similarity between this equation and the LIF equation (1), whereby the main difference lies in the fact that $y_{t_{n-1}}$ in the Θ-function takes the place of the $V_{t_{n-1}}$ in the corresponding LIF equation, which is equivalent.

Despite the apparent similarity between the LIF equations (1)–(2) and the equations derived from the LSTM, (17)–(14), there are also distinct differences which need to be discussed carefully. First of all, the output y_{t_n} of the LSTM based equations is no actual spiking output. Despite the fact that the Θ-function prevents a continuous output, the supra-threshold output is a continuous value in contrast to the binary output of LIF neurons. However, an approximately spiking behavior of the output can be seen, as when $V_{t_{n-1}} > 1$ the output will be $\tanh(V_{t_{n-1}}) \in [\tanh(1) = 0.761, 1[$. Although there are some mathematical deviations from a standard LIF unit, LSTM units (shown for LIF tuning in Fig. 1b) can show spiking behavior (Fig. 2b), if tuned correctly.

Another source of error comes from the practical restrictions when realizing the limits outlined above for application. For the data presented below, we set the parameters to the following values, which realize the limits 1–5 discussed above (the numbers in the following list coincide with the numbers used to label the individual limits):

1. We require that $w_{\text{input}} \cdot x_{t_n} < 1$. For spiking input, $w_{\text{input}} > 0.5$ is not useful as already two input spikes are enough to lift the neuron over the spiking threshold. Nevertheless, it has to be stated that the tanh-function slightly changes the characteristics of LSTM neurons compared to LIF neurons ($\tanh(0.5) \approx 0.46$, error of approx. 10 %).
2. $b_1 = 100 \gg 0$.
4. $f_{\text{output2}} = -100 \ll 0$.
5. $f_{\text{cell3}} = 100 \gg 0$.

The choice of these parameters is to some extent arbitrary, but we have tested their validity (see also paragraph "Supervised Training of Spiking LSTM Units").

D. Feed-Forward Neural Networks for Image Classification

To validate all mathematical assumptions and approximations made above (paragraph "Tuning LSTM Units to Create LIF-Like Behavior"), we trained a feed-forward neural network on image classification. Thus, we used the four most common image data sets (MNIST [27], fashion MNIST [28], EMNIST [29], CIFAR-10 [30]) and applied Poisson rate coding [31], [32] to create spike trains as network input. Therefore, the network input was a sequence of vectors of size N (N is number of pixels). The sequence length was 200 (number of simulated time-steps, t = 1 s, dt = 0.005 s, poisson rate = $\frac{1}{255} \cdot 10$). The network was trained using backpropagation through time with a surrogate gradient approach (see [33]). The LSTM network was trained using simple backpropagation. We used a feed-forward architecture with one hidden layer of 128 spiking units (LSTM, LIF) and an output layer implemented as accumulating layer of 10 units—one for each class. The network was trained for 10 epochs with a batch-size of 256, using the Adam optimizer and the mean squared error loss.

III. Results

A. Spiking Characteristics of LSTM

As shown in detail above peephole LSTM units can be used to approximate LIF units by tuning the parameters in a certain way. Indeed, peephole LSTM units are not very common in AI research and thus we also showed that two connected LSTM units can be tuned so that they show LIF like spiking behavior.

In the following, we numerically validate our mathematical approximations numerically. Thus, we tested the spiking behavior of the peephole LSTM compared with a simple LIF unit for different parameter combinations of f_{output2} and f_{cell3} (see also (10) and (11)). If the variable f_{output2} is not negative enough e.g. just equals -1, the inner state of the LSTM unit is not properly reset (see Fig. 3 a,d,g). Therefore, the inner state surpasses the threshold for each input spike (see Fig. 3 red lines) and thus a wrong choice of f_{output2} causes continuous spiking. The other analyzed parameter is f_{cell3}, which changes the height of the released spike. The spike height for LIF units is always the same (in our case 1). However, in LSTM units the released spike is an analog value ($\tanh(V_{t_{n-1}}) \cdot \sigma(f_{\mathrm{cell3}} \cdot V_{t_{n-1}} - f_{\mathrm{cell3}})$). Thus, for a high value of f_{cell3} the σ-function can be approximated with an Θ-function and thus the released spike has the height $V_{t_{n-1}}$, which is nearly 1 (Fig. 3 g, h, i, see blue spikes for LSTM unit and red spike for LIF unit).

B. Supervised Training of Spiking LSTM Units

To investigate the training characteristics of the spiking LSTM units we trained a simple feed-forward neural network on four common image data sets (MNIST, fashion MNIST, EMNIST, CIFAR-10) and evaluated the classification accuracy. We performed a Poisson rate encoding to convert the images into spike trains and thus to generate a spiking input. (The brightness of a pixel is proportional to the probabilistic spike rate of the neuron. The input dimension is equal to the number of pixels of the images.)

We compared the classification accuracy of the LSTM network to a neural network with the same architecture consisting of LIF units. Trivially, LIF units cannot be trained with a simple gradient descent algorithm as the gradient of these units is always 0. Therefore, we used a simple surrogate gradient approach (cf. [33]). We could show that the accuracy of the LIF-network is slightly better than the accuracy of the LSTM network (cf. Fig. 4a with Fig. 4e). Thus, one major advantage of the LSTM network is the fact that it can be trained by simple gradient descent algorithms with no surrogate gradient adding further arbitrary parameters to the system.

To check if the spiking characteristics is similar for both neuron models, we performed a cross-check. Thus, on the one hand we trained the LIF network and used the weights of the trained network to evaluate the LSTM network and the other way round. Indeed, changing neuron model did not lead to significant accuracy drop, which means that the neuron types can be seen as equivalent. Thus, we show that the usage of the same weights leads to very similar spiking patterns in the LIF respectively LSTM layer of the neural network (see Fig. 4 a, c and Fig 4 e, g).

IV. Discussion

A. Summary

In this study, we showed how LSTM units with peephole connections (see Fig. 1a) can be manually tuned to show similar spiking behaviour as LIF neurons. The fact that LSTM neurons can be tuned, in such a way that they show spiking behavior has already been proven (see [18]). However, it is important to understand the mechanisms that lead to the spiking dynamics and to know the mathematical limitations. Thus, we provide a full mathematical description of these effects and demonstrate that in certain value ranges and under certain assumptions the neuron models are nearly equivalent.

We illustrate that the peephole connections — the connections from the internal state of the LSTM unit to the gates — are essential to achieve properties similar to LIF neurons (see Fig. 1 a, b). However, these peephole connections are often not implemented in standard LSTM neuron models. Thus, these "reduced" LSTM units lack interesting spiking dynamics.

As we provide a complete mathematical description of the spiking peephole LSTM units, we are able to further describe the effects of parameter changes (see Fig. 3) on spiking characteristics of these units. This could be helpful as it is potentially possible to build efficient neural networks by combining peephole LSTM units with different spiking characteristics such as coincidence detectors and integrator neurons (cf. [2], [33]).

Additionally we have shown that it is also possible to create spiking output with non-peephole LSTM units by connecting two of them (see Fig. 1 c, d). Thus, one helper LSTM unit receives the same input as the "spiking" LSTM unit. The connection of the output of the "helper" LSTM unit to the output gate of the "spiking" LSTM unit simulates the

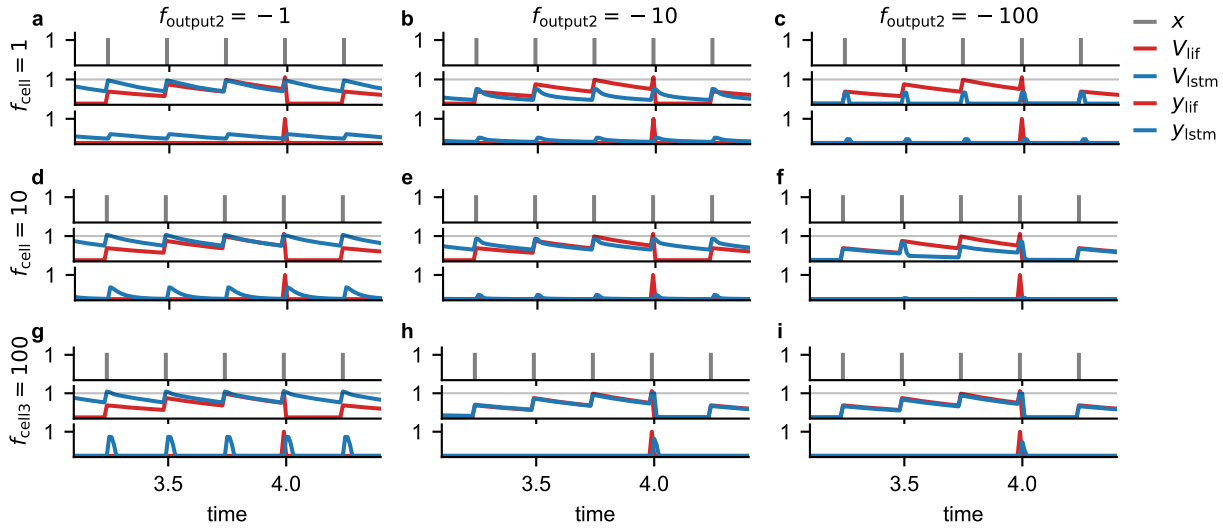

Fig. 3. **Spiking characteristics of restricted peephole LSTM with different parameter combinations of f_{cell3} and f_{output2}.** The figure shows the inner state V as well as the output y of a LIF unit (red curves) and a restricted peephole LSTM unit (blue curves) as a response to a regular spike input x (gray) for different parameter combinations of f_{cell3} and f_{output2}. LSTM units with a low f_{cell3} (scaling connection of inner state and output gate) do not produce sufficiently strong spike outputs (see e.g. **a,b,c** where $f_{\text{cell3}} = 1$). If f_{output2} is not negative enough, which means not far below -10, the inner state V is not properly reset after a spike release (see e.g. **a,d,g** where $f_{\text{output2}} = -1$). **i**, Restricted peephole LSTM units with both values sufficiently large (f_{cell3}) respectively negative (f_{output2}) show proper spiking output.

peephole connection. Indeed, this is an inconvenient way of creating spiking LSTM units, however, these considerations are important to interpret and understand dynamics of neural networks consisting of these reduced resp. standard LSTM units. Therefore, already trained neural networks could become an interesting model system for cognitive science and computational neuroscience (cf. e.g. [40]).

In a next step, learning experiments with spiking peephole LSTM units were performed. Thus, we showed that it is possible to efficiently train spiking peephole LSTM units using backpropagation through time without any surrogate gradient, which is a significant advantage compared to LIF units (Fig. 4). Furthermore, we illustrated that the spiking characteristics of LIF units and restricted peephole LSTM units are very similar (Fig. 4) by replacing the LIF resp. LSTM units in already trained neural networks.

B. Relevance of Spiking LSTM Units for AI and Biology

There are two major reasons for mapping established peephole LSTM units on LIF units. First, one may profit from already existing software frameworks such as Keras [26] or pytorch [41], together with optimized hardware such as GPU devices. Thus, these frameworks could be used to easily create biologically plausible neural networks [42], which are a promising target for theoretical neuroscience and can be trained using already established algorithms [43], [44]. Thus, machine learning approaches can be further developed by applying biological principles [1], such as information encoding by complex spike patterns instead of simple rate codes.

Second, the significant value of artificial neural networks developed by AI researchers to solve complex (cognitive) tasks as model system for neuroscience is indisputable. Thus, AI has triggered a lot of ideas for biological theories during the last decades [45]. These high performance neural networks have two major advantages: These algorithms are already able to solve complex tasks and it is possible to read out all neuron states at all time points in contrast to experimental neuroscience (e.g. [40]), where only a small subset of neurons can be recorded. Consequently, it is for example possible to simulate neural lesions by setting some neurons to zero and to analyze the altered performance and output of these networks (cf. e.g. [46]). LSTM networks are a very interesting model system, as they can be trained with standard techniques on the one hand, and have recurrent connections and thus some kind of memory function on the other hand. Furthermore, as shown above these neuron models can show spiking behavior as shown above.

Spiking neuron networks play an important role in computational and theoretical neuroscience [47]–[52]. Recently, it has been shown that LSTM networks can develop biologically plausible behavior. Thus, one research group trained a deep LSTM network on a navigation task and demonstrated that the network shows similar activity characteristics as the enthorinal cortex [53]. Thus, some LSTM units showed grid cell-like activity being crucial for any navigation tasks [54]. This was the starting point for further investigations on LSTM networks and e.g. biologically plausible vision guided and vision based navigation [55]. These studies illustrate that LSTM units are a valid model system for computational neuroscience. In one further recent study, a research group developed a hybrid analog spiking neural network based on LSTM units. Thus, the researchers developed a conversion method to exploit the advantages of spiking neurons on the one hand and the established training methods for LSTM networks on the other

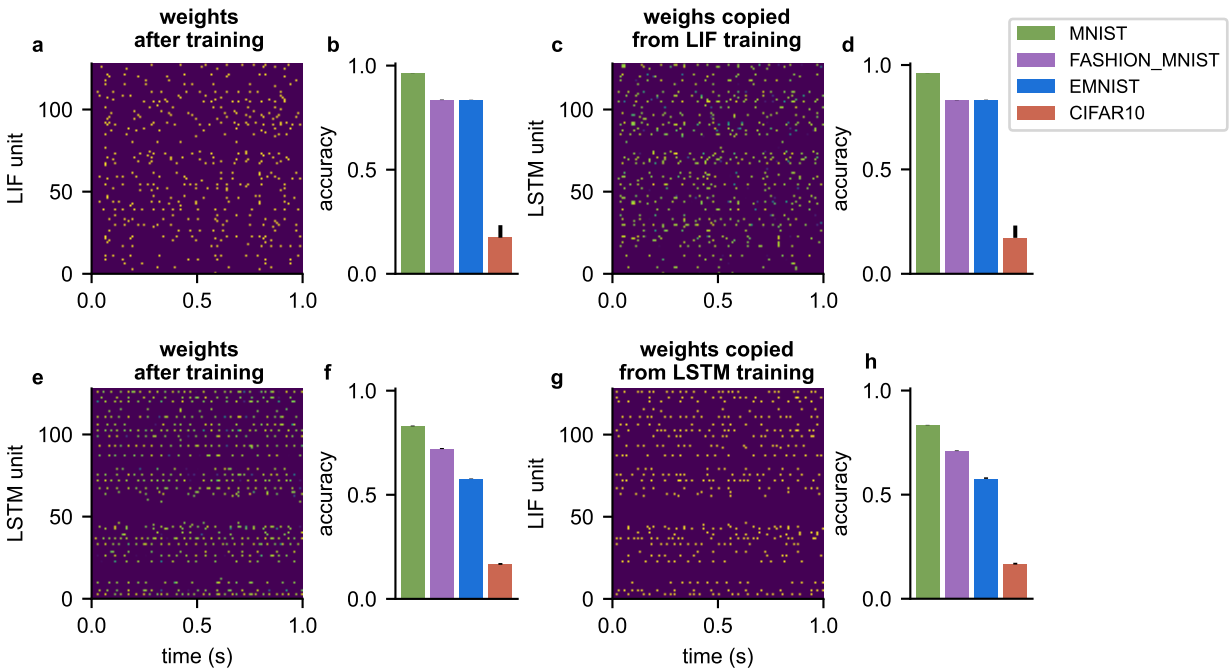

Fig. 4. **Comparison of LIF and LSTM neural networks**
Two simple feed-forward neural networks were trained on image classification using four data sets (MNIST, Fashion MNIST, EMNIST Letter, CIFAR-10) encoded with Poisson rate coding. One neural network consisted of LIF units and was trained with a surrogate gradient approach the other neural network consisted of restricted peephole LSTM neurons and was trained with simple gradient descent. **a**, Output of the hidden LIF layer for one example image input. **b**, Classification accuracy for all four data sets (mean±sem over 3 runs); **e, f** Same as **a, b** for LSTM network; **c, d**, The LIF neurons of the network shown in **a, b** are replaced by LSTM neurons. The spiking patterns are similar (compare **a** to **c**). Thus, spiking characteristics of restricted peephole LSTM units and LIF units is similar. **g, h**, The neural network trained with LSTM units and evaluated with LIF units; The analysis shows both neuron models can be exchanged without changing the network accuracy providing evidence that spiking characteristics of both neuron models is very similar.

hand [56]. In addition, several biologically plausible algorithms to train spiking neural networks have been proposed both supervised [57], [58] and unsupervised [59]. Furthermore, some research groups developed methods to change LSTM networks so that they have the same efficiency as spiking neural networks in storing respectively processing temporal patterns. Thus, they prevent the LSTM units from exponentially losing information [60], [61]. Recently, it has been proven that LSTM units could even be replaced by spiking neuropmorphic hardware without losing their advantages [62].

C. Concluding Remarks

These example studies illustrate that there is a lot of potential in exploiting the advantages of spiking neural networks and to map the characteristics on established recurrent neuron models such as the LSTM unit. We could show that the peephole connections of the LSTM units are necessary to achieve LIF like behavior. We hope that these findings are an inspiration for the community to push forward the fields of neuroscience-inspired AI [1], [63]–[67] and cognitive computational neuroscience [42], [44], [68]–[71].

Acknowledgments

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation): grants KR 5148/2-1 (project number 436456810), KR 5148/3-1 (project number 510395418) and GRK 2839 (project number 468527017) to PK, and grant SCHI 1482/3-1 (project number 451810794) to AS..

References

[1] D. Hassabis, D. Kumaran, C. Summerfield, and M. Botvinick, "Neuroscience-inspired artificial intelligence," *Neuron*, vol. 95, no. 2, pp. 245–258, 2017.

[2] N. Perez-Nieves, V. C. Leung, P. L. Dragotti, and D. F. Goodman, "Neural heterogeneity promotes robust learning," *Nature communications*, vol. 12, no. 1, pp. 1–9, 2021.

[3] M. Ziegler, "Novel hardware and concepts for unconventional computing," 2020.

[4] V. Rybalkin and N. Wehn, "When massive gpu parallelism ain't enough: A novel hardware architecture of 2d-lstm neural network," in *The 2020 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays*, 2020, pp. 111–121.

[5] T. Wang, F. Zhao, J. Wan, and Y. Zhu, "A novel hardware architecture for rapid object detection based on adaboost algorithm," in *International Symposium on Visual Computing*. Springer, 2010, pp. 397–406.

[6] Y. Cao, Y. Chen, and D. Khosla, "Spiking deep convolutional neural networks for energy-efficient object recognition," *International Journal of Computer Vision*, vol. 113, no. 1, pp. 54–66, 2015.

[7] S. K. Esser, R. Appuswamy, P. Merolla, J. V. Arthur, and D. S. Modha, "Backpropagation for energy-efficient neuromorphic computing," in *Advances in neural information processing systems*, 2015, pp. 1117–1125.

[8] E. Stromatias, D. Neil, F. Galluppi, M. Pfeiffer, S.-C. Liu, and S. Furber, "Scalable energy-efficient, low-latency implementations of trained spiking deep belief networks on spinnaker," in *2015 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2015, pp. 1–8.

[9] Y. Wu, L. Deng, G. Li, J. Zhu, Y. Xie, and L. Shi, "Direct training for spiking neural networks: Faster, larger, better," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 1311–1318.

[10] B. Yin, F. Corradi, and S. M. Bohté, "Effective and efficient computation with multiple-timescale spiking recurrent neural networks," *arXiv preprint arXiv:2005.11633*, 2020.

[11] A. Tavanaei, M. Ghodrati, S. R. Kheradpisheh, T. Masquelier, and A. Maida, "Deep learning in spiking neural networks," *Neural Networks*, vol. 111, pp. 47–63, 2019.

[12] A. L. Hodgkin and A. F. Huxley, "A quantitative description of membrane current and its application to conduction and excitation in nerve," *The Journal of physiology*, vol. 117, no. 4, pp. 500–544, 1952.

[13] E. M. Izhikevich and R. FitzHugh, "Fitzhugh-nagumo model," *Scholarpedia*, vol. 1, no. 9, p. 1349, 2006.

[14] A. N. Burkitt, "A review of the integrate-and-fire neuron model: I. homogeneous synaptic input," *Biological cybernetics*, vol. 95, no. 1, pp. 1–19, 2006.

[15] Z. Wu, H. Zhang, Y. Lin, G. Li, M. Wang, and Y. Tang, "Liaf-net: Leaky integrate and analog fire network for lightweight and efficient spatiotemporal information processing," *arXiv preprint arXiv:2011.06176*, 2020.

[16] W. Fang, "Leaky integrate-and-fire spiking neuron with learnable membrane time parameter," *arXiv preprint arXiv:2007.05785*, 2020.

[17] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[18] I. Pozzi, R. Nusselder, D. Zambrano, and S. Bohté, "Gating sensory noise in a spiking subtractive lstm," in *International Conference on Artificial Neural Networks*. Springer, 2018, pp. 284–293.

[19] D. Salaj, A. Subramoney, C. Kraisnikovic, G. Bellec, R. Legenstein, and W. Maass, "Spike-frequency adaptation provides a long short-term memory to networks of spiking neurons," *bioRxiv*, 2020.

[20] A. Lotfi Rezaabad and S. Vishwanath, "Long short-term memory spiking networks and their applications," in *International Conference on Neuromorphic Systems 2020*, 2020, pp. 1–9.

[21] A. Koopman, M. Van Leeuwen, and J. Vreeken, "Dynamic neural networks, comparing spiking circuits and lstm," 2003.

[22] F. A. Gers and J. Schmidhuber, "Recurrent nets that time and count," in *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium*, vol. 3. IEEE, 2000, pp. 189–194.

[23] F. A. Gers and E. Schmidhuber, "Lstm recurrent networks learn simple context-free and context-sensitive languages," *IEEE Transactions on Neural Networks*, vol. 12, no. 6, pp. 1333–1340, 2001.

[24] T. Yang, H. Wang, S. Aziz, H. Jiang, and J. Peng, "A novel method of wind speed prediction by peephole lstm," in *2018 International Conference on Power System Technology (POWERCON)*. IEEE, 2018, pp. 364–369.

[25] A. R. Voelker, D. Rasmussen, and C. Eliasmith, "A spike in performance: Training hybrid-spiking neural networks with quantized activation functions," *arXiv preprint arXiv:2002.03553*, 2020.

[26] F. Chollet, *Deep Learning mit Python und Keras: Das Praxis-Handbuch vom Entwickler der Keras-Bibliothek*. MITP-Verlags GmbH & Co. KG, 2018.

[27] Y. LeCun, L. D. Jackel, L. Bottou, C. Cortes, J. S. Denker, H. Drucker, I. Guyon, U. A. Muller, E. Sackinger, P. Simard *et al.*, "Learning algorithms for classification: A comparison on handwritten digit recognition," *Neural networks: the statistical mechanics perspective*, vol. 261, no. 276, p. 2, 1995.

[28] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms," *arXiv preprint arXiv:1708.07747*, 2017.

[29] G. Cohen, S. Afshar, J. Tapson, and A. Van Schaik, "Emnist: Extending mnist to handwritten letters," in *2017 international joint conference on neural networks (IJCNN)*. IEEE, 2017, pp. 2921–2926.

[30] A. Krizhevsky, G. Hinton *et al.*, "Learning multiple layers of features from tiny images," 2009.

[31] F. Zenke and S. Ganguli, "Superspike: Supervised learning in multilayer spiking neural networks," *Neural computation*, vol. 30, no. 6, pp. 1514–1541, 2018.

[32] G. Datta, S. Kundu, and P. A. Beerel, "Training energy-efficient deep spiking neural networks with single-spike hybrid input encoding," in *2021 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2021, pp. 1–8.

[33] R. C. Gerum and A. Schilling, "Integration of Leaky-Integrate-and-Fire Neurons in Standard Machine Learning Architectures to Generate Hybrid Networks: A Surrogate Gradient Approach," *Neural Computation*, pp. 1–26, 07 2021. [Online]. Available: https://doi.org/10.1162/neco_a_01424

[34] S. v. d. Walt, S. C. Colbert, and G. Varoquaux, "The numpy array: a structure for efficient numerical computation," *Computing in science & engineering*, vol. 13, no. 2, pp. 22–30, 2011.

[35] M. Bostock, V. Ogievetsky, and J. Heer, "D^3 data-driven documents," *IEEE transactions on visualization and computer graphics*, vol. 17, no. 12, pp. 2301–2309, 2011.

[36] J. D. Hunter, "Matplotlib: A 2d graphics environment," *Computing in science & engineering*, vol. 9, no. 3, pp. 90–95, 2007.

[37] R. Gerum, "pylustrator: code generation for reproducible figures for publication," *Journal of Open Source Software*, vol. 5, no. 51, p. 1989, 2020.

[38] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015, software available from tensorflow.org. [Online]. Available: https://www.tensorflow.org/

[39] F. Zenke and T. P. Vogels, "The remarkable robustness of surrogate gradient learning for instilling complex function in spiking neural networks," *Neural computation*, vol. 33, no. 4, pp. 899–925, 2021.

[40] A. Schilling, A. Maier, R. Gerum, C. Metzner, and P. Krauss, "Quantifying the separability of data classes in neural networks," *Neural Networks*, vol. 139, pp. 278–293, 2021.

[41] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," in *Advances in neural information processing systems*, 2019, pp. 8026–8037.

[42] N. Kriegeskorte and P. K. Douglas, "Cognitive computational neuroscience," *Nature neuroscience*, vol. 21, no. 9, pp. 1148–1160, 2018.

[43] C. Hong, X. Wei, J. Wang, B. Deng, H. Yu, and Y. Che, "Training spiking neural networks for cognitive tasks: A versatile framework compatible with various temporal codes," *IEEE transactions on neural networks and learning systems*, vol. 31, no. 4, pp. 1285–1296, 2019.

[44] A. Schilling, R. Gerum, C. Metzner, A. Maier, and P. Krauss, "Intrinsic noise improves speech recognition in a computational model of the auditory pathway," *Frontiers in Neuroscience*, p. 795, 2022.

[45] S. J. Gershman, "What have we learned about artificial intelligence from studying the brain?" 2023. [Online]. Available: https://gershmanlab.com

[46] E. Jonas and K. P. Kording, "Could a neuroscientist understand a microprocessor?" *PLoS computational biology*, vol. 13, no. 1, p. e1005268, 2017.

[47] P. Krauss, A. Zankl, A. Schilling, H. Schulze, and C. Metzner, "Analysis of structure and dynamics in three-neuron motifs," *Frontiers in Computational Neuroscience*, vol. 13, p. 5, 2019.

[48] F. Bönsel, P. Krauss, C. Metzner, and M. E. Yamakou, "Control of noise-induced coherent oscillations in three-neuron motifs," *Cognitive Neurodynamics*, pp. 1–20, 2021.

[49] P. Krauss, M. Schuster, V. Dietrich, A. Schilling, H. Schulze, and C. Metzner, "Weight statistics controls dynamics in recurrent neural networks," *PloS one*, vol. 14, no. 4, p. e0214541, 2019.

[50] C. Metzner and P. Krauss, "Dynamics and information import in recurrent neural networks," *Frontiers in Computational Neuroscience*, vol. 16, 2022.

[51] P. Krauss, K. Prebeck, A. Schilling, and C. Metzner, "Recurrence resonance" in three-neuron motifs," *Frontiers in computational neuroscience*, vol. 13, p. 64, 2019.

[52] C. Metzner, M. E. Yamakou, D. Voelkl, A. Schilling, and P. Krauss, "Quantifying and maximizing the information flux in recurrent neural networks," *arXiv preprint arXiv:2301.12892*, 2023.

[53] A. Banino, C. Barry, B. Uria, C. Blundell, T. Lillicrap, P. Mirowski, A. Pritzel, M. J. Chadwick, T. Degris, J. Modayil *et al.*, "Vector-based navigation using grid-like representations in artificial agents," *Nature*, vol. 557, no. 7705, pp. 429–433, 2018.

[54] D. C. Rowland, Y. Roudi, M.-B. Moser, E. I. Moser *et al.*, "Ten years of grid cells," *Annu Rev Neurosci*, vol. 39, pp. 19–40, 2016.

[55] L. Songlin, D. Yangdong, and W. Zhihua, "Grid cells are ubiquitous in neural networks," *arXiv preprint arXiv:2003.03482*, 2020.
[56] W. Ponghiran and K. Roy, "Hybrid analog-spiking long short-term memory for energy efficient computing on edge devices," in *2021 Design, Automation & Test in Europe Conference & Exhibition (DATE)*. IEEE, 2021, pp. 581–586.
[57] G. Bellec, D. Salaj, A. Subramoney, R. Legenstein, and W. Maass, "Long short-term memory and learning-to-learn in networks of spiking neurons," *arXiv preprint arXiv:1803.09574*, 2018.
[58] G. Bellec, F. Scherr, A. Subramoney, E. Hajek, D. Salaj, R. Legenstein, and W. Maass, "A solution to the learning dilemma for recurrent networks of spiking neurons," *Nature communications*, vol. 11, no. 1, pp. 1–15, 2020.
[59] A. Lazar, G. Pipa, and J. Triesch, "Sorn: a self-organizing recurrent neural network," *Frontiers in computational neuroscience*, p. 23, 2009.
[60] W. Zheng, P. Zhao, G. Chen, H. Zhou, and Y. Tian, "A hybrid spiking neurons embedded lstm network for multivariate time series learning under concept-drift environment," *IEEE Transactions on Knowledge and Data Engineering*, 2022.
[61] Q. Liu, L. Long, Q. Yang, H. Peng, J. Wang, and X. Luo, "Lstm-snp: A long short-term memory model inspired from spiking neural p systems," *Knowledge-Based Systems*, vol. 235, p. 107656, 2022.
[62] A. Rao, P. Plank, A. Wild, and W. Maass, "A long short-term memory for ai applications in spike-based neuromorphic hardware," *Nature Machine Intelligence*, vol. 4, no. 5, pp. 467–479, 2022.
[63] R. C. Gerum, A. Erpenbeck, P. Krauss, and A. Schilling, "Sparsity through evolutionary pruning prevents neuronal networks from overfitting," *Neural Networks*, vol. 128, pp. 305–312, 2020.
[64] Z. Yang, A. Schilling, A. Maier, and P. Krauss, "Neural networks with fixed binary random projections improve accuracy in classifying noisy data," in *Bildverarbeitung für die Medizin 2021*. Springer, 2021, pp. 211–216.
[65] A. Maier, H. Köstler, M. Heisig, P. Krauss, and S. H. Yang, "Known operator learning and hybrid machine learning in medical imaging—a review of the past, the present, and the future," *Progress in Biomedical Engineering*, 2022.
[66] P. Stoewer, C. Schlieker, A. Schilling, C. Metzner, A. Maier, and P. Krauss, "Neural network based successor representations to form cognitive maps of space and language," *Scientific Reports*, vol. 12, no. 1, p. 11233, 2022.
[67] P. Stoewer, A. Schilling, A. Maier, and P. Krauss, "Neural network based formation of cognitive maps of semantic spaces and the putative emergence of abstract concepts," *Scientific Reports*, vol. 13, no. 1, p. 3644, 2023.
[68] P. Krauss, K. Tziridis, C. Metzner, A. Schilling, U. Hoppe, and H. Schulze, "Stochastic resonance controlled upregulation of internal noise after hearing loss as a putative cause of tinnitus-related neuronal hyperactivity," *Frontiers in neuroscience*, vol. 10, p. 597, 2016.
[69] A. Schilling, R. Tomasello, M. R. Henningsen-Schomers, A. Zankl, K. Surendra, M. Haller, V. Karl, P. Uhrig, A. Maier, and P. Krauss, "Analysis of continuous neuronal activity evoked by natural speech with computational corpus linguistics methods," *Language, Cognition and Neuroscience*, vol. 36, no. 2, pp. 167–186, 2021.
[70] A. Schilling, W. Sedley, R. Gerum, C. Metzner, K. Tziridis, A. Maier, H. Schulze, F.-G. Zeng, K. J. Friston, and P. Krauss, "Predictive coding and stochastic resonance: Towards a unified theory of auditory (phantom) perception," *arXiv preprint arXiv:2204.03354*, 2022.
[71] K. Surendra, A. Schilling, P. Stoewer, A. Maier, and P. Krauss, "Word class representations spontaneously emerge in a deep neural network trained on next word prediction," *arXiv preprint arXiv:2302.07588*, 2023.

Neural Networks 128 (2020) 305–312

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

Sparsity through evolutionary pruning prevents neuronal networks from overfitting

Richard C. Gerum [a], André Erpenbeck [b], Patrick Krauss [c,d,e], Achim Schilling [c,d,*]

[a] Biophysics Group, Department of Physics, Friedrich Alexander University Erlangen-Nürnberg (FAU), Germany
[b] The Raymond and Beverley Sackler Center for Computational Molecular and Materials Science, School of Chemistry, Tel Aviv University (TAU), Israel
[c] Neuroscience Lab, Experimental Otolaryngology, University Hospital Erlangen, Germany
[d] Cognitive Computational Neuroscience Group at the Chair of English Philology and Linguistics, Friedrich-Alexander University Erlangen-Nürnberg (FAU), Germany
[e] Department of Otorhinolaryngology/Head and Neck Surgery, University of Groningen, University Medical Center Groningen (UMCG), The Netherlands

ARTICLE INFO

Article history:
Received 8 November 2019
Received in revised form 31 January 2020
Accepted 4 May 2020
Available online 11 May 2020

Keywords:
Evolution
Artificial neural networks
Maze task
Evolutionary algorithm
Overfitting
Biological plausibility

ABSTRACT

Modern Machine learning techniques take advantage of the exponentially rising calculation power in new generation processor units. Thus, the number of parameters which are trained to solve complex tasks was highly increased over the last decades. However, still the networks fail – in contrast to our brain – to develop general intelligence in the sense of being able to solve several complex tasks with only one network architecture. This could be the case because the brain is not a randomly initialized neural network, which has to be trained from scratch by simply investing a lot of calculation power, but has from birth some fixed hierarchical structure. To make progress in decoding the structural basis of biological neural networks we here chose a bottom-up approach, where we evolutionarily trained small neural networks in performing a maze task. This simple maze task requires dynamic decision making with delayed rewards. We were able to show that during the evolutionary optimization random severance of connections leads to better generalization performance of the networks compared to fully connected networks. We conclude that sparsity is a central property of neural networks and should be considered for modern Machine learning approaches.

1. Introduction

Sparsity is a characteristic property of the wiring scheme of the human brain, which consists of about 8.6×10^{10} neurons (Herculano-Houzel, 2009), interconnected by approximately 10^{15} synapses (Hagmann et al., 2008; Sporns, Tononi, & Kötter, 2005). Thus, from almost 10^{22} theoretically possible synaptic connections, only one of 10 million possible connections is actually realized. This extremely sparse distribution of both neural connections and activity patterns is not a unique feature of the human brain (Hagmann et al., 2008) but can also be found in other vertebrate species such as for example mice and rats (Kerr, Greenberg, & Helmchen, 2005; Oh et al., 2014; Perin, Berger, & Markram, 2011; Song, Sjöström, Reigl, Nelson, & Chklovskii, 2005). Even evolutionary very old species with a quite simple nervous system such as the nematode *C. elegans* with only 302 neurons (Jarrell et al., 2012) and over 7000 connections (White, Southgate, Thomson, & Brenner, 1986) show this sparsity. Besides the described sparsity of virtually all nervous systems, many biological neural networks also show small world properties (Amaral, Scala, Barthélémy, & Stanley, 2000; Bassett & Bullmore, 2006; Latora & Marchiori, 2001; Watts & Strogatz, 1998), such as scale free connectivity patterns (Bassett, Meyer-Lindenberg, Achard, Duke, & Bullmore, 2006; Perin et al., 2011; van den Heuvel & Yeo, 2017).

When dealing with the term sparsity in biology and machine learning, it is necessary to distinguish between three different forms of sparsity: "sparse representation", "input sparsity", and "model sparsity" (Kafashan, Nandi, & Ching, 2016). Sparse representation means that only a small amount of neurons respond to certain stimuli. Thus, even a fully connected network can have the property of sparse representation. One famous but controversial example for sparse representation, often called sparse-coding (Zhang, Yang, & Feng, 2011), is the so called "grandmother cell", being a vivid term for the idea that only a single neuron encodes for one highly complex concept (Barwich, 2019; Quiroga, Kreiman, Koch, & Fried, 2008; Rose, 1996). In recent years much effort has been undertaken to achieve sparse coding of certain

* Correspondence to: Neuroscience Group, Experimental Otolaryngology, Friedrich-Alexander University of Erlangen-Nürnberg, Waldstrasse 1, 91054 Erlangen, Germany.
E-mail address: achim.schilling@uk-erlangen.de (A. Schilling).

https://doi.org/10.1016/j.neunet.2020.05.007

input stimuli to improve machine learning (Babadi & Sompolinsky, 2014; Dasgupta, Stevens, & Navlakha, 2017; Jiao et al., 2018; Jin, Zhou, Gao, & Zhang, 2018; Olshausen & Field, 2004; Pehlevan & Sompolinsky, 2014) and on the other hand sparse coding was also investigated in biology (Crochet, Poulet, Kremer, & Petersen, 2011; Zaslaver et al., 2015). Input sparsity, in contrast, means that the input patterns fed to the neural network are sparse. However, in this study we investigate the development of model sparsity in artificial neural networks, being the analogue to a sparse connectome in biology. For reasons of simplicity, we use from now on the term sparsity to refer to model sparsity in the biological sense, as well as in the context of machine learning.

Sparsity in biology is the result of both, phylogenetic and ontogenetic adaptations. Even though, almost all species' immature nervous systems are already very sparse, this sparsity is even further increased during development and maturation of the agents' nervous systems (Low & Cheng, 2006). In fact, the infant human brain contains two times more synapses than the adult brain (Kolb & Gibb, 2011). Analogously, the immature nervous system of *C. elegans* contains more synapses than the adult form (Oren-Suissa, Bayer, & Hobert, 2016).

But pruning is not restricted to axons and synaptic connections. It even extends to the total number of neurons, which also decreases during development. For instance, the immature nervous system of *C. elegans* initially consists of 308 neurons (Chalfie, 1984), whereas the adult form contains only 302 neurons (Jarrell et al., 2012). And also in humans, the number of neurons decreases during development (Yeo & Gautier, 2004). These ontogenetic changes are referred to as pruning (Paolicelli et al., 2011), and it seems to be a universal phenomenon for all species from *C. elegans* to humans. Furthermore, pruning is found to be mandatory for healthy development (Hong, Dissing-Olesen, & Stevens, 2016). In cases where normal synaptic pruning fails, this may even lead to disorders like schizophrenia (Boksa, 2012).

Since sparse connectivity architectures are realized on all scales and in a vast number of agents of different complexity, it can be assumed that sparse connectivity is a general principle in neural information processing systems, leading to advantages compared to densely connected networks. One major advantage of sparse artificial deep neural networks used for image classification in comparison to fully connected networks, is the reduction of computational costs while at the same time boosting the ability to generalize (Anwar, Hwang, & Sung, 2017; Han, Mao, & Dally, 2015; Mocanu et al., 2018; Wen, Wu, Wang, Chen, & Li, 2016).

However, these pure feed forward network architectures show low biological plausibility as they neither have the ability of dealing with time series data, nor have any memory-like features. Efficient processing of time series data in artificial neural networks is a complex task with a bunch of limitations. The technique of training the neural networks by unfolding the data in time is computational expensive and time consuming and leads to effects such as vanishing or exploding gradients (Hochreiter, 1998; Pascanu, Mikolov, & Bengio, 2012), which have been partly overcome by the introduction of Long–Short-Term-Memory Networks (Schmidhuber & Hochreiter, 1997). However, these networks are difficult to interpret from a biological point of view.

To overcome these limitation, novel biologically inspired approaches for processing time series data were introduced called reservoir computing (Lukoševičius, Jaeger, & Schrauwen, 2012; Verstraeten, Schrauwen, D'Haene, & Stroobandt, 2007). A so called reservoir of neurons with fixed (i.e. not adjusted by training) random recurrent connections is used to calculate higher-order correlations of the input signal which serve as input for a feed forward output layer that is trained with error back-propagation. The properties of the reservoir networks were found to be ideal for biologically inspired parameters with a high sparsity (Alexandre, Embrechts, & Linton, 2009). Thus, these reservoirs work best at the edge of chaos, meaning that the parameters have to be chosen so that they are balanced between complete chaos and absolute periodicity (Bertschinger & Natschläger, 2004; Krauss et al., 2019b; Schrauwen, Verstraeten, & Van Campenhout, 2007).

Much effort has been undertaken to apply the technique of reservoir computing on tasks with delayed rewards such as robot navigation in mazes (Antonelo & Schrauwen, 2012; Antonelo, Schrauwen, & Stroobandt, 2007).

However, the technique of reservoir computing is still based on the fact that the output layer has to be trained in a supervised way using back-propagation and, thus, complex tasks with a delayed reward are difficult to realize. Thus, in this study we used an evolutionary approach to train networks in solving a maze task. Although, reinforcement learning techniques – especially deep reinforcement learning – would be suitable for this kind of hidden-state Markov process, this approach lacks biological plausibility. Thus, deep reinforcement learning needs further techniques for hyper-parameter tuning such as Bayesian optimization (Springenberg, Klein, Falkner, & Hutter, 2016). However, the aim of this study is to analyze self-evolving networks and to characterize the architecture to derive basic principles, which are of relevance for biological systems. Nevertheless, evolutionary techniques could be used to optimize reinforcement learning as well (Young, Rose, Karnowski, Lim, & Patton, 2015).

In our evolutionary system we were able to show that the random severing of connections (evolutionary pruning), without explicitly rewarding sparsity, did lead to a general sparsification of the networks and a better generalization performance. Furthermore, we could demonstrate that evolutionary training works best when the probability for destroying connections is higher but in the same range of the probability for recreating connections.

The paper is structured as follows: In the Method section we describe the used software resources, how the maze task is implemented, the evolutionary algorithm, and the initial neural network architecture. The Results section starts with the analysis of the effect of the different sparsification methods on the performance of the neural networks. We furthermore analyze the architecture of the evolved networks and demonstrate the effect of severance probability and reconnection probability. We added a Conclusion section summarizing the main results, and provide a discussion on the limitations of the study and possible future research directions (Discussion).

2. Materials and methods

2.1. Software resources

All simulations were run on a desktop computer equipped with an i9 extreme processor (Intel) with 10 calculation cores. The complete software was written in Python 3.6 using the libraries sys, os, glob, subprocess, json, natsort, pickle, shutil, NumPy (Van Der Walt, Colbert, & Varoquaux, 2011). Data visualization was done using Matplotlib (Hunter, 2007) and plots were arranged using the Pylustrator (Gerum, 2019).

2.2. Maze task

The task for the agents to perform is a maze based on a rectangular grid of 400 × 22 cells (Fig. 1c). The maze/obstacle task we use here, is similar to the maze task described by Sanchez and coworkers (Sanchez, Pérez-Uribe, & Mesot, 2001), which they also use to analyze the performance of evolutionary approaches.

In our maze task there exist two types of cells, free cells and wall cells. Free cells can be entered and wall cells not. The border of the maze consists of walls to prevent agents from leaving the maze. Starting from the left, every 2 to 10 cells a wall with a

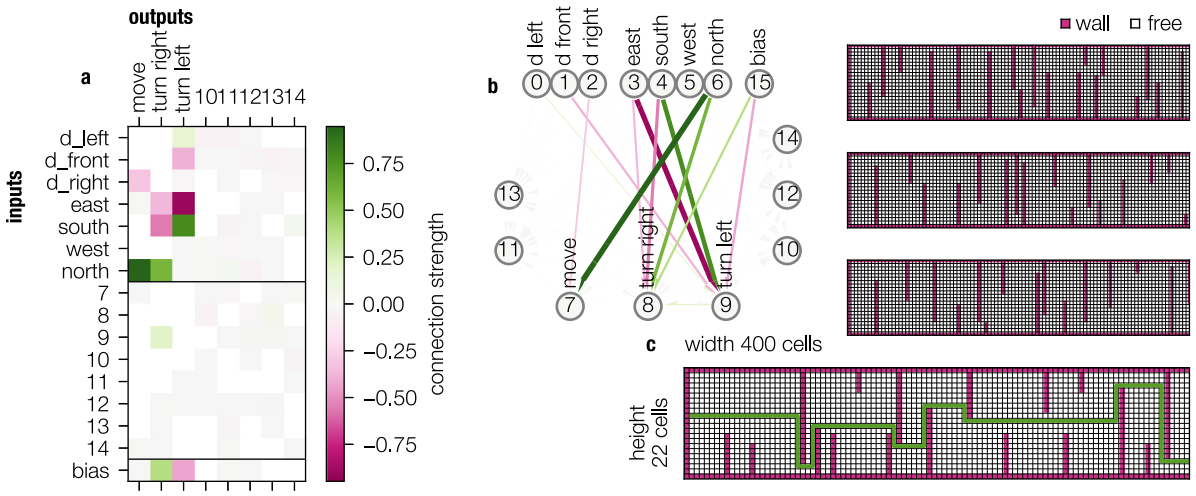

Fig. 1. Example network and mazes. a, Weight matrix of an exemplary network. **b**, The same network displayed as connections. c, Mazes are 400 cells wide and 22 high. Walls (pink) are at all borders and randomly placed in between. The green line depicts one ideal path though the network. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

length between 4 to 20 is inserted. With a probability of 0.25 the wall is inserted from the same side (up or down) as the last wall or with a probability of 0.75 it is inserted from the other side.

Agents start always at the left end of the maze facing to the right. As 'sensory' input each agent receives the distance to the wall in front, to the left and to the right (input neurons 0 to 6, cf. Fig. 1a).

If the distance is larger then 10, it is set to 10 (visual range). It also receives the direction, it is currently looking at, as a one-hot encoded, four-neuron input, i.e. one neuron at a time is in state 1 and the others are in state 0. This input serves as a kind of compass. The seven input neurons do exclusively receive input from the environment, but do not get any input from other neurons, thus, they are reset at each time step.

The agent can output three values for the three possible actions: go straight, turn right, or turn left. The action with the highest value is selected (winner takes all) (output neurons 7 to 9, cf. Fig. 1a). When the agent chooses to go straight and the next field is a wall, it is not moved.

After 400 actions, the covered x-distance of the agent is fed to the fitness function which is proportional to the covered distance (cf. Fig. 1c). Thus, the reward is delayed by 400 time steps.

2.3. Network

The logic of each agent consists of a fully connected network of $N = 16$ neurons (identified as the number of neurons leading to the best performance with fast convergence, cf. Supplements S8, S9, S10), with states s (cf. Fig. 1a,b). The connection weights W are initialized with a random value drawn from a uniformly distribution from the interval $[-\sigma, +\sigma]$ with $\sigma = 4 \cdot \sqrt{\frac{6}{N+N}}/10$. For each time step t in the task, the $3 + 4$ input values (distances (left, front, right) and one-hot encoded direction) are set as the states of the neurons #0 to #6. Then one processing pass through the network is calculated, $s_{t+1} = \mathrm{ReLu}\,(W \cdot s_t)$ (s_t: state of the network at time point t), with $\mathrm{ReLu}(x)$ being the rectified linear function

$$\mathrm{ReLu}(x) = \begin{cases} 0 & x \leq 0 \\ x & x > 0 \end{cases} \tag{1}$$

The states s of the neurons #7 to #9 are used as outputs to choose one of the three possible actions to perform in the maze task (connectivity matrix see Fig. 1a). Thus, the action is determined by a winner-takes all-method (in the case of no activation, the "move" action is chosen). Our approach is a policy based approach, as the output of the network directly is the action to take and no quality assessment of different states is undertaken.

2.4. Evolutionary algorithm

For optimizing the networks to fulfill the maze task, we use an evolutionary algorithm (Fekiac, Zelinka, & Burguillo, 2011). Therefore, a pool of 1000 agents (for simulations with different population sizes cf. Supplements S11) is created with a random initialization and 10 mazes are created for the agents to be trained on.

For each iteration, all agents have to perform the maze task and are assigned a fitness, depending on their score in the task. The best half of the agent pool has now the chance to create offspring. The probability to generate an offspring is proportional to their relative fitness compared to the other agents. Agents with a probability of 10% or more are set to a maximal probability of 10% to retain biodiversity. For each of the old agents to be replaced, a parent agent is selected at random according to their reproduction probabilities. Agents can have multiple offspring or no offspring at all.

After offspring generation, each agent is mutated. We used three different mutation types:

- **Weight mutation**: The connection weights W are each mutated by addition of a Gaussian distributed random variable ($\mu = 0$: mean of distribution, $\sigma = |\sigma_{\mathrm{mut}}|$: standard deviation).
- **Mutation rate change**: The mutation rate σ_{mut} is also mutated by multiplication with a Gaussian distributed random variable ($\mu = 1$, $\sigma = \sigma_{\mathrm{mut}}$).
- **Connection mutation**: Existing connections are removed with a probability $p_{\mathrm{disconnect}}$ and non-existing connections are added with a probability of $p_{\mathrm{connect}} = p_{\mathrm{disconnect}}$. Removed connections have a weight of 0 and are not subject to weight mutations. Thus, a removed connection cannot be recovered by a simple mutation step, but can only be recovered by a reconnection mutation.

The fitness (Eq. (2)) is calculated from the squared mean of the square root distances the agent reached in all 10 training mazes (SMR, Eq. (3), this is done to favor generalizing agents which perform okay in all mazes against a specializing agent which

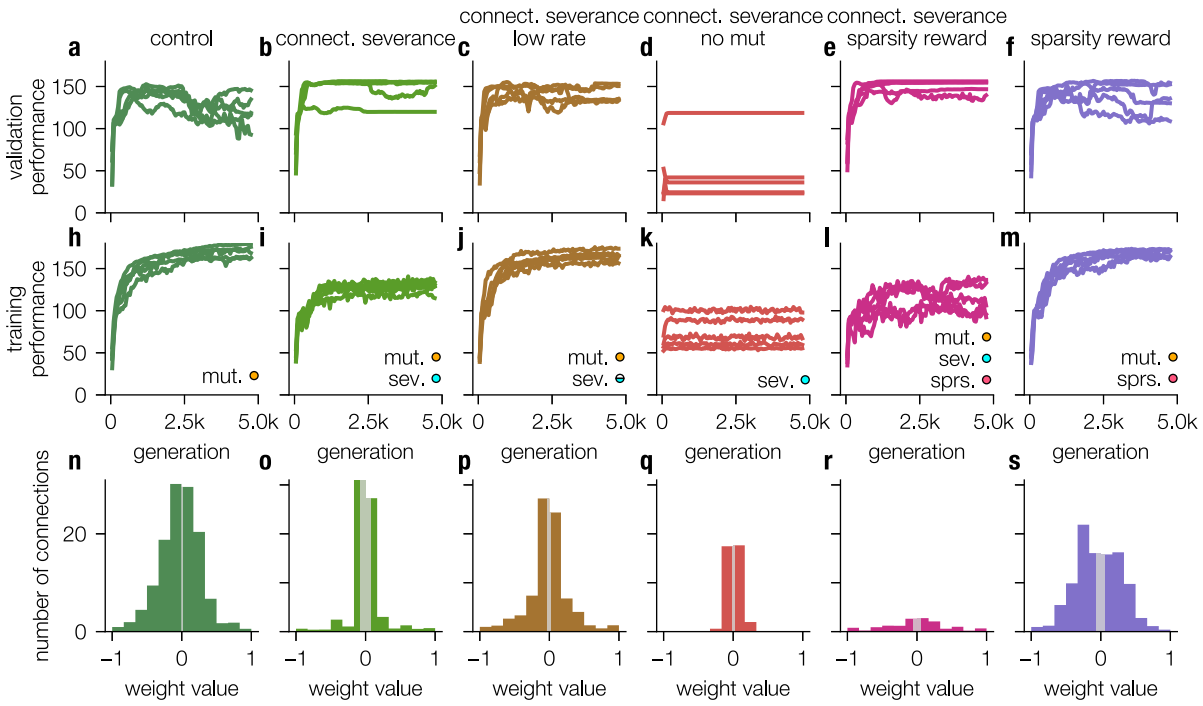

Fig. 2. Performance of networks of the different experiments during training and validation. a–f, Performance over generations for different experiments in 10 validation mazes. Curves show 5 different training seeds, evaluated on the same 10 validation mazes. **h–m**, Performance during training over generations for the 5 different training seeds. Labels stand for the different properties of the condition: "mut." for mutation of the weights, "sev." for severing/restoring connections, "sprs.". for adding a sparsity reward to the fitness function. **n–s**, Histograms of the connection weights for the different experiments. Zero connections are not included in the histograms. Gray shaded areas in the center indicate which weights can be removed without reducing the performance.

Table 1
Settings for the different experiments. The initial mutation rate σ_{mut}, the probability of removing or restoring a connection p_{connect} and the sparsity reward factor f_{sparsity}.

Name	σ_{mut}	p_{connect}	f_{sparsity}
Control	0.01	–	–
Connection severance	0.01	0.01	–
Connection severance	0.01	0.001	–
Connection severance no mut.	–	0.01	–
Connection severance sparsity reward	0.01	0.01	0.1
Sparsity reward	0.01	–	0.1

performs well in only one maze), the maximum mean activation of the neurons and optionally from the mean number of active connections:

$$\text{fitness} = \text{SMR}_{\text{mazes}} - \max_{\text{mazes}}(\text{mean}(\text{activation})) + f_{\text{sparsity}} \cdot \text{sparsity} \quad (2)$$

$$\text{SMR}_{\text{mazes}} = \left(\frac{1}{N_{\text{mazes}}} \cdot \sum_{i=1}^{N_{\text{mazes}}} \sqrt{d_i} \right)^2 \quad (3)$$

($\text{SMR}_{\text{mazes}}$): Root-Mean-Square-Distance, d_i: covered distance in maze i, activation: average activation of neural network, N_{mazes}: number of mazes, f_{sparsity}: sparsity reward factor. All experiments (for the initial parameters see Table 1) are repeated for 5 different seeds of the random number generator that is used to obtain the initial weights and the mutations. The different repetitions were performed on the same 10 training mazes to keep them comparable. The performance of the networks is defined as the average distance covered in the mazes after 400 time-steps. The validation performance is the average distance in unseen validation mazes, whereas the training performance is the average distance in the ten known training mazes.

3. Results

The evolutionary algorithm was able to find solutions enabling the agents to efficiently navigate through the mazes. In all experiments, except the experiment without weight mutations, the agents gradually learned to perform better in the maze tasks over the generations (Fig. 2h–m). The convergence was quite slow, as about 5000 generations were needed to converge to a stable solution. The convergence behavior was mostly independent of the seed of the random number generator, except for the "no mutation" condition, which relied strongly on the initial weights. The performance during training was best for the conditions with no, or low sparsification pressure (Fig. 2h,j,m).

The fitness in validation mazes, that the agents had not seen during training, was more sensitive to the seed than the performance during training. For the experiments with more sparsification pressure (Fig. 2b,e) the validation performance did exceed the fitness during training, showing good generalization, whereas the experiments with lower sparsification pressure (Fig. 2a,c,f) showed more problems with over-fitting, which means that the performance is higher during training than during validation.

Validation performance increased in most runs during training, although some drops in validation performance were observed, which sometimes recovered after a few generations, but in some cases the validation performance continued to fluctuate (see Supplements S2–S6). In general when the mean validation fitness dropped also the variability of the performance/fitness increased, showing that even if some agents found an "overfitting" solution, it was not quickly adopted by all agents, whereas when the solutions were more general, they seemed more stable and were adopted by the whole pool of agents.

The weight distribution of the final generation was different for each experiment. While the experiments with low sparsification pressure, that also showed overfitting, show more connections and more weights with larger absolute values (Fig. 2n,p,s)

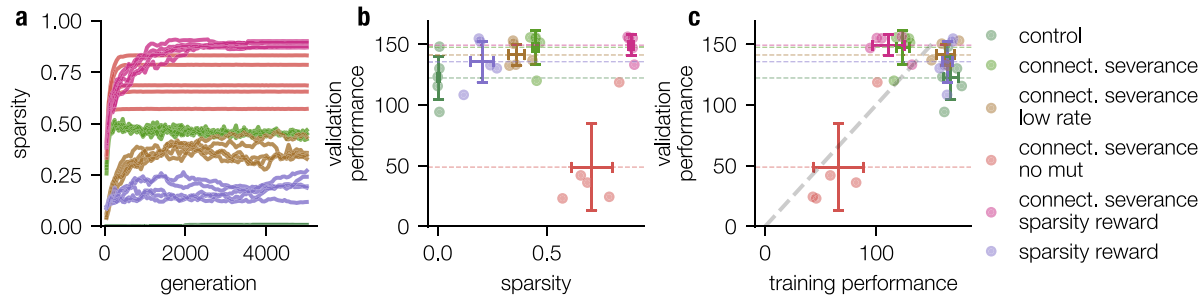

Fig. 3. Correlation of validation performance to sparsity and training performance. a, Sparsity $\left(1-\frac{n_{\text{non-zero}}}{n_{\text{possible}}}\right)$ as a function of the epochs (generations, sparsity is exclusively gained by evolutionary pruning). **b**, Correlation of validation performance to sparsity. Except for the case of "connection severance no mutation" validation performance increases on average with increased sparsity. **c**, Correlation of validation performance to training performance. Higher training performance leads in all experiments also to higher validation performance, indicating that none of the networks runs into severe overfitting.

compared to the other experiments that show more small weight values (Fig. 2o,q). Apart from "connection severance no mut" and "connection severance sparsity reward" (48% and 49% negative weights) all experiments show more negative weights (52%–55%), referring to inhibitory synapses, a fact which indicates more interesting behavior and more efficient information processing (Krauss et al., 2019b).

In most experiments, the sparsity increased over time (Fig. 3a), but in "connection severance" it even slightly decreased after the initial rise and in "sparsity reward" the sparsity fluctuated strongly over time. A higher sparsity was in all cases (except the no mutation case) associated with a higher validation performance (Fig. 3b), showing that sparsity improves the generalization behavior of evolutionary trained networks. A comparison of the training performance to the validation performance also shows that for the sparser cases, the training fitness decreases and in contrast to that, the validation performance increases. Therefore, the sparsification prevents overfitting and enhances the generalization properties of the networks. In addition, it improves computation efficiency in the evolutionary trained networks as the computations, divided between many neurons in the fully connected control group, are, in sparser networks, forced to be carried out on a small subset of neurons.

Furthermore, it could be shown that the networks which perform best in the test mazes develop simple feed-forward structures (cf. Fig. 1a). Additionally, some asymmetry in the connectivity matrix can be observed (cf. Fig. 1b). On the one hand, the bias units prefer the turn towards a certain direction. On the other hand, the connection from the input distance sensor to the turn output (e.g. to turn to the left side d_{left} in example Fig. 4b) is over-represented for one side. Thus, the networks have a preference to go to one side (e.g. to turn left, in the example in Fig. 4b), if there is enough space. The bias unit serves as counterpart, if the agent moves along the upper edge (resp. left side seen when moving along the x-axis) and guarantees that the network can walk away from the wall. The simple network architectures allow for the analysis of the functional tasks of certain neurons. This understanding of the functional tasks of neurons in artificial neural networks could potentially help to understand biological neural networks (Jonas & Kording, 2017; Kriegeskorte & Douglas, 2018). Different experiments develop quite different solutions to solve the maze task (see Supplementary Material S1). Interestingly, in some experiments similar structures emerge, regardless of the seed. This is especially the case in the "connection severance" experiment, where 4 out of 5 solutions are strikingly similar. This hints at the existence of strongly attractive maxima in the space of possible solutions.

Furthermore, we were able to show that training works best, in terms of validation and training performance, when the probability for removing existing connections $p_{\text{disconnect}}$ is in the same range of the probability of recreating removed connections p_{connect} (Fig. 5). Thus, even when $p_{\text{disconnect}} = p_{\text{connect}}$ generalization performance of the networks is increased. Sparsity of the networks can additionally be increased by deleting all unused connections, which do not lead to any changes in performance (combination of pruning and deletion after training, c.f. Fig. 5). This, procedure can be sophisticated as the random severance of connections does not prevent the development of unconnected subnetworks, which should be finally deleted to remove pseudo-complexity of the neural networks. This could be done by simply thresholding the connection weights (c.f. Supplements Fig. S14). However, gradually thresholding the network connections does not help to significantly increase the sparsity of fully connected networks (c.f. control condition in Fig. S14) as these networks were not forced to distribute the information processing over a smaller amount of neurons by the random pruning procedure.

4. Conclusion

In this study we showed that evolutionary pruning of artificial neural networks, evolutionary trained to solve a simple maze task, leads to sparser networks with better generalization properties compared to dense networks trained without pruning. The evolutionary pruning is realized via two mutation mechanisms: First the networks can set a threshold defining the upper limit of the absolute weight value being virtually set to zero. Secondly, some connections are removed (weight set to 0) and also restored at random with given probabilities p_{connect}, $p_{\text{disconnect}}$. The random removal of existing connections leads to more robust networks with better generalization abilities. This effect works best when $p_{\text{connect}} \leq p_{\text{disconnect}}$ and both probabilities are in the same range. In conclusion, evolutionary pruning by random severance of connections can be used as additional mechanism to improve the evolutionary training of neural networks.

5. Discussion

5.1. Limitations

The maze task is still not complex enough to trigger the development of more complex recurrent neural networks which include abilities such as memory. As the agents in the maze are always provided with a "compass" meaning always know in which direction they are pointing, the task can be solved with a Markov like decision process, as the information of one time step is enough to decide the next action. Thus, after 5000 epochs the best performances are achieved by simple feed forward networks (cf. Fig. 4). Nevertheless, it has to be considered that these networks were not forced to develop feed forward architectures but are a result of the Markov properties of the task they were trained

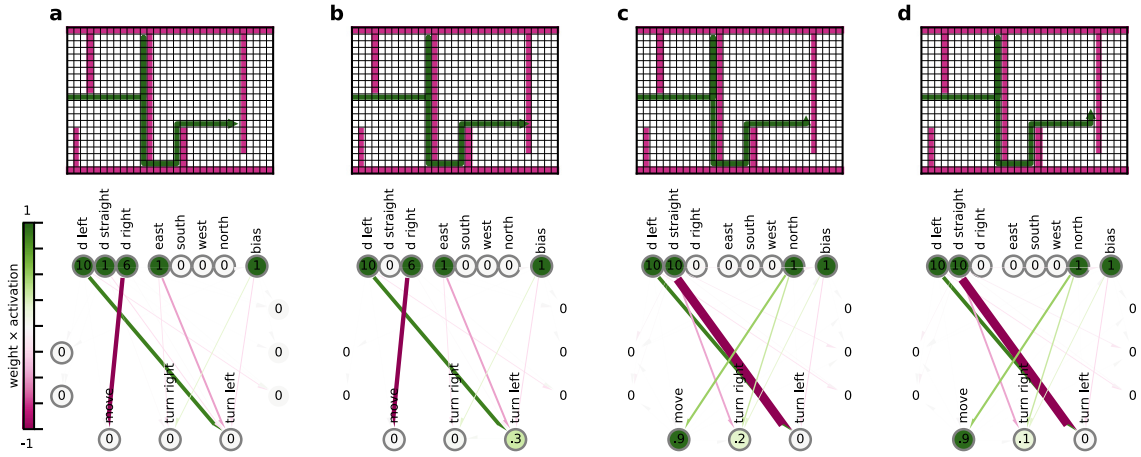

Fig. 4. State of the maze and configuration of the network during a turn. The position and orientation in the maze are denoted by the green triangle. The past trajectory by the green line. Below, the current state of the network is visualized. The values correspond to the current activation of each node and the colored connections to the weight times the activation (note that the weights are static) of the target node (pink negative, green positive). **a**, Network one step before turn. **b**, Network encounters a wall and turns. **c**, Network has just turned and is now heading north. **d**, Network continues walking straight in the new direction. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

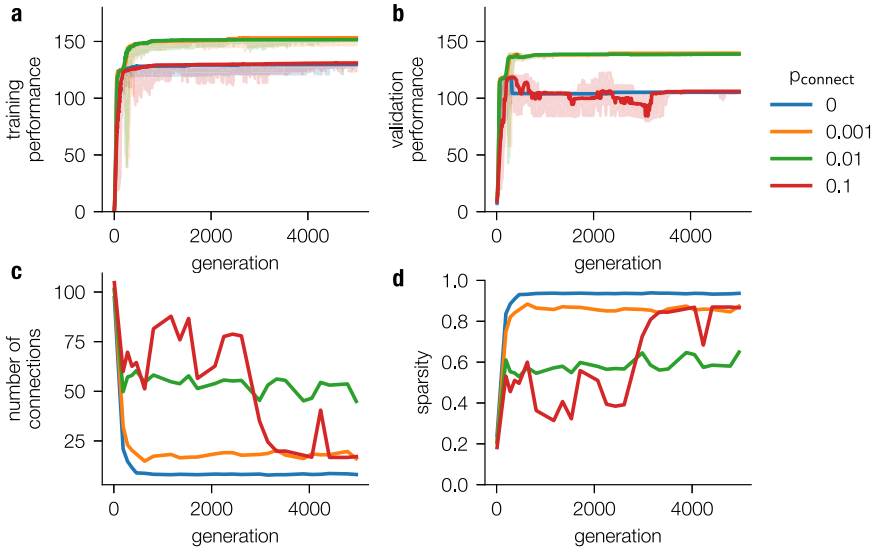

Fig. 5. Training/test performance as a function connection probability p_{connect} and disconnection probability $p_{\text{disconnect}}$ ratio. The training (a) and test (b) performance of the agents as a function of the generation for different values of p_{connect} (0, 0.001, 0.01, 0.1) and a fixed $p_{\text{disconnect}} = 0.01$. Best performance is achieved when $p_{\text{connect}} \leq p_{\text{disconnect}}$ and both are in the same range. (c, d), The sparsity expressed as the number of connections (c) and as a relative count of realized connections (d). A connection is counted active, if the fitness is not decreased when the connection is disabled. Thus, the sparsity of the networks is reduced by deleting sub-networks, which are not used (analogously to Supplements S14).

on. The simulation shows that simple feed forward networks are able to navigate through a maze using only 16 sparsely connected neurons.

5.2. Future research directions

Thus, the here described networks demonstrate that only a small number of neurons is needed to navigate through environment and shows that for example *C. elegans* with its 302 neurons (White et al., 1986) should be indeed able to perform complex tasks. It has already been shown that *C. elegans* shows thermo- and electrotaxis (Gabel et al., 2007) which are not simply a biased random walk in contrast to its' chemotaxis (for certain temperatures thermotaxis also changes to a random walk) (Pierce-Shimomura, Morse, & Lockery, 1999; Ryu & Samuel, 2002). *C. elegans* shows a direct movement along the electric gradient (Gabel et al., 2007). Thus, the amphid sensory neurons of *C. elegans* could be seen as and equivalent to the "compass" neurons in our simulation.

The here described study does not show a biologically inspired neuron configuration, however, demonstrates that 16 neurons are enough to perform relatively simple navigation tasks and gives a hint that the development of sparsity has evolutionary advantages. These findings contrast with current developments in the field of artificial intelligence where the size of the networks due to higher calculation power are scaled up (Xu et al., 2018).

Furthermore, the maze task could be extended so that the task is not a simple Markov process, which means that the next decision cannot be made by simply analyzing the current position in the maze (Littman, 2012; Wierstra & Wiering, 2004). This increase

of complexity can be achieved e.g. by removing the compass neurons. Consequently, a neural network which is able to achieve similar performance than networks with compass neurons need to develop some "memory" features. The networks would have to remember which movements they recently executed and they would have to dynamically recall this information. An even more demanding task would be to force the agents to go directly back to the starting point after they passed the maze. This task requires path integration (Etienne & Jeffery, 2004; Wehner & Wehner, 1986), which in turn requires the ability to flexibly navigate in physical space like insects (Andel & Wehner, 2004; Müller & Wehner, 1988, 1994; Wehner & Wehner, 1986), and at least in mammals episodic memory (Etienne, Maurer, & Séguinot, 1996; McNaughton, Battaglia, Jensen, Moser, & Moser, 2006; Séguinot, Cattet, & Benhamou, 1998). It has been demonstrated that these abilities can only be achieved within highly recurrent networks, as they can be found in the hippocampus (Etienne & Jeffery, 2004). Recently it has been shown that the network architecture of the hippocampus is not limited to spatial navigation, but seems to be domain-general (Aronov, Nevers, & Tank, 2017; Killian & Buffalo, 2018; Nau, Schröder, Bellmund, & Doeller, 2018) and even allows navigation in abstract high-dimensional cognitive feature spaces (Bellmund, Gärdenfors, Moser, & Doeller, 2018; Constantinescu, O'Reilly, & Behrens, 2016; Eichenbaum, 2015; Garvert, Dolan, & Behrens, 2017; Theves, Fernandez, & Doeller, 2019). Future work will have to investigate, whether networks with the above mentioned abilities can also be found by evolutionary algorithms.

Further potential research directions could be to optimize network architectures using evolutionary algorithms to find efficient neural networks, which can be trained supervisely or via reinforcement learning. The counterpart to this approach would be the analysis of evolutionary trained sparse neural networks, and to search for functional units such as motifs (Krauss, Zankl, Schilling, Schulze and Metzner, 2019c), weight statistics (Krauss et al., 2019b), or to analyze complex dynamics like 'recurrence resonance' effects (Krauss, Prebeck, Schilling and Metzner, 2019a), which is up to now often done in untrained networks, that perform no real information processing.

These two strands, would be in line with the philosophy that artificial and biological intelligence are "two sides of the same coin" (Kriegeskorte & Douglas, 2018; Schilling et al., 2018), and that the understanding of brain mechanisms on the one hand and the development of artificial neural network algorithms on the other hand are an iterative process, stimulating each other.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the German Research Foundation (DFG, grant KR5148/2-1 to PK, project number: 436456810), and the Emergent Talents Initiative (ETI) of the University Erlangen–Nuremberg, Germany (grant 2019/2-Phil-01 to PK).

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.neunet.2020.05.007.

References

Alexandre, L. A., Embrechts, M. J., & Linton, J. (2009). Benchmarking reservoir computing on time-independent classification tasks. In *Proceedings of the international joint conference on neural networks* (pp. 89–93). IEEE.

Amaral, L. A., Scala, A., Barthélémy, M., & Stanley, H. E. (2000). Classes of small-world networks. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(21), 11149–11152.

Andel, D., & Wehner, R. (2004). Path integration in desert ants, Cataglyphis: how to make a homing ant run away from home. *Proceedings of the Royal Society of London, Series B*, *271*(1547), 1485–1489.

Antonelo, E., & Schrauwen, B. (2012). Learning slow features with reservoir networks for biologically-inspired robot localization. *Neural Networks*, *25*, 178–190.

Antonelo, E. A., Schrauwen, B., & Stroobandt, D. (2007). Event detection and localization for small mobile robots using reservoir computing. In *Conference on artificial neural networks* (pp. 660–669).

Anwar, S., Hwang, K., & Sung, W. (2017). Structured pruning of deep convolutional neural networks. *ACM Journal on Emerging Technologies in Computing Systems*, *13*(3), 1–18.

Aronov, D., Nevers, R., & Tank, D. W. (2017). Mapping of a non-spatial dimension by the hippocampal–entorhinal circuit. *Nature*, *543*(7647), 719.

Babadi, B., & Sompolinsky, H. (2014). Sparseness and expansion in sensory representations. *Neuron*, *83*(5), 1213–1226.

Barwich, A.-S. (2019). The value of failure in science: The story of grandmother cells in neuroscience. *Frontiers in Neuroscience*, *13*, 1121.

Bassett, D. S., & Bullmore, E. (2006). Small-world brain networks. *Neuroscientist*, *12*(6), 512–523.

Bassett, D. S., Meyer-Lindenberg, A., Achard, S., Duke, T., & Bullmore, E. (2006). Adaptive reconfiguration of fractal small-world human brain functional networks. *Proceedings of the National Academy of Sciences*, *103*(51), 19518–19523.

Bellmund, J. L., Gärdenfors, P., Moser, E. I., & Doeller, C. F. (2018). Navigating cognition: Spatial codes for human thinking. *Science*, *362*(6415), eaat6766.

Bertschinger, N., & Natschläger, T. (2004). Real-time computation at the edge of chaos in recurrent neural networks. *Neural Computation*, *16*(7), 1413–1436.

Boksa, P. (2012). Abnormal synaptic pruning in schizophrenia: Urban myth or reality? *Journal of Psychiatry & Neuroscience: JPN*, *37*(2), 75.

Chalfie, M. (1984). Neuronal development in Caenorhabditis elegans. *Trends in NeuroSciences*, *7*(6), 197–202.

Constantinescu, A. O., O'Reilly, J. X., & Behrens, T. E. (2016). Organizing conceptual knowledge in humans with a gridlike code. *Science*, *352*(6292), 1464–1468.

Crochet, S., Poulet, J. F., Kremer, Y., & Petersen, C. C. (2011). Synaptic mechanisms underlying sparse coding of active touch. *Neuron*, *69*(6), 1160–1175.

Dasgupta, S., Stevens, C. F., & Navlakha, S. (2017). A neural algorithm for a fundamental computing problem. *Science*, *358*(6364), 793–796.

Eichenbaum, H. (2015). The hippocampus as a cognitive map…of social space. *Neuron*, *87*(1), 9–11.

Etienne, A. S., & Jeffery, K. J. (2004). Path integration in mammals. *Hippocampus*, *14*(2), 180–192.

Etienne, A. S., Maurer, R., & Séguinot, V. (1996). Path integration in mammals and its interaction with visual landmarks. *Journal of Experimental Biology*, *199*(1), 201–209.

Fekiac, J., Zelinka, I., & Burguillo, J. C. (2011). A review of methods for encoding neural network topologies in evolutionary computation. In *Proceedings - 25th European conference on modelling and simulation, ECMS 2011* (pp. 410–416).

Gabel, C. V., Gabel, H., Pavlichin, D., Kao, A., Clark, D. A., & Samuel, A. D. (2007). Neural circuits mediate electrosensory behavior in Caenorhabditis elegans. *Journal of Neuroscience*, *27*(28), 7586–7596.

Garvert, M. M., Dolan, R. J., & Behrens, T. E. (2017). A map of abstract relational knowledge in the human hippocampal–entorhinal cortex. *Elife*, *6*, e17086.

Gerum, R. (2019). Pylustrator: Code generation for reproducible figures for publication. arXiv:1910.00279.

Hagmann, P., Cammoun, L., Gigandet, X., Meuli, R., Honey, C. J., Van Wedeen, J., et al. (2008). Mapping the structural core of human cerebral cortex. *PLoS Biology*, *6*(7), 1479–1493.

Han, S., Mao, H., & Dally, W. J. (2015). Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding. (pp. 1–14). arXiv:1510.00149.

Herculano-Houzel, S. (2009). The human brain in numbers: a linearly scaled-up primate brain. *Frontiers in Human Neuroscience*, *3*, 31.

Hochreiter, S. (1998). The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *6*(2), 107–116.

Hong, S., Dissing-Olesen, L., & Stevens, B. (2016). New insights on the role of microglia in synaptic pruning in health and disease. *Current Opinion in Neurobiology*, *36*, 128–134.

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science and Engineering*, *9*, 90–95.

Jarrell, T. A., Wang, Y., Bloniarz, A. E., Brittin, C. A., Xu, M., Thomson, J. N., et al. (2012). The connectome of a decision-making neural network. *Science*, *337*(6093), 437–444.

Jiao, Y., Zhang, Y., Chen, X., Yin, E., Jin, J., Wang, X., et al. (2018). Sparse group representation model for motor imagery EEG classification. *IEEE Journal of Biomedical and Health Informatics*, *23*(2), 631–641.

Jin, Z., Zhou, G., Gao, D., & Zhang, Y. (2018). EEG classification using sparse Bayesian extreme learning machine for brain–computer interface. *Neural Computing and Applications*, 1–9.

Jonas, E., & Kording, K. P. (2017). Could a neuroscientist understand a microprocessor? *PLoS Computational Biology*, *13*(1), e1005268.

Kafashan, M., Nandi, A., & Ching, S. (2016). Relating observability and compressed sensing of time-varying signals in recurrent linear networks. *Neural Networks*, *83*, 11–20.

Kerr, J. N., Greenberg, D., & Helmchen, F. (2005). Imaging input and output of neocortical networks in vivo. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(39), 14063–14068.

Killian, N. J., & Buffalo, E. A. (2018). Grid cells map the visual world. *Nature Neuroscience*, *21*(2), 161.

Kolb, B., & Gibb, R. (2011). Brain plasticity and behaviour in the developing brain. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, *20*(4), 265.

Krauss, P., Prebeck, K., Schilling, A., & Metzner, C. (2019a). Recurrence resonance in three-neuron motifs. *Frontiers in Computational Neuroscience*, *13*.

Krauss, P., Schuster, M., Dietrich, V., Schilling, A., Schulze, H., & Metzner, C. (2019b). Weight statistics controls dynamics in recurrent neural networks. *PLoS ONE*, *14*(4), 1–13.

Krauss, P., Zankl, A., Schilling, A., Schulze, H., & Metzner, C. (2019c). Analysis of structure and dynamics in three-neuron motifs. *Frontiers in Computational Neuroscience*, *13*, 5.

Kriegeskorte, N., & Douglas, P. K. (2018). Cognitive computational neuroscience. *Nature Neuroscience*, *21*(9), 1148–1160.

Latora, V., & Marchiori, M. (2001). Efficient behavior of small-world networks. *Physical Review Letters*, *87*(19), 198701.

Littman, M. (2012). Inducing partially observable Markov decision processes.. In *11th international conference on grammatical inference, Vol. 21* (pp. 145–148).

Low, L. K., & Cheng, H.-J. (2006). Axon pruning: an essential step underlying the developmental plasticity of neuronal connections. *Philosophical Transactions of the Royal Society, Series B (Biological Sciences)*, *361*(1473), 1531–1544.

Lukoševičius, M., Jaeger, H., & Schrauwen, B. (2012). Reservoir computing trends. *KI - Künstliche Intelligenz*, *26*(4), 365–371.

McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I., & Moser, M.-B. (2006). Path integration and the neural basis of the'cognitive map'. *Nature Reviews Neuroscience*, *7*(8), 663.

Mocanu, D. C., Mocanu, E., Stone, P., Nguyen, P. H., Gibescu, M., & Liotta, A. (2018). Scalable training of artificial neural networks with adaptive sparse connectivity inspired by network science. *Nature Communications*, *9*(1), 2383.

Müller, M., & Wehner, R. (1988). Path integration in desert ants, Cataglyphis fortis. *Proceedings of the National Academy of Sciences*, *85*(14), 5287–5290.

Müller, M., & Wehner, R. (1994). The hidden spiral: systematic search and path integration in desert ants, Cataglyphis fortis. *Journal of Comparative Physiology A*, *175*(5), 525–530.

Nau, M., Schröder, T. N., Bellmund, J. L., & Doeller, C. F. (2018). Hexadirectional coding of visual space in human entorhinal cortex. *Nature Neuroscience*, *21*(2), 188.

Oh, S. W., Harris, J. A., Ng, L., Winslow, B., Cain, N., Mihalas, S., et al. (2014). A mesoscale connectome of the mouse brain. *Nature*, *508*(7495), 207–214.

Olshausen, B. A., & Field, D. J. (2004). Sparse coding of sensory inputs. *Current Opinion in Neurobiology*, *14*(4), 481–487.

Oren-Suissa, M., Bayer, E. A., & Hobert, O. (2016). Sex-specific pruning of neuronal synapses in Caenorhabditis elegans. *Nature*, *533*(7602), 206.

Paolicelli, R. C., Bolasco, G., Pagani, F., Maggi, L., Scianni, M., Panzanelli, P., et al. (2011). Synaptic pruning by microglia is necessary for normal brain development. *science*, *333*(6048), 1456–1458.

Pascanu, R., Mikolov, T., & Bengio, Y. (2012). Understanding the exploding gradient problem. arXiv:1211.5063v1.

Pehlevan, C., & Sompolinsky, H. (2014). Selectivity and sparseness in randomly connected balanced networks. *PLoS One*, *9*(2).

Perin, R., Berger, T. K., & Markram, H. (2011). A synaptic organizing principle for cortical neuronal groups. *Proceedings of the National Academy of Sciences*, *108*(13), 5419–5424.

Pierce-Shimomura, J. T., Morse, T. M., & Lockery, S. R. (1999). The fundamental role of pirouettes in Caenorhabditis elegans chemotaxis. *Journal of Neuroscience*, *19*(21), 9557–9569.

Quiroga, R. Q., Kreiman, G., Koch, C., & Fried, I. (2008). Sparse but not 'grandmother-cell'coding in the medial temporal lobe. *Trends in Cognitive Sciences*, *12*(3), 87–91.

Rose, D. (1996). *Some reflections on (or by?) grandmother cells*. London, England: SAGE Publications Sage UK.

Ryu, W. S., & Samuel, A. D. (2002). Thermotaxis in Caenorhabditis elegans analyzed by measuring responses to defined thermal stimuli. *Journal of Neuroscience*, *22*(13), 1–7.

Sanchez, E., Pérez-Uribe, A., & Mesot, B. (2001). Solving partially observable problems by evolution and learning of finite state machines. In *International conference on evolvable systems* (pp. 267–278). Springer.

Schilling, A., Metzner, C., Rietsch, J., Gerum, R., Schulze, H., & Krauss, P. (2018). How deep is deep enough?–Quantifying class separability in the hidden layers of deep neural networks. arXiv preprint arXiv:1811.01753.

Schmidhuber, J., & Hochreiter, S. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.

Schrauwen, B., Verstraeten, D., & Van Campenhout, J. (2007). An overview of reservoir computing: theory, applications and implementations. In *Proceedings of the 15th European symposium on artificial neural networks* (pp. 471–482).

Séguinot, V., Cattet, J., & Benhamou, S. (1998). Path integration in dogs. *Animal Behaviour*, *55*(4), 787–797.

Song, S., Sjöström, P. J., Reigl, M., Nelson, S., & Chklovskii, D. B. (2005). Highly nonrandom features of synaptic connectivity in local cortical circuits. *PLoS Biology*, *3*(3), 0507–0519.

Sporns, O., Tononi, G., & Kötter, R. (2005). The human connectome: A structural description of the human brain. *PLoS Computational Biology*, *1*(4), 0245–0251.

Springenberg, J. T., Klein, A., Falkner, S., & Hutter, F. (2016). Bayesian optimization with robust Bayesian neural networks. In *Advances in neural information processing systems* (pp. 4134–4142).

Theves, S., Fernandez, G., & Doeller, C. F. (2019). The hippocampus encodes distances in multidimensional feature space. *Current Biology*, *29*(7), 1226–1231.

van den Heuvel, M. P., & Yeo, B. T. T. (2017). A spotlight on bridging microscale and macroscale human brain architecture. *Neuron*, *93*(6), 1248–1251.

Van Der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The NumPy array: A structure for efficient numerical computation. *Computing in Science and Engineering*, *13*, 22–30.

Verstraeten, D., Schrauwen, B., D'Haene, M., & Stroobandt, D. (2007). An experimental unification of reservoir computing methods. *Neural Networks*, *20*(3), 391–403.

Watts, D. J., & Strogatz, S. H. (1998). Strogatz - small world network nature. *Nature*, *393*, 440–442.

Wehner, R., & Wehner, S. (1986). Path integration in desert ants. approaching a long-standing puzzle in insect navigation. *Monitore Zoologico Italiano-Italian Journal of Zoology*, *20*(3), 309–331.

Wen, W., Wu, C., Wang, Y., Chen, Y., & Li, H. (2016). Learning structured sparsity in deep neural networks. In *Advances in neural information processing systems* (pp. 2074–2082).

White, J. G., Southgate, E., Thomson, J. N., & Brenner, S. (1986). The structure of the nervous system of the nematode caenorhabditis elegans author (s): J. G. White, E. Southgate, J. N. Thomson, S. Brenner source : Philosophical transactions of the royal society of London. series B, biological published by. *Philosophical Transactions of the Royal Society of London*, *314*(1165), 1–340.

Wierstra, D., & Wiering, M. (2004). Utile distinction hidden Markov models. In *Proceedings, twenty-first international conference on machine learning, ICML 2004* (pp. 855–862).

Xu, X., Ding, Y., Hu, S. X., Niemier, M., Cong, J., Hu, Y., et al. (2018). Scaling for edge inference of deep neural networks. *Nature Electronics*, *1*(4), 216–222.

Yeo, W., & Gautier, J. (2004). Early neural cell death: dying to become neurons. *Developmental Biology*, *274*(2), 233–244.

Young, S. R., Rose, D. C., Karnowski, T. P., Lim, S.-H., & Patton, R. M. (2015). Optimizing deep learning hyper-parameters through an evolutionary algorithm. In *Proceedings of the workshop on machine learning in high-performance computing environments* (pp. 1–5).

Zaslaver, A., Liani, I., Shtangel, O., Ginzburg, S., Yee, L., & Sternberg, P. W. (2015). Hierarchical sparse coding in the sensory system of Caenorhabditis elegans. *Proceedings of the National Academy of Sciences*, *112*(4), 1185–1189.

Zhang, L., Yang, M., & Feng, X. (2011). Sparse representation or collaborative representation: Which helps face recognition? In *2011 international conference on computer vision* (pp. 471–478). IEEE.

NEUROSCIENCE
RESEARCH ARTICLE
A. Schilling et al. / Neuroscience 520 (2023) 39–45

Behavioral Assessment of Zwicker Tone Percepts in Gerbils

Achim Schilling, [†] **Konstantin Tziridis** [†] **Holger Schulze and Patrick Krauss** [*]

Neuroscience Lab, University Hospital Erlangen, Friedrich-Alexander University Erlangen-Nürnberg (FAU), Germany

Abstract—The Zwicker tone illusion – an auditory phantom percept after hearing a notched noise stimulus – can serve as an interesting model for acute tinnitus. Recent mechanistic models suggest that the underlying neural mechanisms of both percepts are similar. To date it is not clear if animals do perceive the Zwicker tone, as up to now no behavioral paradigms are available to objectively assess the presence of this phantom percept. Here we introduce, for the first time, a modified version of the gap pre-pulse inhibition of the acoustic startle reflex (GPIAS) paradigm to test if it is possible to induce a Zwicker tone percept in our rodent model, the Mongolian gerbil. Furthermore, we developed a new aversive conditioning learning paradigm and compare the two approaches. We found a significant increase in the GPIAS effect when presenting a notched noise compared to white noise gap pre-pulse inhibition, which is consistent with the interpretation of a Zwicker tone percept in these animals. In the aversive conditioning learning paradigm, no clear effect could be observed in the discrimination performance of the tested animals. When investigating the first 33% of the correct conditioned responses, an effect of a possible Zwicker tone percept can be seen, i.e. animals show identical behavior as if a pure tone was presented, but the paradigm needs to be further improved. Nevertheless, the results indicate that Mongolian gerbils are able to perceive a Zwicker tone and can serve as a neurophysiological model for human tinnitus generation.

Key words: acoustic startle reflex, aversive conditioning, shuttle box, tinnitus, Zwicker tone, auditory phantom perception.

INTRODUCTION

The Zwicker tone illusion was first described as a so called"auditory afterimage" by Eberhard Zwicker in 1964 (Zwicker, 1964). It is typically evoked by the presentation of white noise with a spectral gap (notched noise, NN) but can also be triggered by e.g. loud low-pass filtered noise (Franosch et al., 2003). Although the name auditory afterimage indicates some analogies of Zwicker tone and visual phantom percepts, the underlying mechanisms are assumed to be completely different, i.e., not caused by a bleaching of the sensing pigment or by some cortical adaptation (Shimojo et al., 2001). Instead, a neuronal mechanism most probably in the early stages of the auditory pathway is supposed to be the cause (Norena et al., 1999). In recent years, the Zwicker tone became a model system for tinnitus research as it can be evoked quickly and reliably and does not cause harm to the auditory system (Leske et al., 2014). Furthermore, as the Zwicker tone is perceived as a pure tone (Fastl & Stoll, 1979) with a pitch within the spectral notch, its frequency can be tuned manually by simply shifting the spectral notch of the presented noise. The spectral notch can have different shapes and bandwidths, ranging from relatively narrow notches of roughly one octave to complete low-pass noises with or without an additional pure tone stimulus (Franosch et al., 2003). Interestingly, real spectral notches – as used in this study – evoke Zwicker tone percepts roughly in the center of the notch, low pass noises evoke percepts somewhat above the upper edge of the cutoff frequency while additionally presented pure tones push the percept to frequencies below that tone. Within this approach, Parra and Pearlmutter have already shown a correlation of the presence of tinnitus and the probability of perceiving a Zwicker tone after noise presentation, indicating that the underlying neural mechanisms might be similar (Parra & Pearlmutter, 2007). Thus, they describe the Zwicker tone as well as tinnitus as an effect of increased neuronal gain along the auditory pathway. There are several mechanistic explanations for increased neuronal gain based on molecular or cell biology (Norena, 2011; Turrigiano, 2012; Tighilet et al., 2016; Gainey & Feldman, 2017). However, all these models have in common that they suppose mechanisms that need hours or

*Corresponding author. Address: Experimental Otolaryngology, ENT-Hospital, Head and Neck Surgery, University Hospital Erlangen, Waldstrasse 1, 91054 Erlangen, Germany.
E-mail address: patrick.krauss@fau.de (P. Krauss).
[†] Both authors contributed equally.
Abbreviations: GPIAS, gap pre-pulse inhibition of the acoustic startle reflex; CR, conditioned response; ASR, acoustic startle response; PPI, pre-pulse inhibition; NN, notched noise; CS+, reinforced conditioning stimulus; CS-, non-reinforced conditioning stimulus; CS, conditioned stimuli; US, unconditioned stimulus; CR+, correct reactions; CR-, incorrect reactions; DP, discrimination performance; WN, white noise.

https://doi.org/10.1016/j.neuroscience.2023.04.011
0306-4522/

days to take effect whereas the Zwicker tone percept emerges within seconds or even faster.

In contrast to the models mentioned above, the recent model by Krauss and colleagues is the only one that proposes a neuronal network based effect, and thereby is capable to operate on very short time scales of seconds or even below. It therefore is the only model that may explain both, the emergence of Zwicker tone and acute tinnitus (Krauss et al., 2016; Schilling et al., 2021; Schilling et al., 2022b). The model predicts a threshold improvement during the perception of a Zwicker tone and/or tinnitus by a physiological mechanism optimizing information transmission from the cochlea into the subsequent auditory pathway. The hearing threshold improvement has already been shown during Zwicker tone perception (Wiegrebe et al., 1996), tinnitus (Krauss et al., 2016; Gollnast et al., 2017), in a new animal model paradigm of long-term notched noise exposure to simulate hearing loss (Krauss & Tziridis, 2021) in gerbils, and in a computational model of the auditory pathway based on deep neural networks (Schilling et al., 2022a). Taken together, all these findings indicate that the Zwicker tone paradigm is a valuable model to investigate and understand the tinnitus phantom percept. However, in order to investigate Zwicker tone perceptions in animal models (such as rodents) a behavioral paradigm for objective assessment of the presence of this phantom percept is crucial. Furthermore, this behavioral correlate of Zwicker tone is a basic requirement for the interpretation of neurophysiological data seeking to describe the neuronal mechanisms leading to auditory phantom percepts in gerbils. The development of such a behavioral paradigm is particularly challenging, as the Zwicker tone percept is transient and fades within just a few seconds after induction.

Here, we introduce and test two potential behavioral paradigms to assess Zwicker tone percepts in gerbils. The first one is based on the gap pre-pulse inhibition of the acoustic startle reflex (GPIAS) paradigm, originally introduced by Turner and coworkers in 2006 (Turner et al., 2006) to assess possible tinnitus percepts in rodents. There, a gap of silence within a moderate band-pass filtered noise is presented before a startle stimulus (e.g. a white noise burst). The gap of silence, if perceived, leads to a suppression (pre-pulse inhibition, PPI) of the acoustic startle reflex. The idea of the paradigm is that the gap of silence is masked ("filling in" hypothesis) by a tinnitus percept, making the gap less salient and hence decreasing the gap-induced pre-pulse inhibition. Due to observations (for details cf. Discussion) indicating that spectrally different pure tones or phantom percepts may serve as an additionally salient pre-pulse within the GPIAS paradigm (Steube et al., 2016; Schilling et al., 2017), we developed the hypothesis that a Zwicker tone induced by a notched noise should also lead to an increased GPIAS. This is because the Zwicker tone is perceived as a pure tone and consequently sounds fundamentally different compared to the notched background noise used, resulting in contrast enhancement. Along these lines, we here propose a behavioral paradigm based on the GPIAS paradigm where the band-pass noise is replaced by broadband, notched noise potentially inducing a Zwicker tone percept. It has been shown that the more salient a pre-pulse stimulus is the more inhibition of the acoustic startle reflex it provides (Carlson & Willott, 1996). Additionally, in gerbils the gap duration has a non-linear effect on the pre-pulse inhibition, with a lower effect strength at durations used in this study (cf. Methods) compared to pure tone or noise burst pre-pulse stimuli (Steube et al., 2016). As we assume the Zwicker tone to serve as a salient pre-stimulus, which is perceived in the gap of silence, we consequently predict an increased GPIAS compared to the control condition of a GPIAS induced by a gap of silence in a white noise background. Therefore, the effect should be opposite compared to the masking of the gap-effect expected in tinnitus testing.

In a second, alternative approach, we use an aversive conditioning GO-NOGO paradigm in the shuttle box where animals are trained to discriminate between a white noise stimulus followed by a pure tone (reinforced conditioning stimulus, CS+) and the white noise stimulus followed by silence (non-reinforced conditioning stimulus, CS-). After the conditioning period, the CS+ is replaced by notched noise - presumably inducing a Zwicker tone of the trained frequency - followed by silence. The idea is that in case the animals indeed perceive the Zwicker tone, they should show the same behavior as for the CS+, thereby indicating the perception of a Zwicker tone.

EXPERIMENTAL PROCEDURES

Animals and housing

39 male Mongolian gerbils, purchased from Janvier (Le Genest-Saint-Isle, France) were housed in standard animal racks (Bio A.S. Vent Light, Zoonlab, Emmendingen, Germany) in groups of 3–4 animals with free access to water and food at a room temperature of 20–25 °C under a 12 h/12 h dark/light circle. The use and care of the animals was approved by the state of Bavaria (Regierungspräsidium Mittelfranken, Ansbach, Germany, No. 54-2532.1-02/13).

Stimulation software

The entire software used to conduct the GPIAS (Gerum et al., 2019) measurements (as well as the evaluation procedures, cf. below) was written in Python 3.6 (Van Rossum & Drake, 1995). For basic numerical operations the Numpy library was used (Van Der Walt et al., 2011); more complex mathematical operations (e.g., signal filter functions) were implemented using the SciPy package (Oliphant, 2007). Data visualization was realized using the Matplotlib library (Hunter, 2007) and the Pylustrator (Gerum, 2019). Efficient storage of the data was achieved by the usage of the Pandas library (McKinney, 2010).

The software for the shuttle box was implemented in Pascal (Wirth, 1971) and is described in detail in earlier work (Schulze & Scheich, 1999). All data evaluations were performed by hand and transferred to Statistica (StatSoft Europe, Hamburg, Germany); for details see below.

GPIAS measurements

For the GPIAS measurements on 24 male gerbils a custom-made open-source setup was used as detailed before (Gerum et al., 2019). In short, the animals were placed in an acrylic glass restrainer tube, closed with a wire mesh at the front side and a cap at the back end, and placed on a sensor platform fixed to a vibration-proof table. Movements of the sensor platform were registered using a 3D acceleration sensor. Two loudspeakers were placed at a distance of 10 cm in front of the animal. They presented, first, the 115 dB SPL startle stimulus (Neo 25 S, SinusLive, noise burst 20 ms, flattened with 5 ms sin^2 ramps) and, second, the 60 dB SPL white/notched noise background (CantonPlus XS.2). Notched (spectral notches centered at 2 kHz or 5 kHz ± half octave; 12 animals each) as well as white noise backgrounds were presented pseudorandomly with and without a gap of silence of 50 ms (flanked by 20 ms sin^2 ramps, 10 ms complete silence) stating 100 ms before the startle stimulus.

Three measurement sessions were performed over the course of 14 days for each animal. During the measurements, animals were allowed 15 min of habituation in darkness in the tube. Before the real stimuli were presented, five habituation stimuli were given to "level" the startle responses. Each noise-type stimulus was repeated 30 times with and without gap, summing up to 120 stimuli (Fig. 1), which took roughly 30 min.

The complete evaluation of the GPIAS measurements were performed using custom-made Python programs.

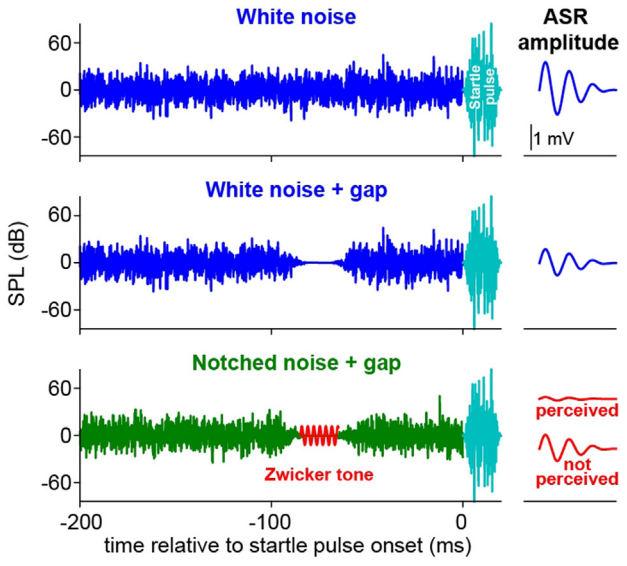

Fig. 1. Stimuli for the modified GPIAS paradigm for Zwicker tone detection. The animals were presented with four different stimuli in a pseudorandomized manner. First, a white noise background stimulus without a gap (**top panel**) leading to a large auditory startle response (ASR) after the startle pulse. Second, a white noise background stimulus with a silent gap (**center panel**), leading to a small ASR response due to pre-pulse inhibition (PPI). Third, a notched noise background without gap (not shown), leading to a similar response as in the corresponding white noise condition and, fourth, a notched noise background with gap (**bottom panel**), leading to either a "normal" PPI, when no Zwicker tone is perceived or an increased PPI if the phantom tone is perceived.

The GPIAS effect was quantified by calculating the median from the full combinatorial startle amplitudes as a response to gap and no gap pre-stimulus (for details see Schilling et al., 2017). Briefly, we calculate the pre-pulse inhibition (PPI) by dividing each gap amplitude by each no-gap amplitude and subtract the result from 1 (i.e., $1 - (\text{amplitude}_{gap}/\text{amplitude}_{no\text{-}gap})$). The median of the resulting distribution is calculated. Positive values indicate inhibited startle responses with the pre-pulse. There was no significant difference between the amplitudes of notched and white noise no-gap conditions (Mann-Whitney U-test, $p > 0.05$). Statistics on the median GPIAS results were performed with Statistica (cf. below).

Shuttle box training and testing

The second behavioral paradigm for Zwicker tone assessment was realized as an aversive conditioning GO-NOGO learning paradigm. Another 15 animals were trained in a two compartment shuttle box (both compartments separated by a 6 cm high hurdle) to distinguish between two auditory stimuli (conditioned stimuli, CS) using an electrical foot shock (which must be aversive but never painful) as unconditioned stimulus (US). The animals were supposed to respond to the CS− by staying within their current compartment while during CS+ presentation they were supposed to cross the hurdle into the other compartment (for details see Depner et al., 2014). First, the animals were trained for up to 18 days (dependent on the performance of the individual animal: median [interquartile range]: 15 d [9 d, 16 d]). The task to learn was to distinguish between 4 s of silence after 8 s of 60 dB SPL white noise (with 2 ms sin2-ramps; CS−) and 8 s of 60 dB SPL white noise followed by a pure tone (2 kHz, 60 dB SPL, CS+) with a duration of 4 s. This loudness should roughly be equivalent to the subsequently evoked possible phantom percept (Fastl et al., 2001). The stimuli were created with Audacity, an open source software tool for audio applications. Both stimuli were repeated 30 times each in a pseudorandomized manner in the daily training sessions. The animals were given 2 s to decide if a CS+ tone was presented and cross the hurdle, else a foot shock was given for maximally 2 s. If they crossed the hurdle during the 4 s of CS− silence presentation, a 500 ms foot shock was given. The difference of correct reactions (CR+) and incorrect reactions (CR−) was defined as discrimination performance (DP) which in our case can have values between −30 and +30, where zero indicates no discrimination between both stimuli. After successful completion of the training phase defined as three consecutive days with p-values smaller than 0.01 in a chi^2 test of jump response results, animals were exposed to a single test session. There, the same CS− (white noise with silence) and an adjusted CS+ (notched noise with spectral notch at 2 kHz center frequency, half octave width, followed by silence) were applied according to the same temporal protocol as during training sessions yet but without foot shocks. Our hypothesis was that the notched noise induces a Zwicker tone around 2 kHz which is then perceived as pure-tone stimulus, similar to the one the ani-

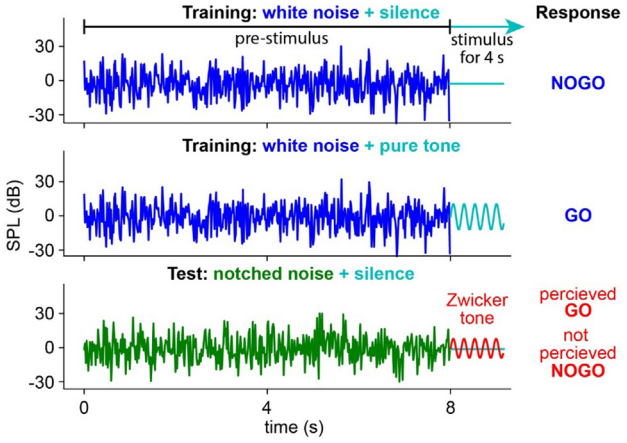

Fig. 2. Stimuli of the conditioning paradigm for Zwicker tone detection in the shuttle box. The animals were trained to discriminate between white noise followed by silence (**top panel**) and white noise followed by a pure tone (**center panel**). After successful training, the pure tone was replaced by a potential Zwicker tone by presenting a notched noise with silence (**bottom panel**) instead of white noise with pure tone for one single test session; the white noise with silence stimulus was kept the same. If the animal perceived the Zwicker tone that roughly sounds similar to the GO stimulus, it should respond accordingly.

mals were trained to. Thus, a significant discrimination between CS+ and CS− within the test session would be an indication of a Zwicker tone percept (Fig. 2).

Statistics

As mentioned above, all statistics were performed with Statistica. For the evaluation of the GPIAS results, the medians of the calculated full combinatorial PPIs of white noise (WN) and notched noise (NN) were taken for each animal and each measurement and analyzed by Friedman-ANOVAs. Pairwise tests of WN and NN results within one measurement were performed by Wilcoxon matched pair tests. For the comparison of the responses of the animals during the three successful training days and the test day, we analyzed the DP by Wilcoxon matched pair tests and the performance during the first and last ten trials of the sessions separately, as the foot shocks were turned off. This was analyzed by chi^2 tests.

RESULTS

GPIAS paradigm

We exposed two groups of Mongolian gerbils (*12 animals in each group*) to the white noise stimuli with gap, Fig. 1, top and center panels) and the notched noise stimuli (gap and no gap, Fig. 1, bottom panel; notch at 2 kHz and 5 kHz, respectively, ±half octave) followed by the startle pulse. This experiment was repeated three times (with one week between successive measurements) to check for re-test reliability. The first result we found in this context was the increase in the PPI in the consecutive measurements for both frequency groups (Fig. 3A, B). In all four cases (WN and NN in 2 kHz and 5 kHz groups), the significant results of the Friedman-ANOVAs indicate this increase (cf. also Table 1). This finding indicates some sensitization of the animals to the gap pre-stimulus. However, the NN-GPIAS compared to WN responses is systematically increased: This is true for both 1st measurements in the two groups (Wilcoxon tests: Fig. 3A: $p = 0.008$, Fig. 3B: $p = 0.01$) and the 3rd measurement in the 5 kHz group (Fig. 3B: $p = 0.008$) while the results in the 2 kHz group were here not different ($p = 0.16$). No significant differences were found also for both comparisons in the 2nd measurements (Fig. 3A: $p = 0.34$; Fig. 3B: $p = 0.27$). In other words, when choosing the appropriate stimulus (in the case of Mongolian gerbils: frequencies around 5 kHz), the behavioral responses of the animals can be interpreted as a Zwicker tone percept (cf. Discussion).

Conditioning paradigm

In conditioning paradigm, we use an alternative approach to elicit and assess Zwicker tone percepts in rodents. In particular, the animals were trained to discriminate between silence and a 2 kHz pure tone after a white noise background stimulus (Fig. 4**A**, left panel). After successful training (defined as three consecutive days of a positive DP with $p < 0.01$ in a chi^2 test), a single test session without any foot shock was performed in order to assess the putative perception of a 2 kHz Zwicker tone. When investigating the DP, the animals did not perceive such a phantom percept, as hardly any animal showed a positive DP with a median of 1 [−3, 3] and a significant performance drop (Wilcoxon test,

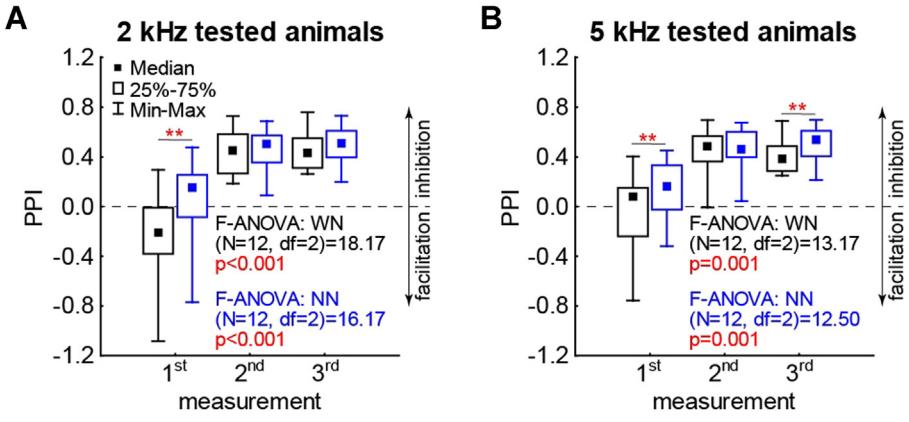

Fig. 3. Results of Zwicker GPIAS paradigm. **(A)** Median of gap induced pre-pulse inhibition (PPI) (full combinatorial, cf. Schilling et al., 2017) for the 2 kHz group. Black symbols indicate the values and Friedman-ANOVA statistics for the white noise (WN) condition, blue symbols indicate the results of the notched noise (NN) condition. Asterisks indicate significant differences in the Wilcoxon tests (cf. Table 1), $^{**}p < 0.01$. Positive PPI values indicate pre-pulse inhibition. **(B)** Median of gap induced GPIAS for the 5 kHz group. Symbols as in **(A)**.

Table 1. Wilcoxon test results for WN – NN comparisons in the three GPIAS measurements.

Frequency	Measurement	Valid N	T score	*p*-value
2 kHz	1	12	5	0.008
	2	12	27	0.35
	3	12	21	0.16
5 kHz	1	12	6	0.01
	2	12	25	0.27
	3	12	5	0.08

$p < 0.001$) compared to the three days of successful training DP with 12 [15, 9] (Fig. 4A, right panel).

In order to reveal the reasons for this lack of measurable perception, we inspected the raw data of the total number of correct (CR+) and false (false rejection) responses to the CS+ in the first 10 (Fig. 4B) and last 10 trials (Fig. 4C) in comparison with the responses of the data from the three days of successful learning. In both cases, the chi^2 tests show differences between the number of responses, with lower correct response rates already in the first 10 trials of the test condition. When comparing the correct responses (i.e., CR+) of the first and last 10 trials separately, we find a significant performance drop in the test condition (Wilcoxon test, $p < 0.001$), while we do not see this drop in the data of the three training days before (Wilcoxon tests, always $p > 0.05$). This indicates that, even though the animals may perceive a Zwicker tone, they do not perform the task, as no foot shocks are applied during the test session.

DISCUSSION

We here present two novel approaches of behavioral paradigms to objectively assess putative Zwicker tone perception in rodents. The first paradigm is based on GPIAS measurements introduced by Turner and coworkers (Turner et al., 2006), while the second is based on an aversive conditioning learning paradigm. In both paradigms, we provide a-proof-of principle that Mongolian gerbils may perceive an assessable Zwicker tone illusion, whereas the adjusted GPIAS paradigm yields to the most promising results.

Additionally, the results of that new GPIAS paradigm may have a major impact on the interpretation of the original GPIAS paradigm for tinnitus assessment. This is because the two paradigms differ in the spectral composition of the applied noise: Whereas the original GPIAS paradigm is based on the gap detection ability in band-pass filtered noise, our new Zwicker tone GPIAS paradigm is based on the modulation of the gap detection ability in notched noise. In the original GPIAS paradigm, it is assumed that a potential tinnitus percept leads to a "filling-in" of the gap of silence, and thus to a decreased ability to detect this gap, consequently leading to a decrease of the gap pre-pulse inhibition. However, this mechanism only works, if the perceived tinnitus sounds similar to the surrounding noise as suggested by the study of Basavaraj and Yan (Basavaraj & Yan, 2012), even though they tested only specific pure tones against each other (background tone vs. pre-pulse stimulus).

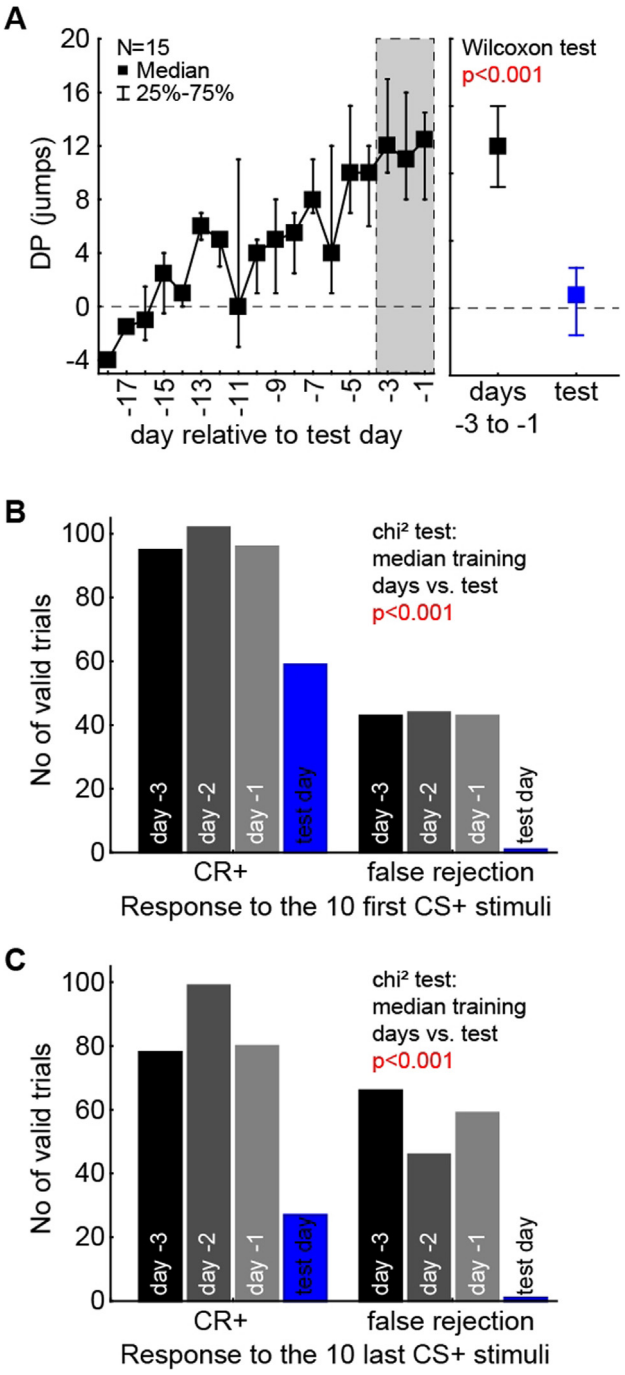

Fig. 4. Shuttle box results for Zwicker stimuli. **(A)** Left panel: Discrimination performance (DP) for training sessions as a function of time relative to test day (day 0). Sufficient learning was defined when the DP was significant on a $p < 0.01$ level in the chi^2 test over three consecutive days (gray area). Right panel: significant difference (Wilcoxon test) of successful training sessions (black) and test session (blue). **(B)** Number of CR+ and false rejections for the first 10 CS+ trials in the successful training sessions and the test session with chi^2 test result. **C** Number of CR+ and false rejections for the last 10 CS+ trials in the successful training sessions and the test session with chi^2 test result.

If the GPIAS paradigm is actually a valid model for the assessment of a frequency specific tinnitus perception, the Zwicker tone paradigm would not predict the"filling in" of the gap of silence (Turner et al., 2006; Turner, 2007) as the perceived sound is significantly different (it has an ''inverse" frequency spectrum) compared to the surrounding noise. In other words, the Zwicker tone acts as a more salient pre-pulse stimulus and should therefore increase the PPI (instead of decreasing it in the case of standard GPIAS tinnitus assessment). Here, also the difference in presented notch frequencies might play a crucial role (Franosch et al., 2003), and may furthermore explain why e.g. the 5 kHz percept is easier to detect compared to a 2 kHz percept. Additionally, it has already been shown that a GPIAS paradigm is at least partially dependent on cortical learning mechanisms (Moreno-Paublete et al., 2017; Lanaia et al., 2021), which may interact (e.g. suppress) the percept after consecutive presentation. These learning mechanisms might explain the strong differences between the first measurement and the two following sessions. Lanaia and coworkers (Lanaia et al., 2021) already described the increase in PPI in non-tinnitus frequencies starting with the second GPIAS session (the session after the induction of a possible tinnitus percept). These findings are consistent with the ones presented here, even though no tinnitus percept is induced here. Independent of the frequency or the timing, one can expect an increase of the gap pre-pulse inhibition for the Zwicker tone GPIAS paradigm, as the notched noise should sound completely different compared to the Zwicker tone which resembles a pure tone (Zwicker, 1964), thereby resulting in a contrast enhancement. This is also suggested by experiments presenting an acoustic noise burst or pure tone within a background noise, leading to an increase in the pre-pulse inhibition effect compared to a gap-of-silence approach of equal background sound pressure level (Steube et al., 2016). This increase of the gap detection ability could indeed been shown in our study (cf. Fig. 3), and thus can be seen as a cross-validation of the original GPIAS paradigm for tinnitus assessment.

On the other hand, the second behavioral paradigm presented in this study, the aversive conditioning learning paradigm in a shuttle box, seemed not suited to assess the perception of a Zwicker tone on the first glance. Only a very small positive DP being not significantly different from zero (sign test, $p > 0.05$) could be found on during the test session. Only when analyzing the responses to the CS+ alone, an effect for the first 10 CS+ trials of the test day could be found. This indicates a possible Zwicker tone percept, but is clearly only a first hint and proof-of-principle that this approach can be used to assess the existence of the phantom percept. This could be due to several reasons. First, the training may have been too short. We used an individualized approach with defining the training successful when three consecutive training days showed a significant DP of $p < 0.01$ in the chi^2 test. In other words, as soon as the animals knew the basic task, we tested them. The median number of training days was 15, which is corresponds to the standard time for discrimination training in the shuttle box (e.g., Ohl et al., 1999; Ohl et al., 2000) but some animals only needed six days to reach this training level, which in retrospective may have been much too early to stop the training. Nevertheless, we did not find a correlation of number of training days and DP on the test day (multiple linear regression analysis, $p > 0.05$). Second, the stimulus may not have been optimal. As we have seen in the Zwicker GPIAS data, the stimulation with 2 kHz stimuli only showed significant results on the first day of testing. The 5 kHz stimulus seems to be much more salient in these kind of paradigms, which we were not aware of when starting the training. It is possible, that a 5 kHz stimulus might be more salient and therefore produce a significantly better result, this remains to be investigated in further studies. Third, even if a strong Zwicker tone was induced by our paradigm, it would only be rated as CS + by the animals if its pitch was close to the 2 kHz tone used in the training. If on the other hand the pitch of the Zwicker tone would be significantly different from 2 kHz, the animal would most probably not generalizing the tone into the CS+ category. In a future study, this problem can be overcome by training the animals to generalizing any tone pitch into the CS+ category. Fourth and last, as described above we did not use a foot shock reinforcement of the behavior during test day. As we have seen in the raw data of the CR+ response, the animals did respond to the stimuli in the first few trials but soon found out that no punishment was given when behaving ''wrong", an effect for which the term ''extinction learning" (e.g., Happel & Frischknecht, 2016) has been coined. This and the above mentioned too early end of training may have resulted in the lack of DP at the test day. Nevertheless, a more thorough approach with this paradigm may result in a comparable positive result as the Zwicker GPIAS approach.

Taken together, we were able to show two proof-of-concept approaches to investigate a possible Zwicker tone percept in rodents. With either the easy to use Zwicker GPIAS paradigm or the more elaborate shuttle box paradigm, the percept can now be quantified and investigated further.

ACKNOWLEDGEMENTS

We are grateful for technical assistance provided by Leonie Stahl, Eva Reingruber, Lea Geissler, Nico Sitzmann and Jana Haag.

FUNDING

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation): grant KR5148/2-1 (project number 436456810) to PK, grant KR5148/3-1 (project number 510395418) to PK, grant TZ100/2-1 (project number 510395418) to KT, and grant SCHI1482/3-1 (project number 451810794) to AS.

REFERENCES

Basavaraj S, Yan J (2012) Prepulse inhibition of acoustic startle reflex as a function of the frequency difference between prepulse and background sounds in mice. PLoS One 7:e45123.

Carlson S, Willott JF (1996) The behavioral salience of tones as indicated by prepulse inhibition of the startle response: relationship to hearing loss and central neural plasticity in C57BL/6J mice. Hear Res 99:168–175.

Depner M, Tziridis K, Hess A, Schulze H (2014) Sensory cortex lesion triggers compensatory neuronal plasticity. BMC Neurosci 15:57.

Fastl, H., Patsouras, D., Franosch, M., van Hemmen, L., 2001. Zwicker-tones for pure tone plus bandlimited noise. In: Proceedings of the 12th International Symposium on Hearing, Physiological and Psychophysical Bases of Auditory. City. pp. 67–74.

Fastl H, Stoll G (1979) Scaling of pitch strength. Hear Res 1:293–301.

Franosch J-M-P, Kempter R, Fastl H, van Hemmen JL (2003) Zwicker tone illusion and noise reduction in the auditory system. Phys Rev Lett 90 178103.

Gainey MA, Feldman DE (2017) Multiple shared mechanisms for homeostatic plasticity in rodent somatosensory and visual cortex. Philos Trans R Soc B: Biol Sci 372:20160157.

Gerum, R., 2019. Pylustrator: code generation for reproducible figures for publication. arXiv preprint arXiv:1910.00279.

Gerum R, Rahlfs H, Streb M, Krauss P, Grimm J, Metzner C, Tziridis K, Günther M, Schulze H, Kellermann W (2019) Open (G) PIAS: An open-source solution for the construction of a high-precision acoustic startle response setup for tinnitus screening and threshold estimation in rodents. Frontiers in Behavioral Neuroscience 13:140.

Gollnast D, Tziridis K, Krauss P, Schilling A, Hoppe U, Schulze H (2017) Analysis of Audiometric Differences of Patients with and without Tinnitus in a Large Clinical Database. Front Neurol 8:31.

Happel M, Frischknecht R (2016) Neuronal plasticity in the juvenile and adult brain regulated by the extracellular matrix. Compos Funct Extracell Matrix Hum Body:143–158.

Hunter JD (2007) Matplotlib: A 2D graphics environment. Comput Sci Eng 9:90–95.

Krauss P, Tziridis K (2021) Simulated transient hearing loss improves auditory sensitivity. Sci Rep 11:14791.

Krauss P, Tziridis K, Metzner C, Schilling A, Hoppe U, Schulze H (2016) Stochastic resonance controlled upregulation of internal noise after hearing loss as a putative cause of tinnitus-related neuronal hyperactivity. Front Neurosci 10:597.

Lanaia V, Tziridis K, Schulze H (2021) Salicylate-Induced Changes in Hearing Thresholds in Mongolian Gerbils Are Correlated With Tinnitus Frequency but Not With Tinnitus Strength. Front Behav Neurosci 15 698516.

Leske S, Tse A, Oosterhof NN, Hartmann T, Müller N, Keil J, Weisz N (2014) The strength of alpha and beta oscillations parametrically scale with the strength of an illusory auditory percept. Neuroimage 88:69–78.

McKinney, W., 2010. Data structures for statistical computing in python. In: Proceedings of the 9th Python in Science Conference. Austin, TX, City, vol. 445. pp. 51–56.

Moreno-Paublete R, Canlon B, Cederroth CR (2017) Differential neural responses underlying the inhibition of the startle response by pre-pulses or gaps in mice. Frontiers in cellular neuroscience 11:19.

Norena, A., Micheyl, C., Chéry-Croze, S., 1999. The Zwicker tone (ZT) as a model of phantom auditory perception. Sixth International Tinnitus Seminar. City, Vol. 429.

Norena AJ (2011) An integrative model of tinnitus based on a central gain controlling neural sensitivity. Neurosci Biobehav Rev 35:1089–1109.

Ohl FW, Schulze H, Scheich H, Freeman WJ (2000) Spatial representation of frequency-modulated tones in gerbil auditory cortex revealed by epidural electrocorticography. J Physiol Paris 94:549–554.

Ohl FW, Wetzel W, Wagner T, Rech A, Scheich H (1999) Bilateral ablation of auditory cortex in Mongolian gerbil affects discrimination of frequency modulated tones but not of pure tones. Learn Mem 6:347–362.

Oliphant TE (2007) Python for scientific computing. Comput Sci Eng 9:10–20.

Parra LC, Pearlmutter BA (2007) Illusory percepts from auditory adaptation. J Acoust Soc Am 121:1632–1641.

Schilling A, Gerum R, Metzner C, Maier A, Krauss P (2022a) Intrinsic noise improves speech recognition in a computational model of the auditory pathway. Front Neurosci 795.

Schilling A, Krauss P, Gerum R, Metzner C, Tziridis K, Schulze H (2017) A New Statistical Approach for the Evaluation of Gap-prepulse Inhibition of the Acoustic Startle Reflex (GPIAS) for Tinnitus Assessment. Front Behav Neurosci 11:198.

Schilling, A., Sedley, W., Gerum, R., Metzner, C., Tziridis, K., Maier, A., Schulze, H., Zeng, F.-G., Friston, K.J., Krauss, P., 2022b. Predictive Coding and Stochastic Resonance: Towards a Unified Theory of Auditory (Phantom) Perception. arXiv preprint arXiv:2204.03354.

Schilling A, Tziridis K, Schulze H, Krauss P (2021) The stochastic resonance model of auditory perception: A unified explanation of tinnitus development, Zwicker tone illusion, and residual inhibition. Progress in Brain Research. Elsevier.

Schulze H, Scheich H (1999) Discrimination learning of amplitude modulated tones in Mongolian gerbils. Neurosci Lett 261:13–16.

Shimojo S, Kamitani Y, Nishida SY (2001) Afterimage of perceptually filled-in surface. Science 293:1677–1680.

Steube N, Nowotny M, Pilz PK, Gaese BH (2016) Dependence of the startle response on temporal and spectral characteristics of acoustic modulatory influences in rats and gerbils. Front Behav Neurosci 10:133.

Tighilet B, Dutheil S, Siponen MI, Noreña AJ (2016) Reactive neurogenesis and down-regulation of the potassium-chloride cotransporter KCC2 in the cochlear nuclei after cochlear deafferentation. Front Pharmacol 7:281.

Turner JG (2007) Behavioral measures of tinnitus in laboratory animals. Prog Brain Res 166:147–156.

Turner JG, Brozoski TJ, Bauer CA, Parrish JL, Myers K, Hughes LF, Caspary DM (2006) Gap detection deficits in rats with tinnitus: a potential novel screening tool. Behav Neurosci 120:188–195.

Turrigiano G (2012) Homeostatic synaptic plasticity: local and global mechanisms for stabilizing neuronal function. Cold Spring Harbor Perspect Biol 4 a005736.

Van Der Walt S, Colbert SC, Varoquaux G (2011) The NumPy array: a structure for efficient numerical computation. Comput Sci Eng 13:22–30.

Van Rossum, G., Drake Jr, F.L., 1995. Python reference manual. CWI Report CS-R9525.

Wiegrebe L, Kossl M, Schmidt S (1996) Auditory enhancement at the absolute threshold of hearing and its relationship to the Zwicker tone. Hear Res 100:171–180.

Wirth N (1971) The programming language Pascal. Acta Informatica 1:35–63.

Zwicker E (1964) "Negative afterimage" in hearing. J Acoust Soc Am 36:2413–2415.

(Received 23 December 2022, Accepted 12 April 2023)
(Available online 18 April 2023)

Cognitive Neurodynamics (2024) 18:1753–1765
https://doi.org/10.1007/s11571-023-10038-0

RESEARCH ARTICLE

Coincidence detection and integration behavior in spiking neural networks

Andreas Stoll[1] · Andreas Maier[1] · Patrick Krauss[1,2] · Richard Gerum[3] · Achim Schilling[1,2]

Received: 8 May 2023 / Revised: 11 September 2023 / Accepted: 9 November 2023 / Published online: 13 December 2023

Abstract

Recently, the interest in spiking neural networks (SNNs) remarkably increased, as up to now some key advances of biological neural networks are still out of reach. Thus, the energy efficiency and the ability to dynamically react and adapt to input stimuli as observed in biological neurons is still difficult to achieve. One neuron model commonly used in SNNs is the leaky-integrate-and-fire (LIF) neuron. LIF neurons already show interesting dynamics and can be run in two operation modes: coincidence detectors for low and integrators for high membrane decay times, respectively. However, the emergence of these modes in SNNs and the consequence on network performance and information processing ability is still elusive. In this study, we examine the effect of different decay times in SNNs trained with a surrogate-gradient-based approach. We propose two measures that allow to determine the operation mode of LIF neurons: the number of contributing input spikes and the effective integration interval. We show that coincidence detection is characterized by a low number of input spikes as well as short integration intervals, whereas integration behavior is related to many input spikes over long integration intervals. We find the two measures to linearly correlate via a correlation factor that depends on the decay time. Thus, the correlation factor as function of the decay time shows a powerlaw behavior, which could be an intrinsic property of LIF networks. We argue that our work could be a starting point to further explore the operation modes in SNNs to boost efficiency and biological plausibility.

Keywords Computational modeling · Neural networks · Artificial intelligence · Leaky-integrate-and-fire neuron · Coincidence detection

Introduction

Biological neural networks and especially the human brain achieve astonishing performance in dynamical information processing and energy efficiency. Thus, the human brain does not consume significantly more power than a 20 W light-bulb (Furber 2012), whereas huge matrix processor units used for machine learning approaches consume far more energy (Wang et al. 2020). How is it possible that general intelligence emerges in the human brain, although there exist significant biological constraints? On the one hand, the answer to this question has the potential to significantly boost artificial intelligence (AI) research by

Richard Gerum and Achim Schilling have contributed equally to this work.

✉ Andreas Stoll
andi.stoll@fau.de

✉ Achim Schilling
achim.schilling@fau.de

Andreas Maier
andreas.maier@fau.de

Patrick Krauss
patrick.krauss@fau.de

Richard Gerum
richard.gerum@protonmail.com

[1] Pattern Recognition Lab, University Erlangen-Nürnberg, Erlangen, Germany

[2] Neuroscience Lab, University Hospital Erlangen, Erlangen, Germany

[3] Department of Physics and Astronomy, York University, Toronto, Canada

implementing the underlying biological principles in artificial neural networks [neuroscience inspired AI, (Hassabis et al. 2017), e.g. (Yang et al. 2021; Schilling et al. 2022)]. On the other hand, better artificial neural networks could help to understand how the brain works, as these networks could serve as a model system which can be analyzed with much more detail compared to their biological counterpart [cognitive computational neuroscience, (Kriegeskorte and Douglas 2018), see also (Gerum et al. 2020; Schilling et al. 2021a; Stoewer et al. 2023a, b; Surendra et al. 2023; Metzner et al. 2023; Schilling et al. 2022)].

Consequently, the interest in neuromorphic computing and especially spiking neural networks (SNNs) increased in recent years, as they offer a promising approach to bridge the gap between the performance achieved by current deep learning methods and the energy efficiency of biological neural networks (Eshraghian et al. 2021; Yamazaki et al. 2022; Xiao et al. 2022; Gerum et al. 2023).

The behaviour of biological neurons has first been mathematically described by Hodgkin and Huxley (1952). Even though the Hodgkin–Huxley model accurately describes the underlying experimental observations, it is computationally complex and therefore often simplified. One simplification is for example the Fitzhugh–Nagumo neuron model, which uses a reduced set of differential equations (FitzHugh 1961; Izhikevich and FitzHugh 2006; Nagumo et al. 1962). However, to simulate large networks based on these neuron models is computationally expensive and unfeasible. Therefore, biological neurons are commonly approximated with phenomenological spiking neuron models (Gerstner and Kistler 2002). Spiking neurons produce identical action potentials (spikes) when certain threshold criteria are met for their internal states (Kandel et al. 2000). As a result, these neurons transmit information via an energy-efficient communication scheme that is binary and sparse.

One prevalent spiking neuron model is the leaky-integrate-and-fire (LIF) neuron. It is computationally efficient and therefore used in many spiking neural network studies (Yamazaki et al. 2022).

SNNs provide a more biologically-inspired approach to artificial neural networks compared to standard deep neural networks used for pattern recognition. However, training SNNs remains challenging and is an active field of research (Xiao et al. 2022; Alonso et al. 2022; Apolinario and Roy 2023; Gerum and Schilling 2021; Gerum et al. 2023). Even though simple models like the LIF neuron provide a wide range of parameters that influence the dynamics of information processing, many recently proposed training methods, e.g. Xiao et al. (2022) and Apolinario and Roy (2023), still only optimize the synaptic weights. This is likely due to a lack of understanding the effects of different parameters on the spiking dynamics of SNNs.

Recently though, multiple publications set a starting point to evaluate how SNNs behave under different conditions and parameter combinations. More so, a special focus was set on the temporal parameters (resp. time constants) of LIF neurons. For example, Perez-Nieves et al. (2021) show that SNNs perform best, when there is a certain heterogeneity in the time constants of LIF neurons and report a special benefit for tasks with a lot of temporal structure in the data. Further studies carried out similar experiments, but the authors set their focus on running SNNs more efficiently on neuromorphic hardware (Fang et al. 2021; Quax et al. 2020; Yin et al. 2020). In Gerum and Schilling (2021), the authors proposed that a single LIF neuron can be run in two operation modes: a LIF neuron with a short membrane decay time can be regarded as coincidence detector whereas a LIF neuron with a long membrane decay time acts as integrator neuron. However, it is still unclear whether these operations modes actually arise, and thus, whether they can be precisely tuned in populations of neurons (resp. SNNs).

Coincidence detection in particular seems to be a desirable operation mode, as it was observed in many different sensory modalities and cognitive processes by multiple studies of biological neural networks. It is known to be involved in e.g. memory formation (Bender et al. 2006; Fino et al. 2010) and decoding motor input or sensory stimuli (Xu et al. 2012; Roome and Kuhn 2020). Coincidence detection was also found to be involved in the auditory (Franken et al. 2015) and visual system (Ran et al. 2020) of mammals, respectively.

Detecting a coincidence refers to the process of extracting information from activity across different neurons which occurs within a short period of time. However, there are differences both in what kind of coincidence a system is trying to detect as well as the mechanisms of detecting such. In this study, coincidence detection refers to a postsynaptic neuron being prone to pre-synaptic activity that arrives over a short period of time.

A biological mechanism for detecting this kind of coincidence has been reported to exist in e.g. the auditory system of the Mongolian gerbil (Franken et al. 2015) and is crucial for sound localization. Intrinsic conductances of neurons in the medial superior olive—a brainstem nucleus of the auditory pathway—that interact with preceding synaptic activity are reported to generate an internal phase delay as part of the coincidence detection process. Both the recent input activity as well as low-voltage-activated Kv1 potassium channels alter the postsynaptic neuron's membrane potential which enables fine tuned responses to different temporal input patterns. This biological mechanism is involved in spatial hearing, as the coincidence detection allows to resolve the small time difference of a sound arriving at the two ears. Further studies on the auditory

system proposed that coincidence detection is also important to generate neural networks, which are able to calculate auto-correlations of complex signals. Thus, Krauss et al. (2016, 2018), Schilling et al. (2021b), Schilling and Krauss (2022) and Schilling et al. (2023) propose that these auto-correlations of auditory signals are used by the auditory system to enhance sensory processing and to compensate the effects of hearing loss.

Despite a central role of coincidence detection in the auditory system, it is important in various other modalities and brain regions as well. Coincidence detection also plays a role in e.g. cortical integration of sensory and motor input (Xu et al. 2012), sub-cortical processing of visual stimuli (Ran et al. 2020) and information processing in the cerebellum (Roome and Kuhn 2020).

Regardless of numerous evidence of this operation mode to exist in biological neural networks, its potential is yet to bet tapped into by a majority of SNN studies. In this study, we therefore explore the connection between the membrane time constant and the proposed LIF operation modes in SNNs that are trained on four commonly used image classification datasets: MNIST, EMNIST/Letters, Fashion-MNIST and CIFAR-10. We propose two measures, that allow to determine a neuron's operation mode with respect to the other neurons in a spiking neural network. Thus, the proposed measures can be used to better understand the contribution of single neurons to the dynamics of entire populations of neurons. Besides supporting the explainability of SNNs, we also demonstrate a clear correlation between the membrane decay time (inverse leak term) and the neuron's spiking dynamics in SNNs optimized with a supervised surrogate-gradient-based training method. We find, that the coincidence detection mechanisms observed in biology can be reproduced in networks of LIF neurons in a simplified manner. This makes tuning the operation modes (resp. membrane decay times) an interesting approach to more biological plausibility in machine learning.

Methods

Computational resources

All simulations were performed on standard Desktop PC hardware. The experiments were run on a modified version of the tf_spiking Python package (Gerum 2020b) which is the backbone of our machine learning approaches and based on Keras (Chollet et al. 2015) and TensorFlow (Abadi et al. 2015). For further evaluations, we used NumPy (Harris et al. 2020) and Pandas (The pandas development team 2020). Thus, all visualizations were created with Matplotlib (Hunter 2007) and Pylustrator (Gerum 2020a). The experiments were conducted with a five-fold cross-validation on four different image classification datasets, namely the MNIST database of handwritten digits (Deng 2012), EMNIST/Letters (Cohen et al. 2017), Fashion-MNIST (Xiao et al. 2017) and CIFAR-10 (Krizhevsky et al. 2009). The image pixels are converted into spike trains using Poisson encoding.

Our fully connected network has one hidden layer of 128 LIF neurons and is supervisedly trained using the surrogate-gradient approach proposed in (Gerum and Schilling 2021). The membrane decay times (resp. leak terms) are initialized either with identical values for all neurons in the hidden layer ("constant") or with 32 bins of four neurons each ("binned uniform"). The neurons of a bin are initialized with the same membrane decay time.

Deep learning with leaky-integrate-and-fire neurons

For our experiments, we build a feed-forward spiking neural network based on leaky-integrate-and-fire (LIF) neurons (Burkitt 2006) (see Fig. 1a). As shown in Gerum and Schilling (2021), LIF neurons can be mathematically described by the following equations:

$$V_{t_n} = \mathrm{ReLu}[w_{\mathrm{input}} \cdot x_{t_n} + (1 - w_{\mathrm{leak}} \cdot \Delta t) \cdot V_{t_{n-1}} \cdot \Theta_2(V_{\mathrm{thresh}} - V_{t_{n-1}})] \quad (1)$$

$$y_{t_n} = \Theta_1(V_{t_n} - V_{\mathrm{thresh}}) \quad (2)$$

$$t_n = t_{n-1} + \Delta t \quad (3)$$

$$V_{t_n}, V_{\mathrm{thresh}}, x_{t_n} \in \mathbb{R}, w_{\mathrm{leak}}, \Delta t, y_{t_n} \in \mathbb{R}^+, n \in \mathbb{N}$$

We simulate the neuron for $n = 1, ..., N$ discrete time steps with a temporal resolution of $\Delta t = 5$ ms. The internal state V_{t_n} of the LIF neuron, also referred to as membrane potential, is computed for every time step t_n. It is the sum of the inputs at this time, x_{t_n}, weighted by trainable input weights w_{input}, and the state of the membrane potential of the previous timestep $V_{t_{n-1}}$, weighted by leakage term w_{leak} that prevents long temporal correlations. As this study investigates the influence of the leakage term on the network dynamics, w_{leak} is set to be non-trainable and therefore remains unchanged for all time steps. With the Heaviside step function Θ_i, we can model the neuron to release a spike (Θ_1 in (2)) and to reset the membrane potential (Θ_2 in (1)) to its resting state, if V_{t_n} surpasses the threshold V_{thresh} at the respective time step. If the threshold is not surpassed, V_{t_n} is multiplied with w_{leak} and fed to the inner state via a recurrent connection. The output of a LIF neuron, y_{t_n}, is 0 if no spike occurs at t_n and 1 otherwise. Without loss of generality, V_{thresh} is set to 1. As we work with spike trains as input

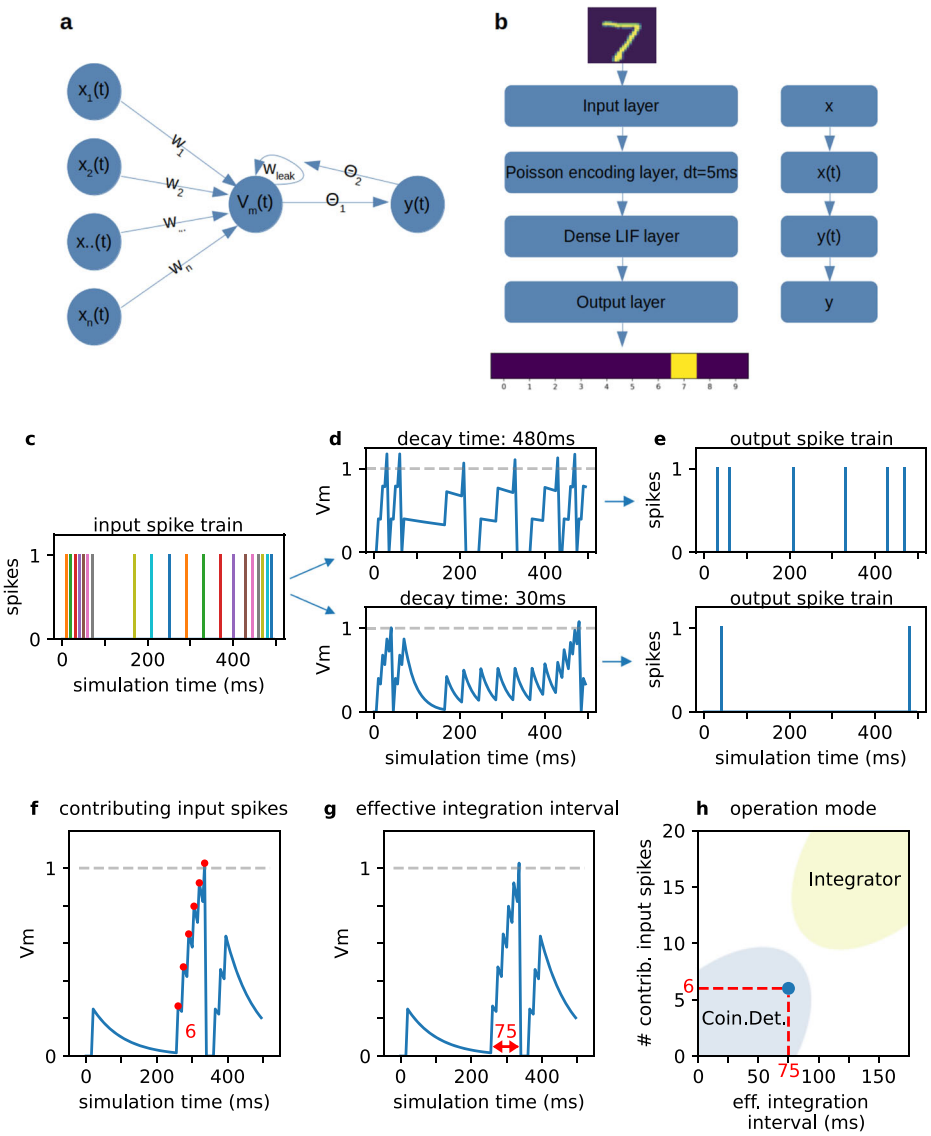

Fig. 1 **a**: The leaky-integrate-and-fire neuron describes the relationship between input currents $x_i(t)$ and the output current $y(t)$. **b**: The feed-forward network architecture used in our study with an exemplary MNIST input image and the resulting classification. Whether the signal is or isn't time-dependent is indicated on the right. Given the input spike train (**c**), a neuron's membrane potential behaves differently according to its decay time (**d**). If it is high (480 ms), V_{m} stays above its resting state for some time and can be integrated over multiple time steps. If it is low (30 ms), the spike threshold is only surpassed when either strong or lots of input activity rapidly stimulates the neuron. The respective output spike trains are visualized in **e**. The number of contributing input spikes (**f**) and the effective integration interval (**g**) are determined by backtracking input spikes in V_{m} after an output spike was elicited. During this process, weight as well as membrane decay effects are being considered. By combining these measures (**h**), we can determine a neuron's operation mode in the context of the total simulation time of the network

signals, both x_{t_n} as well as y_{t_n} implicitly are binary, i.e. $\{0, 1\}$, and independent of Δt. We do not allow negative values for the inner state of the LIF neurons [cf. (2), (Gerum and Schilling 2021)].

For training the SNN, we work with the surrogate gradient-based backpropagation through time approach proposed in Gerum and Schilling (2021). With this learning paradigm, supervised training by minimizing a loss function is possible and classification tasks can be solved. The loss function, in our case the mean squared error loss, is minimized by using a gradient descent algorithm and a step size (learning rate). For the optimization we use the Adam stochastic gradient descent method (Kingma and Ba 2017). Thus, a weight update can be calculated via the chain rule the same way as for multi-layer-perceptron-based artificial neural networks.

As we work with image datasets but the LIF neurons have a time dimension, all image pixels are encoded as spike trains. In this study, we use Poisson rate coding (Zenke and Vogels 2021; Pfeiffer and Pfeil 2018; Lee et al. 2016), where every pixel value is translated to a probability of the neuron to spike in each time step. After the encoding, the inputs are passed to a dense layer consisting of 128 LIF neurons. In order to get a final classification score consistent with the ground truth labels, the output layer sums up the incoming spikes from the hidden layer and maps it to the respective class. We simulate the network for a duration of 500 ms with a temporal resolution of 5 ms, resulting in 100 discrete time steps. An overview of the network architecture is visualized in Fig. 1b.

Tuning the spiking behavior of LIF neurons via their decay times

The decay time is inversely proportional to the leak term ($t_{\mathrm{decay}} = 1/w_{\mathrm{leak}}$) and influences both training dynamics

and spiking behavior of the LIF neurons. For a high decay time (resp. low leakage), a LIF neuron simply sums up the input stimuli over time with little decay of the membrane potential and therefore operates as integrator. On the contrary, if the neuron's decay time is low (resp. high leakage), it can only release output spikes for inputs with small time differences. In such a case, the LIF unit operates as coincidence detector. In Fig. 1c–e, we visualized the membrane potential of a single LIF neuron with a high (low) decay time and the resulting output behavior given an identical synthetic input spike train, respectively.

We refer to these operation modes as integrator and coincidence detector, however, both terms are not correlated to a specific value for the decay time but rather describe the neuron's tendency of operation in the context of the network's simulation time. Neurons with intermediate decay times (resp. an intermediate operation mode) can of course exist and thus the proposed terms for the operation modes strongly depend on the context of the input stimuli.

As our learning rule depends on the membrane potential, the decay time also affects the training behavior of the neuron. An integrator neuron can memorize its inputs over a long duration, so the error gradient also has to be back-propagated over a longer duration than it is the case for coincidence detector neurons. Tuning the decay time therefore allows modeling populations of neurons with different spiking behavior and memorization properties. For our experiments we consider decay times in the range [15, 480] ms. This interval roughly covers the simulation time of the neurons (500 ms) and excludes possible sampling artifacts given our choice of temporal resolution (5 ms). Thus, we can split this range into 32 equidistant bins with 15ms time difference. Via this approach we can distribute the decay times uniformly throughout the network with four neurons sharing a respective decay time. We refer to this initialization scheme as "binned uniform" (visualized in the inset in Fig. 2d).

Results

Coincidence detection and integration behavior in feed-forward spiking networks

In the following, we analyze how coincidence detector and integrator neurons (Gerum and Schilling 2021) perform and behave in a feed forward neural network trained with a surrogate gradient algorithm. Thus, we study the effect of different decay times on the spiking behavior in networks trained on four common datasets (MNIST, EMNIST/Letters, Fashion-MNIST or CIFAR-10 dataset). We investigate networks with constant decay times (i.e. it is equal for all LIF units of the network), and with binned uniform decay times, (i.e. it is uniformly distributed with an equal amount of neurons sharing a respective decay time). Thus, the decay time is identical for all time steps and is not a trainable parameter. For better readability, we only report the results based on MNIST in this Section; the visualizations for the other datasets can be found in Suppl. Fig. 1.

The operation mode of the neurons is quantified by the number of input spikes effectively contributing to the generation of an output spike (see Fig. 1f). A low number of contributing input spikes suggests that either the weights of these input spikes were very high, or that an input spike volley (several spikes coming from arbitrary input neurons with small inter-spike intervals) was present. As we are particularly interested in detecting coincidences, we additionally measure the average time interval in which the input spikes stimulate the LIF unit and cause it to spike. We call this the effective integration interval (see Fig. 1g).

To calculate this measure, we start with the time of the output spike and trace the membrane potential back to the first input spike that actively contributes to the generation of this output spike. During this backtracking, weight effects and membrane decay effects are being taken into account. A long effective integration interval (w.r.t. the network's simulation time) is present in integrator neurons due to little decay of the membrane potential. We expect that coincidence detector neurons have shorter integration intervals compared to integrator neurons and require less input spikes.

We compute both measures for every output spike of every neuron in the hidden layer and use them to determine the operation modes of the neurons: If both measures (effective integration interval and contributing input spikes) are low the neuron operates as coincidence detector. If both measures are high the neuron operates as integrator as indicated in Fig. 1h.

Prior to analyzing whether a low number of contributing input spikes correlates with a short effective integration interval in networks trained on real data, we investigated the decay time's influence on both measures. Low decay times correspond to low measure scores in experiments using constant as well as binned uniformly distributed decay times, as reported in Fig. 2a–d.

Visualizing both measures in the "operation mode" scatter plot (introduced in Fig. 1h), we find that a low (high) number of contributing input spikes correlates with a short (long) effective integration interval (see Fig. 2e, f). However, we find the decay time to not determine the exact behavior of the neuron but instead defines a range of operation. This range gets smaller for low and saturates for high decay times, respectively. These results imply, that we can in fact influence the operation mode of a LIF neuron via its decay time (resp. leak term). Furthermore, we see

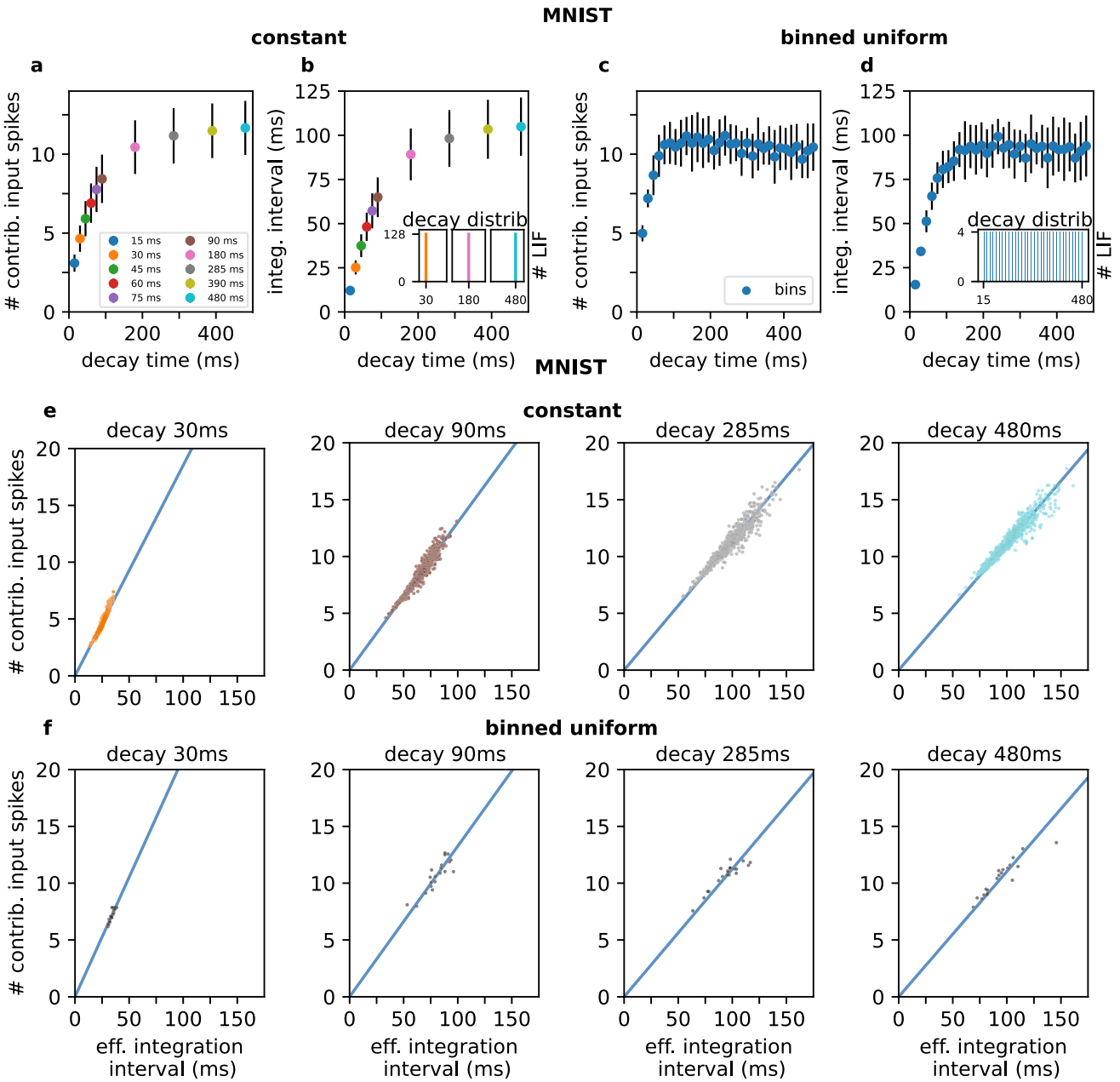

Fig. 2 Both measures are influenced by the decay time. The two measures we introduced to determine the neuron's operation mode are clearly influenced by the decay time in networks trained on MNIST. The results of the experiments using constant decay times are shown in **a** and **b**. Every point denotes an experiment with exclusively using the respective decay time. **c** and **d** show the results of the experiment using binned uniformly distributed decay times. Here, every point denotes one bin of neurons with the respective decay time. The average numbers of contributing input spikes are visualized in plots **a** and **c**; the effective integration intervals are plotted in **b** and **d**. The decay time impacts both measures similarly in experiments with constant as well as binned uniform initialization.When visualizing the number of contributing input spikes and the respective effective integration interval w.r.t. a specific decay time as scatter plot, a linearly-shaped distribution forms which gets steeper and more compact for lower decay times. This trend was observed to be identical in constant (**e**) and binned uniform (**f**) experiments. These results suggest that the decay time determines a neuron's operational range rather than an exact operation mode. The colored lines show linear fits through the respective distributions

that in a population of neurons trained on real data, integration and coincidence detection behavior emerge depending on the decay time. The networks did not simply adjust the weights to counter the effect of the decay time, but instead worked with neurons operating on different time scales.

The slopes of the drawn measurement values shown in Fig. 2e and f become steeper for lower decay times when training on EMNIST/Letters, Fashion-MNIST and CIFAR-

10 data, respectively (see Suppl.Fig. 1 B). We therefore fitted lines through every distribution and the coordinate origin and computed their slopes. When plotting the slopes of these line fits over their respective decay time, we find similar curves for constant and binned uniform experiments and for all four datasets, as shown in Fig. 3a and 3b.

An in-depth analysis of these slopes as a function of the decay times (Fig. 3) indicates that the slopes of the curves are shifted along the y-axis depending on the dataset.

We found that the offset along the y-axis is influenced by image brightness. We therefore adjusted the average brightness of MNIST, EMNIST/Letters and Fashion-MNIST images to approximately match (brightness difference < 0.045 in [0, 255] color space, see Fig. 3c, d) However, we also see that curves from brightness-adjusted MNIST and unmodified EMNIST/Letters are similar, despite an average brightness difference of approx. 10 in [0, 255] color space. This suggests that not only image brightness, but also the structure of the data influences the slopes.

Additionally, when removing the y-offsets and fitting powerlaws to the resulting curves, we can compare the slopes for neural networks trained on the different datasets. The powerlaw fits are presented in double logarithmic scale in Fig. 3e, f. The curves of brightness-adjusted and non-adjusted data are very similar indicating that image brightness influences the y-offset but has little impact on the shape of the curve. In general, the different powerlaw fits are not perfectly aligned, potentially due to an influence of the structure of the data. Furthermore, we observed that the fit functions for networks trained on MNIST and EMNIST/Letters datasets are similar for the binned uniform experiments. Also the fit functions of Fashion-MNIST and CIFAR-10 experiments are similar. We therefore assume that the distribution of the pixel intensities shapes the slope of the curve, as MNIST and EMNIST/Letters are almost binary in intensity, whereas Fashion-MNIST and CIFAR-10 more extensively use the offered value range of the brightness scale.

Impact of decay times on model accuracy

Different decay times lead to changes in spiking dynamics. In the following, we show the influence of the different decay times on classification accuracy. Therefore, we evaluated the five-fold cross-validated mean accuracy, macro f1-score and area under the receiver operating characteristic (AUC) for all experimental conditions. In Fig. 4a–c, the accuracies are reported for MNIST, Fashion-MNIST and EMNIST/Letters (detailed performance overview of accuracy, macro f1 and AUC scores of all datasets see Suppl. Figure 1 C).

The binned uniform distributions lead to similar performances as the best models with constant decay time.

In a next step, we investigated the influence of the decay time on overall classification accuracy by cumulatively ablating neurons either starting with coincidence detectors (starting ablation at low decay times) or integrator neurons (starting ablation at high decay times), respectively. The results for all datasets are reported in Fig. 4d-i. Deleting integrator neurons first leads to an instant accuracy drop for MNIST (4d), Fashion-MNIST (4f) and EMNIST/Letters (4h) datasets. For CIFAR-10 (4i) no preference of operation mode can be detected.

Comparing the ablation curves of MNIST (4d) and brightness-adjusted MNIST (4e), we find that increasing the image brightness leads to a smaller difference between ascending (green) and descending (blue) ablation curves. Consistently, a greater difference is observed when lowering the image brightness of Fashion-MNIST (see 4f, g). This therefore suggests, that the performance difference between integrator and coincidence detector neurons is linked to image brightness.

In summary, integrator neurons seem to be more important for classification performance. However, this result has to be discussed due to the observed dependence on image brightness (resp. spike rate given the Poisson spike encoding).

Discussion

Summary

In the present study, we investigated whether the previously proposed operation modes of single LIF neurons do emerge in spiking neural networks that are trained on real data. We thus studied the influence of tuning the membrane decay time [i.e. membrane time constant (Perez-Nieves et al. 2021) or inverse leak term] on the operation modes of LIF neurons. Furthermore, we analyzed the resulting effects on spiking dynamics and image classification accuracy of SNNs. We found that the proposed operation modes do emerge in SNNs and that they can be tuned via the membrane decay time: Neurons with low decay times operate as coincidence detectors, whereas neurons with high decay times operate as integrators.

We performed experiments with four image datasets (MNIST, EMNIST/Letters, Fashion-MNIST and CIFAR-10), which were transformed to spike sequences by Poisson encoding [Zenke and Vogels 2021; Pfeiffer and Pfeil 2018; Lee et al. 2016]. We deployed feed-forward SNNs with a single hidden layer and used the surrogate-gradient-based training method proposed by Gerum and Schilling (2021). In this study, we experimentally investigated the effect of

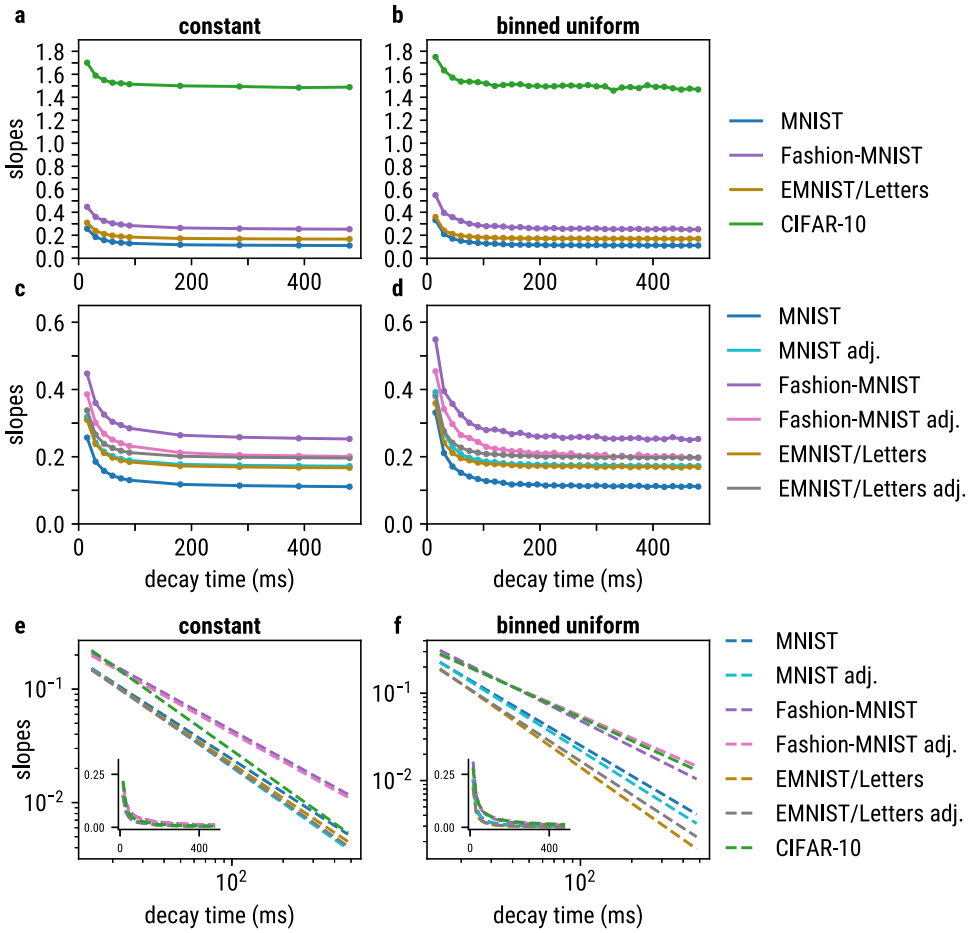

Fig. 3 The slopes of the line fits in the scatter plots correlate with the decay time for both constant (**a**) and binned uniform (**b**) initialization. The impacts of different decay times are similar between the different datasets, however, the brightness of the input data produces an offset along the y-axis. After adjusting the average brightness of MNIST, EMNIST/Letters and Fashion-MNIST to approximately match, the curves are close in constant (**c**) and binned uniform (**d**) experiments. As they are not entirely aligned, the structure of the data also seems to effect the slopes. **e** and **f**: In order to compare the slopes between the different datasets, we subtracted the y-offset and fitted powerlaws to the curves from **a** and **b**, respectively. The powerlaw fits are presented in double logarithmic scale. For adjusted and non-adjusted data, the curves are closely aligned. In contrast, the powerlaw fits of different datasets only approximately match, suggesting an influence of the structure within the data. Remarkably, we observe the curves of MNIST and EMNIST/Letters to closely match in the binned uniform experiments, as is also the case for Fashion-MNIST and CIFAR-10, respectively. This suggests, that the distribution of the pixel intensities shapes the slope of the curve

different membrane decay times and considered two different ways of initializing them: constant decay times and uniformly distributed decay times, respectively.

In order to study the relationship between membrane decay times and operation modes of LIF neurons, we proposed two measures: the number of contributing input spikes and the effective integration interval.

We found the first measure, the number of contributing input spikes, to be low for coincidence detectors, whereas it was found to be high for integrator neurons. The second measure, the effective integration interval, was also found to be higher for integrator neurons.

We analyzed the distribution of the two measures across the SNNs with respect to particular membrane decay times and found the two measures to linearly correlate. Besides investigating the relationship between the measures and a neuron's operation mode, we found that both measures are strongly influenced by the neuron's membrane decay time (see Fig. 2a–d).

Therefore, we can conclude that the operation mode of a LIF neuron can be determined via the membrane decay time. A low decay time correlates to a low number of contributing input spikes and a short effective integration interval. This consequently makes a LIF neuron to operate as coincidence detector. High decay times correspond to high numbers of contributing input spikes as well as long effective integration intervals. This makes such neurons to operate as integrators. However, saturation effects for high decay times were visible in both measures. This suggests a

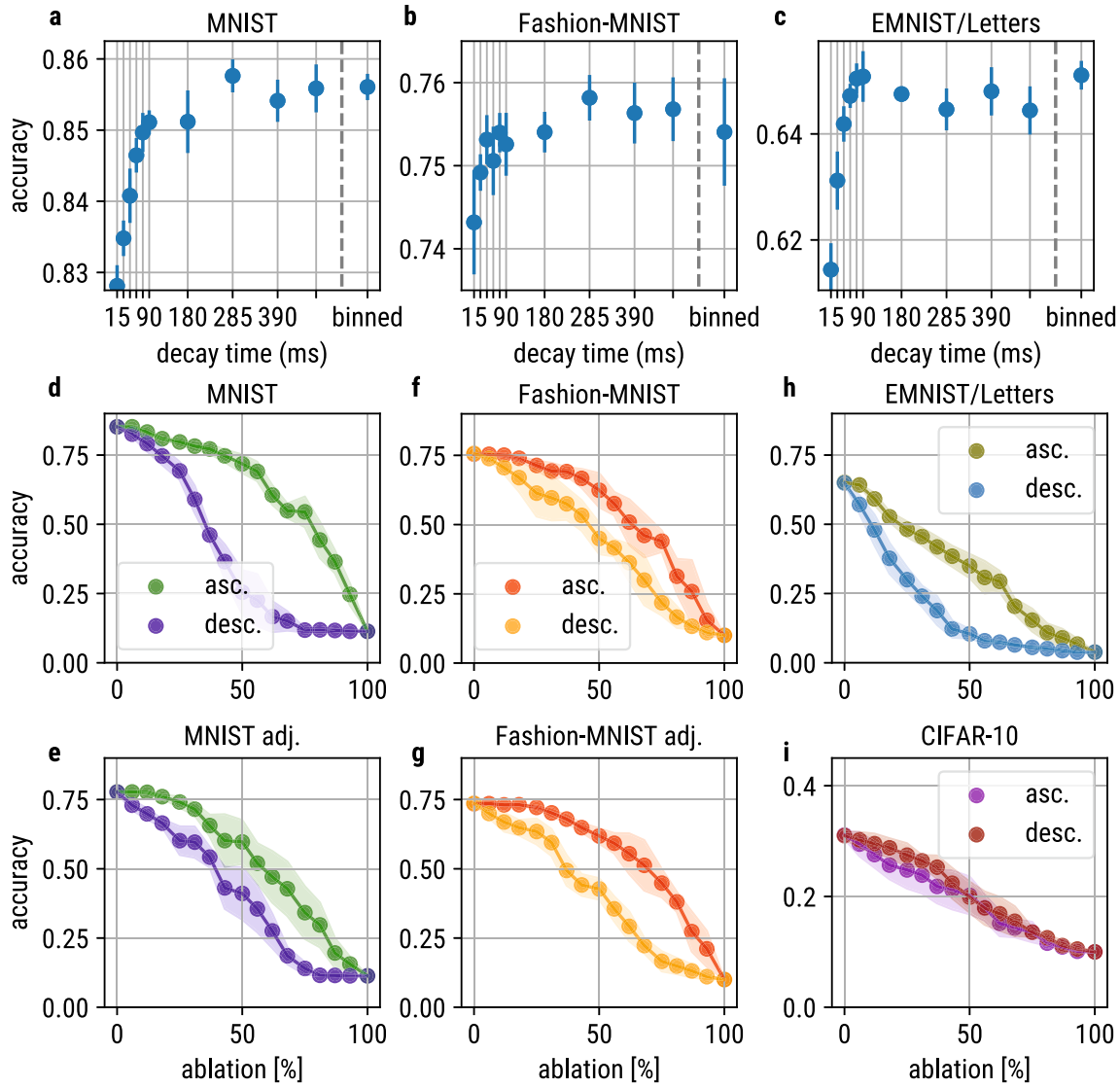

Fig. 4 Accuracy and cumulative decay-based ablation. The accuracy scores of networks trained on MNIST (**a**), EMNIST/Letters (**b**) and Fashion-MNIST (**c**) are visualized w.r.t. the decay time of experiments using constant initialization. In the binned uniform experiment, the decay times were distributed equidistantly over 32 bins in the range [15, 480] ms. Every dot shows the five-fold cross-validated mean accuracy and the respective standard deviation. The accuracy is similar between most decay times within every dataset, however a tendency to decrease for low decay times can be noticed. The experiments using binned uniform initialization achieve approx. equal accuracy to the best models using constant decay times. The impact of different decay times on classification accuracy can be determined by cumulatively ablating neurons from the binned uniform experiments w.r.t. their decay times. We start with ablating coincidence detectors (ascending decay time) or integrators (descending decay time) first, respectively. The results for all datasets are visualized in **d, f, h, i**. For MNIST, EMNIST/Letters and Fashion-MNIST, ablating neurons with high decay times first, i.e. integrators, (desc. decay time) more rapidly results in a major loss of accuracy compared to ablating coincidence detectors first (asc. decay time). Comparing the models trained on MNIST (**d**) and Fashion-MNIST (**f**) to their adjusted versions (**e** and **g**), respectively, the image brightness seems to influence the importance of the different operation modes. A higher brightness in MNIST images decreases the difference of the ascending and descending ablation curves, while reducing the brightness in Fashion-MNIST images increases the difference of the curves

non-linear relationship between the membrane decay time and a neuron's respective operation mode.

Our experiments give strong evidence that LIF neurons can be precisely tuned towards detecting coincidences, whereas integration behavior is not accurately determined: our analyses show, that the effective integration interval and number of contributing input spikes are more strongly determined towards a precise range (i.e. clustered together) for low decay times (see Fig. 2e, f). We can therefore conclude that the membrane decay time defines a neuron's operational range rather than an exact operation mode. This

operational range gets smaller (resp. more precise) for lower decay times.

Additionally, we found that the correlation factor between the two measures as a function of the membrane decay time follows a powerlaw. Therefore, the decay time offers a different way of influencing the spiking dynamics compared to the synaptic weights: while the weights linearly influence the spiking behavior of a LIF layer, the membrane decay time influences the spiking dynamics in a non-linear way.

The powerlaw relation between the measures and the decay time was present in all experiments and all datasets and we therefore argue that it is an intrinsic property of LIF neurons. However, we found it to slightly differ between different datasets and it is still not completely clear to which proportion the powerlaw relation is influenced by the structure of input data. We observed the brightness of the input data to linearly shift the curve of the correlation factor but to have little influence on the slope of the curve (see Fig. 3c–f). The observed similarities of the fit functions between MNIST and EMNIST/Letters and those between Fashion-MNIST and CIFAR-10, respectively, thus suggest that the exact shape of the powerlaw is indeed influenced by the structure of input data, e.g. the distribution of the pixel intensities.

Besides showing the emergence of coincidence detectors and integrators in SNNs and analyzing the resulting effects on the spiking dynamics, we explored the impact of different operation modes on image classification accuracy. For that, we cumulatively ablated neurons according to their decay times. Neurons were ablated either in ascending or descending order, respectively (see Fig. 4d–i).

Ascending ablation refers to deleting coincidence detectors first, which forces the network to use neurons that operate as integrators.

Descending ablation refers to deleting integrator neurons first, which forces the network to use neurons that operate as coincidence detectors.

We found that ablating integrator neurons had a more severe effect on classification accuracy than ablating coincidence detectors, which indicates that integrator neurons are more important in our experiments.

Additionally, we found the differences between the ascending and descending ablation curve to be strongly dependent of image brightness. This is likely due to encoding the images via a Poisson process. A brighter pixel is represented with a higher probability of the respective input neuron to spike. Consequently, the spike rate of such neurons is higher compared to neurons that encode darker pixels. Therefore, when the image brightness decreases, the spiking activity in the SNNs gets more sparse and consequently, fewer spikes coincide. In order to achieve good classification performance in such a case, integrator neurons, i.e. long decay times, are required, as they are able to memorize the information carried by spikes over a longer duration.

Limitations of the study

It has to be noted, however, that the Poisson process encodes all the information of an image pixel via the spike rate of the respective input neuron. As a result, detecting a coincidence in our experiments does not provide more/different but less information than simply integrating over an arbitrary amount spikes. Because of that, we found integrator neurons to be more important than coincidence detectors in our experiments in terms of classification performance.

Even though this limits the results of our ablation study, we can conclude that when working with spike rates, tuning the membrane decay times can be neglected and training the synaptic weights is sufficient in order to achieve good classification performance. However, Perez-Nieves et al. (2021) showed that tuning the time constants can result in improved network performance, when information is not only encoded in the spike rate but in the spike timing as well. We therefore argue that when working with such data, tuning the membrane decay times of LIF neurons should be taken into account. This can be achieved either by considering the membrane decay times as trainable parameters as proposed by Gerum (2020b), or alternatively, by considering the distribution of decay times as hyper-parameter as we did in this study.

We thus want to emphasize that the timings of spikes are important when working with data from neuromorphic sensors like dynamic vision sensors or artificial cochleas (Eshraghian et al. 2021). We will therefore shift our focus to working with neuromorphic sensor data in the future.

Also, we only considered small feed-forward SNNs in this study due to the computational complexity induced by training SNNs [see also (Perez-Nieves et al. 2021)]. This limits our experimental evidence, as more complex effects could potentially emerge in larger—or even recurrent—spiking neural networks. Still, our experimental setup is a reasonable choice as there currently is no viable alternative to training SNNs in time-stepped simulation frameworks when good performance needs to be achieved (Eshraghian et al. 2021).

Discussion and future research directions

In summary, we could demonstrate the rich dynamics of LIF-based spiking neural networks trained with surrogate gradient descent and provide evidence for the validity of defining two operation modes of LIF neurons: the integrator and coincidence detector.

We show that the coincidence detection mechanisms that have been observed in biological neural networks by multiple studies can be reproduced in LIF neurons by tuning the membrane decay times.

A recent study already showed that heterogeneous time constants can improve the performance of LIF-based SNNs (Perez-Nieves et al. 2021). Much work was already spend on investigating the effect of weight matrix heterogeneity on network dynamics [e.g. (Krauss et al. 2019b; Yang et al. 2021; Krauss et al. 2019c, a)]. However, only recently the exact temporal dynamics have moved into the focus of AI research.

As the temporal dynamics of LIF neurons are influenced by multiple parameters (e.g. membrane time constant, spike rate adaptation, data encoding), our aim was to disentangle these parameters and to study the impact of different membrane decay times. With this study, we contribute towards better understanding the dynamics of SNNs by providing experimental evidence for the emergence of different neural operation modes and their dependence on the membrane decay time.

Because the timing of spikes is important when working with neuromorphic sensor data, we strongly encourage the neuromorphic community to consider tuning the operation mode of LIF neurons in future experiments and to consider the membrane decay time in new training methods.

Concluding remarks

To the best of our knowledge this study is the first to investigate the integration and coincidence detection behavior of LIF neurons in spiking neural networks and we thus provide a valid contribution to decode the basis of heterogeneity as fundamental principle of brain dynamics and efficient information processing in SNNs.

As already suggested in Jonas and Kording (2017), the best way to understand a complex system like the brain or artificial neural networks, is to search for already known building blocks (i.e. integrator and coincidence detector). To put it in a nutshell, a mechanistic theory is necessary to make real progress in understanding cognition in biological and artificial neural networks (Jonas and Kording 2017; Schilling et al. 2023).

Author Contributions Conceptualization, ASchi, RG; methodology, ASchi, ASt, RG; software, ASt; visualization, ASt; writing-original draft preparation, ASt, ASchi; internal review and editing, ASchi, AM, PK, RG; supervision, ASchi, RG; project administration, ASchi; All authors have read and agreed to the published version of the manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation): grants KR 5148/2-1 (project number 436456810), KR 5148/3-1 (Project Number 510395418) and GRK 2839 (Project Number 468527017) to PK, and Grant SCHI 1482/3-1 (Project Number 451810794) to ASchi. Furthermore, the research leading to these results has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (ERC Grant No. 810316 to AM).

Data availability The current study was conducted using publicly available datasets.

Code Availability The code generated during the current study is available in the GitHub repository, https://github.com/andistoll/coincidence_detection_and_integration_behavior_in_SNNs

Declarations

Conflict of interest The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Supplementary Information The online version of this article (https://doi.org/10.1007/s11571-023-10038-0) contains supplementary material, which is available to authorized users.

References

Abadi M et al (2015) TensorFlow: large-scale machine learning on heterogeneous systems. Software. https://doi.org/10.5281/zenodo.4724125

Alonso N, Millidge B, Krichmar J, et al. (2022) A theoretical framework for inference learning. In: Koyejo S, Mohamed S, Agarwal A, et al. (eds) Advances in Neural Information Processing Systems, vol 35. Curran Associates, Inc., pp 37335-37348. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/f242c4cba2467637256722cb679642bd-Paper-Conference.pdf

Apolinario MPE, Roy K (2023) S-tllr: Stdp-inspired temporal local learning rule for spiking neural networks. https://doi.org/10.48550/arXiv.2306.15220. Currently under review

Bender VA, Bender KJ, Brasier DJ et al (2006) Two coincidence detectors for spike timing-dependent plasticity in somatosensory cortex. J Neurosc 26(16):4166–4177. https://doi.org/10.1523/JNEUROSCI.0176-06.2006

Burkitt AN (2006) A review of the integrate-and-fire neuron model: I. Homogeneous synaptic input. Biol Cybern 95(1):1–19. https://doi.org/10.1007/s00422-006-0068-6

Chollet F, et al. (2015) Keras. Software available from https://keras.io

Cohen G, Afshar S, Tapson J, et al. (2017) EMNIST: an extension of MNIST to handwritten letters. https://doi.org/10.48550/ARXIV.1702.05373

Deng L (2012) The MNIST database of handwritten digit images for machine learning research. IEEE Signal Process Mag 29(6):141–142. https://doi.org/10.1109/MSP.2012.2211477

Eshraghian JK, Ward M, Neftci E, et al. (2021) Training spiking neural networks using lessons from deep learning. CoRR abs/2109.12894. https://doi.org/10.48550/arXiv.2109.12894

Fang W, Yu Z, Chen Y, et al. (2021) Incorporating Learnable Membrane Time Constant to Enhance Learning of Spiking Neural Networks. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 2661–2671. https://doi.org/10.1109/ICCV48922.2021.00266

Fino E, Paille V, Cui Y et al (2010) Distinct coincidence detectors govern the corticostriatal spike timing-dependent plasticity. J Physiol 588(16):3045–3062. https://doi.org/10.1113/jphysiol.2010.188466

FitzHugh R (1961) Impulses and physiological states in theoretical models of nerve membrane. Biophys J 1(6):445–466. https://doi.org/10.1016/s0006-3495(61)86902-6

Franken TP, Roberts MT, Wei L et al (2015) In vivo coincidence detection in mammalian sound localization generates phase delays. Nat Neurosci 18:444–452. https://doi.org/10.1038/nn.3948

Furber S (2012) To build a brain. IEEE Spectrum 49(8):44–49. https://doi.org/10.1109/MSPEC.2012.6247562

Gerstner W, Kistler WM (2002) Spiking neuron models: single neurons, populations, plasticity. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511815706

Gerum R, Erpenbeck A, Krauss P, et al. (2023) Leaky-integrate-and-fire neuron-like long-short-term-memory units as model system in computational biology. In: 2023 international joint conference on neural networks (IJCNN), pp 1–9. https://doi.org/10.1109/IJCNN54540.2023.10191268

Gerum RC (2020a) pylustrator: code generation for reproducible figures for publication. J Open Source Softw 5(51):1989. https://doi.org/10.21105/joss.01989

Gerum RC (2020b) TensorFlow spiking layer. Software https://github.com/rgerum/tf spiking

Gerum RC, Schilling A (2021) Integration of leaky-integrate-and-fire neurons in standard machine learning architectures to generate hybrid networks: a surrogate gradient approach. Neural Comput 33(10):2827–2852. https://doi.org/10.1162/neco_a_01424

Gerum RC, Erpenbeck A, Krauss P et al (2020) Sparsity through evolutionary pruning prevents neuronal networks from overfitting. Neural Netw 128:305–312. https://doi.org/10.1016/j.neunet.2020.05.007

Harris CR et al (2020) Array programming with NumPy. Nature 585(7825):357–362. https://doi.org/10.1038/s41586-020-2649-2

Hassabis D, Kumaran D, Summerfield C et al (2017) Neuroscience-inspired artificial intelligence. Neuron 95(2):245–258. https://doi.org/10.1016/j.neuron.2017.06.011

Hodgkin AL, Huxley AF (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. J Physiol 117:500–544. https://doi.org/10.1113/jphysiol.1952.sp004764

Hunter JD (2007) Matplotlib: A 2D graphics environment. Comput Sci Eng 9(3):90–95. https://doi.org/10.5281/zenodo.592536

Izhikevich EM, FitzHugh R (2006) Fitzhugh–Nagumo model. Scholarpedia 1(9):1349. https://doi.org/10.4249/scholarpedia.1349

Jonas E, Kording KP (2017) Could a neuroscientist understand a microprocessor? PLoS Comput Biol 13(1):e1005268. https://doi.org/10.1371/journal.pcbi.1005268

Kandel ER, Schwartz JH, Jessell TM et al (2000) Principles of Neural Science, vol 4. McGraw-Hill, New York

Kingma DP, Ba J (2017) Adam: A Method for Stochastic Optimization. https://doi.org/10.48550/arXiv.1412.6980

Krauss P, Tziridis K, Metzner C et al (2016) Stochastic resonance controlled upregulation of internal noise after hearing loss as a putative cause of tinnitus-related neuronal hyperactivity. Front Neurosci. https://doi.org/10.3389/fnins.2016.00597

Krauss P, Tziridis K, Schilling A et al (2018) Cross-modal stochastic resonance as a universal principle to enhance sensory processing. Front Neurosci. https://doi.org/10.3389/fnins.2018.00578

Krauss P, Prebeck K, Schilling A et al (2019) Recurrence resonance'' in three-Neuron Motifs. Front Comput Neurosci 13:64. https://doi.org/10.3389/fncom.2019.00064

Krauss P, Schuster M, Dietrich V et al (2019) Weight statistics controls dynamics in recurrent neural networks. PloS One 14(4):e0214541. https://doi.org/10.1371/journal.pone.0214541

Krauss P, Zankl A, Schilling A et al (2019) Analysis of structure and dynamics in three-neuron motifs. Front Comput Neurosci 13:5. https://doi.org/10.3389/fncom.2019.00005

Kriegeskorte N, Douglas PK (2018) Cognitive computational neuroscience. Nature Neurosci 21(9):1148–1160. https://doi.org/10.1038/s41593-018-0210-5

Krizhevsky A, Nair V, Hinton G (2009) CIFAR-10 (Canadian Institute for Advanced Research). http://www.cs.toronto.edu/kriz/cifar.html

Lee JH, Delbruck T, Pfeiffer M (2016) training deep spiking neural networks using backpropagation. Front Neurosci 10:508. https://doi.org/10.3389/fnins.2016.00508

Metzner C, Yamakou ME, Voelkl D, et al. (2023) Quantifying and maximizing the information flux in recurrent neural networks. arXiv preprint arXiv:2301.12892https://doi.org/10.48550/arXiv.2301.12892

Nagumo J, Arimoto S, Yoshizawa S (1962) An active pulse transmission line simulating nerve axon. Proc IRE 50(10):2061–2070. https://doi.org/10.1109/JRPROC.1962.288235

Perez-Nieves N, Leung VC, Dragotti PL et al (2021) Neural heterogeneity promotes robust learning. Nat Commun 12(1):1–9. https://doi.org/10.1038/s41467-021-26022-3

Pfeiffer M, Pfeil T (2018) Deep learning with spiking neurons: opportunities and challenges. Front Neurosci 12:774. https://doi.org/10.3389/fnins.2018.00774

Quax SC, D'Asaro M, van Gerven MA (2020) Adaptive time scales in recurrent neural networks. Sci Rep 10(1):1–14. https://doi.org/10.1038/s41598-020-68169-x

Ran Y, Huang Z et al (2020) Type-specific dendritic integration in mouse retinal ganglion cells. Nat Commun. https://doi.org/10.1038/s41467-020-15867-9

Roome CJ, Kuhn B (2020) Dendritic coincidence detection in Purkinje neurons of awake mice. eLife 9:e59619. https://doi.org/10.7554/eLife.59619

Schilling A, Krauss P (2022) Tinnitus is associated with improved cognitive performance and speech perception: Can stochastic resonance explain? Front Aging Neurosci. https://doi.org/10.3389/fnagi.2022.1073149

Schilling A, Maier A, Gerum R et al (2021) Quantifying the separability of data classes in neural networks. Neural Netw 139:278–293. https://doi.org/10.1016/j.neunet.2021.03.035

Schilling A, Tziridis K, Schulze H et al (2021) The stochastic resonance model of auditory perception: a unified explanation of tinnitus development, zwicker tone illusion, and residual inhibition. Progress Brain Res 262:139–157. https://doi.org/10.1016/bs.pbr.2021.01.025

Schilling A, Gerum R, Metzner C et al (2022) Intrinsic noise improves speech recognition in a computational model of the auditory pathway. Front Neurosci. https://doi.org/10.3389/fnins.2022.908330

Schilling A, Sedley W, Gerum R et al (2023) Predictive coding and stochastic resonance as fundamental principles of auditory phantom perception. Brain. https://doi.org/10.1093/brain/awad255

Stoewer P, Schilling A, Maier A, et al. (2023a) Conceptual cognitive maps formation with neural successor networks and word embeddings. arXiv preprint arXiv:2307.01577https://doi.org/10.48550/arXiv.2307.01577

Stoewer P, Schilling A, Maier A et al (2023) Neural network based formation of cognitive maps of semantic spaces and the putative emergence of abstract concepts. Sci Rep 13(1):3644. https://doi.org/10.1038/s41598-023-30307-6

Surendra K, Schilling A, Stoewer P, et al. (2023) Word class representations spontaneously emerge in a deep neural network trained on next word prediction. arXiv preprint arXiv:2302.07588https://doi.org/10.48550/arXiv.2302.07588

The Pandas Development Team (2020) pandas-dev/pandas: pandas. Software. https://doi.org/10.5281/zenodo.3509134

Wang Y, Wang Q, Shi S, et al. (2020) Benchmarking the performance and energy efficiency of AI accelerators for AI training. In: 2020 20th IEEE/ACM international symposium on cluster, cloud and internet computing (CCGRID), pp 744–751. https://doi.org/10.1109/CCGrid49817.2020.00-15

Xiao H, Rasul K, Vollgraf R (2017) Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms. CoRR abs/1708.07747. URL http://arxiv.org/abs/1708.07747

Xiao M, Meng Q, Zhang Z, et al. (2022) Online training through time for spiking neural networks. In: Koyejo S, Mohamed S, Agarwal A, et al (eds) Advances in neural information processing systems, vol 35. Curran Associates, Inc., pp 20717–20730. https://proceedings.neurips.cc/paper_files/paper/2022/file/82846e19e6d42ebfd4ace4361def29ae-Paper-Conference.pdf

Xu N, Harnett MT et al (2012) Nonlinear dendritic integration of sensory and motor input during an active sensing task. Nature 492:247–251. https://doi.org/10.1038/nature11601

Yamazaki K, Vo-Ho VK, Bulsara D et al (2022) Spiking neural networks and their applications: a review. Brain Sci 12:863. https://doi.org/10.3390/brainsci12070863

Yang Z, Schilling A, Maier A, et al. (2021) Neural networks with fixed binary random projections improve accuracy in classifying noisy data. In: Palm C, Deserno TM, Handels H, et al (eds) Bildverarbeitung für die Medizin 2021. Springer Fachmedien Wiesbaden, pp 211–216. https://doi.org/10.1007/978-3-658-33198-6_51

Yin B, Corradi F, Bohté SM (2020) Effective and efficient computation with multiple-timescale spiking recurrent neural networks. CoRR abs/2005.11633:1–8. https://doi.org/10.48550/arXiv.2005.11633

Zenke F, Vogels TP (2021) The remarkable robustness of surrogate gradient learning for instilling complex function in spiking neural networks. Neural Comput 33(4):899–925. https://doi.org/10.1162/neco_a_01367

CHAPTER

6

The stochastic resonance model of auditory perception: A unified explanation of tinnitus development, Zwicker tone illusion, and residual inhibition

Achim Schilling[a,b], Konstantin Tziridis[a], Holger Schulze[a], and Patrick Krauss[a,b,c,d,*]

[a]*Neuroscience Lab, Experimental Otolaryngology, University Hospital Erlangen, Erlangen, Germany*

[b]*Cognitive Computational Neuroscience Group at the Chair of English Philology and Linguistics, Friedrich-Alexander University Erlangen-Nürnberg (FAU), Erlangen, Germany*

[c]*FAU Linguistics Lab, Friedrich-Alexander University Erlangen-Nürnberg (FAU), Erlangen, Germany*

[d]*Department of Otorhinolaryngology/Head and Neck Surgery, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands*

**Corresponding author: Tel: +4991318533911, e-mail address: patrick.krauss@uk-erlangen.de*

Abstract

Stochastic resonance (SR) has been proposed to play a major role in auditory perception, and to maintain optimal information transmission from the cochlea to the auditory system. By this, the auditory system could adapt to changes of the auditory input at second or even sub-second timescales. In case of reduced auditory input, somatosensory projections to the dorsal cochlear nucleus would be disinhibited in order to improve hearing thresholds by means of SR. As a side effect, the increased somatosensory input corresponding to the observed tinnitus-associated neuronal hyperactivity is then perceived as tinnitus. In addition, the model can also explain transient phantom tone perceptions occurring after ear plugging, or the Zwicker tone illusion. Vice versa, the model predicts that via stimulation with acoustic noise, SR would not be needed to optimize information transmission, and hence somatosensory noise would be tuned down, resulting in a transient vanishing of tinnitus, an effect referred to as residual inhibition.

Progress in Brain Research, Volume 262, ISSN 0079-6123, https://doi.org/10.1016/bs.pbr.2021.01.025

Keywords

Auditory phantom perception, Somatosensory projections, Dorsal cochlear nucleus, Speech perception, Tinnitus, Zwicker tone, Residual inhibition, Stochastic resonance

1 Stochastic resonance

In engineering, the term *noise*, defined as undesirable disturbances or fluctuations, is considered to be the "fundamental enemy" (McDonnell and Abbott, 2009) for error-free information transmission, processing, and communication. However, a vast and even increasing number of studies show the various benefits of noise in the context of signal detection and processing. Here, the most important phenomena are called stochastic resonance (McDonnell and Abbott, 2009), coherence resonance (Pikovsky and Kurths, 1997), and recurrence resonance (Krauss et al., 2019a).

The term stochastic resonance (SR), which has been introduced by Benzi in 1981 (Benzi et al., 1981), refers to the phenomenon that signals otherwise sub-threshold for a given sensor can be detected by adding a random signal, i.e. noise, of appropriate intensity to the sensor input (Gammaitoni et al., 1998; Moss et al., 2004). Fig. 1 illustrates this principle.

SR has been found ubiquitously in nature in a broad range of systems from physical to biological contexts (Hänggi, 2002; Wiesenfeld and Moss, 1995). In particular in neuroscience, SR has been demonstrated to play an essential role in virtually all kinds of systems (Faisal et al., 2008): from tactile (Collins et al., 1996; Douglass et al., 1993), auditory (Mino, 2014) and visual (Aihara et al., 2008) perception (Ward et al., 2002), through memory retrieval (Usher and Feingold, 2000) and cognition (Chandrasekharan et al., 2005), to behavioral control (Kitajo et al., 2003; Ward et al., 2002). SR explains how the brain processes information in noisy environments at each level of scale from single synapses (Stacey and Durand, 2001), through individual neurons (Kosko and Mitaim, 2003; Nozaki et al., 1999), to complete networks (Gluckman et al., 1996).

In self-adaptive signal detection systems exploiting SR, the optimum intensity of the noise is continuously adjusted so that information transmission is maximized, even if the characteristics and statistics of the input signal change (Fig. 2). For this processing principle, the term adaptive SR has been coined (Krauss et al., 2017; Mitaim and Kosko, 1998, 2004; Wenning and Obermayer, 2003).

2 Tinnitus development

In a number of recently published studies, we demonstrated theoretically and empirically that SR might be a major processing principle of the auditory system that serves to partially compensate for acute or chronic hearing loss (Krauss et al., 2016, 2017, 2018, 2019b; Gollnast et al., 2017). According to our model, the noise required for SR is generated within the brain and then perceived as a phantom sound. We have

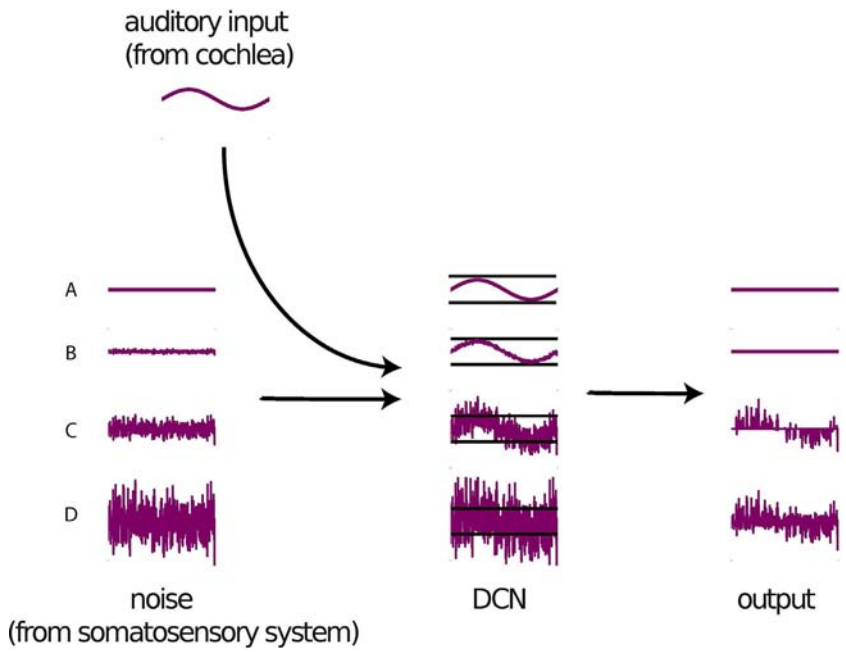

FIG. 1

Principle of stochastic resonance. The auditory input without any added noise is too weak to pass the threshold (A). Also if the intensity of added noise is too weak, the sum of auditory input and noise cannot pass the threshold (B). Both cases result in zero output. In contrast, if the optimal amount of noise is added to the signal before thresholding, the resulting output's envelope resembles the auditory input signal (C). However, if the noise intensity is further increased, the signal vanishes again in the noisy output (D).

proposed that it corresponds to increased spontaneous neuronal firing rates in early processing stages of the auditory brain stem - a phenomenon which is frequently observed in both humans with subjective tinnitus (Ahlf et al., 2012; Tziridis et al., 2015; Wang et al., 1997; Wu et al., 2016) and animal models, where the presence of tinnitus is tested using behavioral paradigms (Gerum et al., 2019, Schilling et al., 2017, Turner et al., 2006). Furthermore, tinnitus is assumed to be virtually always caused by some kind of either apparent (Heller 2003; König et al., 2006; Nelson & Chen 2004; Shore et al., 2016) or hidden hearing loss (Liberman & Liberman 2015; Schaette & McAlpine 2011). From this point of view, auditory phantom perceptions like tinnitus (or even the Zwicker tone, cf. below) seem to be a side effect of an adaptive mechanism within the auditory system whose primary purpose is to compensate for reduced input through continuous optimization of information transmission (Krauss et al., 2016, 2017, 2018, 2019b). This new interpretation may also explain why auditory sensitivity is increased in tinnitus ears (Gollnast et al., 2017; Hébert et al., 2013): the increased amount of neural noise during tinnitus improves auditory sensitivity by means of SR.

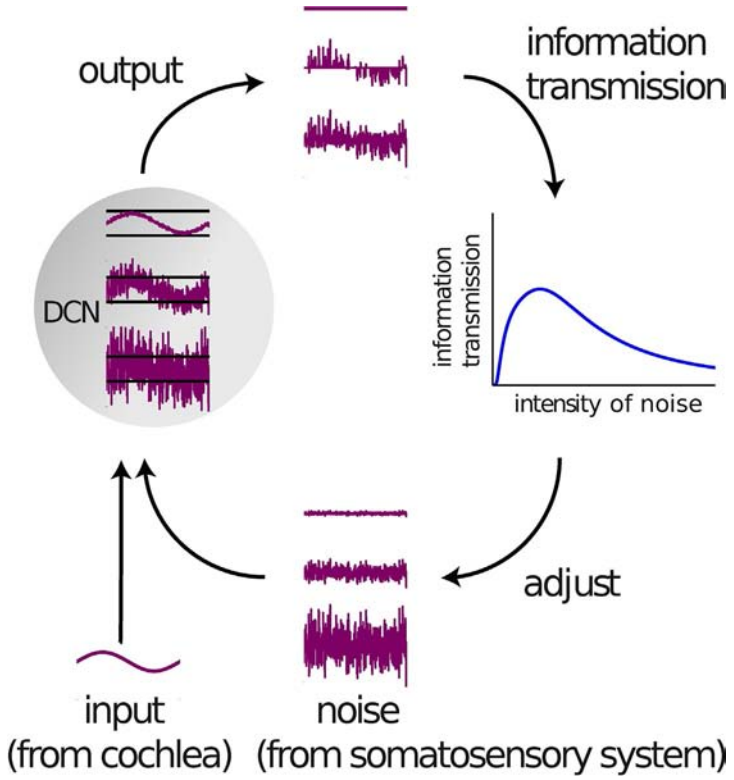

FIG. 2

Adaptive stochastic resonance control circuit in the DCN. In self-adaptive signal detection systems based on SR, the optimum noise level is continuously adjusted via a feedback loop, so that the system's response in terms of information throughput remains optimal, even if the properties of the input signal change. In the SR model of tinnitus development, this process takes place in the DCN. The input signal comes from the cochlea, the noise from the somatosensory system.

According to our model, the noise intensity is adjusted independently in each frequency channel. This is in line with several findings. The dorsal cochlear nucleus (DCN) has been shown to be the earliest processing stage where acoustic trauma, including complete cochlea ablation (Zacharek et al., 2002), causes increased spontaneous firing rates (Kaltenbach & Afman 2000; Kaltenbach et al., 1998; Wu et al., 2016; Zacharek et al., 2002). Interestingly, this increase in spontaneous activity, i.e. neural hyperactivity, is correlated with the strength of the behavioral signs of tinnitus in animal models (Kaltenbach et al., 2004). Furthermore, the hyperactivity is localized in those regions of the tonotopically organized DCN that are innervated by the damaged parts of the cochlea (Kaltenbach et al., 2002). Gao and colleagues (Gao et al., 2016) recently described changes in DCN fusiform cell spontaneous activity

after noise exposure that perfectly supports the proposed SR mechanism. In particular, the time course of spontaneous rate changes shows an almost complete loss of spontaneous activity immediately after loud sound exposure (as no SR is needed due to stimulation that is well above threshold), followed by an overcompensation of spontaneous rates to levels well above pre-exposition rates since SR is now used to compensate for acute hearing loss (Gao et al., 2016).

It is well known that the DCN receives not only auditory input from the cochlea, but also input from the somatosensory system (Ryugo et al., 2003; Shore & Zhou 2006; Wu et al., 2015), and that noise trauma alters long-term somatosensory-auditory processing in the DCN (Dehmel et al., 2008, 2012; Shore 2011; Wu et al., 2016), i.e. somatosensory projections are up-regulated after hearing loss (Zeng et al., 2012). In addition, DCN responses to somatosensory stimulation are enhanced after noise-induced hearing loss (Shore, 2011; Shore et al., 2008; Wu et al., 2016). Therefore, we previously proposed the possibility that the neural noise which is necessary for SR is injected into the auditory system via somatosensory projections to the DCN (Krauss et al., 2016, 2018, 2019b), and that these non-auditory projections into the DCN are the cause of the altered "spontaneous activity" within the DCN after hearing loss described previously (Gao et al., 2016). From an information processing point of view, somatosensory inputs are completely uncorrelated, i.e. have no mutual auditory information. Hence, these somatosensory inputs are perfectly suited to serve as a random signal, i.e. noise, in the context of SR, and this seems to be the reason why the auditory system does not generate the noise needed for SR itself.

Our idea that cross-modal SR, with cochlear inputs being the signal and somatosensory projections being the noise (Fig. 2), is a key processing principle of the auditory system and actually takes place in the DCN (Krauss et al., 2018) is supported by a large number of different findings. For instance, it is well known, that jaw movements lead to a modulation of subjective tinnitus loudness (Pinchoff et al., 1998). This may easily be explained within our framework, as jaw movements alter somatosensory input to the DCN: Since this somatosensory input corresponds to the noise for SR, auditory input to the DCN is modulated through this mechanism, and the altered noise level would then be perceived as modulated tinnitus (Krauss et al., 2016, 2018, 2019b). Along the same line, one may explain why both, the temporomandibular joint syndrome and whiplash, frequently cause so called somatic tinnitus (Levine, 1999; Shore et al., 2007).

Furthermore, the finding of Tang and Trussell that somatosensory input and hence tinnitus sensation may also be modified by serotonergic regulation of excitability of principal cells in the DCN (Tang & Trussell, 2015, 2017) supports the SR model. It even provides a mechanistic explanation of salicylate induced tinnitus, since salicylate affects DCN processing by disinhibition of somatosensory inputs (Koerber et al., 1966; Stolzberg et al., 2012). Thus, it increases the noise in the auditory system, which then may again be perceived as a phantom sound.

Finally, and maybe most remarkable, electro-tactile stimulation of finger tips, i.e. increased somatosensory input, significantly improves both, melody recognition (Huang et al., 2020) and speech recognition (Huang et al., 2017) in patients with

cochlear implants. Very recently, we were able to reproduce and mechanistically explain this finding, using a hybrid-computational model that exploits SR. The model consists of a cochlea model, a DCN model and an artificial deep neural network trained on a speech recognition task representing all further processing stages of the auditory pathway beyond the DCN. Simulated hearing loss, i.e. weakening the input from the cochlea model to the DCN model, reduced accuracy for speech recognition in the deep neural network, as expected. However, subsequent addition of noise, i.e. somatosensory input to the DCN model, results in an improved accuracy for speech recognition (Schilling et al., 2020).

3 Zwicker tone illusion

The Zwicker tone effect was discovered by Eberhard Zwicker in 1964 and is a temporal auditory phantom percept which was originally induced by the presentation of a 60 dB broadband noise with a spectral gap (notched noise) with a gap-width of half an octave (Zwicker, 1964). The Zwicker tone was described as "Negative Auditory After Image," although the underlying mechanisms generating an "After Image" are supposed to be different in the visual system. The Zwicker tone perception is not exclusively induced by a notched noise stimulus, but can also be caused by low-pass noise or white noise with a loud pure tone embedded (Fastl et al., 2001; Franosch et al., 2003).

Several models exist trying to explain the Zwicker tone percept. For example, Franosch and colleagues viewed the Zwicker tone as an asymmetric lateral inhibition effect along the auditory pathway (Franosch et al., 2003). In this view, the neurons in the DCN are disinhibited by surrounding neurons, which receive less stimulus driven activity due to the notch.

Another model suggested the Zwicker tone to be caused by a prediction error within the cortex in combination with an increased spontaneous rate of auditory pathway neurons at frequency ranges deprived by the notch within the presented broadband noise (Hullfish et al., 2019). However, these models have certain shortcomings such as they do not account for all properties of Zwicker tone percepts (described in the following) or do not describe the effect on a neuronal network level.

It has previously been proposed that the Zwicker tone and tinnitus and thus also the neural mechanisms of these two auditory phantom perceptions are closely connected (Hoke et al., 1996; Lummis and Guttmann, 1972; Mohan et al., 2020), and a number of findings support this assumption: For example, Parra and Pearlmutter were able to show that people with a tinnitus percept are also more likely to perceive a Zwicker tone percept (Parra & Pearlmutter, 2007). Additionally, Wiegrebe and coworkers showed that the presence of a Zwicker tone leads to decreased auditory thresholds of 13 dB even in normal hearing subjects (Norena et al. 1999; Wiegrebe et al., 1996), a finding which may easily be explained within our above described model of SR, since a similar effect can be observed in tinnitus patients (Gollnast et al., 2017; Krauss et al. 2016) who have improved hearing thresholds

in comparison to patients without tinnitus, at least within frequency ranges below 3 kHz. In this context, psychoacoustic experiments revealed that notched noise presentation leads to higher sensitivity to tones embedded in noise (Zhou et al., 2010).

Next, human studies using MEG showed that Zwicker tone perception correlates with a reduced alpha activity (Leske et al., 2014) in the auditory cortex. Interestingly, the effect of reduced alpha activity is also correlated to tinnitus perception (Weisz et al. 2007, 2011).

Furthermore, in most models tinnitus is supposed to be caused by hearing loss (Moffat et al., 2009) through e.g. cochlea damage or hidden hearing loss which cannot be detected by pure-tone audiograms but is characterized by a deafferentation of the inner hair cells (Liberman & Liberman, 2015; Paul et al., 2017). Analogously, the induction of the Zwicker tone through notched noise can be viewed as a deprivation of certain inner hair cells, that is, a temporary and reversible hearing loss (Hullfish et al., 2019).

These observations and resemblances support the view that the neural mechanisms of Zwicker tone and acute tinnitus are similar and that therefore the Zwicker tone may be a good model for tinnitus (Franosch et al. 2003; Hullfish et al., 2019; Krauss et al. 2018; Norena et al., 1999, 2000, 2002; Wrzosek et al., 2017). As a result, the investigation of the Zwicker tone has recently attracted further attention. Norena & Eggermont showed that Zwicker tone related neuronal activity changes can be observed on time scales in the range of seconds (Norena & Eggermont, 2003). In particular, cats were implanted with multi-electrode arrays and notched noise stimuli of 1 s duration were presented. It could be shown that neurons in the auditory cortex representing frequencies within the range of the notch show increased firing rates after notched noise presentation (Norena & Eggermont, 2003). This result indicates that the Zwicker tone is correlated with a hyperactivity of neurons along the complete auditory pathway that represent the frequency notch, although to our knowledge systematic studies of activity along the auditory pathway in animals during Zwicker tone induction are missing.

Despite all these similarities between the Zwicker tone and acute tinnitus, there are only few mechanistic explanation approaches on a neural network level (Okamoto et al., 2005). Our stochastic resonance model (Krauss et al., 2016, see above) provides such a mechanistic explanation of Zwicker tone percepts. As stated above the presentation of a notched noise stimulus can be viewed as temporary hearing loss or deprivation of inner hair cells located within the frequency notch within the tonotopic gradient (Hullfish et al., 2019; Krauss et al., 2018). According to our model, this reduced input would cause SR within the auditory system to restore hearing by optimizing information transmission at the level of the DCN via increased neuronal noise (as described above). This increase of the neural noise would take place within the frequency channels of the spectral notch, leading to a hyperactivity of the respective neurons in the DCN (Krauss et al., 2016). This hyperactivity is transmitted along the auditory pathway and causes a Zwicker tone percept at the cortical level.

Our explanation is supported by the observation that notched noise stimulation leads to hyperactivity of auditory cortex neurons representing the notch frequency

(cf. Norena & Eggermont, 2003) via disinhibition (cf. Weisz et al., 2007, 2011). Furthermore, only the SR mechanism may explain improved hearing thresholds for frequencies near the Zwicker tone frequency during Zwicker tone perception (cf. Norena et al. 1999; Wiegrebe et al., 1996): internal noise from the somatosensory system is increased in the deprived frequency ranges (notch frequency range) in order to compensate for reduced auditory input by means of SR. This, in turn, leads as a side effect to improved hearing thresholds for neighboring frequencies above and below the notch. Additionally, the SR feedback control circuit (Fig. 2) operates on time scales in the range of or below a second and thus fits to the observation of Zwicker tone related hyperactivity after 1 s of notched noise presentation (Norena & Eggermont, 2003).

According to our model, the increased neural noise to the DCN which is necessary for SR is supposed to originate from the somatosensory system (Krauss et al., 2016, 2018, 2019b). In analogy to the afore mentioned phenomenon of tinnitus modulation by voluntary jaw movements, our model also predicts a modulation of the Zwicker tone perception by somatosensory stimulation. It has indeed been reported that transcutaneous electrical stimulation has an effect on Zwicker tone perception (Ueberfuhr et al., 2017).

4 Residual inhibition

In 1971 Feldmann found that the presentation of acoustic noise leads to a suppression of the tinnitus precept after noise offset (Feldmann, 1971), for approximately 1 min (Roberts, 2007; Roberts et al., 2006). This effect was named Residual Inhibition (RI; Henry & Meikle, 2000; Vernon, 1977).

RI should not be mixed up with tinnitus masking, where tinnitus is perceived less intense as it is masked by a noise of similar frequency range (Hazell & Wood, 1981; Terry et al., 1983). In contrast, the presentation of masking noise causes RI *after* the end of noise presentation. As RI is a technique to temporarily modulate the tinnitus percept, it is a potential target for experimental studies on tinnitus mechanisms (Deklerck et al., 2019).

Interestingly, it was reported that RI works best when the masking noise covers the range of the hearing loss of the subjects and is related to the tinnitus pitch (Roberts et al., 2006, 2008). The cause of the suppression of the tinnitus percept during RI has been discussed to be a decreased spontaneous neural activity after masking noise offset (Galazyuk et al., 2017). This is in line with the explanation that there is a neural adaptation along the auditory pathway induced by the noise presentation (Fournier et al., 2018).

These findings emphasize the idea that spontaneous activity of spiking neurons or in other words internally generated neural noise are crucial for processing of acoustic stimuli along the auditory pathway (Galazyuk et al., 2019). This internal noise is suppressed after the presentation of external acoustic noise. To understand the basic neural mechanisms of RI as well as auditory phantom perception, it is crucial to gain a better understanding of how the neural noise contributes to auditory processing.

The idea that the neural system exploits the effect of SR to improve hearing (Krauss et al., 2016, 2018, 2019b) provides a putative explanation for the effect of RI. As described above, tinnitus is potentially induced by the deprivation of neurons along the auditory pathway in tonotopic regions where a cochlea damage occurred. Thus, the auditory system tries to compensate for this deprivation, i.e. hearing loss, by adding internally generated neural noise. This internally generated noise potentially produced by the somatosensory system and fed to the DCN is propagated along the auditory pathway to the cortex, where it is perceived as auditory phantom percept. RI is potentially the consequence of replacing internally generated neural noise by external acoustic noise. In this view, the external noise would replace the internal noise, thereby causing its downregulation and thus suppression of the tinnitus percept as already described in previous publications (Krauss et al., 2016, 2019b).

According to our model, the optimal noise is tuned and controlled on time scales of seconds via a control circuit (Krauss et al., 2016; Fig. 2). From this point of view, Zwicker tone and tinnitus are basically the same phenomenon, but on different time scales. Furthermore, the proposed control circuit would work inversely for Zwicker tone and RI. Whereas, the Zwicker tone corresponds to an upregulation of internal neural noise caused by a reduced auditory input (i.e. the notch), RI in contrast corresponds to a downregulation of internal noise, due to increased auditory input (i.e. external acoustic noise). Thus, both phenomena can be considered to be opposite effects that may be explained by exactly the same neural control circuitry proposed by our SR model. To put it in a slogan, the SR model of auditory processing suggests that "RI can be interpreted as an inverse Zwicker tone illusion."

5 Summary and discussion

In summary, our SR model provides a unified explanation for the induction of acute subjective tinnitus, Zwicker tone, and RI. The total duration these phenomena are perceived differs greatly, e.g. the Zwicker tone lasts a few seconds, residual inhibition a few minutes, and tinnitus might even last decades. However, the time scales on which these apparently different perceptions can be induced (e.g. the Zwicker tone, Norena & Eggermont, 2003), or reduced (e.g. tinnitus removal by hearing aids or cochlear implants, McNeill et al., 2012; Ito & Sakakihara, 1994; Baguley & Atlas, 2007), are within a narrow range of some seconds. This indicates that these phenomena cannot be exclusively explained by brain plasticity, which takes place on much longer time scales. The SR model, describing tinnitus as a side effect of the neural system trying to optimize information transmission after hearing loss by exploiting the SR effect, would offer an explanation of how these phantom perceptions can be induced or suppressed so quickly. Thus, the neural system does not need any plasticity in the first place as the SR mechanism is optimized by a simple control circuit (Krauss et al., 2016; Fig. 2).

One may argue that the advantage of the sensory system might be close to zero in individuals suffering from extreme hearing loss or deafness and ask why the injected

somatosensory noise apparently stays at a level that bears no benefit but rather evokes a percept that induces stress for the individual but comes without meaningful information. We argue that the knowledge, that this perception is actually a phantom perception without any physical source in the environment, is only available at the highest processing stages in the brain associated with conscious perception. In contrast, the early processing stages within the auditory pathway, i.e. the DCN, have no access to this knowledge, hence from the point of view of the DCN, a "pure tone" always contains the same amount of information whether its source is actually in the environment or not. Our proposed feedback-loop for the adjustment of noise intensity to maintain optimal information processing is comparable to a reflex arc in the motor system, but without any top-down regulation. Hence, the noise amplification is not readapted, since this would require both, knowledge about the phantom perception, which is only available to higher processing stages, and top-down connections from these higher processing stages to the DCN. Furthermore, results from another study of our group suggest that the information benefit (in this case, accuracy improvement in a speech recognition task) as a function of noise intensity may show, under certain conditions, a second maximum besides the global maximum (Schilling et al., 2020). Therefore, it seems possible that the noise adjusting feedback loop of the auditory system gets "trapped" in this side maximum.

We speculate that in subjects, where the Zwicker tone can be induced by short noise presentation the RI effect should vanish more quickly, because the tuning of the optimal noise level works faster in certain subjects and thus the downregulated neural noise during RI is quickly re-increased. On the other hand, the Zwicker tone is induced faster as the neural noise is quickly upregulated when notched noise is presented. Thus, the duration of notched noise needed to induce the Zwicker tone could potentially correlate with the duration of the RI effect. This would be only the case, if both effects were produced by the same SR control circuit in the DCN (Fig. 2), which could be a characteristic feature of different individuals. The characteristic parameter of this control circuit is the time needed for controlling the noise amplitude.

This is a testable hypothesis derived from the SR model, which has to be verified or falsified in future studies.

However, it is obvious that the SR model has some limitations, such as that -in contrast to homeostatic plasticity models- it does not predict massive structural and functional changes (cf. Noreña, 2011) along the auditory pathway, which is indeed found in several studies (Li et al., 2015; Singer et al., 2013; Yang et al., 2011). These findings are supported by computational models demonstrating the influence of this plasticity (Nagashino et al., 2012; Schaette & Kempter, 2006).

Additionally, our model does not address the question why not all people with hearing loss perceive or even suffer from tinnitus. The influence of stress (Mazurek et al., 2012, 2015) and psychological burden (Landgrebe & Langguth 2011; Langguth et al., 2007, 2011) on tinnitus percepts was shown in several studies. Furthermore, the model does not differentiate between chronic and acute tinnitus.

Despite these limitations, we are convinced that we now have the knowledge to draw a complete picture in the light of preceding studies. Figs. 3 and 4 provide an

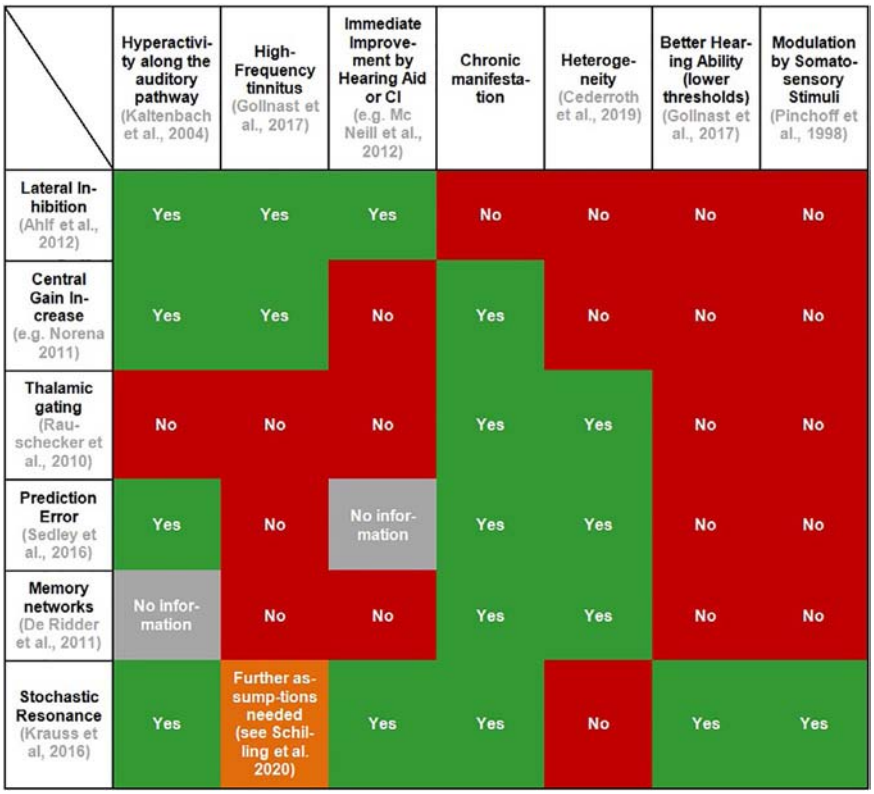

FIG. 3

Explanatory power of different models of tinnitus development. The figure summarizes different models of tinnitus development (rows) and how these models fit to certain observations (columns). For each model and effect, one exemplary paper is cited (e.g. Cederroth et al., 2019).

overview of the main models and their explanatory power for tinnitus development and Zwicker tone perception. The different models work on different time scales, as well as in different brain areas, as illustrated in Fig. 5.

Our SR model provides a mechanistic explanation of the initial cause ("the first seconds") leading to the induction of tinnitus after e.g. a loud acoustic noise presentation, the induction of the Zwicker tone illusion by notched noise, or the suppression of the tinnitus perception by acoustic noise presentation (i.e., residual inhibition). As mentioned above, these phenomena occur within seconds, and thus cannot be explained by any of the models based on brain plasticity. However, as described above, neural plasticity occurs along the auditory pathway (Li et al., 2015; Singer et al., 2013; Yang et al., 2011), and very probably contributes to chronic manifestation of tinnitus, yet after and on top of the initial induction caused by SR.

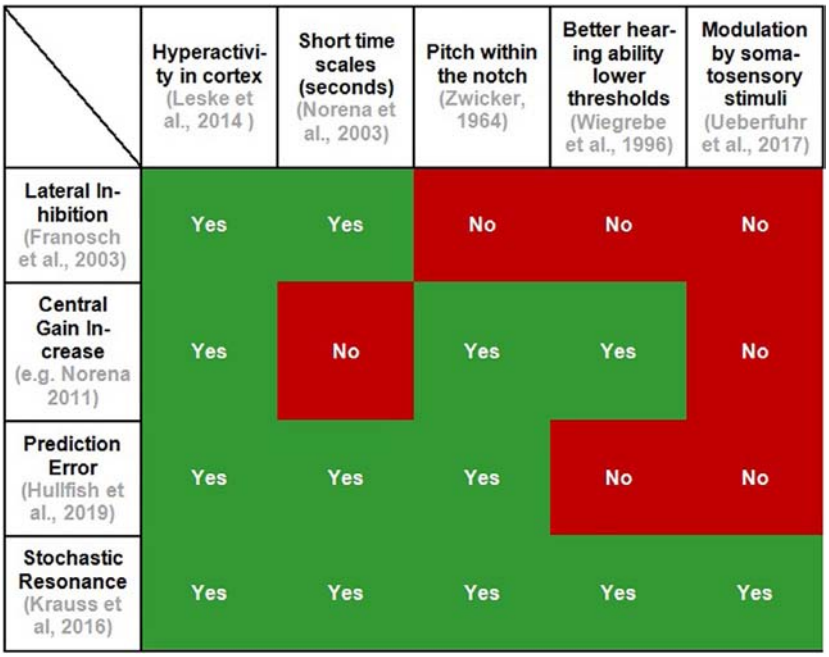

FIG. 4

Explanatory power of different models of the Zwicker tone illusion. The figure summarizes different models of the Zwicker tone illusion (rows) and how these models fit to certain observations (columns). For each model and effect, one exemplary paper is cited.

Furthermore, it is still unclear why the gating function of the thalamus does not prevent the neural hyperactivity from being directly transmitted to the cortex as it does for other unwanted permanent stimuli (McCormick & Bal, 1994). This effect could be explained by the model of Rauschecker and coworkers (Rauschecker et al., 2010). There, the auditory input can be canceled out by the medial geniculate nucleus within the thalamus. This noise cancellation function can be modulated by the limbic system especially the nucleus accumbens, which is indirectly connected to the medial geniculate nucleus. A breakdown of this system impairs the gating function of the medial geniculate nucleus (Rauschecker et al., 2010) and thus brings the neural hyperactivity to consciousness.

De Ridder and coworkers go even one step further and assume a conscious tinnitus percept to be a consequence of different overlapping brain networks including prefrontal areas as well as brain structures responsible for emotional labeling of certain memories such as the amygdala. Thus, learning effects are involved, which generate a connection of the phantom percept and distress (De Ridder et al., 2011). Unfortunately, this model does not provide mechanistic explanations at a neural network level, but it explains the involvement of different brain structures. Nevertheless, the model could provide an explanation why not every hearing loss causes tinnitus, and why not

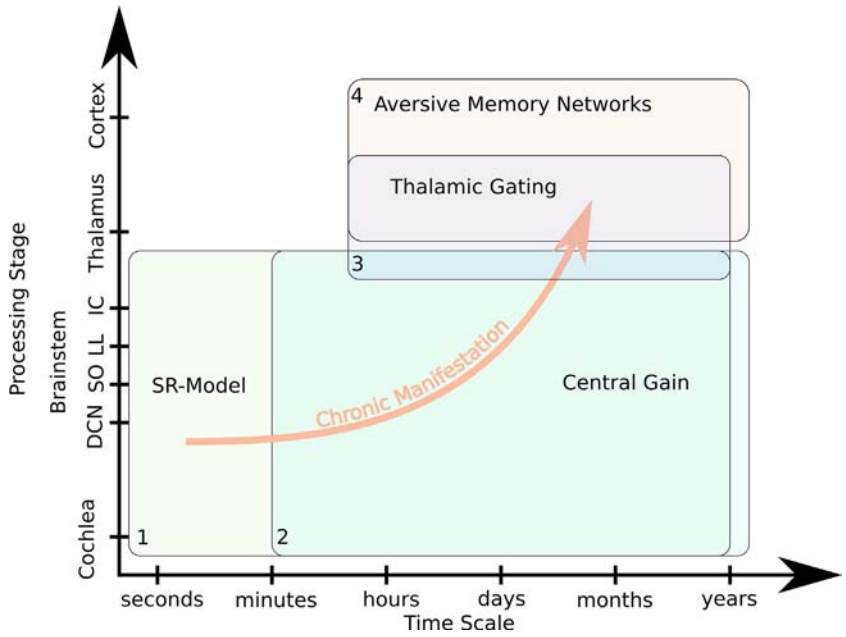

FIG. 5

The space of tinnitus models. Models of tinnitus development can be defined at different levels of description and can vary in time scale of the explained observations (horizontal axis) and in proposed anatomical substrate, i.e. processing stage (vertical axis). The SR model fills the "missing gap" in time scales of minutes and seconds.

everyone perceiving tinnitus also suffers from it. Individual memories and neuronal pathways could lead to different effects in different subjects.

Rather than mutually excluding each other as claimed by Sedley and coworkers (Sedley et al., 2016), the described models complement each other and draw a complete and consistent image of tinnitus development, its chronic manifestation, and heterogeneity. Furthermore, mechanistic explanations for RI, Zwicker tone, and better hearing thresholds of tinnitus patients compared to patients without tinnitus (Gollnast et al., 2017; Krauss et al., 2016) support the model.

Acknowledgments

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation): grant KR5148/2-1 to PK—project number 436456810, and grant SCHU1272/12-1 to HS—project number 332767752, and the Emergent Talents Initiative (ETI) of the University Erlangen-Nuremberg (grant 2019/2-Phil-01 to PK), and the Interdisciplinary Center for Clinical Research (IZKF) at the University Hospital of the University Erlangen-Nuremberg (grant ELAN-17-12-27-1-Schilling to AS).

We thank the reviewers for their valuable comments.

References

Ahlf, S., Tziridis, K., Korn, S., Strohmeyer, I., Schulze, H., 2012. Predisposition for and prevention of subjective tinnitus development. PLoS One 7 (10), e44519.

Aihara, T., Kitajo, K., Nozaki, D., Yamamoto, Y., 2008. Internal noise determines external stochastic resonance in visual perception. Vis. Res. 48 (14), 1569–1573.

Baguley, D.M., Atlas, M.D., 2007. Cochlear implants and tinnitus. Prog. Brain Res. 166, 347–355.

Benzi, R., Sutera, A., Vulpiani, A., 1981. The mechanism of stochastic resonance. J. Phys. A Math. Gen. 14 (11), L453.

Cederroth, C.R., Gallus, S., Hall, D.A., Kleinjung, T., Langguth, B., Maruotti, A., … Searchfield, G., 2019. Towards an understanding of tinnitus heterogeneity. Front. Aging Neurosci. 11, 53.

Chandrasekharan, S., Lebiere, C., Stewart, T.C., West, R.L., 2005. Stochastic resonance in human cognition: ACT-R versus game theory, associative neural networks, recursive neural networks, q-learning, and humans. In: Proceedings of the Annual Meeting of the Cognitive Science Society (Vol. 27, No. 27).

Collins, J.J., Imhoff, T.T., Grigg, P., 1996. Noise-enhanced tactile sensation. Nature 383, 770.

De Ridder, D., Elgoyhen, A.B., Romo, R., Langguth, B., 2011. Phantom percepts: tinnitus and pain as persisting aversive memory networks. Proc. Natl. Acad. Sci. 108 (20), 8075–8080.

Dehmel, S., Cui, Y.L., Shore, S.E., 2008. Cross-modal interactions of auditory and somatic inputs in the brainstem and midbrain and their imbalance in tinnitus and deafness. Am. J. Audiol. 17, S193–S209.

Dehmel, S., Pradhan, S., Koehler, S., Bledsoe, S., Shore, S., 2012. Noise overexposure alters long-term somatosensory-auditory processing in the dorsal cochlear nucleus—possible basis for tinnitus-related hyperactivity? J. Neurosci. 32 (5), 1660–1671.

Deklerck, A.N., Degeest, S., Dhooge, I.J., Keppler, H., 2019. Test–retest reproducibility of response duration in tinnitus patients with positive residual inhibition. J. Speech Lang. Hear. Res. 62 (9), 3531–3544.

Douglass, J.K., Wilkens, L., Pantazelou, E., Moss, F., 1993. Noise enhancement of information transfer in crayfish mechanoreceptors by stochastic resonance. Nature 365 (6444), 337–340.

Faisal, A.A., Selen, L.P., Wolpert, D.M., 2008. Noise in the nervous system. Nat. Rev. Neurosci. 9 (4), 292–303.

Fastl, H., Patsouras, D., Franosch, M., van Hemmen, L., 2001. Zwicker-tones for pure tone plus bandlimited noise. In: Proceedings of the 12th International Symposium on Hearing, Physiological and Psychophysical Bases of Auditory, pp. 67–74.

Feldmann, H., 1971. Homolateral and contralateral masking of tinnitus by noise-bands and by pure tones. Audiology 10 (3), 138–144.

Fournier, P., Cuvillier, A.F., Gallego, S., Paolino, F., Paolino, M., Quemar, A., … Norena, A., 2018. A new method for assessing masking and residual inhibition of tinnitus. Trends Hear. 22, 2331216518769996.

Franosch, J.M.P., Kempter, R., Fastl, H., van Hemmen, J.L., 2003. Zwicker tone illusion and noise reduction in the auditory system. Phys. Rev. Lett. 90 (17), 178103.

Galazyuk, A.V., Voytenko, S.V., Longenecker, R.J., 2017. Long-lasting forward suppression of spontaneous firing in auditory neurons: implication to the residual inhibition of tinnitus. J. Assoc. Res. Otolaryngol. 18 (2), 343–353.

Galazyuk, A.V., Longenecker, R.J., Voytenko, S.V., Kristaponyte, I., Nelson, G.L., 2019. Residual inhibition: from the putative mechanisms to potential tinnitus treatment. Hear. Res. 375, 1–13.

Gammaitoni, L., Hänggi, P., Jung, P., Marchesoni, F., 1998. Stochastic resonance. Rev. Mod. Phys. 70 (1), 223.
Gao, Y., Manzoor, N., Kaltenbach, J.A., 2016. Evidence of activity-dependent plasticity in the dorsal cochlear nucleus, in vivo, induced by brief sound exposure. Hear. Res. 341, 31–42.
Gerum, R.C., Rahlfs, H., Streb, M., Krauss, P., Grimm, J., Metzner, C., … Schilling, A., 2019. Open (G) PIAS: an open-source solution for the construction of a high-precision acoustic startle response setup for tinnitus screening and threshold estimation in rodents. Front. Behav. Neurosci. 13, 140.
Gluckman, B.J., Netoff, T.I., Neel, E.J., Ditto, W.L., Spano, M.L., Schiff, S.J., 1996. Stochastic resonance in a neuronal network from mammalian brain. Phys. Rev. Lett. 77 (19), 4098.
Gollnast, D., Tziridis, K., Krauss, P., Schilling, A., Hoppe, U., Schulze, H., 2017. Analysis of audiometric differences of patients with and without tinnitus in a large clinical database. Front. Neurol. 8, 31.
Hänggi, P., 2002. Stochastic resonance in biology how noise can enhance detection of weak signals and help improve biological information processing. ChemPhysChem 3 (3), 285–290.
Hazell, J.W.P., Wood, S.M., 1981. Tinnitus masking-a significant contribution to tinnitus management. Br. J. Audiol. 15 (4), 223–230.
Hébert, S., Fournier, P., Noreña, A., 2013. The auditory sensitivity is increased in tinnitus ears. J. Neurosci. 33 (6), 2356–2364.
Heller, A.J., 2003. Classification and epidemiology of tinnitus. Otolaryngol. Clin. N. Am. 36 (2), 239–248.
Henry, J.A., Meikle, M.B., 2000. Psychoacoustic measures of tinnitus. J. Am. Acad. Audiol. 11 (3), 138–155.
Hoke, E.S., Hoke, M., Ross, B., 1996. Neurophysiological correlate of the auditory after-image ('ZwickerTone'). Audiol. Neurootol. 1 (3), 161–174.
Huang, J., Sheffield, B., Lin, P., Zeng, F.G., 2017. Electro-tactile stimulation enhances cochlear implant speech recognition in noise. Sci. Rep. 7 (1), 1–5.
Huang, J., Lu, T., Sheffield, B., Zeng, F.G., 2020. Electro-tactile stimulation enhances cochlear-implant melody recognition: effects of rhythm and musical training. Ear Hear. 41 (1), 106–113.
Hullfish, J., Sedley, W., Vanneste, S., 2019. Prediction and perception: insights for (and from) tinnitus. Neurosci. Biobehav. Rev. 102, 1–12.
Ito, J., Sakakihara, J., 1994. Suppression of tinnitus by cochlear implantation. Am. J. Otolaryngol. 15 (2), 145–148.
Kaltenbach, J.A., Afman, C.E., 2000. Hyperactivity in the dorsal cochlear nucleus after intense sound exposure and its resemblance to tone-evoked activity: a physiological model for tinnitus. Hear. Res. 140 (1–2), 165–172.
Kaltenbach, J.A., Godfrey, D.A., Neumann, J.B., McCaslin, D.L., Afman, C.E., Zhang, J., 1998. Changes in spontaneous neural activity in the dorsal cochlear nucleus following exposure to intense sound: relation to threshold shift. Hear. Res. 124 (1–2), 78–84.
Kaltenbach, J.A., Rachel, J.D., Mathog, T.A., Zhang, J., Falzarano, P.R., Lewandowski, M., 2002. Cisplatin-induced hyperactivity in the dorsal cochlear nucleus and its relation to outer hair cell loss: relevance to tinnitus. J. Neurophysiol. 88 (2), 699–714.
Kaltenbach, J.A., Zacharek, M.A., Zhang, J., Frederick, S., 2004. Activity in the dorsal cochlear nucleus of hamsters previously tested for tinnitus following intense tone exposure. Neurosci. Lett. 355 (1–2), 121–125.
Kitajo, K., Nozaki, D., Ward, L.M., Yamamoto, Y., 2003. Behavioral stochastic resonance within the human brain. Phys. Rev. Lett. 90 (21), 218103.

Koerber, K.C., Pfeiffer, R.R., Warr, W.B., Kiang, N.Y.S., 1966. Spontaneous spike discharges from single units in the cochlear nucleus after destruction of the cochlea. Exp. Neurol. 16 (2), 119–130.

König, O., Schaette, R., Kempter, R., Gross, M., 2006. Course of hearing loss and occurrence of tinnitus. Hear. Res. 221 (1–2), 59–64.

Kosko, B., Mitaim, S., 2003. Stochastic resonance in noisy threshold neurons. Neural Netw. 16 (5–6), 755–761.

Krauss, P., Tziridis, K., Metzner, C., Schilling, A., Hoppe, U., Schulze, H., 2016. Stochastic resonance controlled upregulation of internal noise after hearing loss as a putative cause of tinnitus-related neuronal hyperactivity. Front. Neurosci. 10, 597.

Krauss, P., Metzner, C., Schilling, A., Schütz, C., Tziridis, K., Fabry, B., Schulze, H., 2017. Adaptive stochastic resonance for unknown and variable input signals. Sci. Rep. 7 (1), 1–8.

Krauss, P., Tziridis, K., Schilling, A., Schulze, H., 2018. Cross-modal stochastic resonance as a universal principle to enhance sensory processing. Front. Neurosci. 12, 578.

Krauss, P., Schilling, A., Tziridis, K., Schulze, H., 2019a. Models of tinnitus development: from cochlea to cortex. HNO 67, 172–177.

Krauss, P., Prebeck, K., Schilling, A., Metzner, C., 2019b. "Recurrence resonance" in three-neuron motifs. Front. Comput. Neurosci. 13, 64.

Landgrebe, M., Langguth, B., 2011. Tinnitus and psychiatric co-morbidity. In: Møller, A.R., Langguth, B., DeRidder, D., Kleinjung, T. (Eds.), Textbook of Tinnitus. Springer, New York, NY, pp. 491–492.

Langguth, B., Kleinjung, T., Fischer, B., Hajak, G., Eichhammer, P.S.P.G., Sand, P.G., 2007. Tinnitus severity, depression, and the big five personality traits. Prog. Brain Res. 166, 221–225.

Langguth, B., Landgrebe, M., Kleinjung, T., Sand, G.P., Hajak, G., 2011. Tinnitus and depression. World J. Biol. Psychiatry 12 (7), 489–500.

Leske, S., Tse, A., Oosterhof, N.N., Hartmann, T., Müller, N., Keil, J., Weisz, N., 2014. The strength of alpha and beta oscillations parametrically scale with the strength of an illusory auditory percept. NeuroImage 88, 69–78.

Levine, R.A., 1999. Somatic (craniocervical) tinnitus and the dorsal cochlear nucleus hypothesis. Am. J. Otolaryngol. 20 (6), 351–362.

Li, S., Kalappa, B.I., Tzounopoulos, T., 2015. Noise-induced plasticity of KCNQ2/3 and HCN channels underlies vulnerability and resilience to tinnitus. elife 4, e07242.

Liberman, L.D., Liberman, M.C., 2015. Dynamics of cochlear synaptopathy after acoustic overexposure. J. Assoc. Res. Otolaryngol. 16 (2), 205–219.

Lummis, R.C., Guttman, N., 1972. Exploratory studies of Zwicker's "negative afterimage" in hearing. J. Acoust. Soc. Am. 51 (6B), 1930–1944.

Mazurek, B., Haupt, H., Olze, H., Szczepek, A.J., 2012. Stress and tinnitus—from bedside to bench and back. Front. Syst. Neurosci. 6, 47.

Mazurek, B., Szczepek, A.J., Hebert, S., 2015. Stress and tinnitus. HNO 63 (4), 258–265.

McCormick, D.A., Bal, T., 1994. Sensory gating mechanisms of the thalamus. Curr. Opin. Neurobiol. 4 (4), 550–556.

McDonnell, M.D., Abbott, D., 2009. What is stochastic resonance? Definitions, misconceptions, debates, and its relevance to biology. PLoS Comput. Biol. 5 (5), e1000348.

McNeill, C., Távora-Vieira, D., Alnafjan, F., Searchfield, G.D., Welch, D., 2012. Tinnitus pitch, masking, and the effectiveness of hearing aids for tinnitus therapy. Int. J. Audiol. 51 (12), 914–919.

Mino, H., 2014. The effects of spontaneous random activity on information transmission in an auditory brain stem neuron model. Entropy 16 (12), 6654–6666.

Mitaim, S., Kosko, B., 1998. Adaptive stochastic resonance. Proc. IEEE 86 (11), 2152–2183.
Mitaim, S., Kosko, B., 2004. Adaptive stochastic resonance in noisy neurons based on mutual information. IEEE Trans. Neural Netw. 15 (6), 1526–1540.
Moffat, G., Adjout, K., Gallego, S., Thai-Van, H., Collet, L., Norena, A.J., 2009. Effects of hearing aid fitting on the perceptual characteristics of tinnitus. Hear. Res. 254 (1–2), 82–91.
Mohan, A., Bhamoo, N., Riquelme, J.S., Long, S., Norena, A., Vanneste, S., 2020. Investigating functional changes in the brain to intermittently induced auditory illusions and its relevance to chronic tinnitus. Hum. Brain Mapp. 41, 1819–1832.
Moss, F., Ward, L.M., Sannita, W.G., 2004. Stochastic resonance and sensory information processing: a tutorial and review of application. Clin. Neurophysiol. 115 (2), 267–281.
Nagashino, H., Kinouchi, Y., Danesh, A.A., Pandya, A.S., 2012. A neuronal network model with homeostatic plasticity for tinnitus generation and its management by sound therapy. In: 2012 IEEE-EMBS Conference on Biomedical Engineering and Sciences. IEEE, pp. 706–711.
Nelson, J.J., Chen, K., 2004. The relationship of tinnitus, hyperacusis, and hearing loss. Ear Nose Throat J. 83 (7), 472–476.
Noreña, A.J., 2011. An integrative model of tinnitus based on a central gain controlling neural sensitivity. Neurosci. Biobehav. Rev. 35 (5), 1089–1109.
Norena, A.J., Eggermont, J.J., 2003. Neural correlates of an auditory afterimage in primary auditory cortex. J. Assoc. Res. Otolaryngol. 4 (3), 312–328.
Norena, A., Micheyl, C., Chéry-Croze, S., 1999. The Zwicker tone (ZT) as a model of phantom auditory perception. In: Sixth International Tinnitus Seminar, 429.
Norena, A., Micheyl, C., Chery-Croze, S., 2000. An auditory negative after-image as a human model of tinnitus. Hear. Res. 149 (1–2), 24–32.
Norena, A., Micheyl, C., Garnier, S., Chery-croze, S., 2002. Loudness changes associated with the perception of an auditory after-image: cambios en la intensidad asociados a la percepción de una imagen post-auditiva. Int. J. Audiol. 41 (3), 202–207.
Nozaki, D., Mar, D.J., Grigg, P., Collins, J.J., 1999. Effects of colored noise on stochastic resonance in sensory neurons. Phys. Rev. Lett. 82 (11), 2402.
Okamoto, H., Kakigi, R., Gunji, A., Kubo, T., Pantev, C., 2005. The dependence of the auditory evoked N1m decrement on the bandwidth of preceding notch-filtered noise. Eur. J. Neurosci. 21 (7), 1957–1961.
Parra, L.C., Pearlmutter, B.A., 2007. Illusory percepts from auditory adaptation. J. Acoust. Soc. Am. 121 (3), 1632–1641.
Paul, B.T., Bruce, I.C., Roberts, L.E., 2017. Evidence that hidden hearing loss underlies amplitude modulation encoding deficits in individuals with and without tinnitus. Hear. Res. 344, 170–182.
Pikovsky, A.S., Kurths, J., 1997. Coherence resonance in a noise-driven excitable system. Phys. Rev. Lett. 78 (5), 775.
Pinchoff, R.J., Burkard, R.F., Salvi, R.J., Coad, M.L., Lockwood, A.H., 1998. Modulation of tinnitus by voluntary jaw movements. Am. J. Otol. 19 (6), 785–789.
Rauschecker, J.P., Leaver, A.M., Mühlau, M., 2010. Tuning out the noise: limbic-auditory interactions in tinnitus. Neuron 66 (6), 819–826.
Roberts, L.E., 2007. Residual inhibition. Prog. Brain Res. 166, 487–495.
Roberts, L.E., Moffat, G., Bosnyak, D.J., 2006. Residual inhibition functions in relation to tinnitus spectra and auditory threshold shift. Acta Oto-Laryngol. 126 (sup556), 27–33.
Roberts, L.E., Moffat, G., Baumann, M., Ward, L.M., Bosnyak, D.J., 2008. Residual inhibition functions overlap tinnitus spectra and the region of auditory threshold shift. J. Assoc. Res. Otolaryngol. 9 (4), 417–435.

Ryugo, D.K., Haenggeli, C.A., Doucet, J.R., 2003. Multimodal inputs to the granule cell domain of the cochlear nucleus. Exp. Brain Res. 153 (4), 477–485.

Schaette, R., Kempter, R., 2006. Development of tinnitus-related neuronal hyperactivity through homeostatic plasticity after hearing loss: a computational model. Eur. J. Neurosci. 23 (11), 3124–3138.

Schaette, R., McAlpine, D., 2011. Tinnitus with a normal audiogram: physiological evidence for hidden hearing loss and computational model. J. Neurosci. 31 (38), 13452–13457.

Schilling, A., Krauss, P., Gerum, R., Metzner, C., Tziridis, K., Schulze, H., 2017. A new statistical approach for the evaluation of gap-prepulse inhibition of the acoustic startle reflex (GPIAS) for tinnitus assessment. Front. Behav. Neurosci. 11, 198.

Schilling, A., Gerum, R., Zankl, A., Schulze, H., Metzner, C., Krauss, P., 2020. Intrinsic noise improves speech recognition in a computational model of the auditory pathway. bioRxiv, 2020.03.16.993725. https://doi.org/10.1101/2020.03.16.993725.

Sedley, W., Friston, K.J., Gander, P.E., Kumar, S., Griffiths, T.D., 2016. An integrative tinnitus model based on sensory precision. Trends Neurosci. 39 (12), 799–812.

Shore, S.E., 2011. Plasticity of somatosensory inputs to the cochlear nucleus–implications for tinnitus. Hear. Res. 281 (1–2), 38–46.

Shore, S.E., Zhou, J., 2006. Somatosensory influence on the cochlear nucleus and beyond. Hear. Res. 216, 90–99.

Shore, S., Zhou, J., Koehler, S., 2007. Neural mechanisms underlying somatic tinnitus. Prog. Brain Res. 166, 107–548.

Shore, S.E., Koehler, S., Oldakowski, M., Hughes, L.F., Syed, S., 2008. Dorsal cochlear nucleus responses to somatosensory stimulation are enhanced after noise-induced hearing loss. Eur. J. Neurosci. 27 (1), 155–168.

Shore, S.E., Roberts, L.E., Langguth, B., 2016. Maladaptive plasticity in tinnitus—triggers, mechanisms and treatment. Nat. Rev. Neurol. 12 (3), 150.

Singer, W., Zuccotti, A., Jaumann, M., Lee, S.C., Panford-Walsh, R., Xiong, H., … Rohbock, K., 2013. Noise-induced inner hair cell ribbon loss disturbs central arc mobilization: a novel molecular paradigm for understanding tinnitus. Mol. Neurobiol. 47 (1), 261–279.

Stacey, W.C., Durand, D.M., 2001. Synaptic noise improves detection of subthreshold signals in hippocampal CA1 neurons. J. Neurophysiol. 86 (3), 1104–1112.

Stolzberg, D., Salvi, R.J., Allman, B.L., 2012. Salicylate toxicity model of tinnitus. Front. Syst. Neurosci. 6, 28.

Tang, Z.Q., Trussell, L.O., 2015. Serotonergic regulation of excitability of principal cells of the dorsal cochlear nucleus. J. Neurosci. 35 (11), 4540–4551.

Tang, Z.Q., Trussell, L.O., 2017. Serotonergic modulation of sensory representation in a central multisensory circuit is pathway specific. Cell Rep. 20 (8), 1844–1854.

Terry, A.M.P., Jones, D.M., Davis, B.R., Slater, R., 1983. Parametric studies of tinnitus masking and residual inhibition. Br. J. Audiol. 17 (4), 245–256.

Turner, J.G., Brozoski, T.J., Bauer, C.A., Parrish, J.L., Myers, K., Hughes, L.F., Caspary, D.M., 2006. Gap detection deficits in rats with tinnitus: a potential novel screening tool. Behav. Neurosci. 120 (1), 188.

Tziridis, K., Ahlf, S., Jeschke, M., Happel, M.F., Ohl, F.W., Schulze, H., 2015. Noise trauma induced neural plasticity throughout the auditory system of Mongolian gerbils: differences between tinnitus developing and non-developing animals. Front. Neurol. 6, 22.

Ueberfuhr, M.A., Braun, A., Wiegrebe, L., Grothe, B., Drexl, M., 2017. Modulation of auditory percepts by transcutaneous electrical stimulation. Hear. Res. 350, 235–243.

Usher, M., Feingold, M., 2000. Stochastic resonance in the speed of memory retrieval. Biol. Cybern. 83 (6), L011–L016.
Vernon, J., 1977. Attempts to relieve tinnitus. J. Am. Audiol. Soc. 2, 124–131.
Wang, J., Powers, N.L., Hofstetter, P., Trautwein, P., Ding, D., Salvi, R., 1997. Effects of selective inner hair cell loss on auditory nerve fiber threshold, tuning and spontaneous and driven discharge rate. Hear. Res. 107 (1–2), 67–82.
Ward, L.M., Neiman, A., Moss, F., 2002. Stochastic resonance in psychophysics and in animal behavior. Biol. Cybern. 87 (2), 91–101.
Weisz, N., Müller, S., Schlee, W., Dohrmann, K., Hartmann, T., Elbert, T., 2007. The neural code of auditory phantom perception. J. Neurosci. 27 (6), 1479–1484.
Weisz, N., Hartmann, T., Müller, N., Obleser, J., 2011. Alpha rhythms in audition: cognitive and clinical perspectives. Front. Psychol. 2, 73.
Wenning, G., Obermayer, K., 2003. Activity driven adaptive stochastic resonance. Phys. Rev. Lett. 90 (12), 120602.
Wiegrebe, L., Kössl, M., Schmidt, S., 1996. Auditory enhancement at the absolute threshold of hearing and its relationship to the Zwicker tone. Hear. Res. 100 (1–2), 171–180.
Wiesenfeld, K., Moss, F., 1995. Stochastic resonance and the benefits of noise: from ice ages to crayfish and SQUIDs. Nature 373 (6509), 33–36.
Wrzosek, M., Szymiec, E., Obrebowska, Z., Norena, A., 2017. Continuous Zwicker tone illusion imitates tonal tinnitus-could Zwicker tone generators imitate different types of hearing loss? J. Hear. Sci. 7 (2), 168.
Wu, C., Stefanescu, R.A., Martel, D.T., Shore, S.E., 2015. Listening to another sense: somatosensory integration in the auditory system. Cell Tissue Res. 361 (1), 233–250.
Wu, C., Stefanescu, R.A., Martel, D.T., Shore, S.E., 2016. Tinnitus: maladaptive auditory–somatosensory plasticity. Hear. Res. 334, 20–29.
Yang, S., Weiner, B.D., Zhang, L.S., Cho, S.J., Bao, S., 2011. Homeostatic plasticity drives tinnitus perception in an animal model. Proc. Natl. Acad. Sci. U. S. A. 108 (36), 14974–14979.
Zacharek, M.A., Kaltenbach, J.A., Mathog, T.A., Zhang, J., 2002. Effects of cochlear ablation on noise induced hyperactivity in the hamster dorsal cochlear nucleus: implications for the origin of noise induced tinnitus. Hear. Res. 172 (1–2), 137–144.
Zeng, C., Yang, Z., Shreve, L., Bledsoe, S., Shore, S., 2012. Somatosensory projections to cochlear nucleus are upregulated after unilateral deafness. J. Neurosci. 32 (45), 15791–15801.
Zhou, X., Henin, S., Thompson, S.E., Long, G.R., Parra, L.C., 2010. Sensitization to masked tones following notched-noise correlates with estimates of cochlear function using distortion product otoacoustic emissions. J. Acoust. Soc. Am. 127 (2), 970–976.
Zwicker, E., 1964. “Negative afterimage” in hearing. J. Acoust. Soc. Am. 36 (12), 2413–2415.

HNO

Originalien

HNO 2021 · 69:891–898
https://doi.org/10.1007/s00106-020-00963-5
Angenommen: 21. August 2020
Online publiziert: 13. November 2020

A. Schilling[1] · P. Krauss[1] · R. Hannemann[2] · H. Schulze[1] · K. Tziridis[1]

[1] Experimentelle HNO-Heilkunde, HNO-Klinik, Kopf- und Halschirurgie, Universitätsklinikum Erlangen, Erlangen, Deutschland

[2] WSAudiology, Sivantos GmbH, R&D AAA SA ERL, Erlangen, Deutschland

Reduktion der Tinnituslautstärke

Pilotstudie zur Abschwächung von tonalem Tinnitus mit schwellennahem, individuell spektral optimiertem Rauschen

Nach gängigen Modellen [13] resultiert Tinnitus zumeist aus einem Hörverlust (HV), wobei die tinnitusbedingte Belastung mitunter schwerwiegender sein kann als der eigentliche HV. Die Behandlung beschränkt sich oft auf Bewältigungsstrategien, da Mechanismen der Tinnitusentstehung noch immer umstritten sind. In dieser Pilotstudie testen wir an Patienten einen neuartigen Ansatz einer möglichen zukünftigen Therapie, den wir auf der Grundlage unseres Modells der Tinnitusentstehung durch stochastische Resonanz (SR) [14] entwickelt haben und der an den von uns postulierten Ursachen derselben ansetzt.

Neben einigen Erfolgen bei der Linderung von Tinnitus in jüngerer Zeit [1, 19] ist ein wesentlicher Grund dafür, dass es bis heute kein an den Ursachen der Entstehung ansetzendes Heilverfahren zur Behandlung von chronischem Tinnitus gibt, die Uneinigkeit der Forschung über die zu seiner Entstehung führenden neurophysiologischen Mechanismen. Gängige Modelle zur Entstehung der verschiedenen, teilweise sehr heterogenen Tinnitusperzepte [5] gehen zwar nahezu alle davon aus, dass Tinnitus infolge eines (sehr geringen) Hörschadens entsteht, konnten bislang aber lediglich Teilaspekte des Phänomens erklären [2, 6, 9, 16, 18, 20, 21]. Unlängst haben wir ein neuartiges, mechanistisches Modell der Tinnitusentstehung, basierend auf tierexperimentellen Daten und der Modellierung neuronaler Netze, entwickelt [10, 12, 14, 15]. Es macht überprüfbare Vorhersagen und impliziert eine völlig neuartige Behandlungsstrategie gegen tonalen bzw. schmalbandigen Tinnitus. Ziel der vorliegenden Pilotstudie ist ein Proof of Concept dieses neuartigen Ansatzes.

Im Modell wird angenommen, dass die Tinnitusentstehung eine Begleiterscheinung des Versuchs des Gehirns ist, einen entstandenen Hörverlust auszugleichen. Dabei bedient sich das Hörsystem laut Modell des Phänomens der SR: Hierbei kann ein primär unterschwelliges Signal durch Beimischen von Rauschen geeigneter Intensität über die Sensor-Detektionsschwelle gehoben und so messbar gemacht werden [3, 8]. Wir konnten zeigen, dass sich die optimale Intensität des beizumischenden Rauschens mittels Autokorrelation des Sensoroutputs bestimmen lässt [12]. Das Modell [14] nimmt daher an, dass das Hörsystem durch Erhöhung von internem, neuronalem Rauschen SR auslöst und so das Hörvermögen nach Hörschaden sekundär wieder verbessert. Das hierfür nötige interne Rauschen wäre dann als Tinnitus wahrnehmbar.

An audiometrischen Daten von fast 40.000 Patienten konnten wir zeigen, dass Tinnituspatienten im sprachrelevanten Frequenzbereich bis 3 kHz tatsächlich signifikant bessere Hörschwellen aufwiesen als Patienten ohne Tinnitus [10]. Da das Modell somit eine konkrete Ursache für die als Tinnitus wahrgenommene neuronale Aktivität postuliert – das intern generierte neuronale Rauschen – haben wir eine neuartige Behandlungsstrategie gegen Tinnitus entworfen. Der Grundgedanke besteht darin, das laut Modell vom Hörsystem zur Verbesserung des Hörvermögens generierte und als Tinnitus wahrgenommene interne neuronale Rauschen durch extern beigemischtes akustisches Rauschen zu ersetzen. Dass es tatsächlich möglich ist, die Hörschwellen normalhörender Probanden durch extern beigemischtes Rauschen zu verbessern, konnte bereits gezeigt werden [22].

Das Neuartige an unserer Idee ist, dass – im Gegensatz z. B. zu herkömmlichen Rauschgeneratoren zur Tinnitusmaskierung – das SR-auslösende beigemischte Rauschen selbst nur knapp über- oder sogar unterschwellig sein muss, um das interne neuronale Rauschen überflüssig zu machen. Dieses externe Rauschen, so unser Ansatz, könnte dann anstelle des internen Rauschens die Hörschwellen verbessern, sodass für das Hörsystem die Notwendigkeit entfiele, internes Rauschen zu nutzen, wodurch in der Folge auch das darauf beruhende Tinnitusperzept verschwinden oder zumindest abgeschwächt werden sollte.

Methoden

Probanden

Es wurden 22 erwachsene Tinnitusprobanden (4 Frauen) mit Einverständnis (Ethikkommission des UK Erlangen: AZ 159_18) untersucht. Ihr mittleres Alter (±Standardabweichung) war 46,6 (±16,3) Jahre, der Median (unteres, oberes Quartil) der bestimmten Tinnitusfrequenz (TF) betrug 6 (4, 8) kHz.

Tab. 1 HV (dB) und Tinnitusparameter der 22 Probanden

Prob. Nr.	Seite (Ohr)	Frequenz (kHz)												MW HV (dB)	TF (kHz)	TL (dB)	TL (dB SL)	SG	Alter
		0,125	0,25	0,5	0,75	1	1,5	2	3	4	6	8	10						
1	R	5	6	5	4	3	9	8	2	3	8	10	–	5,7	2	12	4	4	22
	L	8	7	4	4	5	5	4	1	8	10	22	–	7,1	2	6	2		
2	R	9	10	9	7	10	22	38	56	71	71	75	–	34,4	8	74	3	2	79
	L	12	13	14	11	10	20	29	53	58	65	70	–	32,3	–	–	–		
3	R	6	3	3	5	5	4	4	3	3	7	7	–	4,5	3	14	11	3	22
	L	13	9	5	5	6	4	4	5	5	5	10	–	6,5	2	13	9		
4	R	10	5	5	3	3	4	4	4	6	6	9	–	5,4	6	11	5	1	26
	L	5	5	3	3	3	1	2	3	2	7	8	–	3,8	4	10	8		
5	R	4	0	1	2	3	2	2	14	19	10	18	–	6,8	–	–	–	3	54
	L	9	3	3	4	5	7	9	39	32	58	53	62	23,7	10	76	14		
6	R	11	8	8	7	7	7	7	8	15	10	10	–	8,9	–	–	–	1	27
	L	11	8	5	5	4	5	5	8	12	38	37	46	15,3	10	46	0		
7	R	8	7	5	6	5	8	8	8	18	19	41	59	16,0	6	21	2	4	38
	L	9	9	8	10	8	13	18	37	44	48	28	21	21,1	–	–	–		
8	R	8	3	3	5	6	13	10	27	33	59	57	72	24,7	6	60	1	–	53
	L	9	8	11	14	7	7	13	29	43	61	70	78	29,2	6	64	3		
9	R	11	10	10	13	20	25	29	46	48	44	48	–	27,6	8	59	11	3	50
	L	12	10	11	14	22	34	39	49	46	40	40	–	28,8	8	46	6		
10	R	11	7	3	3	6	11	15	14	2	79	81	–	21,1	8	82	1	2	54
	L	10	5	5	5	7	15	22	29	21	38	50	–	18,8	8	50	0		
11	R	12	9	8	8	7	7	7	9	18	35	32	–	13,8	0,125	15	3	4	49
	L	10	8	4	4	3	7	7	12	20	44	35	–	14,0	8	65	30		
12	R	3	2	1	3	4	3	0	15	24	28	17	32	11,0	–	–	–	2	40
	L	10	4	7	10	10	11	19	47	61	62	62	66	30,8	6	63	1		
13	R	9	9	10	18	25	44	58	59	54	64	62	–	37,5	6	66	2	2	53
	L	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–		
14	R	8	7	7	10	13	18	9	8	30	11	40	–	14,6	–	–	–	3	59
	L	11	12	17	19	19	25	18	19	29	58	72	–	27,2	0,125	20	9		
15	R	15	16	11	17	15	10	17	19	24	17	12	–	15,7	4	25	1	1	47
	L	21	20	12	17	19	14	18	21	21	27	13	–	18,5	4	22	1		
16	R	–	11	13	10	10	10	10	9	11	14	10	–	10,8	13,8	10	–	1	46
	L	–	17	13	12	11	11	14	14	14	20	14	–	14,0	13,8	12	–		
17	R	15	8	8	8	9	8	10	10	11	12	10	–	9,9	6	13	1	3	40
	L	9	7	7	6	9	9	18	9	20	18	9	–	11,0	4	26	6		
18	R	18	22	21	24	30	27	18	29	41	68	96	–	35,8	6	74	6	1	63
	L	15	16	12	13	18	25	34	64	67	86	90	–	40,0	6	87	1		
19	R	6	6	7	7	8	9	8	7	7	7	12	–	7,6	8	9	–3	1	24
	L	8	5	5	6	6	6	6	6	6	10	10	–	6,7	8	17	7		
20	R	13	12	10	11	15	20	30	41	58	63	75	73	35,1	–	–	–	4	72
	L	9	8	10	10	16	16	31	52	50	71	82	76	35,9	10	53	–23		
21	R	13	13	10	9	11	19	21	20	20	24	27	–	17,0	6	28	4	1	36
	L	19	19	15	17	15	12	16	20	57	78	60	–	29,8	–	–	–		
22	R	19	11	12	9	9	9	7	30	28	31	37	41	18,4	6	32	1	1	73
	L	13	12	13	9	9	9	27	31	30	31	35	38	19,9	6	34	3		

HNO 2021 · 69:891–898 https://doi.org/10.1007/s00106-020-00963-5

A. Schilling · P. Krauss · R. Hannemann · H. Schulze · K. Tziridis

Reduktion der Tinnituslautstärke. Pilotstudie zur Abschwächung von tonalem Tinnitus mit schwellennahem, individuell spektral optimiertem Rauschen

Zusammenfassung

Hintergrund. Tinnitus betrifft ca. 15 % der Bevölkerung, jedoch existiert noch immer kein echtes Heilverfahren. Ein von uns entwickeltes neuartiges Erklärungsmodell erlaubt nun die Erprobung einer gezielten, an den Ursachen der Tinnitusentstehung ansetzenden Behandlung. Diese basiert auf stochastischen Resonanzphänomenen an bestimmten synaptischen Verbindungen im Hörsystem, welche gezielt durch extern zugeführtes schwellennahes Rauschen induziert werden sollen.
Fragestellung. Die vorliegende Pilotstudie soll zeigen, ob ein spektral individuell angepasstes Rauschen erfolgreich chronischen tonalen/schmalbandigen Tinnitus während der Stimulation abschwächen kann.
Material und Methoden. Bei 22 volljährigen Tinnituspatienten (46.6±16.3 Jahre; 4 Frauen) wurden Hörverlust (HV) sowie Tinnitusfrequenzen (TF) und -lautstärken (TL) audiometrisch bestimmt. Darauf basierend wurden bis zu 8 verschiedene Rauschstimuli (RS) mit je 5 Lautstärken (–20 bis +20 dB SL) erzeugt. Diese wurden über audiologische Kopfhörer in einer Schallkammer für jeweils 40 s präsentiert. Nach jeder Präsentation wurde mithilfe einer 5-stufigen Bewertungsskala (–2 bis +2) ermittelt, ob sich die TL verändert hat.
Ergebnisse. Es fanden sich Patienten ohne Verbesserung der TL ($n = 6$) und solche mit Verbesserung ($n = 16$), wobei hier RS um die TF besonders effektiv waren. Die Gruppen zeigten post hoc deutliche Unterschiede in den Audiogrammen: Offenbar ist das hier getestete Verfahren insbesondere bei normalhörenden Tinnituspatienten und solchen mit geringgradigem HV effektiv.
Schlussfolgerung. Die subjektiv wahrgenommene TL war bei 16 von 22 Probanden für die Dauer der Stimulation reduziert. Für den möglichen Erfolg einer zukünftigen Therapie scheint der HV relevant zu sein.

Schlüsselwörter

Stochastische Resonanz · Reintonaudiometrie · Tinnitusfragebögen · Hörverlust · Individualtherapie

Reducing tinnitus intensity. Pilot study to attenuate tonal tinnitus using individually spectrally optimized near-threshold noise

Abstract

Background. Around 15% of the general population is affected by tinnitus, but no real cure exists despite intensive research. Based on our recent causal model for tinnitus development, we here test a new treatment aimed at counteracting the perception. This treatment is based on the stochastic resonance phenomenon at specific auditory system synapses that is induced by externally presented near-threshold noise.
Objective. This pilot study will investigate whether individually spectrally adapted noise can successfully reduce chronic tonal/narrowband tinnitus during stimulation.
Materials and methods. Hearing loss (HL) as well as tinnitus pitch (TP) and loudness (TL) were audiometrically measured in 22 adults (46.6±16.3 years; 4 women) with tinnitus. Based on these measurements, up to eight different noise stimuli with five intensities (–20 to +20 dB SL) were generated. These were presented for 40 s each via audiologic headphones in a soundproof chamber. After each presentation, the change in TL was rated on a five-level scale (–2 to +2).
Results. We found patients ($n = 6$) without any improvement in their TL perception as well as patients with improvement ($n = 16$), where stimulation around the TP was most effective. The groups differed in post-hoc analysis of their audiograms: the effectiveness of our new therapeutic strategy obviously depends on the individual HL, and was most effective in normal-hearing tinnitus patients and those with mild HL.
Conclusion. Subjective TL could be reduced in 16 out of 22 patients during stimulation. For a possible success of a future therapy, the HL seems to be of relevance.

Keywords

Stochastic resonance · Pure tone audiometry · Tinnitus questionnaires · Hearing loss · Individualized therapy

Alle Probanden hatten einen Reinton- bzw. schmalbandigen Tinnitus. Zwölf von 22 Probanden (vgl. auch **Tab. 1**) zeigten eine binaurale, meist nahezu symmetrische, 4 Probanden eine monaurale Mittel- bis Hochtonschwerhörigkeit (Hörverlust ≥20 dB), ein Proband war einseitig taub und monaural schwerhörig und 5 Probanden waren klinisch normalhörend. Alle Tinnitusschweregrade (SG; Mini-Tinnitus-Fragebogen, Mini TF12: SG I: 8; SG II: 4; SG III: 5; SG IV: 4; ein Fragebogen wurde nicht ausgefüllt; vgl. **Tab. 1**) waren vertreten.

Audiometrie und Rauschanpassung

Die Probanden wurden in der Audiologie der HNO-Klinik, Kopf- und Halschirurgie des UK Erlangen binaural audiologisch untersucht und sowohl die Reinton-Hörschwellen sowie die TF zwischen 0,125 und 10 kHz nach ISO 8253-1 bestimmt (Ausnahme Proband 16, TF von externem Audiologen bestimmt). Basierend auf diesen Daten wurden zwischen 6 und 8 (abhängig von ihrer TF) Rauschstimuli (RS) mit jeweils 5 Lautstärken (–20 bis +20 dB SL, 10-dB-Schritte) mittels eines selbst geschriebenen Python-Programms (Python 3.6, Numpy Bibliothek; Anaconda Distribution, Anaconda, Berlin, Deutschland) erzeugt. Zusätzlich wurde ein Durchlauf ohne Stimulation (Stille) erzeugt. Die verschiedenen RS waren: weißes Rauschen (WR), ein tiefpassgefiltertes (TPR) sowie ein hoch-

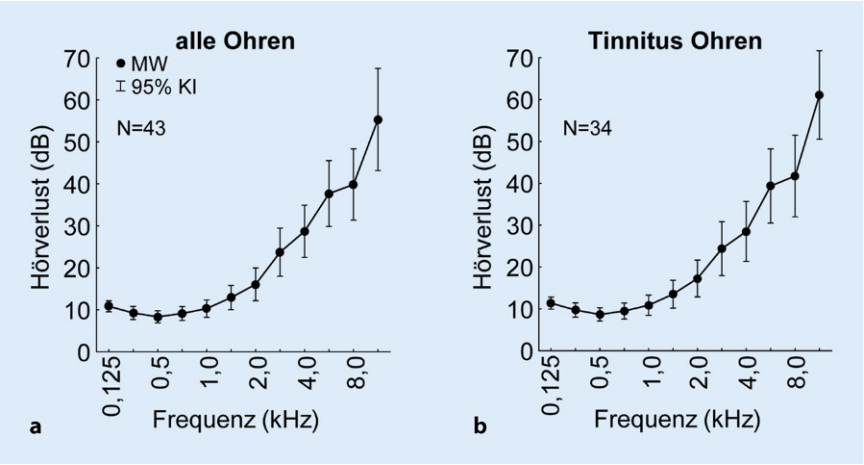

Abb. 1 ▲ HV der 22 Probanden. **a** Mittelwerte (MW) und 95%-Konfidenzintervalle (95%-KI) aller 43 gemessenen Ohren. **b** Mittlerer HV der 34 Tinnitus-Ohren

passgefiltertes Rauschen (HPR), jeweils mit der Grenzfrequenz bei der TF. Bis zu 5 verschiedene Schmalband gefilterte Rauschen (SBR) wurden verwendet mit jeweils einer Breite von ±½ Oktave. Die Mittenfrequenzen dieser SBR reichten von einer Oktave unterhalb der TF bis eine Oktave oberhalb der TF mit einer Schrittweite von ½ Oktave. In seltenen Ausnahmen wurden RS mit Mittenfrequenzen außerhalb dieser Spezifikationen verwendet (z. B. TF >8 kHz). Die Lautstärke der RS wurde entsprechend der Audiogramme der Ohren der Probanden gewählt, wobei bei unterschiedlichen Audiogrammen das Tinnitus-Ohr gewählt wurde; bei beidseitigem Tinnitus wurde das Audiogramm des besseren Ohrs verwendet. Für die Lautstärke des WR wurde der Audiogramm-Mittelwert aller Frequenzen verwendet, bei den HPR und TPR die Audiogramm-Mittelwerte der beinhalteten Frequenzen und bei den SBR wurde die Hörschwelle der Mittenfrequenz als 0 dB SL Referenz gewählt.

Experimentelle Durchführung

Die Stimuli wurden den Probanden an einem gesonderten Messtag über audiologische Kopfhörer in einer Schallkammer für jeweils 40 s einmal präsentiert, nach jeder Präsentation wurden sie vom Versuchsleiter gefragt, ob sich die TL verändert habe. Zur Eingewöhnung wurden die WR-Stimuli immer zuerst in absteigender Lautstärke präsentiert, mit Stille am Ende dieser Reihe. Danach wurden die individuell gefilterten RS in aufsteigender Lautstärke und Frequenz präsentiert. Die Probanden waren instruiert, nur auf ihren Tinnitus zu achten und konnten mit einer von 5 möglichen gewerteten Antworten antworten: „Tinnitus war deutlich lauter“ (−2), „Tinnitus war etwas lauter“ (−1), „TL hat sich nicht verändert“ (0), „Tinnitus war etwas leiser“ (+1) und „Tinnitus war deutlich leiser“ (+2). Zusätzlich konnten sie auch weitere Angaben machen (z. B. empfundene Maskierungseffekte) und wurden vor dem Versuch auch entsprechend instruiert. Die Messung aller Stimuli dauerte zwischen 45 und 60 min, Pausen waren möglich, wurden aber selten genutzt. Jeder Proband erhielt € 50,- Aufwands- sowie eine Fahrtkostenerstattung nach Abschluss der Experimente.

Statistische Auswertung

Die Daten wurden mittels Statistica 8 (Fa. StatSoft, Hamburg) ausgewertet. Zunächst wurden die Daten nach den individuellen Berichten der Probanden klassifiziert. Für alle Analysen der Hörschwellen wurden nur die von Tinnitus betroffenen Seiten (Ohren) ausgewertet. Es fanden sich 2 Gruppen von Probanden: Zum einen die „Nichtresponder“ (NR). Diese Probanden zeigten bei keiner der Stimulationen eine subjektive Verbesserung ihres Tinnitusperzepts ($n = 6$; 10 Tinnitus-Ohren, 27,3 % [29,4 %]). Bei ihnen kam es also bestenfalls zu Maskierungen. Zum anderen die „Responder“ (R; $n = 16$; 24 Tinnitus-Ohren, 72,7 % [70,6 %]), also die Probanden, die mindestens bei einem RS eine Antwort mit dem Wert +1 gaben, also eine subjektive Verbesserung ihres Tinnitusperzepts ohne Maskierung. Basierend auf dieser Klassifikation wurden die Hörschwellen der Tinnitus-Ohren (22 Probanden: 43 Ohren gesamt, 34 Tinnitus-Ohren, 79,1 %, ausgewertet) sowie die Antworten der Probanden auf die verschiedenen Rauschreize mittels ANOVA bzw. Kruskal-Wallis(KW)-ANOVA statistisch ausgewertet.

Ergebnisse

Hörverlust

In ◘ **Tab. 1** ist der audiologisch gemessene Hörverlust (HV) in dB für jeden Probanden angegeben. Der Mittelwert des HV (MW HV) ist über alle gemessenen Frequenzen für jedes Ohr berechnet und die Grundlage für die Lautstärke des WR. In den letzten Spalten ist die TF in kHz sowie die TL in dB und dB SL angegeben. Als zusätzliche Information sind in den letzten beiden Spalten der SG sowie das Alter der Probanden angegeben. Die mittleren HV (±95%-Konfidenzintervall, 95%-KI) aller 22 Probanden (43 Ohren) sind in ◘ **Abb. 1a** gezeigt; ◘ **Abb. 1b** zeigt den HV der 34 Tinnitus-Ohren. Nur diese Ohren wurden für die weiteren Analysen verwendet.

Tinnitus und Klassifizierung der Probanden

Die TF der 22 Probanden lag im Median (±Interquartilsabstand) bei 6 (4, 8) kHz. Die TF zeigten keine signifikanten Unterschiede in Abhängigkeit vom SG (KW-ANOVA: $H(3, 29) = 2{,}26$; $p = 0{,}52$). Die TL (dB SL) dagegen zeigte eine signifikante Abhängigkeit von diesem (KW-ANOVA: $H(3, 29) = 9{,}32$; $p = 0{,}025$) mit der größten TL bei SG III im Vergleich zu SG II (multiple Vergleiche der mittleren Ränge, $p = 0{,}045$), wobei die TL bei allen anderen SG nicht signifikant unterschiedlich voneinander waren. TF und TL (dB SL) zeigten keinen Zusammenhang (KW-ANOVA: $H(12, 33) = 4{,}95$; $p = 0{,}55$).

Tab. 2 Übersicht der Antworten der 22 Probanden während der Präsentation der RS

Rauschtyp	Probandennummer																					
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Stille	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
WR	0 (0,0)	0	4 (0,2)	4 (0,2)	1 (0,1)	1 (0,1)	0 (0,0)	0 (0,0)	0 (0,0)	1 (0,1)	−1 (−1,0)	0 (0,0)	0 (0,0)	2 (0,1)	2 (0,2)	0 (0,0)	−2 (−1,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)	0 (0,0)
HPR	−1 (−1,0)	0 (0,0)	−1 (−1,0)	2 (0,1)	4 (0,2)	3 (0,1)	1 (0,1)	0 (0,0)	0 (0,0)	0 (0,0)	−1 (−1,0)	0 (0,0)	0 (0,0)	2 (0,1)	5 (0,2)	–	2 (0,2)	0 (0,0)	0 (0,0)	–	1 (0,1)	2 (0,1)
TPR	–	0 (0,0)	4 (0,1)	–	0 (0,0)	0 (0,0)	1 (0,1)	0 (0,0)	0 (0,0)	−1 (−1,0)	−2 (−1,0)	5 (0,1)	–	–	–	0 (0,0)	–	–	3 (0,1)	0 (0,0)	–	–
SBR-2	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0 (0,0)	–	–	–	0 (0,0)	–	–
SBR-1,5	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0 (0,0)	–	–	–	0 (0,0)	–	–
SBR-1	0 (0,0)	0 (0,0)	2 (0,1)	0 (0,0)	1 (0,1)	2 (0,1)	−1 (−1,0)	0 (0,0)	0 (0,0)	2 (0,1)	−1 (−1,0)	1 (0,1)	0 (0,0)	–	4 (0,2)	0 (0,0)	1 (0,1)	0 (0,0)	−3 (−1,0)	0 (0,0)	1 (0,1)	2 (0,1)
SBR-0,5	0 (−1,1)	0 (0,0)	3 (0,1)	1 (−1,1)	6 (0,2)	2 (0,1)	1 (0,1)	0 (0,0)	0 (0,0)	4 (0,1)	0 (0,0)	2 (0,1)	0 (0,0)	–	0 (0,0)	0 (0,0)	2 (0,1)	0 (0,0)	0 (0,0)	0 (0,0)	4 (0,2)	4 (0,2)
SBR 0	0 (−1,1)	0 (0,0)	4 (0,2)	0 (0,0)	4 (0,2)	2 (0,1)	5 (0,2)	0 (0,0)	0 (0,0)	5 (0,2)	0 (−1,1)	1 (0,1)	3 (0,2)	4 (0,2)	0 (0,0)	–	3 (0,1)	0 (0,0)	−1 (−1,0)	0 (0,0)	2 (0,2)	4 (0,2)
SBR +0,5	0 (0,0)	0 (0,0)	0 (0,0)	1 (0,1)	–	–	–	0 (0,0)	0 (0,0)	0 (0,0)	0 (−1,1)	2 (0,1)	2 (0,2)	2 (0,2)	6 (1,2)	–	3 (0,2)	0 (0,0)	0 (0,0)	–	0 (0,0)	1 (0,1)
SBR+1	−1 (−1,0)	–	–	4 (0,2)	–	–	–	–	–	–	–	1 (0,1)	0 (0,0)	1 (0,1)	4 (0,2)	–	4 (0,2)	0 (0,0)	–	–	0 (0,0)	4 (0,1)
SBR +1,5	–	–	–	–	–	–	–	–	–	–	–	–	–	1 (0,1)	–	–	–	–	–	–	–	–
Maskiert	j	j	j	j	j	*n*	j	j	j	*n*	j	*n*	*n*	*n*	*n*	j	*n*	j	j	j	j	*n*
Klasse	R	NR	R	R	R	R	R	NR	NR	R	R	R	R	R	R	NR	R	NR	R	NR	R	R

Die **Tab. 2** zeigt Übersichtsdaten der Probandenantworten: individuelle Antwort-Summe über alle 5 Lautstärken (Minimum: −10; Maximum: +10), sowie in Klammern Minimum und Maximum der gegebenen Antworten (−2 bis +2) bei den verschiedenen RS (bei den SBR entspricht die Zahl dahinter dem Abstand der Mittenfrequenz von der TF in Oktaven). Eine eventuelle Maskierung und die entsprechende Klassifizierung in NR und R (vgl. Methoden) folgt. Bei Stille war nur eine Antwort möglich, sodass hier die Summe von −2 bis +2 geht und es kein Minimum und Maximum gibt.

Klassifizierungsabhängige Analysen der Hörschwellen

Basierend auf der Klassifizierung aus den Antworten der Probanden wurde der Hörverlust der Tinnitus-Ohren analysiert. Zunächst wurde eine 2-faktorielle ANOVA (Faktoren *Frequenz* und *Gruppe*) der auf die Frequenzen alignierten individuellen HV-Daten durchgeführt. Die Probanden zeigten im Mittel eine Hochtonschwerhörigkeit (*Frequenz*: $F(11, 380) = 34{,}98$; $p < 0{,}001$). Der Faktor *Gruppe* zeigte ebenfalls einen Einfluss auf den Hörverlust ($F(1, 380) = 69{,}98$; $p < 0{,}001$) mit den Mittelwerten (±95%-KI) der Gruppe NR 31,3 (±4,4) dB und R 18,2 (±1,9) dB. Die signifikante Interaktion ($F(11, 380) = 3{,}85$; $p = 0{,}001$) der beiden Faktoren (**Abb. 2a**) zeigt die Staffelung des HV der beiden Gruppen frequenzabhängig auf. Erst ab den Frequenzen oberhalb von 3 kHz zeigen sich die HV in den Tukey-Post-hoc-Tests ($p < 0{,}05$) zwischen NR- und R-Probanden signifikant.

In einer weiteren 2-faktoriellen ANOVA (Faktoren: *Abstand zur TF* sowie *Gruppe*) wurden die HV-Daten auf die individuelle TF aligniert und der Abstand der Frequenzen in Halboktaven dazu berechnet. Es zeigte sich auch hier eine Abhängigkeit des HV vom Abstand zur TF ($F(10, 276) = 22{,}3$; $p < 0{,}001$), wobei sich ein Plateau von −1 okt TF bis +1 okt TF erstreckt, das einen signifikant größeren HV im Bereich der TF im Vergleich zu den tieferen Frequenzen zeigt (Tukey-Post-hoc-Tests, $p < 0{,}05$). Innerhalb des Plateaus zeigen

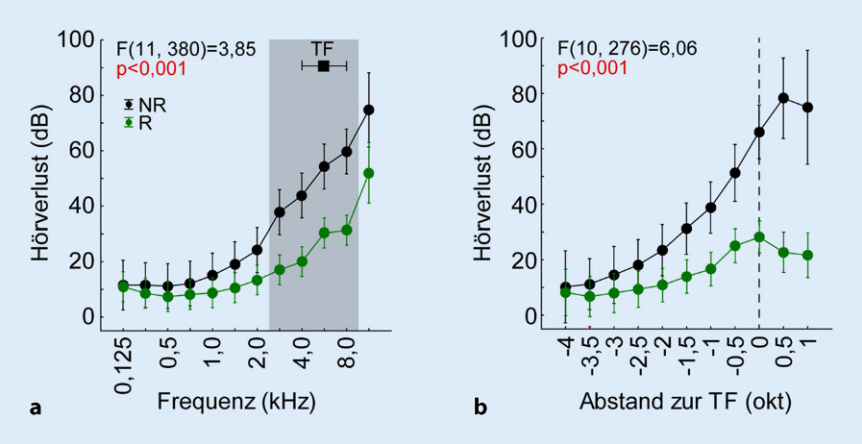

Abb. 2 ▲ Mittlerer HV der Probandengruppen. **a** MW und 95%-KI aller 34 Tinnitus-Ohren aligniert auf die Frequenzen, aufgeteilt nach Gruppen. Median der TF (±Interquartilsabstand, IQ): *schwarzes Quadrat*; sign. Unterschied der Gruppen: *graues Areal*. **b** MW und 95%-KI aller 34 Tinnitus-Ohren aligniert auf die individuelle TF (*gestrichelte Linie*)

sich keine signifikanten Unterschiede im HV. Auch hier unterscheiden sich die mittleren HV der 2 Gruppen signifikant ($F(2, 276) = 115{,}2$; $p < 0{,}001$) mit NR 38,0 (±5,5) dB und R 15,6 (±2,4) dB. Auch die Interaktion der beiden Faktoren (**Abb. 2b**) zeigt einen signifikanten frequenzabhängigen Einfluss der Gruppen auf den HV ($F(10, 276) = 6{,}06$; $p < 0{,}001$). Die beiden Gruppen unterscheiden sich dabei in der Kurvenform deutlich (Tukey-Post-hoc-Tests), die NR-Probanden zeigen ab +0,5 okt TF ein Plateau mit größtem HV und die R-Probanden zeigen eine HV-Spitze zwischen −0,5 okt TF und der TF. Mit anderen Worten, die NR-Probanden haben einen großen HV, der vor allem oberhalb der TF zu finden ist, und die R-Probanden zeigen einen moderaten HV, der knapp unterhalb und bei der TF am größten ist.

Subjektive Veränderung der TL bei Respondern

Die Responder berichteten Verbesserungen der TL durch bestimmte RS. Dies ist in **Abb. 3** zusammengefasst. Die Analyse der Antworten innerhalb der einzelnen RS wurde mit KW-ANOVA mit dem Faktor *Rauschlautstärke* durchgeführt. Die Antworten aller 16 R-Probanden (**Abb. 3a**) zeigte in 2 der 8 RS eine signifikante Abhängigkeit von der Rauschlautstärke und in einem der 8 RS einen Trend dazu. Es fällt auf, dass die Mediane der Antworten erst ab +10 dB SL zu steigen scheinen (also eine subjektive Verringerung der TL anzeigen). Die Post-hoc-Statistik kann aber aufgrund der geringen Anzahl der Messungen (maximal 16 Datenpunkte pro Lautstärke) und der relativ hohen Streuung nur einen einzigen Trend (Median Tests für multiple Vergleiche) aufzeigen: SBR −0,5 okt TF: 0 dB vs. 20 dB; $p = 0{,}066$. Wie schon beim HV wurden die Antworten der Probanden auf ihre TF aligniert: der Rauschreiz mit der niedrigsten Cut-off- oder Mittenfrequenz und größten subjektiven Antwort wurde in Relation zur TF gesetzt. Dieses „optimale Rauschen" lag zwischen −1 okt und +1 okt zur TF (**Abb. 3b**), wobei 14 der 24 möglichen positiven Antworten (58,3 %) direkt bei der TF lagen und insgesamt ebenfalls 14 der 24 Antworten mit einer Wertigkeit von +2 angegeben wurden.

Diskussion

In dieser Pilotstudie konnte der Proof of Concept erbracht werden, dass spektral an den HV von Tinnituspatienten angepasstes, schwellennahes Rauschen in der Lage ist, ein Tinnitusperzept mindestens teilweise abzuschwächen, ohne es zu maskieren. Diese Stimulation scheint einen signifikanten Effekt zu haben, im Gegensatz zu Studien mit amplitudenmodulierten Reintönen [17]. Im Gegensatz zur herkömmlichen Maskierung, die durch ein lautes unspezifisches Rauschen den Tinnitus übertönen soll [4], zielt unser therapeutischer Ansatz darauf ab, die Entstehungsursache von Tinnitus im Gehirn, die laut unserem Modell primär der Verbesserung des Hörens (genauer, der Optimierung der Informationsübertragung zwischen Ohr und Gehirn) dient [14], überflüssig zu machen.

Einige Probanden der Responder-Gruppe 8/16 (50 %) berichteten bei bestimmten RS auch von Maskierungseffekten (R_{mask}), meist beim WR (26 %), deutlich seltener bei anderen RS (zwischen 4 und 10 % je Stimulus). Von diesen Maskierungen traten 83,6 % bei Stimuli mit +10 oder +20 dB SL, also bei vergleichsweise hohen Lautstärken, auf. Die R_{mask}-Probanden (vgl. **Tab. 1**) hatten einen signifikant geringeren mittleren HV (7,3 dB; t-Test, $p < 0{,}001$) als die Probanden, die niemals über einen Maskierungseffekt berichteten. Wir schließen daraus, dass die sehr guten Hörschwellen der überwiegend normalhörenden Probanden dieser Subgruppe (R_{mask}) dazu führten, dass bereits Stimuli mit +10 oder +20 dB SL eine deutliche Maskierung des Tinnitus bewirkten, dies aber nicht frequenzspezifisch, sondern vor allem bei WR-Stimulation auftrat. Alle Responder zeigten demgegenüber ihre besten Ergebnisse bei RS, welche spektral im Bereich der TF lagen, was den Vorhersagen aus unserem Modell entspricht [13]. Die NR-Probanden zeigten in der Audiometrie dagegen den größten HV. Offenbar stieß hier das von uns entwickelte akustische Stimulationsverfahren schlicht an seine Grenzen.

Dies zeigt auch eine der Limitationen dieser Pilotstudie. Wir untersuchten ein kleines Kollektiv mit breiter Altersverteilung und sehr unterschiedlichen Hörverlusten, Tinnitusfrequenzen und Belastungsgraden. Das Rauschen wurde in relativ groben 10-dB-Schritten und relativ breiten Frequenzspektren einmalig dargeboten. Allgemeine Aussagen zu allen Tinnitustypen lassen sich mit einer solchen Pilotstudie also nicht treffen, allerdings haben fast 73 % der Probanden mit Reinton-/Schmalband-Tinnitus von einem positiven Effekt der Stimulation berichtet.

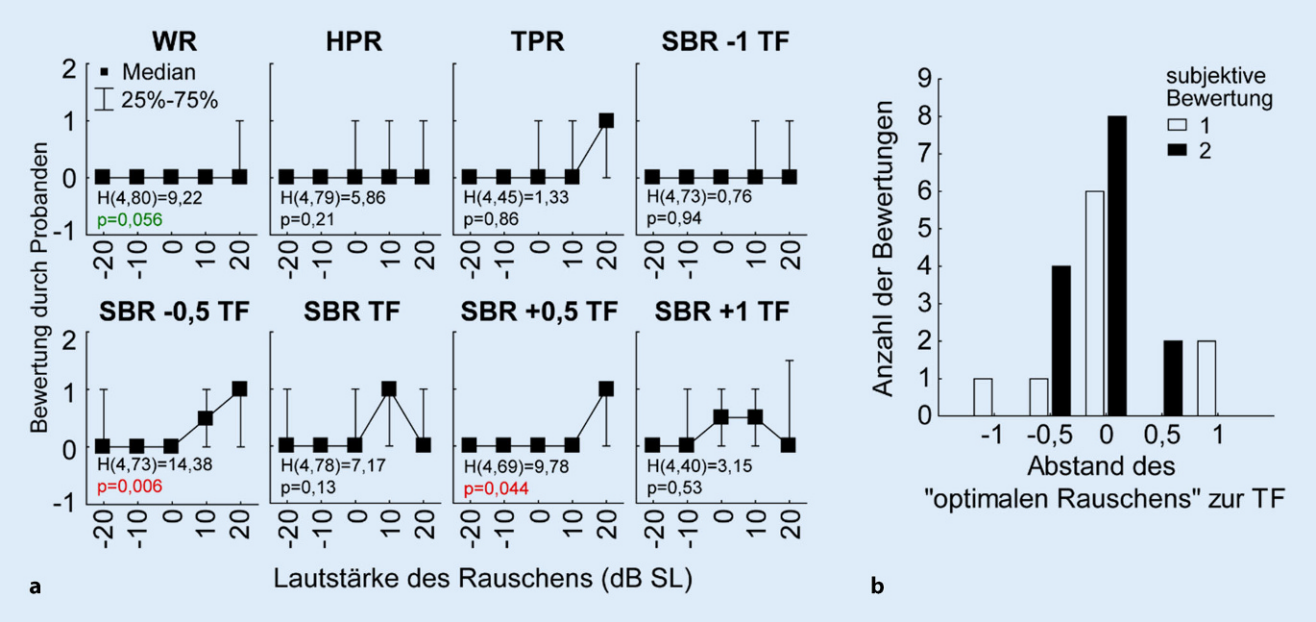

Abb. 3 ▲ Mediane (±IQ) der Antworten der 16 R-Probanden. **a** Aufgetragen nach RS. Ergebnisse der KW-ANOVA in jeder Teilabbildung. **b** Histogramm der subjektiven Bewertung über den Abstand des „optimalen Rauschstimulus" relativ zur individuellen TF

Unser Ansatz unterscheidet sich von herkömmlicher Maskierung und auch von sog. Residual-Inhibition-Ansätzen [7] vor allem durch die Lautstärke des dargebotenen RS (vgl. auch [11]). Das spektral individuell angepasste Rauschen liegt im Bereich der Hörschwelle oder wenige dB darüber. Die Probanden nehmen auch nur während der Stimulation eine Unterdrückung ihres Tinnitus wahr. Dies entspricht ebenfalls genau den Vorhersagen des Modells und ermöglicht es uns als nächsten Schritt, Lautstärke und Spektralzusammensetzung der RS noch genauer anzupassen, um dann perspektivisch Geräte mit dieser Technologie auszustatten.

Fazit für die Praxis

Insgesamt 16/22 Probanden profitierten von der Stimulation, d. h. zeigten bei optimaler spektraler Zusammensetzung und Lautstärke des Rauschreizes eine Verringerung der Tinnituslautstärke, die nicht auf simpler Maskierung beruhte.
Wir schließen aus diesen Befunden, dass für einige Tinnituspatienten mit ausreichendem Hörvermögen ein kontinuierlich dargebotener Rauschstimulus so optimiert werden kann, dass das Tinnitusperzept im Idealfall ganz unterdrückt wird.

Korrespondenzadresse

Dr. K. Tziridis
Experimentelle HNO-Heilkunde, HNO-Klinik, Kopf- und Halschirurgie, Universitätsklinikum Erlangen
Waldstraße 1, 91054 Erlangen, Deutschland
Konstantin.tziridis@uk-erlangen.de

Danksagung. Wir danken Prof. Dr. Hoppe für die Ermöglichung der Messungen in der Audiologie der HNO-Klinik Erlangen.

Funding. Open Access funding enabled and organized by Projekt DEAL.

Einhaltung ethischer Richtlinien

Interessenkonflikt. A. Schilling, P. Krauss, R. Hannemann, H. Schulze und K. Tziridis geben an, dass kein Interessenkonflikt besteht.

Alle beschriebenen Untersuchungen am Menschen oder an menschlichem Gewebe wurden mit Zustimmung der zuständigen Ethikkommission, im Einklang mit nationalem Recht sowie gemäß der Deklaration von Helsinki von 1975 (in der aktuellen, überarbeiteten Fassung) durchgeführt. Von allen beteiligten Patienten liegt eine Einverständniserklärung vor.

Literatur

1. Adamchic I, Toth T, Hauptmann C et al (2017) Acute effects and after-effects of acoustic coordinated reset neuromodulation in patients with chronic subjective tinnitus. Neuroimage Clin 15:541–558
2. Ahlf S, Tziridis K, Korn S et al (2012) Predisposition for and prevention of subjective tinnitus development. Plos One 7:e44519
3. Benzi R, Sutera A, Vulpiani A (1981) The mechanism of stochastic resonance. J Phys A: Math Gen 14:L453
4. Cai Y, Zhou Q, Yang H et al (2017) Logistic regression analysis of factors influencing the effectiveness of intensive sound masking therapy in patients with tinnitus. BMJ Open 7:e18050

5. Cederroth CR, Gallus S, Hall DA et al (2019) Towards an understanding of tinnitus heterogeneity. Front Aging Neurosci 11:53
6. Eggermont JJ, Roberts LE (2004) The neuroscience of tinnitus. Trends Neurosci 27:676–682
7. Fournier P, Cuvillier A-F, Gallego S et al (2018) A new method for assessing masking and residual inhibition of tinnitus. Trends Hear 22:2331216518769996
8. Gammaitoni L, Hänggi P, Jung P et al (1998) Stochastic resonance. Rev Mod Phys 70:223
9. Gerken GM (1996) Central tinnitus and lateral inhibition: an auditory brainstem model. Hear Res 97:75–83
10. Gollnast D, Tziridis K, Krauss P et al (2017) Analysis of audiometric differences of patients with and without Tinnitus in a large clinical database. Front Neurol 8:31
11. Henry JA, Jastreboff MM, Jastreboff PJ et al (2002) Assessment of patients for treatment with tinnitus retraining therapy. J Am Acad Audiol 13:523–544
12. Krauss P, Metzner C, Schilling A et al (2017) Adaptive stochastic resonance for unknown and variable input signals. Sci Rep 7:2450
13. Krauss P, Schilling A, Tziridis K et al (2019) Modelle der Tinnitusentstehung. HNO 67:172–177
14. Krauss P, Tziridis K, Metzner C et al (2016) Stochastic resonance controlled upregulation of internal noise after hearing loss as a putative cause of tinnitus-related neuronal hyperactivity. Front Neurosci 10:597
15. Krauss P, Tziridis K, Schilling A et al (2018) Cross-modal stochastic resonance as a universal principle to enhance sensory processing. Front Neurosci 12:578
16. Leaver AM, Turesky TK, Seydell-Greenwald A et al (2016) Intrinsic network activity in tinnitus investigated using functional MRI. Hum Brain Mapp 37:2717–2735
17. Neff P, Zielonka L, Meyer M et al (2019) Comparison of amplitude modulated sounds and pure tones at the tinnitus frequency: residual tinnitus suppression and stimulus evaluation. Trends Hear 23:2331216519833841
18. Schaette R, Mcalpine D (2011) Tinnitus with a normal audiogram: physiological evidence for hidden hearing loss and computational model. J Neurosci 31:13452–13457
19. Stein A, Wunderlich R, Lau P et al (2016) Clinical trial on tonal tinnitus with tailor-made notched music training. BMC Neurol 16:38
20. Tziridis K, Ahlf S, Jeschke M et al (2015) Noise trauma induced neural plasticity throughout the auditory system of Mongolian gerbils: differences between Tinnitus developing and non-developing animals. Front Neurol 6:22
21. Vanneste S, De Ridder D (2016) Deafferentation-based pathophysiological differences in phantom sound: tinnitus with and without hearing loss. Neuroimage 129:80–94
22. Zeng F-G, Fu Q-J, Morse R (2000) Human hearing enhanced by noise1. Brain Res Brain Res Protoc 869:251–255

https://doi.org/10.1093/brain/awad255 BRAIN 2023: 146; 4809–4825 | 4809

BRAIN
REVIEW ARTICLE

Predictive coding and stochastic resonance as fundamental principles of auditory phantom perception

Achim Schilling,[1,2] William Sedley,[3] Richard Gerum,[2,4] Claus Metzner,[1] Konstantin Tziridis,[1] Andreas Maier,[5] Holger Schulze,[1] Fan-Gang Zeng,[6] Karl J. Friston[7] and Patrick Krauss[1,2,5]

Mechanistic insight is achieved only when experiments are employed to test formal or computational models. Furthermore, in analogy to lesion studies, phantom perception may serve as a vehicle to understand the fundamental processing principles underlying healthy auditory perception. With a special focus on tinnitus—as the prime example of auditory phantom perception—we review recent work at the intersection of artificial intelligence, psychology and neuroscience. In particular, we discuss why everyone with tinnitus suffers from (at least hidden) hearing loss, but not everyone with hearing loss suffers from tinnitus.

We argue that intrinsic neural noise is generated and amplified along the auditory pathway as a compensatory mechanism to restore normal hearing based on adaptive stochastic resonance. The neural noise increase can then be misinterpreted as auditory input and perceived as tinnitus. This mechanism can be formalized in the Bayesian brain framework, where the percept (posterior) assimilates a prior prediction (brain's expectations) and likelihood (bottom-up neural signal). A higher mean and lower variance (i.e. enhanced precision) of the likelihood shifts the posterior, evincing a misinterpretation of sensory evidence, which may be further confounded by plastic changes in the brain that underwrite prior predictions. Hence, two fundamental processing principles provide the most explanatory power for the emergence of auditory phantom perceptions: predictive coding as a top-down and adaptive stochastic resonance as a complementary bottom-up mechanism.

We conclude that both principles also play a crucial role in healthy auditory perception. Finally, in the context of neuroscience-inspired artificial intelligence, both processing principles may serve to improve contemporary machine learning techniques.

1 Neuroscience Lab, University Hospital Erlangen, 91054 Erlangen, Germany
2 Cognitive Computational Neuroscience Group, University Erlangen-Nürnberg, 91058 Erlangen, Germany
3 Translational and Clinical Research Institute, Newcastle University Medical School, Newcastle upon Tyne NE2 4HH, UK
4 Department of Physics and Astronomy and Center for Vision Research, York University, Toronto, ON M3J 1P3, Canada
5 Pattern Recognition Lab, University Erlangen-Nürnberg, 91058 Erlangen, Germany
6 Center for Hearing Research, Departments of Anatomy and Neurobiology, Biomedical Engineering, Cognitive Sciences, Otolaryngology–Head and Neck Surgery, University of California Irvine, Irvine, CA 92697, USA
7 Wellcome Centre for Human Neuroimaging, Institute of Neurology, University College London, London WC1N 3AR, UK

Received October 26, 2022. Revised June 27, 2023. Accepted July 15, 2023. Advance access publication July 28, 2023

Correspondence to: Achim Schilling
ENT Clinic, Head and Neck Surgery, University Hospital Erlangen
Waldstrasse 1, 91054 Erlangen, Germany
E-mail: achim.schilling@fau.de

Keywords: artificial intelligence; Bayesian brain; phantom perception; predictive coding; stochastic resonance; tinnitus

Introduction

The ultimate goal of neuroscience is to gain a mechanistic understanding of how information is processed in the brain. Since the early beginnings of the scientific study of the brain, lesions or more broadly anatomical damages and their physiological effects have provided pivotal insights into brain function. Analogously, phantom perception may serve as a vehicle to understand the fundamental processing principles underlying normal perception. The prime example of an auditory phantom perception is tinnitus, which is believed to be caused by anatomical damage along the auditory pathway. Here we provide a mechanistic explanation of how tinnitus emerges in the brain: namely, how the neural and mental processes underlying perception, cognition and behaviour contribute to and are affected by the development of tinnitus. These insights may not only point to strategies how tinnitus may be reversed or at least mitigated, but also how auditory perception is implemented in the brain in general.

While there is broad agreement in the auditory neuroscience community on these goals, there is far less agreement on the way to achieve them. There is still a popular belief among neuroscientific and psychological tinnitus researchers that we are largely data-driven. In other words, generating large, multi-modal and complex datasets—analysed with advanced data science methods—are believed to lead to fundamental insights into how tinnitus emerges. Indeed, in the last decades we have assembled a broad database, which has inspired models that make quantitative predictions. These predictions scaffold new experimental paradigms that aim to unravel the mechanisms of tinnitus perception. In the following, we summarize some of the main findings in tinnitus research, over the last decades, and then turn to strategic questions about how to leverage these advances, from the perspective of formal modelling.

Some universal correlations between hearing loss, tinnitus and neural hyperactivity in the auditory system have been found in both animal and human studies. These reproducible findings can be considered as the common denominator of tinnitus research and could offer the minimal starting point for theoretical considerations. Tinnitus is a phenomenon arising somewhere along the auditory pathway, but not in the inner ear.[1] Thus, it can be shown that the spontaneous activity of neurons along the auditory pathway is increased after hearing loss,[2–4] whereas the damaged cochlea transmits less information to the higher auditory nuclei.[5,6] However, it has been argued that not all alterations in neural activity in animal models, which were caused by an acoustic trauma, are necessarily related to tinnitus.[1,7] Although there exist some behavioural tests to check for the putative presence of a tinnitus percept based on conditioning[8,9] or startle responses,[10–12] the reliability of these paradigms remains controversial.[7,13] Thus, studies on human subjects complement these findings. In several recent studies, it was shown that the tinnitus pitch lies within the frequency range of the hearing loss and thus it is an obvious assumption that tinnitus can be regarded as a within frequency channel phenomenon.[14–17] Potentially, it is sufficient to assume that the mechanisms causing tinnitus occur in each impaired frequency channel individually and that crosstalk between the different frequency channels along the tonotopic map is not crucial to explain the basic principles behind tinnitus development.[18]

This assumption is supported by recent findings e.g. by Dalligna and coworkers,[14] who report that the tinnitus is directly centred at the frequency of the largest hearing loss. For the sake of completeness, it should be mentioned that other studies on tinnitus and its relation to hearing loss found a special emphasis of the edges of the impaired frequency range on the tinnitus pitch.[19-21] However, recently, Keppler and coworkers[15] contradicted these findings and stated that there is no correlation between tinnitus frequency and the edges of the impaired frequency ranges.

Indeed, the above neural correlates of tinnitus and hearing loss are just a small distillation of all studies that aspire to unravel mechanisms that underpin tinnitus, but these findings are robust and constitute the basis of most theoretical and computational models of tinnitus. In the 1990s the first computational models of tinnitus emerged. These models considered decreased lateral inhibition—due to deficient auditory input (i.e. cochlear damage)—as the main cause of tinnitus. Gerken[22] created a feed-forward brainstem model and suggested the inferior colliculus to be the crucial structure for tinnitus development. Kral and Majernik,[23] as well as Langner and Wallhäuser-Franke[24] pursued computational models, based on decreased lateral inhibition. Bruce and coworkers[25] developed these models further and implemented lateral inhibition in a spiking recurrent neural network. In a subsequent step, the principles were implemented in a model of the auditory cortex based on spiking neurons.[26]

Besides lateral inhibition, homeostatic plasticity[27] and central gain changes are hypothesized to be the cause for tinnitus emergence and manifestation. These hypotheses are based on the idea that incoming neuronal signals are amplified, in order to compensate the decreased input from the damaged cochlea. Thus, Parra and Pearlmutter[28] implemented that principle in an 'abstract' model, where they simply defined several frequency channels with a certain input. The output was scaled with the average, which means that a decreased input leads to a higher scaling or amplification factor, respectively. However, they did not consider a plausible neural implementation of their mathematical model. Schaette and Kempter[29-31] further developed several computational models, investigating the effects of central gain increase on tinnitus emergence. Finally, Chrostowski and coworkers[32] developed a cortex model to investigate central gain changes in the cortex (for detailed review on computational tinnitus models see Schaette and Kempter[1]).

In 2013, Zeng[33] introduced a model that argues that tinnitus is not caused by increased central gain, which means a multiplicative amplification of the signal, but by increased central noise, which means an additive neural noise, that is intrinsically generated. The idea of an additional intrinsic or extrinsic noise as an explanation for tinnitus has gained some popularity in recent years e.g. Koops and Eggermont.[34] However, Zeng raised the question why the brain should increase central noise levels. This question was

addressed in 2016 by Krauss and coworkers,[18] who showed that internally generated neural noise could partially restore hearing ability after hearing loss through the effect of stochastic resonance.[35-37] Stochastic resonance is a phenomenon in which the addition of noise to a non-linear system can improve its sensitivity to weak signals. It occurs when a system—which is normally unable to detect weak signals—features an optimal level of noise that lifts the weak signals above the detection threshold. This is because the noise serves to 'jiggle' the system, making it easier for weak signals to cross a response threshold. However, the effect only works in a narrow range, as the noise amplitude has to be tuned to an optimal level. Noise amplitudes that are too low would not lift the subthreshold signal above the detection threshold. Conversely, noise amplitudes that are too high would significantly worsen the signal-to-noise ratio up to a level at which the signal disappears completely in the noise. Stochastic resonance has been observed in a variety of physical, biological and neural systems, (for overview *cf*. Koops and Eggermont[34] and Krauss *et al*.[35]).

The idea behind the stochastic resonance model of auditory phantom perception (Erlangen model of tinnitus development) is that a subthreshold signal—from an impaired cochlea—is lifted stochastically above the detection threshold by adding uncorrelated neural noise. In earlier studies it has been shown that human hearing may be enhanced beyond the absolute threshold of hearing by adding acoustic white noise to a subthreshold acoustic stimulus.[38] The Erlangen model hypothesizes that this mechanism is also implemented in the dorsal cochlear nucleus (DCN) and that —instead of acoustic white noise—internally generated neural noise is added to the cochlea output, to lift it above the detection threshold.[37] Recently, several studies have provided evidence that cross-modal stochastic resonance is a universal principle for enhancing sensory perception.[36,39,40]

This stochastic resonance hypothesis is further supported by the finding that, on average, hearing thresholds are better in patients suffering from hearing loss with tinnitus compared to a control group of patients suffering from hearing loss but without tinnitus.[19,41,42] Along the same line, the stochastic resonance effect as add-on to the central noise model may explain the Zwicker tone illusion,[43-45] i.e. the perception of a phantom sound, which occurs after stimulation with notched noise, and why auditory sensitivity for frequencies adjacent to the Zwicker tone are improved beyond the absolute threshold of hearing during Zwicker tone perception.[46] Furthermore, recently, a crucial prediction of the stochastic resonance model of tinnitus development was confirmed experimentally by using brainstem audiometry[47] and assessing behavioural signs of tinnitus[10] in an animal model: simulated transient hearing loss improves auditory thresholds and leads, as a side effect, to the perception of tinnitus.[48] Both the model from Zeng and the model from Krauss *et al*.,[35] are not based on a particular or specified neural network architecture. However, in 2020, Schilling and coworkers developed a hybrid model based on a biophysically realistic model of the cochlea and the DCN combined with a deep neural network representing all further processing stages along the auditory pathway. In this model, intrinsically generated noise could indeed significantly increase speech perception via SR.[49] Recently, a similar hybrid neural network model has led to further insights into the mechanisms of impaired speech recognition caused by hearing loss.[50]

In parallel to the intrinsic neural noise models from Zeng and Krauss and colleagues, Sedley and coworkers developed a conceptual model, which describes tinnitus as arising from a prediction error of the brain.[51,52] This model is based on the idea that the brain is a Bayesian prediction machine, trying to minimize prediction errors or free energy,[53,54] a principle also known as predictive coding. According to the theoretical framework of predictive coding, the brain's main function is to generate and test predictions about incoming sensory information. In particular, the brain is constantly generating hypotheses or predictions about what is happening in the environment, based on past experiences, and then comparing these predictions with incoming sensory data. The ensuing prediction error is then thought to drive representations about states of affairs generating sensations towards better predictions; thereby resolving prediction errors.

This predictive coding model of tinnitus addresses the issue of whether or not an individual perceives tinnitus as an interplay between existing auditory predictions (which, by default, do not feature tinnitus) and spontaneous activity (i.e. noise) in the central auditory pathway (considered a 'tinnitus precursor'). Whether the posterior (i.e. percept) crosses the threshold for perception depends on both of these factors, including their mean values (e.g. firing rate) and their precision. More recently, by using a hierarchical Gaussian filter, a computational instantiation of this model has been able to explain phenomenology in individual tinnitus subjects and predict their residual inhibition characteristics.[55] Despite the fact that in recent years tinnitus research converged to the three main models described above (central noise, central gain, predictive coding), it has to be stated that there exist several further computational simulations and approaches trying to explain and characterize tinnitus development based e.g. on information theoretical considerations (see Dotan and Shriki[56] and Gault *et al*.[57]).

Besides the computational models that rest upon a mathematical formulation, there exist several phenomenological models, such as the thalamo-cortical dysrhythmia model,[58,59] the thalamic low-threshold calcium spike model,[60] the fronto-striatal gating hypothesis[61,62] and the overlapping subnetwork theory.[63,64] Finally, there exists another Bayesian brain/predictive coding model of tinnitus, which is somewhat the polar opposite to what Sedley and Friston were arguing for. There, tinnitus is not believed to arise from spontaneous noise increase, which higher predictions go on to accept, but on the contrary that tinnitus arises from reduced input to the auditory cortex, leading it to 'make up' or 'fill in' an auditory percept from auditory memory.[65] However, this assumption contradicts the findings that spontaneous neural activity is increased along the entire auditory pathway starting from the DCN after hearing loss.[2-4]

As there exist various models of tinnitus development, which are far too numerous to be treated in detail in this study, criteria are needed to define which models are apt to understand tinnitus development. In their review paper, Schaette and Kempter[1] define three major criteria for the quality of a model: first—and in line with Popper's ideas[66]—a model should be falsifiable, which means there should be experimental paradigms, which could be used to test a certain candidate model. Second, a model should make quantitative predictions, as opposed to purely qualitative, often vague, predictions, *cf*. also Lazebnik.[67] Third, a model should be as simple as possible, i.e. contain the smallest number of parameters and assumptions as possible, a principle called Ockham's razor.[68] Hence, if two models explain experimental data equally well, the simpler one has to be considered the better one.

With the huge progress of artificial intelligence (AI) during the last decade, which is mainly due to increased computing power, a new discipline has been founded, called Cognitive Computational Neuroscience (CCN) as an integrative endeavour at the intersection of AI, cognitive science and neuroscience.[69,70]

Here, we first discuss the opportunities and limitations of this new research agenda. In particular, we present key thought experiments that highlight the major challenges on the road towards a CCN of tinnitus. In the light of these considerations, we subsequently review current models of tinnitus and assess their explanatory power. Finally, we present an integration of those models that we consider most promising and point towards a unified theory of tinnitus development.

Three challenges ahead

The challenge of developing a common formal language

In 2002, Yuri Lazebnik compared the biologists' endeavour—of trying to understand the building blocks and processes of living cells—with the problems that engineers typically deal with. In his opinion paper '*Can a biologist fix a radio?—Or, what I learned while studying apoptosis*', Lazebnik argued that many fields of biomedical research at some point reach

> 'a stage at which models, that seemed so complete, fall apart, predictions that were considered so obvious are found to be wrong, and attempts to develop wonder drugs largely fail. This stage is characterized by a sense of frustration at the complexity of the process'.[67]

Subsequently, Lazebnik[67] discussed a number of intriguing analogies between the physical and life sciences. In particular, he identified formal language as the most important difference between the two. Lazebnik argues that biologists and engineers use quite different languages for describing phenomena. On the one hand, biologists draw box-and-arrow diagrams, which are—even if a certain diagram makes overall sense—difficult to translate into quantitative assumptions, and hence limits its predictive or investigative value.

Indeed, these thoughts fit to the criterion for a 'good model' as pointed out by Schaette and Kempter,[1] i.e. that a model should make quantitative predictions. However, on the other hand a model should be as simple as possible and understandable, which means that it is important to find a compromise between too fine-grained and too coarse-grained descriptions of the system to be explained (see Marr's levels of analysis in Fig. 1A, based on Marr and Poggio[71]).

Lazebnik also remarks that scientific assumptions and conversations are often 'vague' and 'avoid clear, quantifiable predictions'. A freely adapted example drawn from Lazebnik's paper[67] would be a statement like

> 'an imbalance of excitatory and inhibitory neural activity after hearing-loss appears to cause an overall neural hyperactivity, which in turn seems to be correlated with the perception of tinnitus'.

Descriptions of electrophysiological findings are an important starting point for hypothesis generation, but they are no more than a first step. Description needs to be complemented with explanation and prediction (compare also the four main goals of psychology as described in Holt *et al.*[72]). Furthermore, Lazebnik urges a more formal common language for biological sciences, in particular a language that has the precision and expressivity found in engineering, physics or computer science. Any engineer trained in electronics for instance, is able to unambiguously understand a diagram describing a radio or any other electronic device. Thus, engineers can discuss a radio using terms that are common ground in the community. Furthermore, this commonality enables engineers to identify familiar functional architectures or motifs; even in a diagram of a completely novel device. Finally, due to the mathematical underpinnings of the language used in engineering, it is perfectly suited for quantitative analyses and computational modelling. For instance, a description of a certain radio includes all key parameters of each component like the capacity of a capacitor, but not irrelevant parameters—that do not 'matter'—like its colour, shape or size.

We emphasize that this does not mean that anatomical descriptions are useless in order to understand brain function, especially since there is a close correlation between structure and function in the brain. However, also in neurobiology there exist both kinds of detail: those that are crucial for understanding neural processing, and those that are not relevant variables.

Lazebnik concludes that 'the absence of such language is the flaw of biological research that causes David's paradox', i.e. the paradoxical phenomenon frequently observed in biology and neuroscience that 'the more facts we learn the less we understand the process we study'.[67]

Some conclusions for tinnitus research can be drawn from Lazebniks' thoughts on a more formal approach in biological sciences. The 'central gain' and 'homeostatic plasticity' theory on tinnitus emergence is a good example how the communication on tinnitus research can be improved. For example, Roberts stated in 2018 that the increase of central gain is 'increase of input output functions by forms of homeostatic plasticity', which means that homeostatic plasticity is necessarily connected to central gain adaptations.[73] In contrast to that, Schaette and Kempter[1] state that central gain changes can occur within seconds and thus are not necessarily caused by homeostatic plasticity. Only on longer timescales, both effects can be regarded as 'functionally equivalent'.[1] Indeed, tinnitus research would profit from a unified terminology for the different concepts, in the best case, a mathematical formulation.

The challenge of developing a unified mechanistic theory

In 2014, Joshua Brown built on Lazebnik's ideas and published the opinion article 'The tale of the neuroscientists and the computer: why mechanistic theory matters'.[74] In this thought experiment, a group of neuroscientists finds an alien computer and tries to figure out its function.

First, the MEG/EEG researcher tried to investigate the computer. She found that every time 'when the hard disk was assessed, the disk controller showed higher voltages on average, and especially more power in the higher frequency bands'.[74]

Subsequently, the cognitive neuroscientist, i.e. the functional MRI researcher argued that MEG/EEG has insufficient spatial resolution to see what is going on inside the computer. He carried out a large number of experiments, the results of which can be summarized with the realization that during certain tasks, certain regions seem to be more activated and that none of these components could be understood properly in isolation. Thus, the researcher analysed the interactions of these components, showing that there is a vast variety of different task-specific networks in the computer.

Finally, the electrophysiologist noted, critically, that his colleagues may have found coarse-grained patterns of activity, but it is still unclear what the individual circuits are doing. He starts to implant microelectrode arrays into the computer and probes

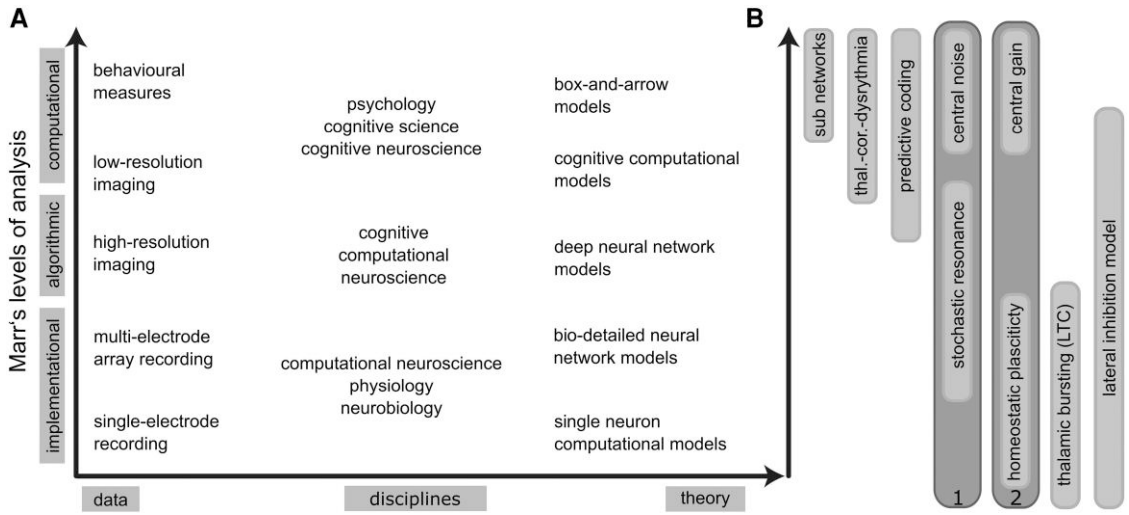

Figure 1 Marr's levels of analysis. (**A**) The scheme illustrates how measurement methods (such as MEG, EEG etc.), neuroscientific disciplines, as well as theoretical models can be structured in three different levels of analysis (according to Marr and Poggio[71]). (**B**) Tinnitus models in the light of the three levels of analysis. The grey bars illustrate how the different models cover the different levels of analysis (implementational, algorithmic, computational). The central noise model and the stochastic resonance model can be unified (1). The stochastic resonance model is at the algorithmic level as there exists a neural network model,[38] which could reproduce the stochastic resonance effect in tinnitus context. The exact molecular mechanism, such as the specific neurotransmitter, are unknown and therefore it is not at the implementational level. The mathematical formulation of the predictive coding model cannot be fully translated to a neural network model and therefore it is at the computational model. A neural network implementation of the predictive coding model would be algorithmic. Homeostatic plasticity is a collection of the molecular and thus implementational mechanisms behind the central gain model (2).

individual circuit points by measuring voltage fluctuations. With careful observation, the electrophysiologist identifies units responding stochastically when certain inputs are presented, and that nearby units seem to process similar inputs. Furthermore, each unit seems to have characteristic tuning properties.

Brown's tale ends with the conclusion that even though they performed a multitude of different empirical investigations, yielding a broad range of interesting results, it is still highly questionable whether 'the neuroscientists really understood how the computer works'. [74]

This provocative thought experiment speaks to some ideas that are relevant for tinnitus research.

In 2021, four leading scientists in tinnitus research discussed different tinnitus models at the Annual-Mid-Winter Meeting of the Association for Otolaryngology and diagnosed a 'lack of consistency of concepts about the neural correlate of tinnitus'.[75] Thus, a clearly defined theoretical framework is needed, which helps empirical groups to develop experimental paradigms suited to confirm or falsify different candidate models. To achieve that, interdisciplinary teams or at least an inter-disciplinary approach is needed.[76]

The challenge of developing appropriate analysis methods

In 2017 Jonas and Kording implemented the thought experiment of Brown in an experimental study. In their study 'Could a neuroscientist understand a microprocessor?'[77] the authors address this question by emulating a classical microprocessor, the MOS 6502, which was implemented as the central processing unit (CPU) in the Apple I, the Commodore 64, and the Atari Video Game System, in the 1970s and 1980s. In contrast to contemporary CPUs, like Intel's i9-9900K, that consist of more than three billion transistors, the MOS 6502 only consisted of 3510 transistors. It served as a 'model organism' in the mentioned study, and performed three different 'behaviors', i.e. three classical video games (*Donkey Kong*, *Space Invaders* and *Pitfall*).

The idea behind this approach is that the microprocessor, as an artificial information processing system, has three decisive advantages compared to natural nervous systems. First, it is fully understood at all levels of description and complexity, from its gross architecture and the overall data flow, through logical gate primitives, to the dynamics of single transistors. Second, its internal state is fully accessible without any restrictions to temporal or spatial resolution. And third, it offers the ability to perform arbitrary invasive experiments on it, which are impossible in living systems due to ethical or technical reasons. Using this framework, the authors applied a wide range of popular data analysis methods from neuroscience to investigate the structural and dynamical properties of the microprocessor. The methods used included—but were not restricted to—Granger causality for analysing task-specific functional connectivity, time-frequency analysis as a hallmark of MEG/EEG research, spike pattern statistics, dimensionality reduction, lesioning and tuning curve analysis.

The authors concluded that although each of the applied methods yielded results strikingly similar to what is known from neuroscientific or psychological studies, none of them could actually elucidate how the microprocessor works, or more broadly speaking, was appropriate to gain a mechanistic understanding of the investigated system.

Of course, there are potential criticisms of this study; for example, the brain is no computer and thus the drawn parallels are insufficient. Nevertheless, the idea to use a known model system to check for the validity of the evaluation procedures and common methods is a seminal principle. In 2009 Bennett and coworkers[78] performed an even stranger experiment, when they used standard functional MRI and statistics techniques to analyse the brain activity of a dead salmon, and indeed found a blood oxygenation level-

dependent (BOLD) signal due to stimulation. At first glance, this experiment seemed to be at least useless if not even funny, but it was a wake-up call and indeed changed the way functional MRI data are evaluated. Nowadays there exist strict rules how to correct for multiple testing in functional MRI research, to prevent pseudo-effects being a result of wrong statistical testing.[78,79] In computational neuroscience and AI research, newly developed methods are always applied to standard datasets such as the MNIST (Modified National Institute of Standards and Technology) database consisting of 60 000 images of hand-written digits[80,81] or artificially generated datasets with known properties, e.g. Schilling *et al.*,[47,82] Zenke and Vogels[83] and Krauss *et al.*[84] The principle of using fully known—even trivial systems—to test the validity of tools, methods or even theories could be an important approach in tinnitus research. Even in computational modelling, simply implementing a system in all details without an underlying theory, which serves as a solid base, will not lead to a real understanding. Indeed, theory needs computational modelling, but the statement is also true the other way around.[85] Therefore, it is crucial that computational models meet a basic standard—they should be capable of accurately explaining well established and simple phenomena. This serves as a basis to verify their validity before drawing more complex conclusions.

Towards a cognitive computational neuroscience of tinnitus

What does it mean to understand a system?

If popular analysis methods fail to deliver mechanistic understanding, what are the alternative approaches? Most obviously, narrative hypotheses about the structure and function of the system under investigation will help. Instead of simply describing data features with correlations, coherence, Granger causality etc.—in the hope of learning something about the functioning of the system under investigation—it would be much more effective to have a concrete hypothesis about the structure or function architecture of the system and then search for empirical evidence for that and alternative hypotheses.

Note that this does not exclude explorative analysis of existing data, in order to generate new hypotheses. However, as we pointed out in a previous publication,[86] to avoid statistical errors due to 'HARKing' ('hypothesizing after results are known' is defined as generating scientific statements exclusively based on the analysis of huge datasets without previous hypotheses[87,88] and to guarantee consistency of the results, it is necessary to apply e.g. resampling techniques such as subsampling.[47] Alternatively, the well established machine learning practice of cross-validation: i.e. splitting the dataset into multiple parts before the beginning of the evaluation can be used. There, one data part is used for generating new hypotheses and another part for subsequently statistically testing these hypotheses. Accumulation of such data-driven knowledge may finally lead to a new theory.

Ideally, the verbally defined (narrative) hypotheses to be experimentally tested would be derived from such an underlying theory. As Kurt Lewin, the father of modern experimental psychology, pointed out: 'There is nothing so practical as a good theory'.[89] If we had theorized that the microprocessor from the thought experiment above performs arithmetic calculations, we could have, e.g. derived the hypothesis that there must be something like 1-bit adders, and could have searched for them specifically.

Conversely, Allan Newell, one of the fathers of artificial intelligence, stated that 'You can't play 20 questions with nature and win'.[90] This suggests that testing one narrative hypothesis after another will never lead to a mechanistic understanding. Therefore, this raises the fundamental question of what it actually means to 'understand' a system.

Yuri Lazebnik argued that understanding of a system is achieved when one could fix a broken implementation:

> 'Understanding of a particular region or part of a system would occur when one could describe so accurately the inputs, the transformation, and the outputs that one brain region could be replaced with an entirely synthetic component'.[67]

In engineering terms, this understanding can be simply described as $y = f(x)$, where x is the input, y is the output and f is the transformation.

According to David Marr, one can seek to understand a system at (at least) three complementary levels of analysis.[71] He distinguished the computational, the algorithmic and the implementational level of analysis (Fig. 1A). The computational level is the most coarse-grained level of analysis. It asks what computational problem is the system seeking to solve, that results in the observed phenomena; in our context, phantom perceptions like tinnitus. This level of analysis is addressed by the fields of psychology and cognitive neuroscience. In contrast, the implementational level represents the most fine-grained description of a system. Here, the system's concrete physical layout is analysed. In computer science and engineering, this corresponds to the exact hardware architecture and the individual software realization, with a particular programming language. In the brain, where there exists no clear distinction between software and hardware (or wetware), this level of description corresponds to the structural design of ion channels, synapses, neurons, local circuits and larger systems, and the physiological processes these components are subject to. This level of analysis can be considered as the hallmark of physiology and neurobiology. Finally, the algorithmic level takes an intermediate position between the previously described levels. It is about which algorithms—that are physically realized at the implementational level—the system employs to manipulate its internal representations, in order to solve the tasks and problems identified at the computational level. In computer science, the algorithmic level would be described independently of a specific programming language by abstract pseudocode.

Indeed, there are ways of moving between the different levels of description, afforded by 'cognitive computational models'[91] and 'cognitive computational neuroscience'.[92] Thus, in both fields, cognitive processes are simulated or recapitulated *in silico*, however, cognitive computational neuroscience uses—in contrast to 'cognitive computational models'—neural networks as basis of the simulations. Therefore, cognitive computational neuroscience gives us an idea how processing might work algorithmically in the brain. Note that the similar terms (cognitive computational neuroscience and cognitive computational models) reflect the long—and not always straight-forward—history of science of mind. Indeed, very recently the term cognitive computational neuroscience is more and more replaced by the term neuroAI.[93,94]

We argue that analysis at the algorithmic level is most crucial to understand auditory phantom perceptions like tinnitus or Zwicker tone. Only by knowing the algorithms that underlie normal auditory perception, we will gain a detailed understanding of what exactly happens under certain pathological conditions such as

hearing loss, and which processes eventually cause the development of tinnitus, so that we can mitigate or reverse these processes.

Which discipline addresses this level of analysis in tinnitus research? Computational neuroscience comes to mind immediately. However, in 'good old-fashioned' computational neuroscience, great efforts have been made to model the physiological and biophysical processes at the level of single neurons, dendrites, axons, synapses or even ion channels, leading to increasingly complex computational models. These models, mostly based on systems of coupled differential equations, can mimic experimental data in great detail. Perhaps the most popular among these models is the famous Hodgkin-Huxley model,[95] which reproduces the temporal course of the membrane potential of a single neuron with impressive accuracy. These types of models are of great importance to deepen our understanding of fundamental physiological processes. However, in our opinion, they also must be considered as belonging to the implementational level of analysis, since they merely describe the physical realization of the algorithms, rather than the algorithms themselves.

In the following section, we will discuss emerging research directions that speak to the algorithmic level of analysis in the context of tinnitus research.

The integration of artificial intelligence in tinnitus research

As we argued above, hypothesis testing alone does not lead to a mechanistic understanding. Instead, it needs to be complemented by the construction of task-pointing computational models, since only synthesis in a computer simulation can reveal the interaction of proposed components entailed by a mechanistic explanation, i.e. which algorithms are realized, and whether they can account for the perceptual, cognitive or behavioural function in question. As Nobel laureate and theoretical physicist Richard Feynman pointed out: 'What I cannot create, I do not understand'.

Along these lines, one may consider extending the four goals of psychology, i.e. to describe, explain, predict and change cognition and behaviour,[72] by adding a fifth one: to build synthetic cognition and behaviour. This is in the tradition of 'Walter's tortoises',[96-99] one major attempt to build synthetic cognition and behaviour using analogue electronics. This approach could be revisited in the 21st century, using artificial deep neural networks.

As pointed out in previous publications,[69,100-103] these computational models can be based on constructs from AI, for example deep learning.[104,105] A related development in AI rests upon the explicit use of generative models, leading to formulations of action and perception, in terms of predictive coding and active inference. Examples of their application to auditory processing and hallucinations range from examining the role of certain oscillatory frequencies in message passing, through to simulations of active listening and speech perception.[106-112]

Artificial deep neural networks are designed to solve problems clearly defined at the computational level of analysis, in our case auditory perception tasks like, e.g. speech recognition. These models are precisely defined at an algorithmic level, which is completely independent from any individual programming language or specific software library, i.e. the implementational level of analysis. Hence, these algorithms could, at least in principle, also be realized in the brain as biological neural networks. Once we have built such models and algorithms in computer simulations, we can subsequently compare their dynamics and internal representations with brain—and behavioural—data in order to reject or adjust putative models, thereby successively increasing biological fidelity.[69] Vice versa, the ensuing models may also serve to generate new testable hypotheses about cognitive and neural processing in auditory neuroscience.

As mentioned above, this research approach—combining AI, cognitive science and neuroscience—has been coined as CCN.[69] Furthermore, besides the advantages discussed above, this approach furnishes the opportunity for *in silico* testing of new, putative treatment interventions for conditions like tinnitus, prior to *in vivo* experiments. In this way, CCN may even serve to reduce the number of animal experiments.

However, we note that CCN of auditory perception is not only beneficial for neuroscience. As noted in Hassabis *et al.*,[113] understanding biological brains could play a vital role in building intelligent machines, and that current advances in AI have been inspired by the study of neural computation in humans and animals. Thus, CCN of auditory perception may contribute to the development of neuroscience-inspired AI systems in the domain of natural language processing.[114] Finally, neuroscience may even serve to investigate machine behaviour,[115] i.e. illuminate the black box of deep learning.[116,117] However, so far, most AI research does not even attempt to mimic or understand the brain or biology in general.

In other neuroscientific strands, such as research on spatial navigation, the fusion of classical neuroscience and AI has already led to major breakthroughs and still promises further advances in the future.[118] For example, Stachenfeld and colleagues developed a mathematical framework for the function of place and grid cells in the entorhinal-hippocampal system based on predictive coding.[119,120] On the other hand, researchers from Google DeepMind developed artificial agents based on Long-Short-Term-Memory (LSTM)[121,122] neurons, in which place and grid cells emerged automatically.[123] In another AI model, Gerum and coworkers[124] showed that spatial navigation in a maze could be achieved by very small neural networks, which are trained with an evolutionary algorithm and are evolutionary pruned.

Towards a unified model of tinnitus development

The hierarchy of the different tinnitus models

In the following section, we describe a path towards a CCN of tinnitus research. Thus, in a first step we have to go back to Labzebnik[67] and find a way to communicate efficiently and formally about various tinnitus models. Extant tinnitus models can be sorted by the different levels of analysis according to Marr and Poggio.[71] This means that each model can be assigned to one or more of the three categories (Fig. 1B): implementational level (molecular mechanisms, synapses etc.), algorithmic level (how neural signals are translated to information processing) and computational level (what are the basic mathematical imperatives for processing; see also 'What does it mean to understand a system?').

The three levels of analysis can be easily illustrated with the Lateral Inhibition Model of tinnitus, which describes tinnitus as a result of decreased lateral inhibition[23,125] due to decreased cochlear input; e.g. caused by a noised-induced cochlear synaptopathy.[6] Thus, the lateral inhibition model explains tinnitus on all different levels of description. The implementational level (see Marr's level of analysis in Fig. 1A), which corresponds to the molecular mechanisms of lateral inhibition, is nearly fully understood. For example, in the DCN cartwheel cells release glycine to inhibit fusiform cells,

which are excitatory.[126-128] The computational role of inhibition is to narrow the input range of the fusiform cells.[128] To provide such contrast enhancement via lateral inhibition, neurons surrounding a certain excitatory neuron, which receives auditory input, are inhibited. This wiring scheme 'sharpens' the tuning curves of neurons along the auditory pathway. The wiring scheme corresponds to the algorithmic level of analysis. The computational level of description is the mathematical description of decreased lateral inhibition. Thus, hearing loss leads to decreased input from the cochlea, which causes a decreased firing rate of the inhibitory neurons and thus to disinhibition of subsequent excitatory neurons. These properties can be easily written down in simple mathematical formulas. This means that the underlying mechanisms of the lateral inhibition model of tinnitus are fully understood from specific neurotransmitter processes to an abstract mathematical formulation. This is the goal of cognitive computational neuroscience. However, the fact that the model explains tinnitus manifestation on all scales does not say anything about the correctness of the model's predictions. Indeed, a good model should be understood on all scales (implementational to computational), but it must also fit experimental observations, which is not the case in the Lateral Inhibition Model. Other models trying to explain tinnitus do not provide full explanatory power.

The thalamic bursting theory—which proposes that bursting neurons in the thalamus cause tinnitus—has a valid explanation for the origin of the spike bursts (low threshold calcium spikes, for details see Jeanmonod *et al.*[60]). However, it remains elusive, in terms of how these bursts cause tinnitus. Other top-down models—such as the predictive coding model,[52] based on the Bayesian Brain theory—provide a valid mathematical description of the proposed mechanisms, but do not provide a full explanation of how the Bayesian statistics can be implemented in a neural network and thus in the brain.[129] However, there exist some first approaches toward neural networks for Bayesian inference which will ultimately prove possible, but are still not fully developed.[130-132] Other tinnitus models describe macro-phenomena such as the thalamo-cortical dysrhythmia,[59] or describe tinnitus as a result of overlapping neural circuits.[63] Those models are phenomenological, but do not provide a mathematical description and thus are difficult to falsify or test *in silico*.

A critical role of stochastic resonance and central noise

In the following paragraph we provide an in-depth discussion of central noise and central gain, as possible causes for tinnitus, and consider how to adjudicate between—or combine—these two theories. To discuss these two models and their relationship, it is necessary to introduce a proper nomenclature. Thus, in the following we refer to the mathematical description of Zeng, who describes central gain as a linear amplification factor g, which increases the input signal I (s = subjectively perceived loudness respectively evoked neural activity; *cf.* Chrostowski *et al.*[32]). Central noise N is a further additive term[33,133] (Eq. 1).

$$s = g \cdot I + N \quad (1)$$

Central gain increase is a collective term summarizing all mechanisms that lead to an increased amplification of the input signal (I) along the auditory pathway (for an extensive review, see Auerbach *et al.*[134]). Therefore, the term central gain increase can refer to a decrease in inhibitory synaptic responses, an increase in excitatory synaptic responses, as well as enhanced intrinsic neuronal excitability.[134] All of these mechanisms cause a multiplicative amplification of the input signal (amplification factor: g). To sharpen the scientific language, it is necessary to distinguish between the observable effect (central gain increase) and the underlying neuronal principles (e.g. homeostatic plasticity[27,134]). Central gain increase could be caused by homeostatic plasticity, which means that the average spike rate of affected neurons—after a decrease of neuronal input due to a hearing loss—is kept constant by plastic changes of the system (e.g. enhanced intrinsic excitability, synaptic scaling, meta-plasticity[134]). Central gain and homeostatic plasticity are often used as synonyms in the context of tinnitus models, although they describe the problem on different levels.

The central noise model—in contrast to the central gain model—describes tinnitus as a consequence of increased spontaneous activity, which is added to the input signal (additive term, N).[33] In analogy to the relation of central gain and homeostatic plasticity, the underlying principle of the central noise model is stochastic resonance.[135-137] The original central noise model of Zeng from 2013 was exclusively based on psychophysical considerations and measurements, which means that the term s in Eq. 1 was meant to be the subjectively perceived loudness.[33] Furthermore, the original model makes no statements on the nature of the central noise, or on higher brain functions such as thalamic gating or predictive coding etc., and thus provides no explanation what the neuronal signal looks like and why an addition of noise causes an ongoing conscious percept.

The novelty of the stochastic resonance model is based on the idea that the abstract concept of an additive central noise can be interpreted as real intrinsically generated neural noise, which increases hearing ability by exploiting the stochastic resonance effect. Thus, the term s in Eq. 1 is re-interpreted as actual neural activity.

We categorized the stochastic resonance model[18,35,37] as an algorithmic level model on Marr's scale (Fig. 1), which means that the calculations (Eq. 1 and the calculation of the autocorrelation function, *cf.* below) necessary to leverage the stochastic resonance effect should be linked to the neural substrate of the auditory pathway. This corresponds to the implementational level according to Marr. The stochastic resonance model (Fig. 2) is based on the idea that the auditory system continuously optimizes sensitivity via a feedback loop, which adapts the amplitude of the additive noise (central noise) to maximize information transmission. The information transmission is quantified via the autocorrelation of the signal.[18,35] Thus, one might call the inverse autocorrelation function the cost-function to be minimized. However, to calculate the autocorrelation of the signal, so-called neuronal delay lines are needed, which are prominent in two brain regions the cerebellum and the DCN.[138] The mechanism behind the auto-correlation calculation based on delay lines is based on the fact that the signal transmission is slowed down by the delay line through inter-neurons or unmyelinated nerve fibres.[42] Thus, the delayed signal is then compared to the same signal at a later time point, which was not delayed. The delay serves the purpose to generate a time shift (in mathematical formulation of auto-correlation function commonly termed as lag-time τ), which allows us to compare one signal stream at several time points with itself.[139,140]

Another strong argument for the validity of the stochastic resonance model is the fact that otherwise there is no plausible explanation for the cross-modal input from the somatosensory system to the DCN,[141,142] except the notion that the somatosensory

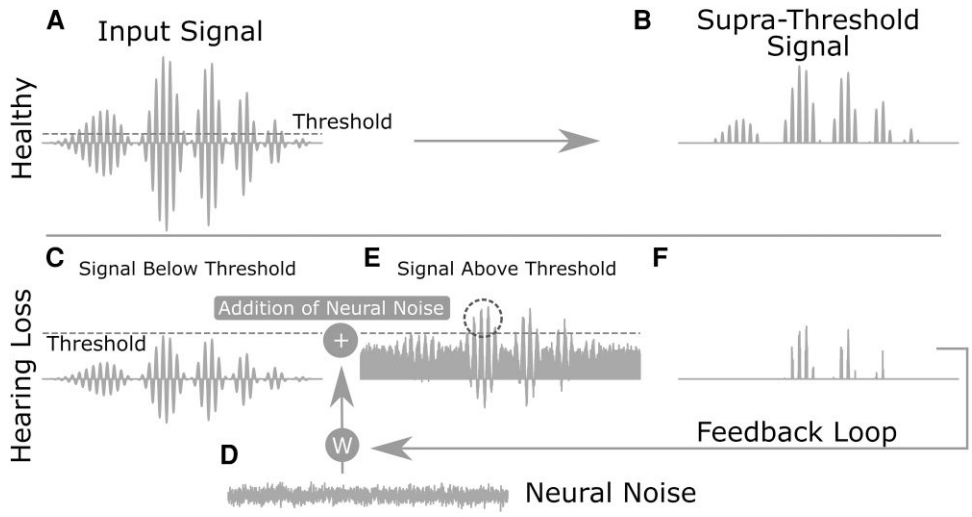

Figure 2 Stochastic resonance model of tinnitus induction. In the healthy auditory system, the input signal (**A**) can pass the detection threshold resulting in a supra-threshold signal as output (**B**). In case of hearing loss, the input signal remains below the threshold (**C**), resulting in zero output. However, if the optimum amount neural noise (**D**) is added to the weak signal, then signal plus noise can pass the threshold again (**E**), making a previously undetectable signal, detectable again (**F**). The optimum amount of noise depends on the momentary statistics of the input signal and is continuously adjusted via a feedback loop. This processing principle is called adaptive stochastic resonance.

system serves as the noise-generator of the stochastic resonance feedback loop.[18] It is common knowledge that—for the stochastic resonance effect—the exact spectral composition of noise is irrelevant.[92,143] This suggests spontaneously firing neurons in the somatosensory system are sufficient to trigger the stochastic resonance effect. In summary, the theoretical construct of an information transmission maximizing feedback can be mapped onto certain neuronal structures, with an architecture that is sufficient to perform the requisite calculations.

The whole stochastic resonance model is an intra-frequency channel model, which means that cross-talk between different frequency channels is not necessary to explain the emergence of tinnitus. As already described above, tinnitus is highly related to frequency channels, which are impaired by e.g. synaptopathy in the cochlea and a resulting (hidden) hearing loss.[14] Frequencies are represented tonotopical along the whole auditory pathway up to the auditory cortex.[144] Thus, it seems plausible that the amplitude of the neural noise added to each frequency channel of the DCN is tuned individually. Such a channel-wise optimization of the noise amplitude is the simplest explanation according to Occam's razor and provides a plausible explanation for the fact that the tinnitus pitch is highly correlated to impaired frequency channels.[37]

Tinnitus as a result of multiplicative central gain or additive central noise?

Central gain increase and central noise increase cannot be fully decoupled, for example, an increased excitability of neurons along the auditory pathway caused by homeostatic plasticity automatically leads to an amplification of neural noise. The fact that the additive neural noise (central noise) is amplified (central gain) along the auditory pathway is a direct consequence of the neuroanatomy of the auditory system. As the neural noise is already added in the DCN[18] being the first processing stage of the auditory pathway, multiplicative amplification (central gain) has necessarily an effect on the noise.[145]

Thus, Eq. 1 could be altered so that the amplification factor also has an effect on the central noise:

$$s = g \cdot (I + N) \quad (2)$$

As described above, the homeostatic plasticity mechanisms mediating central gain increase have been implicated in tinnitus generation,[146] however, these mechanisms are simply too slow to explain acute tinnitus phenomena after a noise trauma caused by a sudden loud stimulus.[147]

In contrast, neural circuits operating on faster time scales can explain acute tinnitus: namely, tinnitus is caused by a subcortical feedback loop adapting neural noise input into the auditory system.[35] As described above and illustrated in Fig. 2 we suggest that stochastic resonance plays a critical role in not only generating tinnitus but also restoring hearing to a certain degree.[18,34,36]

To illustrate this role, we interpret Eq. 2 in a classical signal detection task, in which the neural signal(s) has to reach a threshold for the input signal I to be detected.

In cases of hearing loss, the input I is effectively reduced. Therefore, to reach the same neural threshold, one could increase either the central noise N, or the central gain g, or both. Because increasing gain results in a squared increase in variance,[133] which increases the difficulty of signal detection, it is not the most economical means of compensating for hearing loss in cognitive neural computation (e.g. Occam's razor). Instead, it makes sense to add internal neural noise to lift weak input signals above the sensory threshold by means of stochastic resonance.[135-137] In traditional stochastic resonance, a non-linear device such as hard thresholding and periodic signals, are needed.[136] Recently, it has been shown that autocorrelation can serve as an estimator for the information content of the signal, even if it is non-periodic.[35]

The critical role of stochastic resonance is supported by broad empirical evidence: first, additional intrinsic neural noise[18,41] as well as external acoustic noise[38] can improve pure-tone hearing thresholds by ~5 dB. However, this 5 dB threshold decrease (i.e. improvement) does not explain why this mechanism is evolutionary

advantageous, as the cost of a potentially annoying and morbid tinnitus perception may be high. In a computational model, it has been shown that frequency-specific intrinsic neural noise has the potential to significantly improve speech recognition by a far larger amount (up to a factor of 2).[49] This improvement in speech comprehension and the perception of complex sounds—which could also be important for orienting animals as warning sounds—could be an explanation for the emergence of this mechanism in our auditory system during evolution. Furthermore, a significantly improved speech perception could have major positive effects and might contribute to a decreased cognitive decline in elderly people.[148,149]

In recent studies, the fact that different modalities exploit stochastic resonance to improve the signal has been proven.[40,150] It seems that stochastic resonance and especially cross-modal stochastic resonance is a universal principle of sensory processing.[36]

Second, central noise is needed to stabilize a biological system. Zeng showed that 'mathematically, the loudness at threshold is infinite when the internal noise is zero (c = 0), and vice versa. This is a fundamental argument for why the brain has or needs internal noise because infinite loudness is clearly biologically unacceptable'.[151]

Third, as described above, the central noise model based on the stochastic resonance mechanism provides a mechanistic explanation for the purpose of the somatosensory projections to auditory nuclei such as the DCN.[142,152,153] In fact, very recently, Koops and Eggermont argued that 'increased and uncorrelated noise, potentially the result from a noise source outside of the auditory pathway'[34] might play a major role in tinnitus development. Potentially, this somatosensory input is nothing else than intrinsically generated neural noise, which is modulated in the DCN to leverage stochastic resonance in the auditory system. This theory accords with the finding that tinnitus can be modulated by somatosensory input like, e.g. jaw movement.[154-156] Furthermore, tinnitus development can be prevented[157-160] or suppressed[158-160] by the presentation of external acoustic noise, which works best when the noise spectrum covers the impaired frequencies and the tinnitus pitch.[158-160] In a recent study, a novel approach was developed combining somatosensory stimulation with auditory stimulation, to modulate the tinnitus loudness.[161] Finally, it has been demonstrated that electrotactile stimulation of the fingertips enhances cochlear implant speech recognition in noise,[162] Mandarin tone recognition[163] and melody recognition.[164] While the authors did not make any mention of stochastic resonance or internal noise, it is a reasonable assertion that the observed effect might have acted via cross-modal stochastic resonance.[36]

These arguments suggest that tinnitus is indeed caused by additive neural noise (central noise) instead of a multiplicative gain. Central gain induced tinnitus would be characterized by increased evoked activity along the auditory pathway in tinnitus patients. Thus, auditory brainstem responses should have higher amplitudes in tinnitus patients compared to control patients. However, an increased evoked activity in tinnitus was refuted in several recent human patient as well as in animal studies.[165-168] Increased evoked neural activity is related to hypersensitivity against mild sounds, the so-called hyperacusis. Thus, increased central gain is potentially a better fit to explain hyperacusis rather than tinnitus.[167,168]

Hyperacusis could be one missing key to disentangle central noise and central gain adaptations of the auditory system.

As described above, central gain and central noise cannot be fully disambiguated (Eq. 2) as both auditory input and added neural noise is amplified via homeostatic plasticity along the auditory pathway. Therefore, tinnitus severity should correlate with amplification along the auditory pathway, which means that tinnitus severity should be highly correlated with the hyperacusis severity. This correlation was found in 2020 by Cederroth and coworkers.[169] In summary, the three findings that first tinnitus patients without hyperacusis show no increase in evoked activity,[165,167] second hyperacusis patients show increased evoked activity[167] and third tinnitus severity correlates with hyperacusis,[169] are a strong indication for the theory described above. To put the theory in a nutshell: central noise increase causes tinnitus, central gain increase causes hyperacusis, and central gain increase does not just cause hyperacusis but also amplifies the neural noise perceived as tinnitus.

Tinnitus and the Bayesian brain

The combined central gain and the central noise model provides a sophisticated and mathematically well developed explanation for the tinnitus-related neural hyperactivity in the brainstem. However, these theories do not explain why this hyperactivity is transmitted through the thalamus and induces a conscious experience. Indeed, there exist several mechanisms in the brain that are supposed to prevent ongoing neural activity to be transmitted to the cortex[170] and becoming a conscious and disturbing auditory percept. Up to now it is unclear why these mechanisms fail to do so. Furthermore, the stochastic resonance model does not make predictions on tinnitus heterogeneity. In particular, tinnitus is probably always caused by hearing loss, but hearing loss does not necessarily lead to tinnitus.[171] Additionally, the stochastic resonance model predicts that hearing aids should at least milden tinnitus, due to a downregulation of added neural noise in the DCN control circuit. However, while this is true for some patients, hearing aids do not milden tinnitus in all patients,[172] and this heterogeneity is not covered by the stochastic resonance model.

In the following, we provide an explanation of how and why the added central noise bypasses the filter mechanism of the brain and how this might also deliver solution approaches to the problem of tinnitus heterogeneity.

The only model with a solid mathematical background dealing with these issues is the sensory precision model from Sedley *et al.*,[52] which is based on the algorithmic formulation of predictive coding within the computational 'Bayesian brain' hypothesis.[53,65,129,173-175] Bayesian formulations of predictive processing are based on the Bayes theorem (Eq. 3)[176,177] that describes, mathematically, how to update beliefs in the light of new incoming information. Furthermore, this account also proposes a solution to other paradoxical evidence from the tinnitus literature, including that certain types of brain activity linked to perception (gamma oscillations) can show both positive and negative correlations with perceived tinnitus loudness, depending on how tinnitus loudness is manipulated.[178]

$$p(x|o) \propto p(o|x)p(x) \tag{3}$$

Here, o corresponds to observations (e.g. sensorineural responses) and x to their inferred causes (e.g. auditory loudness). In this context, the brain is continuously updating its posterior belief distribution $p(x|o)$ about actual sound intensity x, given auditory afferents or observations o. This update is achieved by combining the prior expectations $p(x)$ (Fig. 3A), descending from the higher regions of the

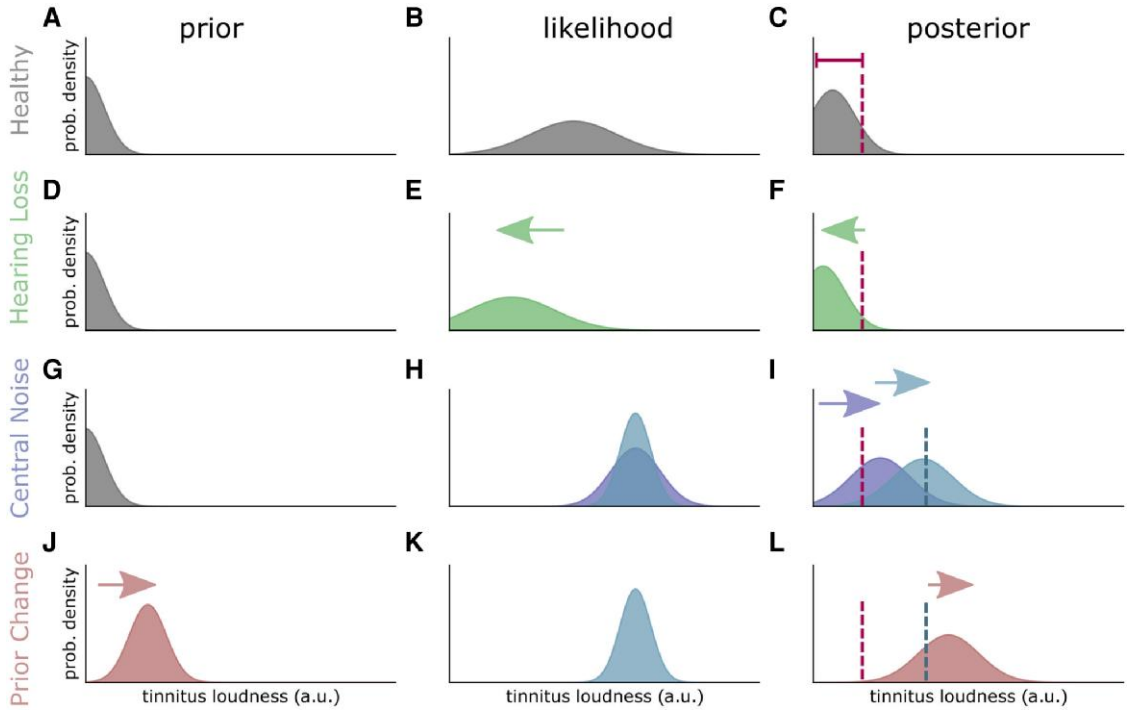

Figure 3 Predictive coding model of tinnitus induction. The posterior (represents the percept) is the product of the likelihood (bottom-up neuronal signal) and the prior (top-down prediction of the auditory input). The predictive coding hypothesis of tinnitus development is formalized in the Bayesian brain framework. The brain predicts the likelihood of the occurrence of a certain auditory input loudness [$p(x)$, x: model] in one certain frequency channel. The prior (prediction) is based on the experiences on how often certain auditory stimuli occurs and has nothing to do with the present neuronal signal coming from the cochlea. In the healthy case the prior distribution has a low mean (standard auditory input is zero, **A**). The likelihood $p(o|x)$ (**B**) represents the bottom-up signal, respectively, the measurements of the sensor (cochlea and brainstem). The posterior is the probability that under the condition of one particular neuronal signal (spike rate) a certain stimulus loudness is the cause of that neural activity. In the healthy case, the low spontaneous activity (**B**) is most probably the consequence of the absence of an auditory input. The effect that low spontaneous activity (with low precision) is assumed to be the consequence of no auditory input is called sensory attenuation (**C**, *left* side of the dashed curve). Decreased cochlear input due to hearing loss (**D**) shifts the likelihood (**E**) and consequently the posterior (**F**) to lower values, which means that a hearing loss does not directly cause tinnitus. Central noise (**G**) increases the spontaneous activity and thus increases the mean of the likelihood (**H**, dark blue distribution). The product of $p(x)$ and $p(o|x)$ is shifted to higher values (**I**, dark blue). Potentially, the precision of the likelihood is also increased (lower variance) through the central noise effect (**H**, cyan distribution), which further shifts the posterior to higher values (**I**, cyan), as the mean of the product of the probabilities is weighted with the precision. The increased neuronal activity is interpreted as auditory input, which means that there is a tinnitus percept. This effect can be amplified, as the continuous change of neural activity (through central gain and central noise) leads to continuous miss predictions. The prediction error between prior [$p(x)$] and likelihood [$p(o|x)$] is decreased by adapting the prior (**J**). Therefore, tinnitus becomes the standard prediction, which further manifests the phantom percept (**L**). The effect might be the correlate of chronic manifestation of tinnitus.

processing hierarchy, with sensory input—reporting the likelihood or sensory evidence—ascending from below [$p(o|x)$, Fig. 3B]. 'Likelihood' refers to the probability that the pattern of sensory input indicates a particular underlying sensory event or cause. In the healthy system (no hearing loss, Fig. 3A–C) the default prediction (prior, Fig. 3A) would be that there is no auditory input. In silence, the likelihood is a broad distribution with a low mean (Fig. 3B), as there is exclusively spontaneous activity. This spontaneous activity has been termed a 'tinnitus precursor', which usually has a low precision and is therefore not interpreted as auditory input. Reducing sensory precision is also called sensory attenuation. However, if the precision of the tinnitus precursor increases (or, sensory attenuation is insufficient) then the posterior shifts to the perception of a sound, and tinnitus occurs.

The occurrence of an external sound (evoked response) shifts the likelihood to higher values and the precision of the likelihood rises, as neuronal activity encoding a certain loudness level is generated, with a high precision. Therefore, the posterior belief is—although the prior predicts no input—that there is an auditory input, as precise sensory evidence shifts the posterior to higher values.

Starting from this configuration, the predictive coding model of tinnitus development can be structured in three main steps: (i) hearing loss (Fig. 3D); (ii) compensation of hearing loss through stochastic resonance and central gain; and (iii) increased precision of this spontaneous central noise (tinnitus precursor). A fourth step is thought to result in tinnitus becoming chronic, which is adjustment of auditory priors (shifting away from 'silence' as the default, to expecting a tinnitus-like sound); this allows the tinnitus precursor to be perceived even at relatively low precision levels, as it shows some concordance with auditory priors. The first step is hearing loss, which means that there is loss of precise input from the cochlea. Thus, the activity of the neurons along the auditory pathway is attenuated, which means that the likelihood becomes less precise in relation to the posterior (Fig. 3E). Thus, the posterior is shifted to lower values (Fig. 3F), and things are perceived as quieter or silent. This means that hearing loss and predictive coding alone are not sufficient to explain tinnitus. In a next step, the decreased input through hearing loss is compensated by adding neural noise by means of stochastic resonance (Fig. 3G). This means that the mean of the likelihood [$p(o|x)$ in Eq. 4, Fig. 3H, dark blue distribution] is increased as neural activity (N in Eq. 2) is added to auditory

input through the mechanism of stochastic resonance. This effect is further increased by the central gain amplification along the auditory pathway (g in Eq. 2, likelihood Fig. 3H, cyan distribution), further amplifying tinnitus loudness (posterior belief: Fig. 3I). There are good reasons to assume that the precision of the neuronal signal is increased through subcortical phenomena: as described above, internal neural noise is not comparable with the pressure fluctuations (white uncorrelated acoustic noise) used to lift subthreshold auditory signal above the detection threshold, as shown be Zeng and coworkers.[38] Thus, noise increase might entail the addition of regular spike trains. Therefore, as the stochastic resonance feedback loop optimizes for a certain noise amplitude with low variance the precision of the likelihood might be increased (note that stochastic resonance is not limited to any particular noise[92,143]). Nevertheless, it is not obvious that central noise increases the sensory precision. The addition of regular spike trains or patterns to the cochlear signal might cause the system to run in an attractor. The number of possible neural states is limited as the neural noise causes a continuous activity and makes low-activity states very unlikely. This is in line with the therapeutic approaches of Tass and Popovych,[179] who tried to get out of this neuronal attractor by presenting acoustic stimuli. An amplification through central gain in contrast to an additive noise might have the opposite effect, as a multiplicative term would increase the number of possible neural activity patterns. This fact indicates that central noise is a better complement to the predictive coding model of tinnitus development. It is an upcoming challenge and important milestone to unravel the exact neuronal patterns that fulfil the properties described above.

Besides the fact that the increased spontaneous activity through central noise and central gain mechanisms changes the likelihood, it also leads to continuous prediction errors. Therefore, the final part of the model is an update of the prior (Fig. 3J). Thus, the prior is shifted to higher input loudness values to minimize the error between likelihood and predictions (Fig. 3K). Physiologically, any accompanying increase in the precision of these priors is usually associated with an increase in the postsynaptic gain or excitability of neuronal populations reporting prediction errors (usually superficial pyramidal cells in the cortex). See Benrimoh *et al.*,[107] Bastos *et al.*,[173] Adams *et al.*,[180] Friston *et al.*,[181] Kanai *et al.*,[182] Shipp[183] and Sterzer *et al.*[184] for a predictive coding account of neuronal message passing and the role of precision weighted prediction errors in hallucinatory phenomena.

In short, the result is that the presence of auditory input becomes the new default prediction and shifts the posterior to higher values (Fig. 3L). This final step could be the correlate of tinnitus and might explain why—in some patients—the restoration of hearing through, e.g. hearing aids does not cure tinnitus.

An important question is why divergent behaviour should occur in optimally functioning systems such as those involved in stochastic resonance and predictive coding; i.e. why should similar conditions, such as hearing loss, result in accepting central noise as a percept in some cases but not others. To address this, we must consider that what is 'optimal' varies according to the hierarchical level concerned, and the situational context. With regard to hierarchical level, accepting central noise as a percept reduces prediction error at the lower hierarchical level where the noise is generated, but (at least initially) results in the introduction of a prediction error at higher hierarchical levels by introducing an unexpected percept. Thus, the balance of priority between hierarchical levels may help to explain the emergence (or non-emergence) of tinnitus in different instances. Regarding wider context, we consider here stress as one example; in certain stressful situations, one is hypervigilant to a broad range of sensory inputs, particularly those which might indicate potential threats, which can include novel or previously unanticipated ones. Such stress can be considered a relative shift of precision away from sensory priors, towards sensory likelihoods. This might explain the initial onset of tinnitus during stress, which has been reported.[185] However, once default sensory priors have adjusted to accept tinnitus, the conflict between hierarchical levels disappears, as low-level likelihood and high-level prediction become concordant.

Conclusion and outlook

In conclusion, the combination of the process theory of central noise increase and adaptive stochastic resonance—as a bottom-up mechanism—together with the computational model of predictive coding—as a complementary top-down mechanism—provides an integrated explanation of tinnitus emergence. Here, bottom-up refers to the overall information flow, i.e. modification of signals originating from lower brain structures, like the cochlear nucleus and primary auditory cortex. It is important to note that this does not imply that the predictive coding framework is solely top-down or that the stochastic resonance model is solely bottom-up.

Furthermore, the models provide a mathematical framework, which can be used to make quantitative predictions that can be tested through novel experimental paradigms, e.g. for the calculation of the autocorrelation function, specific neuronal delay-lines are needed. As the neuroanatomy along the auditory pathway is mostly known, one could calculate how long specific delay lines are and if they fit this hypothesis. Furthermore, one can look specifically for the noise generator—most probably in the somatosensory system—and characterize, e.g. the spectral composition of its output. Independent of these predictions, since both stochastic resonance and predictive coding as universal processing mechanisms are ubiquitous in the brain, we speculate that the presented integrative framework may extend to the perception of other sensory modalities and even beyond to certain aspects of cognition and behaviour in general.

A current challenge is a network theory of predictive coding, which explains how these computations are implemented in the brain.[129] Several studies have attempted to place predictive coding in the larger context of Bayesian belief updating in the brain.[173,181,183,186-188] Furthermore, to unravel the exact characteristics of the neural noise necessary to significantly decrease sensory precision, is an important challenge that needs to be addressed in future studies.

Our integrated model of auditory (phantom) perception demonstrates that the fusion of computational neuroscience, AI and experimental neuroscience leads to innovative ideas and paves the way for further advances in neuroscience and AI research. For instance, novel evaluation techniques for neurophysiological data based on AI and Bayesian statistics were recently established,[189-192] the role of noise in neural networks and other biological information processing systems was considered[193-196] and the benefit and application of noise and randomness in machine learning approaches was further investigated.[49,197,198] On the one hand, the fusion of these complementary fields may evince the neural mechanisms of tinnitus (CCN[69]) and information processing principles that underwrite functional brain architectures. On the other hand, neuroscience-inspired AI[113] may accelerate research in machine learning. We hope that the four major steps towards a CCN

of tinnitus, i.e. (i) finding an exact language; (ii) developing a mechanistic theory; (iii) testing the methods in fully specified test systems; and (iv) merging AI with computational and experimental neuroscience, will afford novel opportunities in tinnitus research.

Acknowledgement

We wish to thank Arnaud Norena for useful discussion.

Funding

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation): grant KR5148/2-1 (project number 436456810) to PK, grant KR5148/3-1 (project number 510395418) to P.K., grant GRK2839 (project number 468527017) to P.K., grant SCHI 1482/3-1 (project number 451810794) to A.S., grant TZ100/2-1 (project number 510395418) to K.T. Additionally, P.K. was supported by the Emerging Talents Initiative (ETI) of the University Erlangen-Nuremberg (grant 2019/2-Phil-01), and K.F. is supported by funding for the Wellcome Centre for Human Neuroimaging (Ref: 205103/Z/16/Z) and a Canada-UK Artificial Intelligence Initiative (Ref: ES/T01279X/1). Furthermore, the research leading to these results has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (ERC Grant No. 810316 to A.M.).

A.M.: ERC Grant No. 810316. A.S.: DFG grant SCHI 1482/3-1 (project number 451810794). K.F.: Wellcome Centre for Human Neuroimaging (Ref: 205103/Z/16/Z); Canada-UK Artificial Intelligence Initiative (Ref: ES/T01279X/1). K.T.: DFG grant TZ100/2-1 (project number 510395418). P.K.: DFG grant KR5148/2-1 (project number 436456810); DFG grant KR5148/3-1 (project number 510395418); DFG grant GRK2839 (project number 468527017); Emerging Talents Initiative (ETI) of the University Erlangen-Nürnberg (grant 2019/2-Phil-01).

Competing interests

The authors report no competing interests.

References

1. Schaette R, Kempter R. Computational models of neurophysiological correlates of tinnitus. *Front Syst Neurosci*. 2012;6:34.
2. Kaltenbach JA. The dorsal cochlear nucleus as a participant in the auditory, attentio naland emotional components of tinnitus. *Hear Res*. 2006;216:224-234.
3. Kaltenbach JA, Afman CE. Hyperactivity in the dorsal cochlear nucleus after intense sound expos ure and its resemblance to tone-evoked activity: A physiological model for tinnitus. *Hear Res*. 2000;140(1-2):165-172.
4. Kaltenbach JA, Rachel JD, Mathog TA, Zhang J, Falzarano PR, Lewandowski M. Cisplatin-induced hyperactivity in the dorsal cochlear nucleus and its relation to outer hair cell loss: Relevance to tinnitus. *J Neurophysiol*. 2002;88:699-714.
5. Moore BCJ. Perceptual consequences of cochlear hearing loss and their implications for the design of hearing aids. *Ear Hear*. 1996;17:133-161.
6. Tziridis K, Forster J, Buchheidt-Dörfler I, *et al*. Tinnitus development is associated with synaptopathy of inner hair cells in Mongolian gerbils. *Eur J Neurosci*. 2021;54:4768-4780.
7. Eggermont JJ. Hearing loss, hyperacusis, or tinnitus: What is modeled in animal research? *Hear Res*. 2013;295:140-149.
8. Jastreboff PJ, Brennan JF, Coleman JK, Sasaki CT. Phantom auditory sensation in rats: An animal model for tinnitus. *Behav Neurosci*. 1988;102:811.
9. Lobarinas E, Sun W, Cushing R, Salvi R. A novel behavioral paradigm for assessing tinnitus using schedule-induced polydipsia avoidance conditioning (SIP-AC). *Hear Res*. 2004;190(1-2):109-114.
10. Gerum R, Rahlfs H, Streb M, *et al*. Open (G) PIAS: An open-source solution for the construction of a high- precision acoustic startle response setup for tinnitus screening and threshold estimation in rodents. *Front Behav Neurosci*. 2019;13:140.
11. Schilling A, Krauss P, Gerum R, Metzner C, Tziridis K, Schulze H. A new statistical approach for the evaluation of gap-prepulse inhibiti on of the acoustic startle reflex (GPIAS) for tinnitus assessment. *Front Behav Neurosci*. 2017;11:198.
12. Turner JG, Brozoski TJ, Bauer CA, *et al*. Gap detection deficits in rats with tinnitus: A potential novel screening tool. *Behav Neurosci*. 2006;120:188.
13. Eggermont JJ, Roberts LE. Tinnitus: Animal models and findings in humans. *Cell Tissue Res*. 2015;361:311-336.
14. Dalligna C, Rosito LS, Dalligna DP, *et al*. Is there an association between tinnitus pitch and hearing loss? *Otolaryngol Head Neck Surg*. 2014;151(1_suppl):P213-P213.
15. Keppler H, Degeest S, Dhooge I. The relationship between tinnitus pitch and parameters of audiometry and distortion product otoacoustic emissions. *J Laryngol Otol*. 2017;131:1017-1025.
16. Schecklmann M, Vielsmeier V, Steffens T, Landgrebe M, Langguth B, Kleinjung T. Relationship between audiometric slope and tinnitus pitch in tinnitus patients: Insights into the mechanisms of tinnitus generation. *PLoS One*. 2012;7:e34878.
17. Yakunina N, Nam E-C. Does the tinnitus pitch correlate with the frequency of hearing loss? *Acta Otolaryngol*. 2021;141:163-170.
18. Krauss P, Tziridis K, Metzner C, Schilling A, Hoppe U, Schulze H. Stochastic resonance controlled upregulation of internal noise after hearing loss as a putative cause of tinnitus-related neuronal hyperactivity. *Front Neurosci*. 2016;10:597.
19. König O, Schaette R, Kempter R, Gross M. Course of hearing loss and occurrence of tinnitus. *Hear Res*. 2006;221(1-2):59-64.
20. Moore BCJ. The relationship between tinnitus pitch and the edge frequency of the audiogram in individuals with hearing impairment and tonal tinnitus. *Hear Res*. 2010;261(1-2):51-56.
21. Pan T, Tyler RS, Ji H, Coelho C, Gehringer AK, Gogel SA. The relationship between tinnitus pitch and the audiogram. *Int J Audiol*. 2009;48:277-294.
22. Gerken GM. Central tinnitus and lateral inhibition: An auditory brainstem model. *Hear Res*. 1996;97(1-2):75-83.
23. Kral A, Majernik V. On lateral inhibition in the auditory system. *Gen Physiol Biophys*. 1996;15:109-128.
24. Langner G, Wallhäusser-Franke E. Computer simulation of a tinnitus model based on labelling of tinnitus activity in the auditory cortex. In: *Proceedings of the Sixth International Tinnitus Seminar*. Cambridge The Tinnitus; 1999:20-25.
25. Bruce IC, Bajaj HS, Ko J. Lateral-inhibitory-network models of tinnitus. *IFAC Proc Vol*. 2003;36:359-363.
26. Dominguez M, Becker S, Bruce I, Read H. A spiking neuron model of cortical correlates of sensorineural hearing loss: Spontaneous firing, synchrony, and tinnitus. *Neural Comput*. 2006;18:2942-2958.
27. Turrigiano GG. Homeostatic plasticity in neuronal networks: The more things change, the more they stay the same. *Trends Neurosci*. 1999;22:221-227.

28. Parra LC, Pearlmutter BA. Illusory percepts from auditory adaptation. *J Acoust Soc Am*. 2007;121:1632-1641.
29. Schaette R, Kempter R. Development of tinnitus-related neuronal hyperactivity through homeostatic plasticity after hearing loss: A computational model. *Eur J Neurosci*. 2006;23: 3124-3138.
30. Schaette R, Kempter R. Development of hyperactivity after hearing loss in a computational mod el of the dorsal cochlear nucleus depends on neuron response type. *Hear Res*. 2008; 240(1-2):57-72.
31. Schaette R, Kempter R. Predicting tinnitus pitch from patients' audiograms with a computation al model for the development of neuronal hyperactivity. *J Neurophysiol*. 2009; 101:3042-3052.
32. Chrostowski M, Yang L, Wilson HR, Bruce IC, Becker S. Can homeostatic plasticity in deafferented primary auditory cortex lead to travelling waves of excitation? *J Comput Neurosci*. 2011; 30:279-299.
33. Zeng F-G. An active loudness model suggesting tinnitus as increased central noise and hyperacusis as increased nonlinear gain. *Hear Res*. 2013;295:172-179.
34. Koops EA, Eggermont JJ. The thalamus and tinnitus: Bridging the gap between animal data and findings in humans. *Hear Res*. 2021;407:108280.
35. Krauss P, Metzner C, Schilling A, *et al*. Adaptive stochastic resonance for unknown and variable input signals. *Sci Rep*. 2017;7: 2450.
36. Krauss P, Tziridis K, Schilling A, Schulze H. Cross-modal stochastic resonance as a universal principle to enhances ensory processing. *Front Neurosci*. 2018;12:578.
37. Schilling A, Tziridis K, Schulze H, Krauss P. The stochastic resonance model of auditory perception: A unified explanation of tinnitus development, Zwicker tone illusion, and residual inhibition. *Prog Brain Res*. 2021;262:139-157.
38. Zeng F-G, Fu Q-J, Morse R. Human hearing enhanced by noise. *Brain Res*. 2000;869(1-2):251-255.
39. Voros JL, Sherman SO, Rise R, *et al*. Galvanic vestibular stimulation produces cross-modal improvements in visual thresholds. *Front Neurosci*. 2021;15:640984.
40. Yashima J, Kusuno M, Sugimoto E, Sasaki H. Auditory noise improves balance control by cross-modal stochastic resonance. *Heliyon*. 2021;7:e08299.
41. Gollnast D, Tziridis K, Krauss P, Schilling A, Hoppe U, Schulze H. Analysis of audiometric differences of patients with and without tinnitus in a large clinical database. *Front Neurol*. 2017;8:31.
42. Tziridis K, Schulze H. Is phase locking crucial to improve hearing thresholds in tinnitus patients? authoreacom. 2023.
43. Zwicker E. "Negative afterimage" in hearing. *J Acoust Soc Am*. 1964;36:2413-2415.
44. Schilling A, Choi B, Parameshwarappa V, Norena AJ. Offset responses in primary auditory cortex are enhanced after notched noise stimulation. *J Neurophysiol*. 2023;129:1114-1126.
45. Schilling A, Tziridis K, Schulze H, Krauss P. Behavioral assessment of Zwicker tone percepts in gerbils. *Neuroscience*. 2023; 520:39-45.
46. Wiegrebe L, Kössl M, Schmidt S. Auditory enhancement at the absolute threshold of hearing and its relationship to the Zwicker tone. *Hear Res*. 1996;100(1-2):171-180.
47. Schilling A, Gerum R, Krauss P, Metzner C, Tziridis K, Schulze H. Objective estimation of sensory thresholds based on neurophysiological parameters. *Front Neurosci*. 2019;13:481.
48. Krauss P, Tziridis K. Simulated transient hearing loss improves auditory sensitivity. *Sci Rep*. 2021;11:14791.
49. Schilling A, Gerum R, Metzner C, Maier A, Krauss P. Intrinsic noise improves speech recognition in a computational model of the auditory pathway. *Front Neurosci*. 2022;16:908330.
50. Haro S, Smalt CJ, Ciccarelli GA, Quatieri TF. Deep neural network model of hearing-impaired speech-in-noise perception. *Front Neurosci*. 2020;14:588448.
51. Sedley W, Alter K, Gander PE, Berger J, Griffiths TD. Exposing pathological sensory predictions in tinnitus using auditory intensity deviant evoked responses. *J Neurosci*. 2019;39: 10096-10103.
52. Sedley W, Friston KJ, Gander PE, Kumar S, Griffiths TD. An integrative tinnitus model based on sensory precision. *Trends Neurosci*. 2016;39:799-812.
53. Friston K. The free-energy principle: A unified brain theory? *Nat Rev Neurosci*. 2010;11:127-138.
54. Friston K. Does predictive coding have a future? *Nat Neurosci*. 2018;21:1019-1021.
55. Hu S, Hall DA, Zubler F, *et al*. Bayesian Brain in tinnitus: Computational modeling of three perceptual phenomena using a modified Hierarchical Gaussian Filter. *Hear Res*. 2021; 410:108338.
56. Dotan A, Shriki O. Tinnitus-like "hallucinations" elicited by sensory deprivation in an entropy maximization recurrent neural network. *PLoS Comput Biol*. 2021;17:e1008664.
57. Gault R, McGinnity TM, Coleman S. Perceptual modeling of tinnitus pitch and loudness. *IEEE Trans Cogn Dev Syst*. 2020;12: 332-343.
58. De Ridder D, Vanneste S, Langguth B, Llinas R. Thalamocortical dysrhythmia: A theoretical update in tinnitus. *Front Neurol*. 2015;6:124.
59. Llinás RR, Ribary U, Jeanmonod D, Kronberg E, Mitra PP. Thalamocortical dysrhythmia: A neurological and neuropsychiatric syndrome characterized by magnetoencephalography. *Proc Natl Acad Sci*. 1999;96:15222-15227.
60. Jeanmonod D, Magnin M, Morel A. Low–threshold calcium spike bursts in the human thalamus: Common physiopathology for sensory, motor and limbic positive symptoms. *Brain*. 1996;119:363-375.
61. Knipper M, Van Dijk P, Schulze H, *et al*. The neural bases of tinnitus: Lessons from deafness and cochlear implants. *J Neurosci*. 2020;40:7190-7202.
62. Rauschecker JP, May ES, Maudoux A, Ploner M. Frontostriatal gating of tinnitus and chronic pain. *Trends Cogn Sci*. 2015;19: 567-578.
63. De Ridder D, Elgoyhen AB, Romo R, Langguth B. Phantom percepts: Tinnitus and pain as persisting aversive memory networks. *Proc Natl Acad Sci*. 2011;108:8075-8080.
64. Vanneste S, De Ridder D. The auditory and non-auditory brain areas involved in tinnitus. An emergent property of multiple parallel overlapping subnetworks. *Front Syst Neurosci*. 2012;6:31.
65. De Ridder D, Vanneste S, Freeman W. The Bayesian brain: Phantom percepts resolve sensory uncertainty. *Neurosci Biobehav Rev*. 2014;44:4-15.
66. Popper KR. Science as falsification. *Conjectures Refutations*. 1963;1:33-39.
67. Lazebnik Y. Can a biologist fix a radio?—Or, what I learned while studying apoptosis. *Cancer Cell*. 2002;2:179-182.
68. Lazar N. Ockham's razor. *Wiley Interdiscip Rev Comput Stat*. 2010; 2:243-246.
69. Kriegeskorte N, Douglas PK. Cognitive computational neuroscience. *Nat Neurosci*. 2018;21:1148-1160.
70. Naselaris T, Bassett DS, Fletcher AK, *et al*. Cognitive computational neuroscience: A new conference for an emerging discipline. *Trends Cogn Sci*. 2018;22:365-367.

71. Marr D, Poggio T. A computational theory of human stereo vision. *Proc R Soc Lond Biol Sci*. 1979;204:301-328.
72. Holt N, Bremner A, Sutherland E, Vliek M, Passer M, Smith R. *EBOOK: Psychology: The science of mind and behaviour*. 4th ed. McGraw Hill; 2019.
73. Roberts LE. Neural plasticity and its initiating conditions in tinnitus. *HNO*. 2018;66:172-178.
74. Brown JW. The tale of the neuroscientists and the computer: Why mechanistic theory matters. *Front Neurosci*. 2014;8:349.
75. Knipper M, Mazurek B, Dijk P, Schulze H. Too blind to see the elephant? Why neuroscientists ought to be interested in tinnitus. *J Assoc Res Otolaryngol*. 2021;22:609-621.
76. Silver R, Boahen K, Grillner S, Kopell N, Olsen KL. Neurotech for neuroscience: Unifying concepts, organizing principles, and emerging tools. *J Neurosci*. 2007;27:11807-11819.
77. Jonas E, Kording KP. Could a neuroscientist understand a microprocessor? *PLoS Comput Biol*. 2017;13:e1005268.
78. Bennett CM, Miller MB, Wolford GL. Neural correlates of interspecies perspective taking in the post-mortem atlantic salmon: An argument for multiple comparisons correction. *Neuroimage*. 2009;47(Suppl 1):S125.
79. Bennett CM, Wolford GL, Miller MB. The principled control of false positives in neuroimaging. *Soc Cogn Affect Neurosci*. 2009; 4:417-422.
80. Gerum RC, Schilling A. Integration of leaky-integrate-and-fire neurons in standard machine learning architectures to generate hybrid networks: A surrogate gradient approach. *Neural Comput*. 2021;33:2827-2852.
81. LeCun Y, Jackel LD, Bottou L, *et al*. Learning algorithms for classification: A comparison on handwritten digit recognition. *Neural Netw Stat Mech Perspect*. 1995;261:2.
82. Schilling A, Maier A, Gerum R, Metzner C, Krauss P. Quantifying the separability of data classes in neural networks. *Neural Netw*. 2021;139:278-293.
83. Zenke F, Vogels TP. The remarkable robustness of surrogate gradient learning for instilling complex function in spiking neural networks. *Neural Computat*. 2021;33:899-925.
84. Krauss P, Metzner C, Lange J, Lang N, Fabry B. Parameter-free binarization and skeletonization of fiber networks from confocal image stacks. *PLoS One*. 2012;7:e36575.
85. Gerstner W, Sprekeler H, Deco G. Theory and simulation in neuroscience. *Science*. 2012;338:60-65.
86. Schilling A, Tomasello R, Henningsen-Schomers MR, *et al*. Analysis of continuous neuronal activity evoked by natural speech with computational corpus linguistics methods. *Lang Cogn Neurosci*. 2021;36:167-186.
87. Kerr NL. HARKing: Hypothesizing after the results are known. *Pers Soc Psychol Rev*. 1998;2:196-217.
88. Munafò MR, Nosek BA, Bishop DVM, *et al*. A manifesto for reproducible science. *Nat Hum Behav*. 2017;1:1-9.
89. Lewin K. Field theory in social science: Selected theoretical papers (Edited by Dorwin Cartwright.). 1951
90. Newell A. You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. 1973
91. Peters JM. A cognitive computational model of risk hypothesis generation. *J Account Res*. 1990;28:83-103.
92. Nozaki D, Mar DJ, Grigg P, Collins JJ. Effects of colored noise on stochastic resonance in sensory neurons. *Phys Rev Lett*. 1999; 82:2402.
93. Wang Z, She Q, Smeaton AF, Ward TE, Healy G. Synthetic-Neuroscore: Using a neuro-AI interface for evaluating generative adversarial networks. *Neurocomputing*. 2020;405: 26-36.
94. Zador A, Escola S, Richards B, *et al*. Catalyzing next-generation Artificial Intelligence through NeuroAI. *Nat Commun*. 2023; 14:1597.
95. Hodgkin AL, Huxley AF. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J Physiol*. 1952;117:500.
96. Walter WG. An imitation of life. *Sci Am*. 1950;182:42-45.
97. Holland O. The first biologically inspired robots. *Robotica*. 2003; 21:351-363.
98. Schlimm D. Learning from the existence of models: On psychic machines, tortoises, and computer simulations. *Synthese*. 2009;169:521-538.
99. Braitenberg V. *Vehicles: Experiments in synthetic psychology*: MIT Press; 1986.
100. Barak O. Recurrent neural networks as versatile tools of neuroscience research. *Curr Opin Neurobiol*. 2017;46:1-6.
101. Marblestone AH, Wayne G, Kording KP. Toward an integration of deep learning and neuroscience. *Front Comput Neurosci*. 2016; 10:94.
102. Van Gerven M. Computational foundations of natural intelligence. *Front Comput Neurosci*. 2017;11:112.
103. Van Gerven M, Bohte S. Artificial neural networks as models of neural information processing. *Front Comput Neurosci*. 2017; 11:114.
104. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521: 436-444.
105. Schmidhuber J. Deep learning in neural networks: An overview. *Neural Netw*. 2015;61:85-117.
106. Arnal LH, Giraud A-L. Cortical oscillations and sensory predictions. *Trends Cogn Sci*. 2012;16:390-398.
107. Benrimoh D, Parr T, Vincent P, Adams RA, Friston K. Active inference and auditory hallucinations. *Comput Psychiatr*. 2020; 2:183.
108. Friston KJ, Sajid N, Quiroga-Martinez DR, Parr T, Price CJ, Holmes E. Active listening. *Hear Res*. 2021;399:107998.
109. Hovsepyan S, Olasagasti I, Giraud A-L. Combining predictive coding with neural oscillations optimizes on-line speech processing. bioRxiv. 2018:477588.
110. Isomura T, Parr T, Friston K. Bayesian Filtering with multiple internal models: Toward a theory of social intelligence. *Neural Comput*. 2019;31:2390-2431.
111. Koelsch S, Vuust P, Friston K. Predictive processes and the peculiar case of music. *Trends Cogn Sci*. 2019;23:63-77.
112. Powers AR, Mathys C, Corlett PR. Pavlovian conditioning–induced hallucinations result from overweighting of perceptual priors. *Science*. 2017;357:596-600.
113. Hassabis D, Kumaran D, Summerfield C, Botvinick M. Neuroscience-inspired artificial intelligence. *Neuron*. 2017;95: 245-258.
114. Cambria E, White B. Jumping NLP curves: A review of natural language processing research. *IEEE Comput Intell Mag*. 2014;9: 48-57.
115. Rahwan I, Cebrian M, Obradovich N, *et al*. Machine behaviour. *Nature*. 2019;568:477-486.
116. Hutson M. *Artificial intelligence faces reproducibility crisis*. vol 359. American Association for the Advancement of Science; 2018.
117. Voosen P. *The AI detectives*. vol 357. American Association for the Advancement of Science; 2017.
118. Bermudez-Contreras E, Clark BJ, Wilber A. The neuroscience of spatial navigation and the relationship to artificial intelligence. *Front Comput Neurosci*. 2020;14:63.
119. McNamee DC, Stachenfeld KL, Botvinick MM, Gershman SJ. Flexible modulation of sequence generation in the entorhinal–hippocampal system. *Nat Neurosci*. 2021;24:851-862.

120. Stachenfeld KL, Botvinick MM, Gershman SJ. The hippocampus as a predictive map. *Nat Neurosci*. 2017;20:1643-1653.
121. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput*. 1997;9:1735-1780.
122. Hochreiter S, Schmidhuber J. LSTM Can solve hard long time lag problems. *Adv Neural Inf Process Syst*. 1997:473-479.
123. Banino A, Barry C, Uria B, *et al*. Vector-based navigation using grid-like representations in artificial agents. *Nature*. 2018;557:429-433.
124. Gerum RC, Erpenbeck A, Krauss P, Schilling A. Sparsity through evolutionary pruning prevents neuronal networks from overfitting. *Neural Netw*. 2020;128:305-312.
125. Eggermont JJ. Central tinnitus. *Auris Nasus Larynx*. 2003;30:7-12.
126. Caspary DM, Hughes LF, Schatteman TA, Turner JG. Age-related changes in the response properties of cartwheel cells in rat dorsal cochlear nucleus. *Hear Res*. 2006;216:207-215.
127. Golding NL, Oertel D. Physiological identification of the targets of cartwheel cells in the dorsal cochlear nucleus. *J Neurophysiol*. 1997;78:248-260.
128. Roberts MT, Trussell LO. Molecular layer inhibitory interneurons provide feedforward and lateral inhibition in the dorsal cochlear nucleus. *J Neurophysiol*. 2010;104:2462-2473.
129. Friston K. The history of the future of the Bayesian brain. *NeuroImage*. 2012;62:1230-1233.
130. Hawkins J, Blakeslee S. *On intelligence*. Times Books; 2004.
131. Kadmon J, Timcheck J, Ganguli S. Predictive coding in balanced neural networks with noise, chaos and delays. *Adv Neural Inf Process Syst*. 2020;33:16677-16688.
132. Choksi B, Mozafari M, Biggs O'May C, Ador B, Alamia A, VanRullen R. Predify: Augmenting deep neural networks with brain-inspired predictive coding dynamics. *Adv Neural Inf Process Syst*. 2021;34:14069-14083.
133. Zeng F-G. Tinnitus and hyperacusis: Central noise, gain and variance. *Curr Opin Physiol*. 2020;18:123-129.
134. Auerbach BD, Rodrigues PV, Salvi RJ. Central gain control in tinnitus and hyperacusis. *Front Neurol*. 2014;5:206.
135. Benzi R, Sutera A, Vulpiani A. The mechanism of stochastic resonance. *J Phys A Math Gen*. 1981;14:L453.
136. Gammaitoni L, Hänggi P, Jung P, Marchesoni F. Stochastic resonance. *Rev Mod Phys*. 1998;70:223.
137. McDonnell MD, Abbott D. What is stochastic resonance? Definitions, misconceptions, debates, and its relevance to biology. *PLoS Comput Biol*. 2009;5:e1000348.
138. Nelken I, Young ED. Why do cats need a dorsal cochlear nucleus? *J Basic Clin Physiol Pharmacol*. 1996;7:199-220.
139. Cariani PA. Neural timing nets. *Neural Netw*. 2001;14:737-753.
140. Cariani PA. Temporal codes and computations for sensory representation and scene analysis. *IEEE Trans Neural Netw*. 2004;15:1100-1111.
141. Wu C, Martel DT, Shore SE. Increased synchrony and bursting of dorsal cochlear nucleus fusiform cells correlate with tinnitus. *J Neurosci*. 2016;36:2068-2073.
142. Shore SE, Zhou J. Somatosensory influence on the cochlear nucleus and beyond. *Hear Res*. 2006;216:90-99.
143. Gingl Z, Kiss L, Moss F. Non-dynamical stochastic resonance: Theory and experiments with white and arbitrarily coloured noise. *Europhys Lett*. 1995;29:191.
144. Kandler K, Clause A, Noh J. Tonotopic reorganization of developing auditory brainstem circuits. *Nat Neurosci*. 2009;12:711-717.
145. Vale C, Sanes DH. The effect of bilateral deafness on excitatory and inhibitory synaptic strength in the inferior colliculus. *Eur J Neurosci*. 2002;16:2394-2404.
146. Yang S, Weiner BD, Zhang LS, Cho S-J, Bao S. Homeostatic plasticity drives tinnitus perception in an animal model. *Proc Natl Acad Sci*. 2011;108:14974-14979.
147. Axelsson A, Hamernik RP. Acute acoustic trauma. *Acta Otolaryngol*. 1987;104(3-4):225-233.
148. Schilling A, Krauss P. Tinnitus is associated with improved cognitive performance and speech perception–can stochastic resonance explain? *Front Aging Neurosci*. 2022;14:1073149.
149. Hamza Y, Zeng F-G. Tinnitus is associated with improved cognitive performance in non-hispanic elderly with hearing loss. *Front Neurosci*. 2021;15:735950.
150. Plater EB, Seto VS, Peters RM, Bent L. Remote subthreshold stimulation enhances skin sensitivity in the lower extremity. *Front Hum Neurosci*. 2021;15:789271.
151. Zeng F-G. A unified theory of psychophysical laws in auditory intensity perception. *Front Psychol*. 2020;11:1459.
152. Dehmel S, Pradhan S, Koehler S, Bledsoe S, Shore S. Noise overexposure alters long-term somatosensory-auditory processing in the dorsal cochlear nucleus—Possible basis for tinnitus-related hyp eractivity? *J Neurosci*. 2012;32:1660-1671.
153. Wu C, Stefanescu RA, Martel DT, Shore SE. Tinnitus: Maladaptive auditory–somatosensory plasticity. *Hear Res*. 2016;334:20-29.
154. Lanting CP, De Kleine E, Eppinga RN, Van Dijk P. Neural correlates of human somatosensory integration in tinnitus. *Hear Res*. 2010;267(1-2):78-88.
155. Pinchoff RJ, Burkard RF, Salvi RJ, Coad ML, Lockwood AH. Modulation of tinnitus by voluntary jaw movements. *Am J Otol*. 1998;19:785-789.
156. Won JY, Yoo S, Lee SK, *et al*. Prevalence and factors associated with neck and jaw muscle modulation of tinnitus. *Audiol Neurotol*. 2013;18:261-273.
157. Sturm JJ, Zhang-Hooks Y-X, Roos H, Nguyen T, Kandler K. Noise trauma-induced behavioral gap detection deficits correlate with reorganization of excitatory and inhibitory local circuits in the inferior colliculus and are prevented by acoustic enrichment. *J Neurosci*. 2017;37:6314-6330.
158. Schaette R, König O, Hornig D, Gross M, Kempter R. Acoustic stimulation treatments against tinnitus could be most effective when tinnitus pitch is within the stimulated frequency range. *Hear Res*. 2010;269(1-2):95-101.
159. Schilling A, Krauss P, Hannemann R, Schulze H, Tziridis K. Reduktion der Tinnituslautstärke : Pilotstudie zur Abschwächung von tonalem Tinnitus mit schwellennahem, individuell spektral optimiertem Rauschen [Reducing tinnitus intensity: Pilot study to attenuate tonal tinnitus using individually spectrally optimized near-threshold noise]. *HNO*. 2021;69:891-898.
160. Tziridis K, Brunner S, Schilling A, Krauss P, Schulze H. Spectrally matched near-threshold noise for subjective tinnitus loudness attenuation based on stochastic resonance. *Front Neurosci*. 2022;16:831581.
161. Conlon B, Langguth B, Hamilton C, *et al*. Bimodal neuromodulation combining sound and tongue stimulation reduces tinnitus symptoms in a large randomized clinical study. *Sci Transl Med*. 2020;12:eabb2830.
162. Huang J, Sheffield B, Lin P, Zeng F-G. Electro-tactile stimulation enhances cochlear implant speech recognition in noise. *Sci Rep*. 2017;7:1-5.
163. Huang J, Chang J, Zeng F-G. Electro-tactile stimulation (ETS) enhances cochlear-implant Mandarin t one recognition. *World J Otorhinolaryngol Head Neck Surg*. 2017;3:219-223.
164. Huang J, Lu T, Sheffield B, Zeng F-G. Electro-tactile stimulation enhances cochlear-implant melody recogniti on: Effects of rhythm and musical training. *Ear Hear*. 2020;41:106-113.

165. Hofmeier B, Wertz J, Refat F, *et al*. Functional biomarkers that distinguish between tinnitus with and without hyperacusis. *Clin Transl Med*. 2021;11:e378.
166. Möhrle D, Hofmeier B, Amend M, *et al*. Enhanced central neural gain compensates acoustic trauma-induced cochlear impairment, but unlikely correlates with tinnitus and hyperacusis. *Neuroscience*. 2019;407:146-169.
167. Koops E. *Neuroimaging correlates of hearing loss, tinnitus, and hyperacusis*. University of Groningen; 2021.
168. Koops EA. A closer fit to hyperacusis than to tinnitus? 2023
169. Cederroth CR, Lugo A, Edvall NK, *et al*. Association between hyperacusis and tinnitus. *J Clin Med*. 2020;9:2412.
170. McCormick DA, Bal T. Sensory gating mechanisms of the thalamus. *Curr Opin Neurobiol*. 1994;4:550-556.
171. Tan CM, Lecluyse W, McFerran D, Meddis R. Tinnitus and patterns of hearing loss. *J Assoc Res Otolaryngol*. 2013;14:275-282.
172. Shekhawat GS, Searchfield GD, Stinear CM. Role of hearing aids in tinnitus intervention: A scoping review. *J Am Acad Audiol*. 2020;24:747-762.
173. Bastos AM, Usrey WM, Adams RA, Mangun GR, Fries P, Friston KJ. Canonical microcircuits for predictive coding. *Neuron*. 2012; 76:695-711.
174. Clark A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav Brain Sci*. 2013;36:181-204.
175. Knill DC, Pouget A. The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends Neurosci*. 2004;27:712-719.
176. Stigler SM. The true title of Bayes's essay. *Stat Sci*. 2013;28:283-288.
177. Vilares I, Kording K. Bayesian Models: The structure of the world, uncertainty, behavior, and the brain. *Ann N Y Acad Sci*. 2011;1224:22.
178. Sedley W, Teki S, Kumar S, Barnes GR, Bamiou D-E, Griffiths TD. Single-subject oscillatory gamma responses in tinnitus. *Brain*. 2012;135:3089-3100.
179. Tass PA, Popovych OV. Unlearning tinnitus-related cerebral synchrony with acoustic coordinated reset stimulation: Theoretical concept and modelling. *Biol Cybern*. 2012;106:27-36.
180. Adams RA, Stephan KE, Brown HR, Frith CD, Friston KJ. The computational anatomy of psychosis. *Front Psychiatry*. 2013;4:47.
181. Friston KJ, Parr T, Vries B. The graphical brain: Belief propagation and active inference. *Netw Neurosci*. 2017;1:381-414.
182. Kanai R, Komura Y, Shipp S, Friston K. Cerebral hierarchies: Predictive processing, precision and the pulvinar. *Philos Trans R Soc B Biol Sci*. 2015;370:20140169.
183. Shipp S. Neural elements for predictive coding. *Front Psychol*. 2016;7:1792.
184. Sterzer P, Adams RA, Fletcher P, *et al*. The predictive coding account of psychosis. *Biol Psychiatry*. 2018;84:634-643.
185. Mazurek B, Boecking B, Brueggemann P. Association between stress and tinnitus—New aspects. *Otol Neurotol*. 2019;40: e467-e473.
186. Adams RA, Shipp S, Friston KJ. Predictions not commands: Active inference in the motor system. *Brain Struct Funct*. 2013; 218:611-643.
187. Da Costa L, Parr T, Sengupta B, Friston K. Neural dynamics under active inference: Plausibility and efficiency of information processing. *Entropy*. 2021;23:454.
188. Friston K, FitzGerald T, Rigoli F, Schwartenbeck P, Pezzulo G. Active inference: A process theory. *Neural Comput*. 2017;29: 1-49.
189. Krauss P, Metzner C, Joshi N, *et al*. Analysis and visualization of sleep stages based on deep neural networks. *Neurobiol Sleep Circadian Rhythms*. 2021;10:100064.
190. Krauss P, Metzner C, Schilling A, *et al*. A statistical method for analyzing and comparing spatiotemporal cortic al activation patterns. *Sci Rep*. 2018;8:1-9.
191. Krauss P, Schilling A, Bauer J, *et al*. Analysis of multichannel EEG patterns during human sleep: A novel approach. *Front Hum Neurosci*. 2018;12:121.
192. Metzner C, Schilling A, Traxdorf M, Schulze H, Krauss P. Sleep as a random walk: A super-statistical analysis of EEG data across sleep stages. *Commun Biol*. 2021;4:1-11.
193. Bönsel F, Krauss P, Metzner C, Yamakou ME. Control of noise-induced coherent oscillations in three-neuron motifs. *Cogn Neurodyn*. 2022;16:941-960.
194. Krauss P, Prebeck K, Schilling A, Metzner C. "Recurrence resonance" in three-neuron motifs. *Front Comput Neurosci*. 2019;13:64.
195. Krauss P, Schulze H, Metzner C. A chemical reaction network to generate random, power-law-distributed time intervals. *Artif Life*. 2017;23:518-527.
196. Metzner C, Krauss P. Dynamical phases and resonance phenomena in information-processing recurrent neural networks. arXiv preprint arXiv:210802545. 2021.
197. Harikrishnan NB, Nagaraj N. When noise meets chaos: Stochastic resonance in neurochaos learning. *Neural Netw*. 2021;143:425-435.
198. Yang Z, Schilling A, Maier A, Krauss P. Neural networks with fixed binary random projections improve accuracy in classifying noisy data. In: *Bildverarbeitung für die Medizin 2021*. Springer; 2021:211-216.

Neurobiology of Sleep and Circadian Rhythms 10 (2021) 100064

Contents lists available at ScienceDirect

Neurobiology of Sleep and Circadian Rhythms

journal homepage: www.elsevier.com/locate/

Analysis and visualization of sleep stages based on deep neural networks

Patrick Krauss [a,b,c,*], Claus Metzner [a,d], Nidhi Joshi [a], Holger Schulze [a], Maximilian Traxdorf [e], Andreas Maier [f], Achim Schilling [a,b,g]

[a] Neuroscience Lab, Experimental Otolaryngology, University Hospital Erlangen, Germany
[b] Cognitive Computational Neuroscience Group at the Chair of English Philology and Linguistics, Friedrich-Alexander University Erlangen-Nürnberg (FAU), Germany
[c] Cognitive Neuroscience Center, University of Groningen, the Netherlands
[d] Biophysics, Friedrich-Alexander University Erlangen-Nürnberg (FAU), Germany
[e] Department of Otolaryngology, Head and Neck Surgery, University Hospital Erlangen, Germany
[f] Machine Intelligence, Friedrich-Alexander University Erlangen-Nürnberg (FAU), Germany
[g] Laboratory of Sensory and Cognitive Neuroscience, Aix-Marseille University, Marseille, France

ARTICLE INFO

Keywords:
Sleep stage scoring
Hypnodensity graphs
Multidimensional scaling (MDS)
Electroencephalography (EEG)
Artificial neural networks
Deep learning
Polysomnography (PSG)
Sleep cycle analysis

ABSTRACT

Automatic sleep stage scoring based on deep neural networks has come into focus of sleep researchers and physicians, as a reliable method able to objectively classify sleep stages would save human resources and simplify clinical routines. Due to novel open-source software libraries for machine learning, in combination with enormous recent progress in hardware development, a paradigm shift in the field of sleep research towards automatic diagnostics might be imminent. We argue that modern machine learning techniques are not just a tool to perform automatic sleep stage classification, but are also a creative approach to find hidden properties of sleep physiology. We have already developed and established algorithms to visualize and cluster EEG data, facilitating first assessments on sleep health in terms of sleep-apnea and consequently reduced daytime vigilance. In the following study, we further analyze cortical activity during sleep by determining the probabilities of momentary sleep stages, represented as hypnodensity graphs and then computing vectorial cross-correlations of different EEG channels. We can show that this measure serves to estimate the period length of sleep cycles and thus can help to find disturbances due to pathological conditions.

1. Introduction

Sleep stage scoring is a standard procedure and part of every polysomnographic analysis (Bradley and Peterson, 2008; Burns et al., 2008). Up to now, sleep stage scoring based on physiological signals (EEG: electroencephalography, EMG: electromyography, EOG: electrooculugraphy) is performed by experienced clinicians, which do the classification by hand according to the AASM guidelines (Berry et al., 2012). However, this procedure is time consuming and highly prone to errors resulting in high inter-rater variability (Danker-Hopfe et al., 2004, 2009; Wendt et al., 2015). To overcome these limitations, machine learning algorithms were applied. These algorithms are still not fully trusted by clinicians as they act as a black-box, from which no one knows what the internal criteria for the different sleep stages really are. Thus, one core-problem of modern machine learning research has met clinical routines, the so called black-box problem (KrauseAdam and Ng, 2016) which in other scientific fields is also called the opacity debate (De Laat, 2018). This problem can be tackled by simply using hand-crafted features such as time tags of K-complexes or sleep spindles as neural network input (Phan et al., 2018; Boostani et al., 2017), (for an example see (PhanQuan et al., 2013)). However, this procedure is very elaborate and the results are still not satisfying. In times of rising computing power and increasing storage capacities, a Big Data approach where neural networks are finding the best features in a self-organized way seems to be more promising. In order to analyze these features and the emerging internal representations, novel concepts for data visualization are needed. A sophisticated dimensionality reduction method for visualising high dimensional data is a demanding task, as certain popular methods can lead to pseudo clustering of data points caused by the initial conditions and the used hyper-parameters of the projection algorithm. A famous example is t-distributed stochastic neighbor embedding (t-SNE (van der Maaten and Hinton, 2008),), which is highly unstable and extremely dependent on the initialization and choice of parameters (Martin et al., 2016).

* Corresponding author. Neuroscience Lab, Experimental Otolaryngology, University Hospital Erlangen, Germany.
E-mail address: patrick.krauss@uk-erlangen.de (P. Krauss).

https://doi.org/10.1016/j.nbscr.2021.100064
Received 16 July 2020; Received in revised form 27 February 2021; Accepted 1 March 2021
Available online 12 March 2021

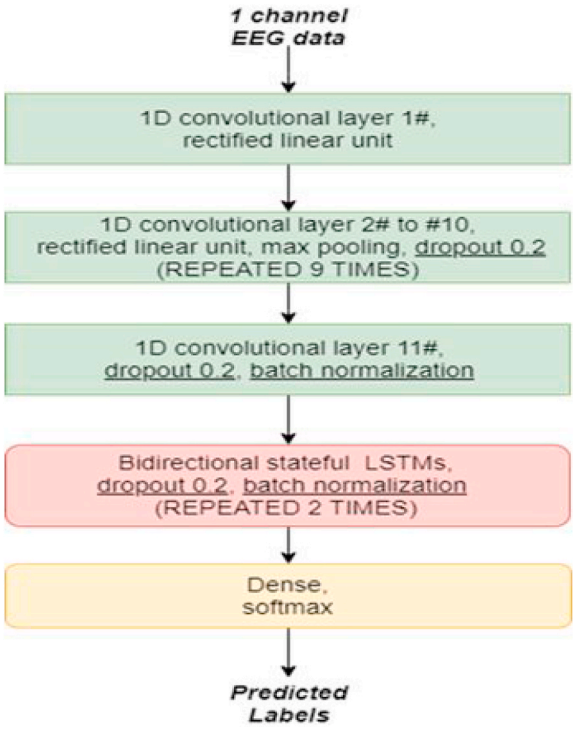

Fig. 1. Network architecture.
Building blocks of the deep neural network trained on sleep stage classification of EEG data. The network consists of eleven 1D convolutional layers, nine max pooling layers, 2 layers of bidirectional stateful LSTMs, and a fully connected classification layer with softmax output.

In previous studies, we developed several approaches to statistically analyze and visualize high-dimensional neural data (Krauss et al., 2018a; Schilling et al., 2018). We developed a statistical method for analyzing and comparing high-dimensional spatio-temporal cortical activation patterns for different auditory and somatosensory stimulus conditions in rodents and humans (Krauss et al., 2018a). The cortical activity patters were represented by amplitude vectors calculated via a sliding window method (for the exact procedure see (Krauss et al., 2018a)). We could already demonstrate that this method can discriminate different sleep stages in human EEG recordings (Krauss et al., 2018b). Furthermore, we could analyze the microstructure of cortical activity during sleep and found that it reflects respiratory events and the state of daytime vigilance (Traxdorf et al., 2019). Recently, our method has been generalized, and can now be used to analyze and compare representations of artificial neural networks (Schilling et al., 2018). In the following study, we first illustrate how to visualize the representations of EEG data gained from different layers of artificial neural networks (sleep stage embeddings). Furthermore, we demonstrate that these complex representations cluster better in higher layers of the artificial neural networks, quantified by the generalized discrimination value (GDV, see also (Schilling et al., 2018)). Subsequently, we visualize the output of the last layer of the neural network, and thus the momentary probabilities of the predicted sleep stages, as so called hypnodensity graphs (as introduced by (Jens et al., 2018)). Finally, we use these probability vectors to calculate vectorial cross-correlations, in order to analyze the period length of sleep cycles.

2. Methods

2.1. Data base

The study was conducted in the Department of Otorhinolaryngology, Head Neck Surgery, of the Friedrich-Alexander University Erlangen-Nürnberg (FAU), following approval by the local Ethics Committee (323–16 Bc). All 68 participating subjects, 46 male and 22 female, mean age 32.5 ± 11.5 years, were recruited by the Department of Otorhinolaryngology, Head and Neck Surgery. Written informed consent was obtained from the participants before the cardiorespiratory polysomnography (PSG). Inclusion criterion for this study was age between 21 and 80 years. Exclusion criteria were a positive history of misuse of sedatives, alcohol or addictive drugs and untreated sleep disorders. Data analysis was carried out during time in bed of the subjects, accumulating to a total recording time of approximately 510 h.

Each of the 68 data sets comprising a full night of sleep consisted of 3 channels of EEG recordings (F4-M1, C4-M1, O2-M1) and the corresponding sleep stages. Sleep stages were analyzed and scored visually in 30 s epochs according to the AASM criteria (Version 2.1, 2014) by a sleep specialist accredited by the German Sleep Society (DGSM) (ConradIber, 2007; Berry et al., 2012). Thereby, typical artifacts (Tatum et al., 2011) have been marked and removed subsequently for further analysis and processing steps.

2.2. Sleep stage classification with neural networks

A deep neural network was trained on single channel sleep EEG data. Therefore, we used the total number of 68 data sets and split them randomly in 54 training data sets (80%) and 14 test data sets (20%), as this is the standard approach in machine learning and pattern recognition (Bishop, 2006; Goodfellow et al., 2016). For each data set, the three different EEG channels were concatenated, yielding a corresponding single channel data set with three times the duration of the original data set. As labels for supervised training and as ground truth for the test data set we used the sleep stages from visual analysis and scoring.

The network consisting of several convolutional layers (LeCun Bengioet al., 1995) and two bidirectional stateful LSTM layers (Hochreiter and Schmidhuber, 1997) was trained with error back propagation to classify the sleep stages (for exact network architecture see Fig. 1). The convolutional layers allow to extract relevant features for classification in an unsupervised way (LeCun et al., 2015), i.e. without the need for manually extracting features like frequency power spectra or amplitudes. Recurrent network layers like the here used LSTM layers allow to extract and capture temporal information [] from the sequences of sleep stages.

2.3. Software resources

The software is written using the programming language python 3.6, together with the KERAS AI library (Chollet, 2018) and tensorFlow backend. The data is pre-processed using Numpy (Stéfan van der Walt and Varoquaux, 2011). For multi-dimension-scaling (MDS) the scikit-learn library (Pedregosa et al., 2011) is used and the visualization of the results is done using the matplotlib library (Hunter, 2007).

2.4. Hypnodensity correlations

The hypnodensity cross-correlation for lag-time τ is defined as

$$r_{\mathbf{x},\mathbf{y}}(\tau) = \frac{\frac{1}{T-\tau}\sum_{t=1}^{T-\tau}(\mathbf{x}_t - \overline{\mathbf{x}})\circ(\mathbf{y}_{t+\tau} - \overline{\mathbf{y}})}{\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left|\left|\mathbf{x}_t - \overline{\mathbf{x}}\right|\right|^2}\cdot\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left|\left|\mathbf{y}_t - \overline{\mathbf{y}}\right|\right|^2}} \quad (1)$$

and the hypnodensity auto-correlation for lag-time τ is defined as

$$r_{\mathbf{x},\mathbf{x}}(\tau) = \frac{\frac{1}{T-\tau}\sum_{t=1}^{T-\tau}(\mathbf{x}_t - \overline{\mathbf{x}})\circ(\mathbf{x}_{t+\tau} - \overline{\mathbf{x}})}{\frac{1}{T}\sum_{t=1}^{T}\left|\left|\mathbf{x}_t - \overline{\mathbf{x}}\right|\right|^2} \quad (2)$$

with $\mathbf{x}_t = (x_1, x_2, x_3, x_4, x_5)_t^T$ and $\mathbf{y}_t = (y_1, y_2, y_3, y_4, y_5)_t^T$ being 5-dimensional vectors that contain the momentary probabilities of all sleep

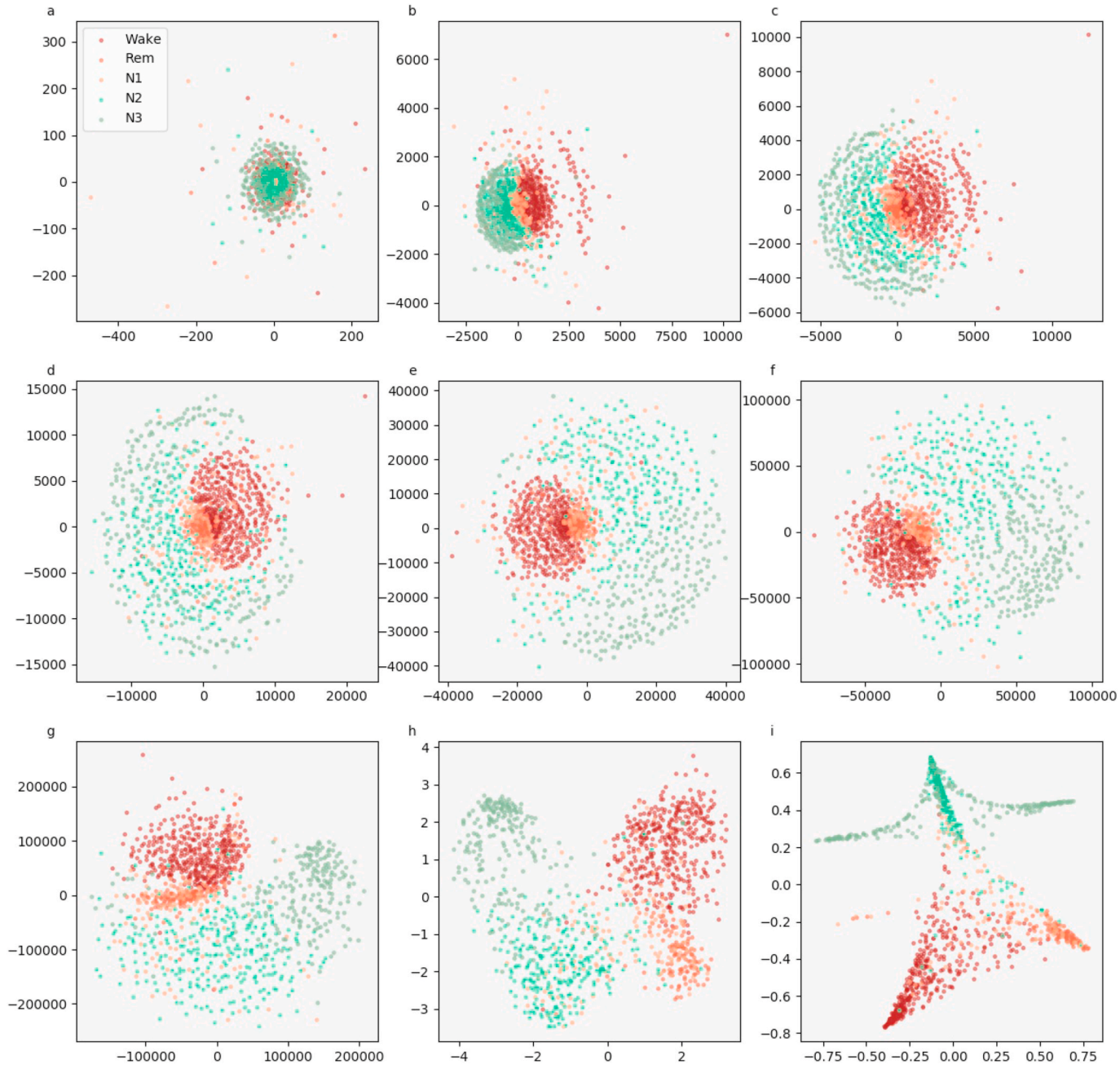

Fig. 2. Sleep stage embeddings.
MDS visualization of EEG data representations (embeddings) within the hidden layers of the deep neural network: input layer (a), max pooling layers (b–g), last LSTM layer (h), and softmax layer (i). The hidden layers (b–g) lead to increasingly better separability of sleep stages compared to z-scored raw EEG data (a). The softmax classification layer (i) normalizes the data to a value range of [0,1]. Note that absolute coordinates of points in MDS projections have no particular meaning other than scaling relative distances between any pair of points.

stages (indices 1 to 5) at time t for EEG channels x and y, respectively, and the mean probability vectors $\overline{\mathbf{x}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}_t$ and $\overline{\mathbf{y}} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{y}_t$. Here, $\circ$ is the dot product, and $||\ ||$ the vector norm.

3. Results

3.1. Sleep stage embeddings

We could show that even a single EEG channel contains enough information to classify different sleep stages. We used several convolutional blocks, which transform the input EEG data into high-dimensional complex features. A procedure referred to as *feature space embedding*. Thus, in contrast to handcrafted features we make the neural network to find its own features. This higher order features lead to a good separability of the transformed EEG vectors (neural network layer output or representation) belonging to different sleep stages. In the following, we refer to these self-organized representations as *sleep stage embeddings*.

We visualize the sleep stage embeddings by a dimensionality reduction into 2D, using multidimensional scaling (MDS) (Krauss et al., 2018a) (Fig. 2). Moreover, we evaluate the generalized discrimination value (GDV), which quantifies separability of data classes in high-dimensional state spaces (Schilling et al., 2018). The MDS plots show that the convolutional layers lead to better separability (Fig. 2b–g) compared to z-scored raw EEG data (network input, Fig. 2a).

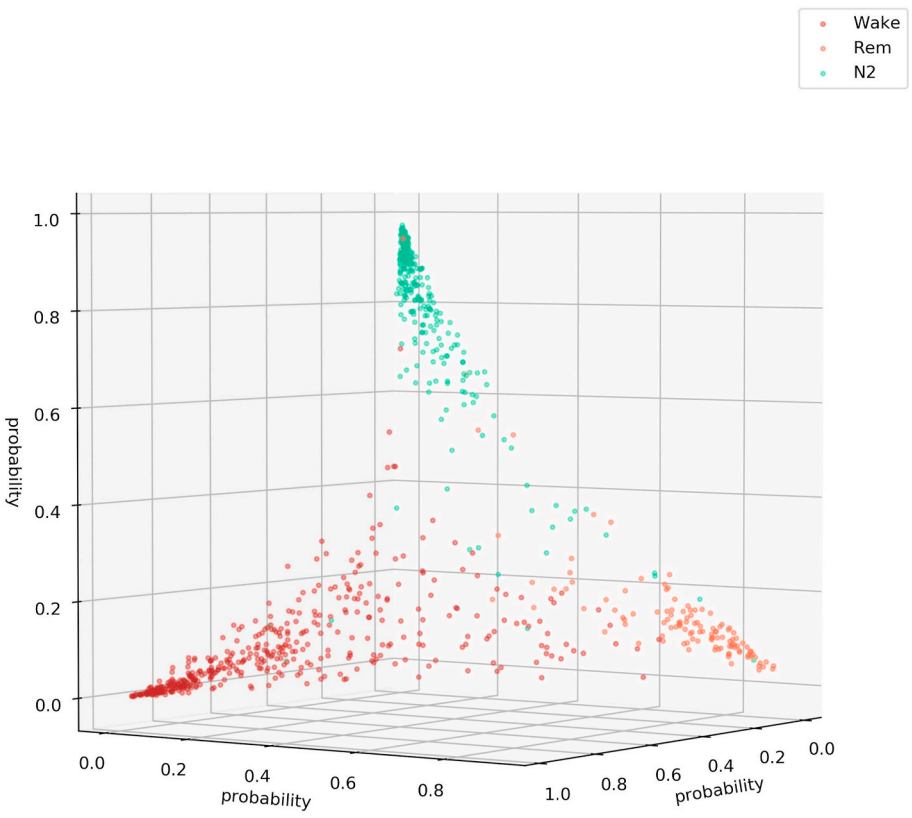

Fig. 3. Softmax output for three sleep stages (Wake, REM, N2).
This figure illustrates that the softmax output spans a hyper-plane (in 3D a 2D hyper-plane with 3 corners). Note that the possible outputs lie in the volume below the hyper-plane. Not all points lie on the plane as confusions with sleep stage N3 or N1 cannot be shown in a 3D plot. Thus, if the softmax output would for example point to sleep stage N3 the point would lie near the origin (0,0,0). Note that absolute coordinates of points in MDS projections have no particular meaning other than scaling relative distances between any pair of points.

Furthermore, the LSTM layer transforms the data so that the different classes are linearly separable (Fig. 2h). The softmax classification layer normalizes the data (value range [0,1]). The shape of the clusters in Fig. 2i can be explained by the fact that the softmax layers places the different classes at the 5 edges of a 4D-hyper-plane (example for 3D projections shown in Figs. 3 and 4). As the probabilities of all 5 sleep stages have to sum up to a value of 1, the range of possible softmax outputs is located on this hyper-plane. The MDS projection of the hyper-plane in 2D leads to the shown patterns (Fig. 2i). Note that, Figs. 2–4) were generated using one sample training data set. Labels are neither used during embedding in the neural network layers, nor for projection, but they serve only for color coding the data points after embedding and projection.

The quantification of the class separability via the GDV (Fig. 2, red dots: input, 6 max-pooling layers, the second LSTM layer, the softmax output layer), shows that the convolutional layers perform data preprocessing. In a previous study, we could demonstrate that the GDV has to overcome an "energy barrier", when the data structure is complex. Thus, the number of layers where the GDV only slightly decreases depends on data complexity. As complex features are needed to classify the data, there are relatively many initial layers with small GDV decrease (cf. Fig. 5). The LSMT layer leads to a clear increase of separability, i.e. drop in the GDV value (8th red point in Fig. 5, compare also Fig. 2 i). The softmax layer causes a further decrease of the GDV value.

3.2. Hypnodensity graphs

An efficient illustration of the softmax layer output are so called hypnodensity graphs (Jens et al., 2018). Here, the probabilities for all sleep stages at each time point are shown instead of only the most probable sleep stage, as it is done in traditional hypnograms. This is an elegant approach, since the probabilities (output of softmax layer) can be compared with the uncertainties, i.e. different assessments of different somnologists, which cause the relatively high inter-rater-variability (Danker-Hopfe et al., 2004, 2009; Wendt et al., 2015).

Additionally, the automatic sleep stage classification can be used to evaluate the data with higher temporal resolution. Even though, the network has been trained on labeled data with a temporal resolution of 30 s which is in accordance with AASM scoring rules (Berry et al., 2012), we evaluate sleep stage probabilities predicted by the neural network with a temporal resolution of 5 s, using a sliding window approach (Fig. 6 and Fig. S1 – S10). Thus, we still analyzed 30 s windows, yet with 5 s steps. The predicted sleep stage probabilities were assigned to the respective center of each window of 30 s width.

The hypnodensity plots show that each EEG channel is sufficient to perform sleep stage classification, as the hypnodensity graphs are very similar (Fig. 6) for all the channels. Interestingly, the sleep stage *wake* seems to be slightly over-represented in channel C4 (6a), whereas in channel O2, *N3* seems to slightly dominate (6c). The similarity between channels can be quantified, for example using pairwise cross-correlations.

3.3. Hypnodensity based sleep cycle period length analysis

On the basis of these EEG channel specific hypnodensity graphs, sleep stage probabilities across and within channels can be compared (see Methods). For this purpose we calculate generalized vector auto- and cross-correlations of the 5-dimensional probability vectors for the 3 EEG channels of each data set. This correlations provide valuable information on neuronal activity patterns during sleep. In particular, all auto-as well as cross-correlations have a global maximum, which is near to the value 1, at zero lag time. This means that there is no time shift of the neuronal patterns between the different channels. Furthermore, the high correlation coefficients at lag time zero indicate that all channels are sufficient to perform sleep scoring, which could already been predicted by the hypnodensity graphs. Local side-maxima indicate repetitions of neuronal patterns as they occur in repeating sleep-cycles.

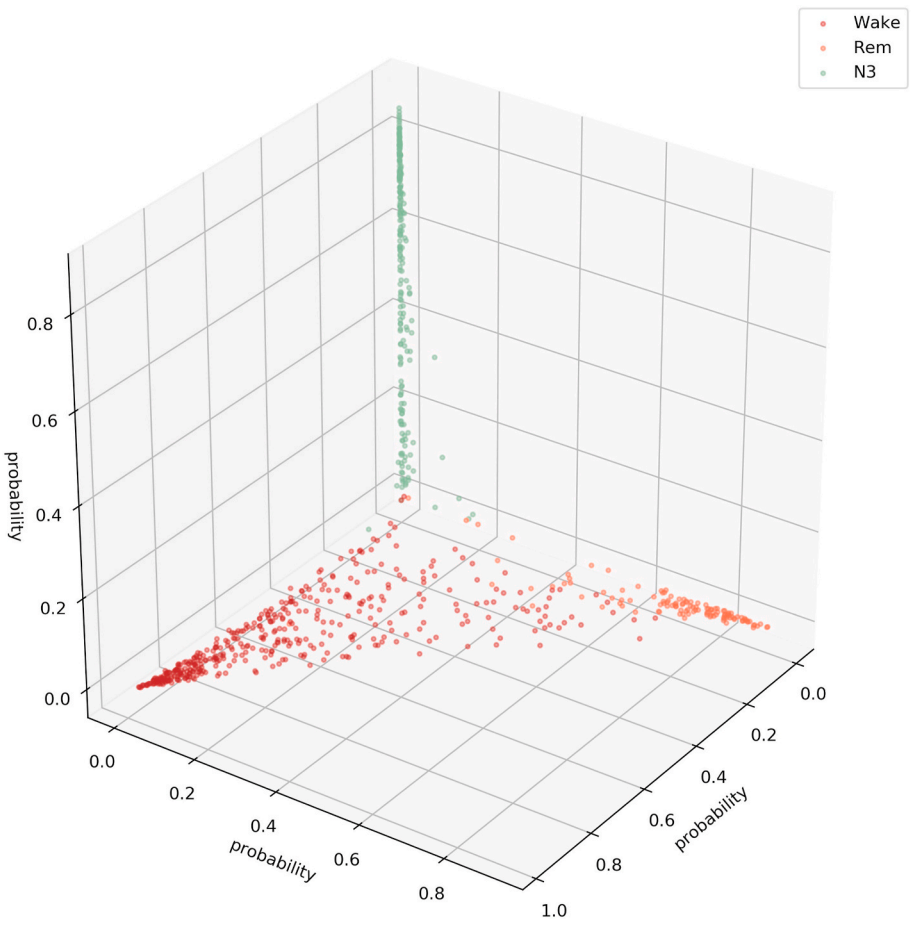

Fig. 4. Softmax output of for three sleep stages (Wake, REM, N3).
This figure illustrates that the softmax output spans a hyper-plane (in 3D a 2D hyper-plane with 3 corners). It can be seen that the sleep stage N3 is not confused with the Wake or the REM state. Note that absolute coordinates of points in MDS projections have no particular meaning other than scaling relative distances between any pair of points.

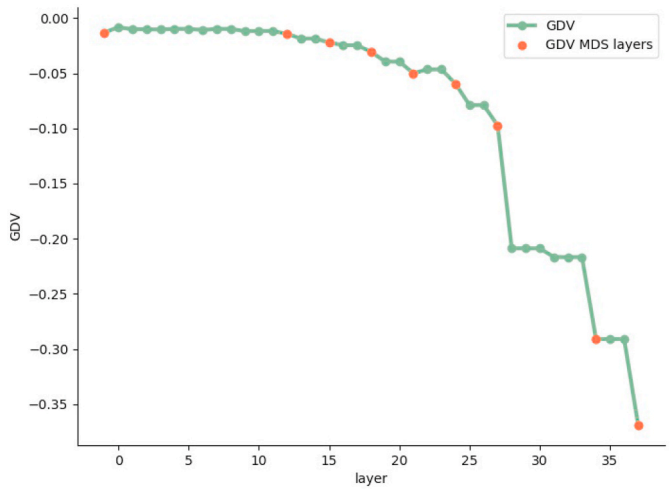

Fig. 5. Separability of sleep stages.
Separability increases with increasing layer depth. Note that, a GDV of 0 corresponds to non-separable data classes, whereas a GDV of −1 corresponds to a perfect data class separation. Red dots refer to the layer outputs shown in 2 (first red dot: input, 2nd-7th: max pooling layers, 8th: 2nd LSTM layer, 9th: softmax output layer). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

For subject 55, the auto and cross-correlations show a local maximum at a lag-time of about 100 min (Fig. 7). This means that after this period of time the sleep stages repeat. In contrast, data of subject 59 show two local maxima at lag-times of 75 and 150 min, respectively (Fig. 8). Hence, this subject's sleep stage period length is about 75 min. These estimated period lengths have been confirmed by the sleep stage scoring based on visual analysis performed by sleep specialists. This means that our novel method helps to reliably estimate the individual period length of sleep cycles. Further data are shown in Figs. S11 – S19.

4. Discussion

In this study, we present novel approaches for the evaluation and visualization of neural data recorded during human sleep. In contrast to numerous previous studies and network architectures used to perform automatic sleep stage scoring (e.g. (Ebrahimi et al., 2008; Koley and Dey, 2012; Zhang and Wu, 2017; Amiya et al., 2018; Alickovic and Subasi, 2018; Zhang et al., 2016)), we provide approaches helping to interpret and understand the network output.

Thus, we illustrate how deep neuronal networks automatically find features that help to separate different sleep stages based on single channel EEG data. That this representations actually lead to a good separability of the different sleep stages is visualized using MDS plots and quantified using the GDV value.

The softmax output, i.e. the predicted probabilities for each sleep stage at every time step, is visualized with hypnodensity graphs (Jens et al., 2018) a sophisticated way to show the uncertainties in sleep stage classification. The calculation of the vectorial auto- and cross-correlations within and between channels, as introduced in this study, is a tool to gain more insight into the architecture of sleep, e.g. sleep cycles. We propose the auto-correlation plots as novel tool to estimate the average period length of sleep stage cycles, and to find disturbances in the period length of sleep cycles, e.g. as an indicator for

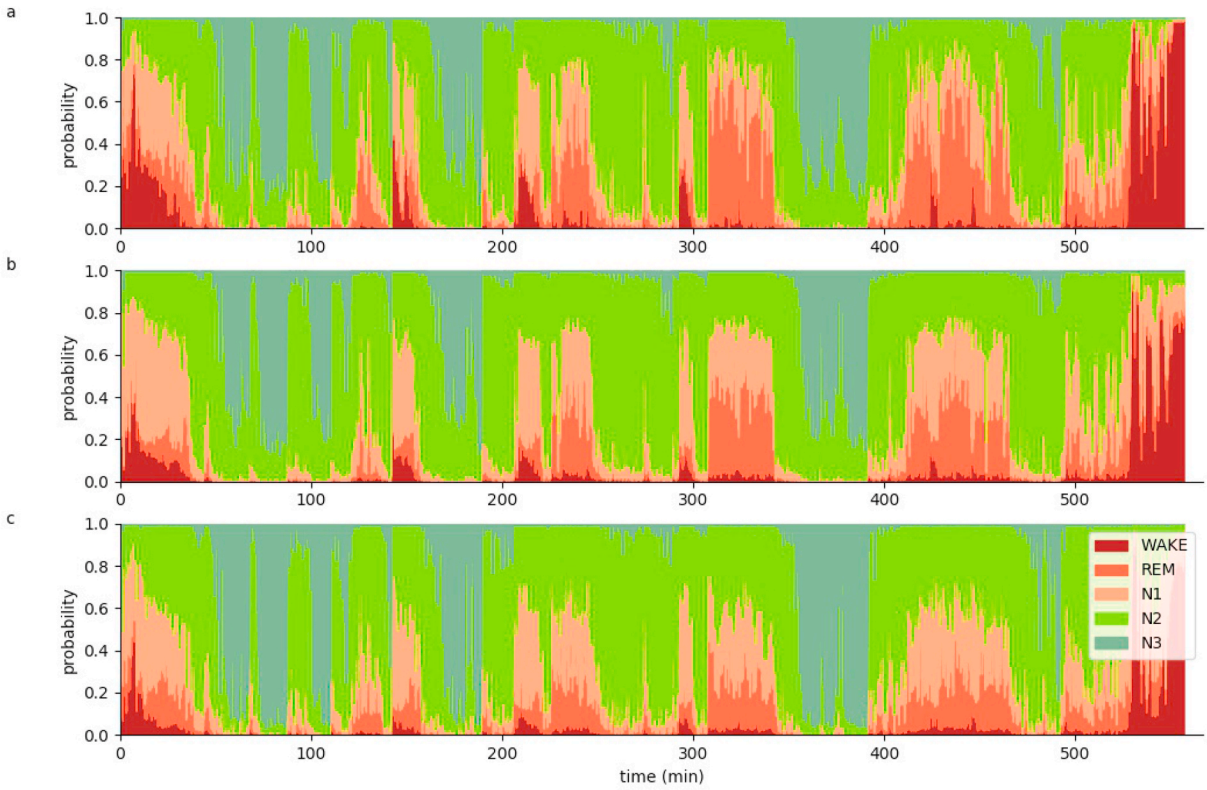

Fig. 6. Hypnodensity graph.
Hypnodensity graph of subject 55 with a temporal resolution of 5 s separately evaluated for the three different EEG channels C4 (a), F4 (b) and O2 (c). The probabilities for all sleep stages at each time point are shown. Each EEG channel is sufficient to perform sleep stage classification, as the hypnodensity graphs are similar.

pathological conditions, for example in the case of patients with obstructive sleep apnea (OSAS). Here, sleep is dominated by events like arousals and short periods of wakefulness. These effects "destroy" the ordered sequence of sleep stages of normal sleep cycles. The cross-correlation plots between different recording channels may serve to shed light into phenomena like local or fragmented sleep. The global maximum of cross-correlation plots, especially if the corresponding lag-time is different from zero, indicates travelling waves across the cerebral cortex, or the order at which the different regions change sleep stages.

Another interesting extension of the correlation method for future studies would be the application to non-stationary cases, i.e. changing period lengths over night. For instance, calculating the correlation only for subsequent parts of the entire recording time would yield different lag-times at maximum correlation, i.e. period length estimates, for each different part. If correlation analysis is initially performed over the complete night, as shown above, this yields the average cycle time, and hence gives a hint on how to choose the duration of the different parts to be analyzed. Alternatively, a sliding window approach could be applied, yielding a time-dependent period length, i.e. lag-time at maximum correlation. Even more sophisticated would be the application of super-statistical methods (Metzner et al., 2015), which are designed to detect the temporal change of statistical parameters in complex systems.

This study is in line with numerous previous studies trying to use artificial neural networks as a model (KriegeskortePamela, 2018; Schilling et al., 2020a; Gerum et al., 2020; Krauss et al., 2019a, 2019b; Krauss and Maier, 2020), and as a tool (Schilling et al., 2020b) to analyze and understand brain activity. This combined approach at the intersection of neuroscience and artificial intelligence is crucial to make further progress in the field, as traditional popular analysis methods have been proven to be not sufficient to understand brain dynamics (Jonas and Paul Kording, 2017; Brown, 2014; Lazebnik, 2002).

CRediT authorship contribution statement

Patrick Krauss: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Supervision, Visualization, Writing – original draft, Writing – review & editing. **Claus Metzner:** Formal analysis, Investigation, Methodology, Software, Validation, Writing – review & editing. **Nidhi Joshi:** Formal analysis, Software, Visualization. **Holger Schulze:** Resources. **Maximilian Traxdorf:** Data curation, Resources. **Andreas Maier:** Investigation, Resources, Supervision, Validation. **Achim Schilling:** Conceptualization, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft.

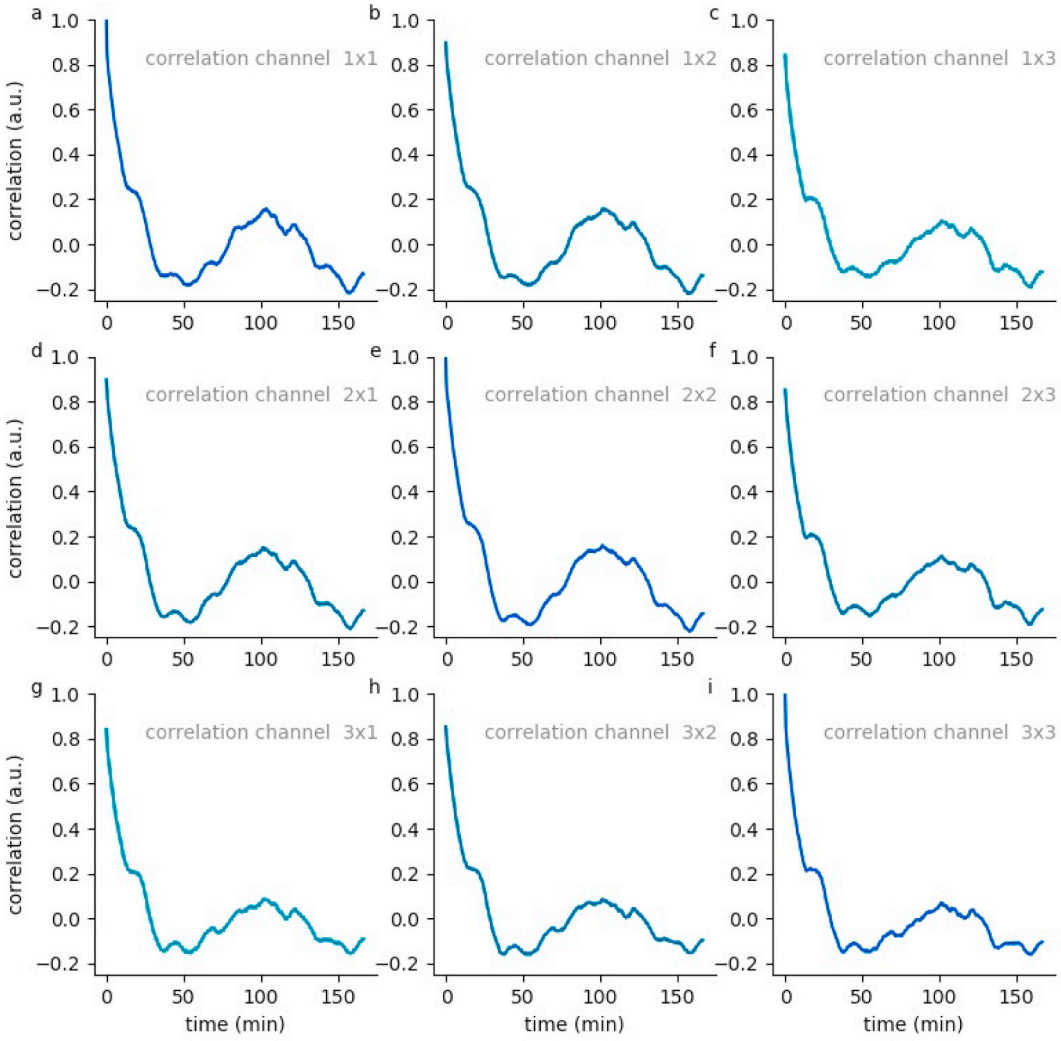

Fig. 7. Hypnodensity cycles.
Temporal auto- and cross correlations of 5-dimensional hypnodensity probability vectors of sleep stages. In subject 55, a local maximum at a lag-time of about 100 min can be observed, indicating an individual period length of sleep cycles of 100 min.

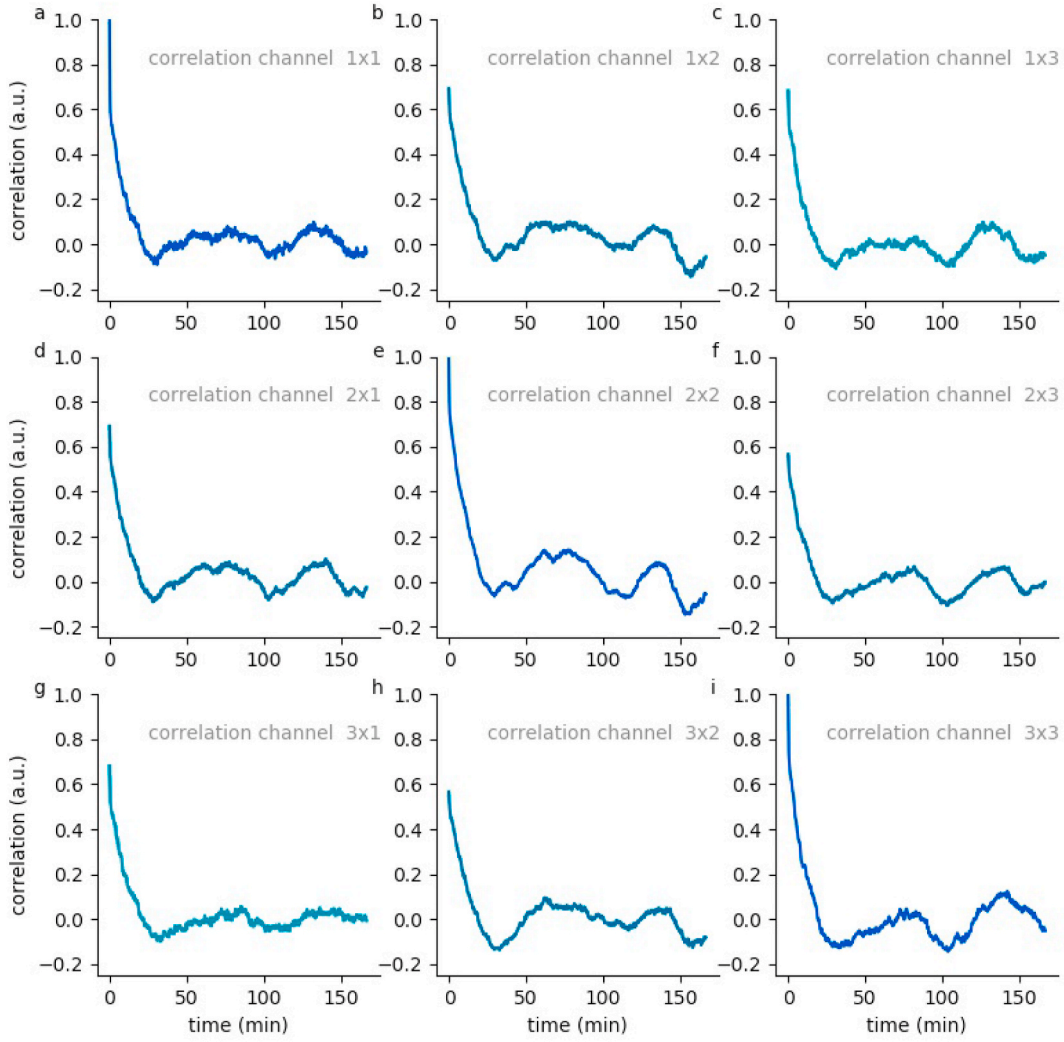

Fig. 8. Hypnodensity cycles.
Temporal auto- and cross correlations of 5-dimensional hypnodensity probability vectors of sleep stages. In subject 59, two local maxima at lag-times of about 75 and 150 min can be observed, indicating an individual period length of sleep cycles of 75 min.

Declaration of competing interest

All authors declare no competing financial interests.

Acknowledgments

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation): grant KR5148/2-1 to PK – project number 436456810, and the Emerging Talents Initiative (ETI) of the University Erlangen-Nuremberg (grant 2019/2-Phil-01 to PK). We thank NVidia for the donation of GPU devices for Machine Learning purposes.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.nbscr.2021.100064.

References

Alickovic, Emina, Subasi, Abdulhamit, 2018. Ensemble svm method for automatic sleep stage classification. IEEE Trans. Instrument. Meas. 67 (6), 1258–1265.
Amiya, Patanaik, Ong, Ju Lynn, J Gooley, Joshua, Ancoli-Israel, Sonia, Chee, Michael WL., 2018. An end-to-end framework for real-time automatic sleep stage classification. Sleep 41 (5), zsy041.
Berry, Richard B., Brooks, Rita, Charlene, E Gamaldo, Harding, Susan M., Marcus, C., Bradley, V Vaughn, et al., 2012. The aasm manual for the scoring of sleep and associated events. Rules Terminol. Techn. Specifi. Darien, Illinois, Am. Acad. Sleep Med. 176, 2012.
Bishop, Christopher M., 2006. Pattern Recognition and Machine Learning. springer.
Boostani, Reza, Karimzadeh, Foroozan, Nami, Mohammad, 2017. A comparative review on sleep stage classification methods in patients and healthy individuals. Comput. Methods Progr. Biomed. 140, 77–91.
Bradley, V Vaughn, Peterson, Giallanza, 2008. Technical review of polysomnography. Chest 134 (6), 1310–1319.
Brown, Joshua W., 2014. The tale of the neuroscientists and the computer: why mechanistic theory matters. Front. Neurosci. 8, 349.
Burns, Joseph W., Crofford, Leslie J., Chervin, Ronald D., 2008. Sleep stage dynamics in fibromyalgia patients and controls. Sleep Med. 9 (6), 689–696.
Chollet, François, et al., 2018. Keras: the python deep learning library. ascl. S. ascl: 1806.022.
Conrad Iber, 2007. The Aasm Manual for the Scoring of Sleep and Associated Events: Rules. Terminology and Technical Specification.
Danker-Hopfe, Heidi, Kunz, Dieter, Gruber, Georg, Klösch, Gerhard, José, L Lorenzo, Sari-Leena, Himanen, Kemp, Bob, Penzel, Thomas, Röschke, Joachim, Dorn, Hans, et al., 2004. Interrater reliability between scorers from eight european sleep laboratories in subjects with different sleep disorders. J. Sleep Res. 13 (1), 63–69.
Danker-Hopfe, Heidi, Anderer, Peter, Zeitlhofer, Josef, Boeck, Marion, Dorn, Hans, Gruber, Georg, Heller, Esther, Loretz, Erna, Moser, Doris, Parapatics, Silvia, et al., 2009. Interrater reliability for sleep scoring according to the rechtschaffen & kales and the new aasm standard. J. Sleep Res. 18 (1), 74–84.
De Laat, Paul B., 2018. Algorithmic decision-making based on machine learning from big data: can transparency restore accountability? Philos. Technol. 31 (4), 525–541.
Ebrahimi, Farideh, Mikaeili, Mohammad, Estrada, Edson, Nazeran, Homer, 2008. Automatic sleep stage classification based on eeg signals by using neural networks

and wavelet packet coefficients. In: 2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, pp. 1151–1154.
Gerum, Richard C., Erpenbeck, André, Krauss, Patrick, Schilling, Achim, 2020. Sparsity through evolutionary pruning prevents neuronal networks from overfitting. Neural Network.
Goodfellow, Ian, Bengio, Yoshua, Courville, Aaron, Bengio, Yoshua, 2016. Deep Learning, vol. 1. MIT press Cambridge.
Hochreiter, Sepp, Schmidhuber, Jürgen, 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.
Hunter, John D., 2007. Matplotlib: a 2d graphics environment. Comput. Sci. Eng. 9 (3), 90–95.
Jens, B Stephansen, Olesen, Alexander N., Olsen, Mads, Ambati, Aditya, Eileen, B Leary, Moore, Hyatt E., Carrillo, Oscar, Lin, Ling, Han, Fang, Han, Yan, et al., 2018. Neural network analysis of sleep stages enables efficient diagnosis of narcolepsy. Nat. Commun. 9 (1), 1–15.
Jonas, Eric, Paul Kording, Konrad, 2017. Could a neuroscientist understand a microprocessor? PLoS Comput. Biol. 13 (1), e1005268.
Koley, B., Dey, Debangshu, 2012. An ensemble system for automatic sleep stage classification using single channel eeg signal. Comput. Biol. Med. 42 (12), 1186–1195.
Krause, Josua, Adam, Perer, Ng, Kenney, 2016. Interacting with predictions: visual inspection of black-box machine learning models. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 5686–5697.
Krauss, Patrick, Maier, Andreas, 2020. Will we ever have conscious machines? Front. Comput. Neurosci. 14.
Krauss, Patrick, Metzner, Claus, Schilling, Achim, Tziridis, Konstantin, Traxdorf, Maximilian, Wollbrink, Andreas, Rampp, Stefan, Pantev, Christo, Schulze, Holger, 2018a. A statistical method for analyzing and comparing spatiotemporal cortical activation patterns. Sci. Rep. 8 (1), 1–9.
Krauss, Patrick, Schilling, Achim, Bauer, Judith, Tziridis, Konstantin, Metzner, Claus, Schulze, Holger, Traxdorf, Maximilian, 2018b. Analysis of multichannel eeg patterns during human sleep: a novel approach. Front. Hum. Neurosci. 12, 121.
Krauss, Patrick, Zankl, Alexandra, Schilling, Achim, Schulze, Holger, Metzner, Claus, 2019a. Analysis of structure and dynamics in three-neuron motifs. Front. Comput. Neurosci. 13 (5).
Krauss, Patrick, Schuster, Marc, Dietrich, Verena, Schilling, Achim, Schulze, Holger, Metzner, Claus, 2019b. Weight statistics controls dynamics in recurrent neural networks. PloS One 14 (4), e0214541.
Kriegeskorte, Nikolaus, Pamela, K Douglas, 2018. Cognitive computational neuroscience. Nat. Neurosci. 21 (9), 1148–1160.
Lazebnik, Yuri, 2002. Can a biologist fix a radio?—or, what i learned while studying apoptosis. Canc. Cell 2 (3), 179–182.
LeCun, Yann, Bengio, Yoshua, et al., 1995. Convolutional networks for images, speech, and time series. Handb. Brain Theor. Neural Network. 3361 (10), 1995.
LeCun, Yann, Bengio, Yoshua, Hinton, Geoffrey, 2015. Deep learning. Nature 521 (7553), 436–444.
Martin, Wattenberg, Viégas, Fernanda, Johnson, Ian, 2016. How to Use t-Sne Effectively. Distill.
Metzner, Claus, Mark, Christoph, Steinwachs, Julian, Lautscham, Lena, Stadler, Franz, Fabry, Ben, 2015. Superstatistical analysis and modelling of heterogeneous random walks. Nat. Commun. 6 (1), 1–8.
Pedregosa, Fabian, Varoquaux, Gaël, Gramfort, Alexandre, Michel, Vincent, Bertrand, Thirion, Grisel, Olivier, Blondel, Mathieu, Peter, Prettenhofer, Weiss, Ron, Vincent, Dubourg, et al., 2011. Scikit-learn: machine learning in python. J. Mach. Learn. Res. 12, 2825–2830.
Phan, Huy, Andreotti, Fernando, Cooray, Navin, Chén, Oliver Y., De Vos, Maarten, 2018. Automatic sleep stage classification using single-channel eeg: learning sequential features with attention-based recurrent neural networks. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC),. IEEE, pp. 1452–1455.
Phan, Huy, Quan, Do, The-Luan, Do, Vu, Duc-Lung, 2013. Metric learning for automatic sleep stage classification. In: In 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, pp. 5025–5028.
Schilling, Achim, Metzner, Claus, Rietsch, Jonas, Gerum, Richard, Schulze, Holger, Krauss, Patrick, 2018. How Deep is Deep Enough?–Quantifying Class Separability in the Hidden Layers of Deep Neural Networks arXiv preprint arXiv:1811.01753.
Schilling, Achim, Gerum, Richard, Zankl, Alexandra, Schulze, Holger, Metzner, Claus, Krauss, Patrick, 2020a. Intrinsic Noise Improves Speech Recognition in a Computational Model of the Auditory Pathway. bioRxiv.
Schilling, Achim, Tomasello, Rosario, Henningsen-Schomers, Malte R., Surendra, Kishore, Haller, Martin, Karl, Valerie, Peter, Uhrig, Maier, Andreas, Krauss, Patrick, 2020b. Analysis of Ongoing Neuronal Activity Evoked by Continuous Speech with Computational Corpus Linguistics Methods bioRxiv.
Stéfan van der Walt, S Chris Colbert, Varoquaux, Gael, 2011. The numpy array: a structure for efficient numerical computation. Comput. Sci. Eng. 13 (2), 22–30.
Tatum, William O., Dworetzky, Barbara A., Schomer, Donald L., 2011. Artifact and recording concepts in eeg. J. Clin. Neurophysiol. 28 (3), 252–263.
Traxdorf, Maximilian, Krauss, Patrick, Schilling, Achim, Schulze, Holger, Tziridis, Konstantin, 2019. Microstructure of cortical activity during sleep reflects respiratory events and state of daytime vigilance. Somnologie 23 (2), 72–79.
van der Maaten, Laurens, Hinton, Geoffrey, 2008. Visualizing data using t-sne. J. Mach. Learn. Res. 9 (Nov), 2579–2605.
Wendt, Sabrina L., Welinder, Peter, Sorensen, Helge BD., Paul, E Peppard, Jennum, Poul, Perona, Pietro, Mignot, Emmanuel, Warby, Simon C., 2015. Inter-expert and intra-expert reliability in sleep spindle scoring. Clin. Neurophysiol. 126 (8), 1548–1556.
Zhang, Junming, Wu, Yan, 2017. A new method for automatic sleep stage classification. IEEE Trans. Biomed. Circuits Syst. 11 (5), 1097–1110.
Zhang, Junming, Wu, Yan, Bai, Jing, Chen, Fuqiang, 2016. Automatic sleep stage classification based on sparse deep belief net and combination of multiple classifiers. Trans. Inst. Meas. Contr. 38 (4), 435–451.

NeuroImage 297 (2024) 120696

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Deep learning based decoding of single local field potential events

Achim Schilling [a,b], Richard Gerum [b,c], Claudia Boehm [a,b], Jwan Rasheed [a,b], Claus Metzner [b,d], Andreas Maier [d], Caroline Reindl [e], Hajo Hamer [e], Patrick Krauss [b,d,*]

[a] Neuroscience Lab, University Hospital Erlangen, Germany
[b] Cognitive Computational Neuroscience Group, University Erlangen-Nürnberg, Germany
[c] Department of Physics and Center for Vision Research, York University, Toronto, Canada
[d] Pattern Recognition Lab, University Erlangen-Nürnberg, Germany
[e] Epilepsy Center, Department of Neurology, University Hospital Erlangen, Germany

ARTICLE INFO

Keywords:
Deep learning
Local field potentials
Auditory neuroscience
Auditory cortex
Speech perception
Auto-encoder
Embeddings
Intracranial EEG (iEEG)
Stereotactic EEG (sEEG)

ABSTRACT

How is information processed in the cerebral cortex? In most cases, recorded brain activity is averaged over many (stimulus) repetitions, which erases the fine-structure of the neural signal. However, the brain is obviously a single-trial processor. Thus, we here demonstrate that an unsupervised machine learning approach can be used to extract meaningful information from electro-physiological recordings on a single-trial basis. We use an auto-encoder network to reduce the dimensions of single local field potential (LFP) events to create interpretable clusters of different neural activity patterns. Strikingly, certain LFP shapes correspond to latency differences in different recording channels. Hence, LFP shapes can be used to determine the direction of information flux in the cerebral cortex. Furthermore, after clustering, we decoded the cluster centroids to reverse-engineer the underlying prototypical LFP event shapes. To evaluate our approach, we applied it to both extra-cellular neural recordings in rodents, and intra-cranial EEG recordings in humans. Finally, we find that single channel LFP event shapes during spontaneous activity sample from the realm of possible stimulus evoked event shapes. A finding which so far has only been demonstrated for multi-channel population coding.

1. Introduction

How is sensory input and information processed in the cerebral cortex? To answer this question, great efforts have been made to measure brain activity with increasing temporal and spatial resolution. Thanks to the enormous advances in fMRI, we now know very precisely where certain processes take place. It is now possible to scan the whole brain in a matter of seconds on a millimeter scale (Bollmann and Barth, 2021).

A proper understanding of what exactly happens in the brain when sensory stimuli are processed, or what the difference between real sensory perception and phantom perception (e.g. tinnitus, Schilling et al. (2023c)) really is, requires a detailed analysis of the temporal cues of neural activity in the brain. It is still not fully clear how to take advantage of the growing temporal resolution of modern neuro-imaging techniques such as MEG/EEG, and intra-cranial multi-unit recordings. For example, surface-based methods like EEG and MEG provide a poor spatial resolution and signal-to-noise ratio, as the signal source and the recording site are separated by the skull. The surface electrodes, or SQUIDs respectively, are used to record the sum-activity of millions of neurons (Li et al., 2014; Kaiser, 2007). Furthermore, the poor signal-to-noise ratio results in the issue that meaningful data can only be extracted, by averaging over many trials for evoked activity (for auditory evoked potentials $\geq$ 100, see Başar et al. (1987), Gajraj et al. (1998) and Koelbl et al. (2023)), or by extracting spectral cues through Fourier-filters (Tonner and Bein, 2006; Newson and Thiagarajan, 2019). However, as Waterstraat and co-workers state the brain is a *"single trial processor"*, that has been shaped by evolutionary processes (Waterstraat et al., 2021).

The only way to increase the signal-to-noise ratio in order to extract meaningful data from single-trials is with intra-cranial recordings. However, these experiments are limited to animal studies as the intra-cranial recording is an invasive method (see e.g. Schilling et al. (2023a, 2019) and Krauss et al. (2018a)). However, in epilepsy diagnostics, intracranial electrodes are sometimes implanted in the brains of human patients in order to precisely localize the epileptic focus (iEEG, Kovac et al. (2017)). Usually, these rare recordings are additionally used to tackle scientific questions (Staresina et al., 2015; Buzsáki et al., 2012; Mormann et al., 2017). In contrast, to surface recordings on top of the

* Correspondence to: Friedrich-Alexander-University, Erlangen-Nürnberg (FAU), Chair of Computer Science 5 (Pattern Recognition), Martensstr. 3, 91058 Erlangen, Germany.
E-mail address: patrick.krauss@fau.de (P. Krauss).

https://doi.org/10.1016/j.neuroimage.2024.120696
Received 18 January 2023; Received in revised form 12 June 2024; Accepted 18 June 2024
Available online 21 June 2024

skull such as EEG or MEG, LFP events from intra-cranial recordings have a far better resolution since they are mainly produced by three orders of magnitude smaller neuron ensembles (Hagen et al., 2016). Thus, the radius within which neurons contribute significantly to the LFP is only up to one millimeter. Kajikawa and Schroeder (2011) and Lindén et al. (2011). Furthermore, LFP shapes contain information on the proportions of the contributing sources of extra-cellular currents as well as on the conductive properties of the surrounding brain tissue (Buzsáki et al., 2012). The low frequency parts of the LFPs are induced by synaptic activity, inducing trans-membrane currents that result in electrical dipoles (Buzsáki et al., 2012). Although fast action potentials contribute to the high-frequency components of LFPs, the main contribution comes from dendritic currents (Kraskov et al., 2007; Kreiman et al., 2006; Logothetis, 2003). Therefore, LFPs are typically regarded as a measure of the input signal to the neurons to be measured (Kreiman et al., 2006). It should be noted that further effects such as calcium spiking, gap junctions, and neuron oscillations play a role for LFP events (Buzsáki et al., 2012).

Since the spectral cues as well as the shape of LFP events contain a lot of information about the geometry, properties and activity of the neuron ensembles, the investigation of LFPs in terms of sensory processing and cognition is a promising approach. Thus, LFPs are for example used to implement brain computer interfaces for locked-in patients (Jackson and Hall, 2016).

Despite the fact that single trial LFP events are a valid measure for local neural population analysis (Constantinou et al., 2016), studies on single-trial LFP events analysis are rare. Thus, a better understanding of LFP events could e.g. help to make major progress in understanding cognitive processes during sleep (Bukhtiyarova et al., 2022). Unfortunately, in contrast to multi-unit activity (MUA), the shapes of LFP events are more versatile. Therefore, it is much more difficult to automatically cluster or classify, the different LFP shapes compared to spiking activity (for spike clustering see e.g. Keshtkaran and Yang (2017)). However, it has already been demonstrated that wavelet analysis could be used to extract meaningful features from single trial LFP events (Wang et al., 2007). Furthermore, an algorithm to automatically detect and calculate three pre-defined features (latency, amplitude, and rebound) of single trial LFP events has been developed (Mahmud et al., 2016).

Nevertheless so far, more advanced evaluation techniques such as machine learning based pattern recognition, especially using deep neural networks, have rarely been used to detect and characterize single trial LFP events. Thus, more coarse-grained features such as spectral cues have been analyzed using machine learning (Golshan et al., 2016). Nurse and co-workers used a top-down approach to classify LFP data by training a convolutional neural network on raw data (Nurse et al., 2016). They argue that the bottleneck in processing LFPs using machine learning is to find the right hand-crafted features (Nurse et al., 2016).

However, such top-down approaches like e.g. training classification networks on raw LFP data, respectively events, also have major drawbacks compared to parameter-free and data-driven bottom-up approaches (Krauss et al., 2012). Although, classifier approaches approach are promising for the development of e.g. brain computer interfaces, the neuroscientific value of these networks is poor: Whereas we can derive statements on certain features of the LFP signal regarding our pre-defined labels, all the other information, which is not needed for classification according to the pre-defined labels is lost.

Therefore, a valid approach to deal with high-dimensional electrophysiological data sets, is to apply unsupervised (resp. self-supervised) machine learning algorithms. These algorithms are used to extract recurring patterns from the signal and to reduce the dimensionality of the signal for visualization purposes, making it easier for humans to interpret. Pang et al. (2016), Cunningham and Byron (2014) and Wang et al. (2016). Indeed, recently some studies have been published, which applied unsupervised machine learning methods on electrophysiological data. For instance, Mackevicius and coworkers developed an unsupervised training algorithm to extract non-redundant sequences of the neurophysiological data (Mackevicius et al., 2019). In addition, Hardcastle and coworkers (Hardcastle et al., 2019) applied an auto-encoder network on peripheral nervous system recordings (for further example see also (Zhou and Wei, 2020)). Note that, some scientific approaches used the unsupervised trained networks the other way around, namely as models for the nervous system (see e.g. Ran et al. (2021)).

In the present study, we use an auto-encoder network to cluster different LFP-event shapes. Therefore, we extract the LFP events by searching for local minima in the signal stream and applying advanced thresholding and filtering techniques. We used an undercomplete auto-encoder network with space expansion (according to Bourlard and Kabil (2022)) to reduce the dimensionality of the LFP events to three, in order to identify certain clusters representing different LFP shapes. Note that for the resulting low-dimensional representations the term ‘embeddings’ has been coined (Schilling et al., 2021; Metzner et al., 2021; Krauss et al., 2021; Metzner et al., 2023; Surendra et al., 2023). To increase interpretability, we decode the embeddings after we have identified different clusters, thereby reverse-engineering the prototypical LFP shape of each cluster. We provide evidence that the LFP-event shape may be used to distinguish between different stimulus conditions for evoked activity and is a highly reproducible marker for the direction of information flow within the cerebral cortex for both evoked and spontaneous activity. Finally, we show that single channel LFP event shapes during spontaneous activity sample from the realm of possible stimulus evoked event shapes. This is a finding which so far has only been demonstrated for multi-channel population coding.

2. Methods

2.1. Ethics

2.1.1. Animal experiments

Mongolian gerbils (Meriones unguiculatus) were housed in standard cages (Bio A. S. Vent Light, EHRET Labor- und Pharmatechnik, Emmendingen, Germany) with 2–3 animals per cage. They had continuous access to food and water in a controlled environment at 20-24 °C and a constant 12/12 h dark/light cycle. The care and use of these animals was officially approved by the Bavarian authorities (TVA Nr.: 55.2.2-2532-2-540, Regierungspräsidium Unterfranken, Würzburg, Germany).

2.1.2. Human iEEG data

In this study we used intracranial EEG data from a single patient with pharmaco-resistant epilepsy who were implanted with intracranial electrodes for diagnostic purposes. The data were originally recorded at the University Hospital Erlangen for diagnostic purposes only. All procedures involving the patient were performed according to the highest ethical standards and in full compliance with all relevant institutional and national guidelines. The German law requires no declaration of consent.

2.2. Rodent data acquisition

2.2.1. Surgery

For the craniotomy, the Mongolian gerbils were put under deep Ketamine–Xylazine anesthesia and kept on a controlled heating pad to guarantee a constant body temperature. After hair removal using an electric razor the scalp was removed using sharp micro-scissor. The skull is cleaned with a drill and tweezers. Four screws (M2, 2 mm) are implanted frontally and caudally from bregma, which serve as base for the head fixation. As head-fixation a small elbow is glued with dental cement to the skull and the fixation screws. A rectangular trepanation of approximately 4 mm x 4 mm is opened located between ear and the eye contra-laterally to the ear to be measured. The dura is removed using a fine needle and micro-tweezers (see also Schilling et al. (2023a)). The trepanation is covered with NaCl solution to prevent the brain from drying out.

2.2.2. Extracellular recordings

After the craniotomy the animal is placed in an anechoic chamber on a controlled heating pad and the head is fixed by screwing the vertical part of the elbow to a aluminium rod. The distance of the loudspeaker and the ear of the animal is 20 cm. For extracellular recording we use 16 electrode arrays (Clunbury Scientific (Bloomfield Hills), 0.5 MΩ, layout:4 x 4, spacing; 500 µm). The electrode is inserted in the auditory cortex of the animal using a manual micro-controller. The insertion depth of the electrode array is between 500 µm and 1 mm to record signals from the granular and infra-granular layers, where we expect significant spiking activity and high field potential amplitudes. The neural signal is recorded using the CerePlex system from Blackrock Neurotech (Salt Lake City, USA). The neural signal is digitized directly at the recording site with the Cereplex μ headstage. We recorded all signals using the maximum sampling rate of 30 kHz. To divide between LFP and spike signal during recording, for LFPs a 250 Hz low-pass filter was applied, whereas spiking activity was assumed to occur in the range between 250 Hz and 2.5 kHz. The spikes were sorted online by manually set thresholds.

2.2.3. Auditory stimulation

For auditory stimulation we used a 3 way loudspeaker (MAC, Racer 320). We calibrated the loudspeaker by measuring the output loudness for frequencies between 500 Hz and 19 kHz (half octave steps). For each frequency a correction factor was calculated. The loudspeaker is driven by a 24 bit external sound card (ASUS Xonar MKII 7.1) connected to the computer via USB. The soundcard output is amplified using an audio amplifier (Amp75). To ensure exact alignment of the auditory stimuli with the recorded neural signal, the soundcard is also used as trigger pulse generator. Thus, a second soundcard channel sends trigger pulses to an analog input channel of the recording setup (cf. also Schilling et al. (2021), Garibyan et al. (2022), Schüller et al. (2023), Koelbl et al. (2023) and Schüller et al. (2024)). To analyze the tuning of the neurons, we presented 50 ms pure tone stimuli with inter-stimulus intervals of 300 ms. Click sounds were prevented by adding sine^2-ramps of 5 ms duration to the pure tone stimuli. The presented stimulus frequencies lie in the range of 500 Hz and 20 kHz (half octave steps). Stimulus intensities and frequencies were presented pseudo-randomly. Depending on the hearing ability of the animals, 5 different sound pressure levels were presented (10 dB steps) starting from 110 dB SPL (90 dB, 70 dB respectively). To analyze changes in auditory processing caused by hearing damage, we applied a pure-tone noise trauma (2 kHz, 115 dB SPL, 75 min, loudspeaker: Canton) under deep anesthesia (same as in surgery section). This is a well established method to induce a hearing loss and potentially tinnitus (for further information see Schilling et al. (2017) and Schilling et al. (2019)).

2.3. Human data acquisition

The intracranial EEG (iEEG as EEG recorded via stereotactically implanted depth electrodes, SEEG) originated from a patient suffering from epilepsy, who underwent presurgical epilepsy evaluation due to a drug refractory focal epilepsy. Depths electrodes were implanted in the auditory cortex of the right temporal lobe solely because of clinical reasons. In this study, we show the data of 4 platin electrodes (RTB-1-4) with an impedance lower than 10 kΩ, which were measured against a reference electrode placed on top of the scalp (CPz). Due to the fact that there are not many patients, who receive electrodes in the auditory cortex we are data limited.

2.4. Data analysis

2.4.1. Computational resources

The complete software used for the project was written in Python 3.8 using the NumPy library (Van Der Walt et al., 2011). The graphical user interface was designed and implemented using PyQt5 (Meier, 2019). The neuronal recordings were imported and converted using the 'brpylib'-library provided by Blackrock Neurotech (BlackrockNeurotech, 2021). Basic data processing and filtering was done using the SciPy-library and especially the 'signal'-module (Virtanen et al., 2020). Machine learning was implemented in Keras (Chollet, 2018) with the TensorFlow backend (Pang et al., 2020). All simulations and evaluations were performed on a standard desktop computer. Plotting was done using the Matplotlib library (Hunter, 2007) in combination with the pylustrator add-on (Gerum, 2019).

Table 1
Auto-encoder network (epochs = 1000, batch-size = 500).

Layer	Activation	Output shapes	Parameters
Input layer	ReLu	(None, 50)	0
Dense	ReLu	(None, 700)	35 700
Dropout (0.3)		(None, 700)	0
Dense	ReLu	(None, 3)	2103
Dense	ReLu	(None, 700)	2800
Dropout (0.3)		(None, 700)	0
Dense	ReLu	(None, 50)	35 050

2.4.2. LFP event detection

Local field potential (LFP) events are identified through a computational algorithm designed to detect local minima within the continuous data flow. Subsequently, a predefined threshold criterion is employed to exclude events exhibiting excessively brief inter-event latencies. This approach ensures the consideration of only significant minima, which are represented by black crosses in Fig. 2a, thereby optimizing the reliability of event detection.

2.4.3. Neural networks

The used neural networks in this study were all implemented using Keras with the tensorflow backend (Chollet, 2018; Pang et al., 2020). The auto-encoder was trained on single channel LFP events from a 1 h spontaneous activity recording of only one animal. Thus, first local minimum technique (described above) was used to find the LFP events. These LFP events were then used to train the auto-encoder. Test-data from other recordings and other animals were used to prove the validity of the training. Before used for training the LFP-event were down-sampled to 100 Hz and cut into 500 ms chunks (50 values). The 50-dimensional input vector is expanded by a factor of 14 to 700 dimensions in the first hidden layer of the network. It has be shown that dimensionality expansion can lead to a huge benefit for data processing in neural networks (Dasgupta et al., 2017; Yang et al., 2021). The bottleneck layer has only 3 dimensions. This choice is guided by the objective of simplifying visualization and facilitating initial interpretability, consistent with standard practices in exploratory data analysis (Van Der Maaten et al., 2009; Kriegeskorte and Kievit, 2013). (for details see Table 1).

The batch-size for training was 500 and the training iterations were 1000. During training the data was randomly shuffled. As optimizer, we used Adam with a learning rate of 0.001, as loss function we used the mean-squared error.

In order to measure the loss of meaningful information caused by the auto-encoding procedure, we trained an additional feed-forward network. In particular, we used a 1D convolutional neural network consisting of two convolutional layers, one max-pooling layer, one fully connected (dense) layer, and finally a softmax layer for classification (for details see Table 2). This network was trained on the classification of LFP events according to 4 different stimulus frequencies, hence the output vector is 4-dimensional. The number of iteration was set to 10,000, but for the classifier the early stopping technique was applied.

2.4.4. Cross-correlation analysis

For correlation analysis we used the normalized cross correlation (Yoo and Han, 2009). Thus, the LFP event of the reference

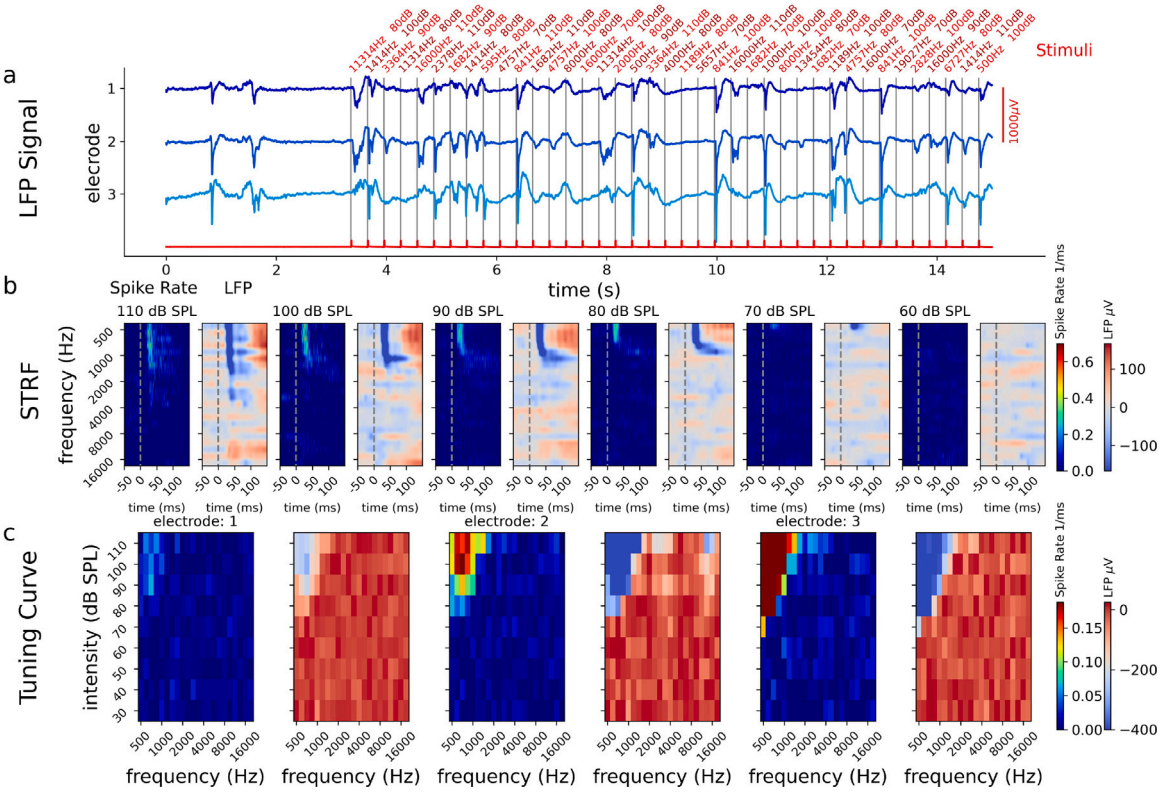

Fig. 1. Tuning curve: a: LFP stream of 3 electrodes (shades of blue) during 50 ms pure tone stimuli (pure tone intensity: 110 dB SPL–70 dB SPL, frequencies: 500 Hz–19 027 Hz in half octave steps). The red curve shows the trigger channel (trigger: black vertical lines). b: Spectro-Temporal-Receptive-Fields (STRFs) of position/electrode 3 in terms of spiking activity and field potentials (sound intensities 110 dB SPL–60 dB SPL). c: Tuning-curves calculated from STRFs in b.

Table 2
Classifier network.

Layer	Activation	Output shapes	Parameters
Input layer		(None, 50, 1)	0
Convolution 1D	ReLu	(None, 41, 60)	330
Dropout (0.3)		(None, 41, 60)	0
Convolution 1D	ReLu	(None, 32, 30)	18 030
Max pooling 1D		(None, 1, 30)	0
Flatten		(None, 30)	0
Dropout (0.3)		(None, 30)	0
Dense	ReLu	(None, 20)	620
Dense	Softmax	(None, 4)	84

channel and the LFP-events of the neighboring channels to be tested were z-scored individually (subtract mean and divide by standard-deviation). After that, the cross-correlation was calculated using the correlate function from numpy (Van Der Walt et al., 2011). The sum is then divided by n_o representing the number of summands. Thus, this definition results in a normed cross-correlation and an auto-correlation of one for no lag time (time delay) $\tau = 0$.

$$C(\tau) = \frac{1}{n_o}\frac{1}{\sigma_r \sigma_e}\sum_{i=0}^{n_o}(r(i\Delta t + \tau) - \mu_r)\cdot(e(i\Delta t) - \mu_e) \quad (1)$$

($C(\tau)$: zero-normed cross correlation, n_o: number of summands which is the number of overlapping samples, $r(i\Delta t)$: reference channel value at time point $i\Delta t$, $i \in \{-n_o, n_o + 1 ..., n_o - 1, n_o\}$, Δt =1 ms, μ_r: mean of all values only for one LFP event in reference channel, τ: time delay, $\tau \in$ [-50 ms, 50 ms], $e(i\Delta t + \tau)$: value of tested channel at time point $e(i\Delta t + \tau)$, μ_e: mean of test channel)

3. Results

In the following paragraphs, we describe how to identify and group LFP events in a continuous LFP data stream.

3.1. Detection of LFP events

Standard evoked response evaluations average LFP responses across many trials to minimize noise, but this can mask inter-trial differences. With high signal-to-noise ratios, single-trial intracranial recordings can provide sufficient data, bypassing the need for averaging. Thus, as shown in Fig. 1a, single pure-tone stimuli of varying frequency and loudness do already evoke clear LFP responses. In most studies, the LFP responses as well as the spike rates are averaged to calculate spectro-receptive fields (STRFs, Fig. 1b) for different stimulus intensities, which then are used to calculate tuning curves for the respective neurons (examples of tuning curves for 3 neighboring electrodes are shown in Fig. 1). As described above, this approach is only possible in cases where the stimulus onset is known. However, similar LFP events also occur spontaneously without any external stimulus (see Fig. 1a before stimulus onset). These LFP events are usually ignored. Nevertheless, the analysis of spontaneous activity plays a crucial role in several neuroscientific fields such as e.g. the investigation of ongoing phantom perception (e.g. tinnitus, see e.g. Schilling et al. (2023c), Krauss et al. (2018b) and Krauss et al. (2016)). Indeed, LFP events also occur spontaneously, without any external stimulus (spontaneous activity measured at three neighboring positions shown in Fig. 2a). Detection of significant LFP events involves algorithm-based identification of local minima and filtering out invalid events based on latency thresholds, as detailed in the Methods section. Only clear minima are considered, indicated by black crosses in Fig. 2a. A huge advantage of the local-minima-search-algorithm is the fact that the detected minimum already defines an exact time-point, which could for example be used to align and average multiple events. Furthermore, the extracted LFP events can be analyzed in terms of shape, size, time of occurrence, etc. (for an example of extracted LFP events see Fig. 2). Thus, the data of the exemplary animal indicates that the inter-event-intervals (time between occurrence of two subsequent events) are distributed in

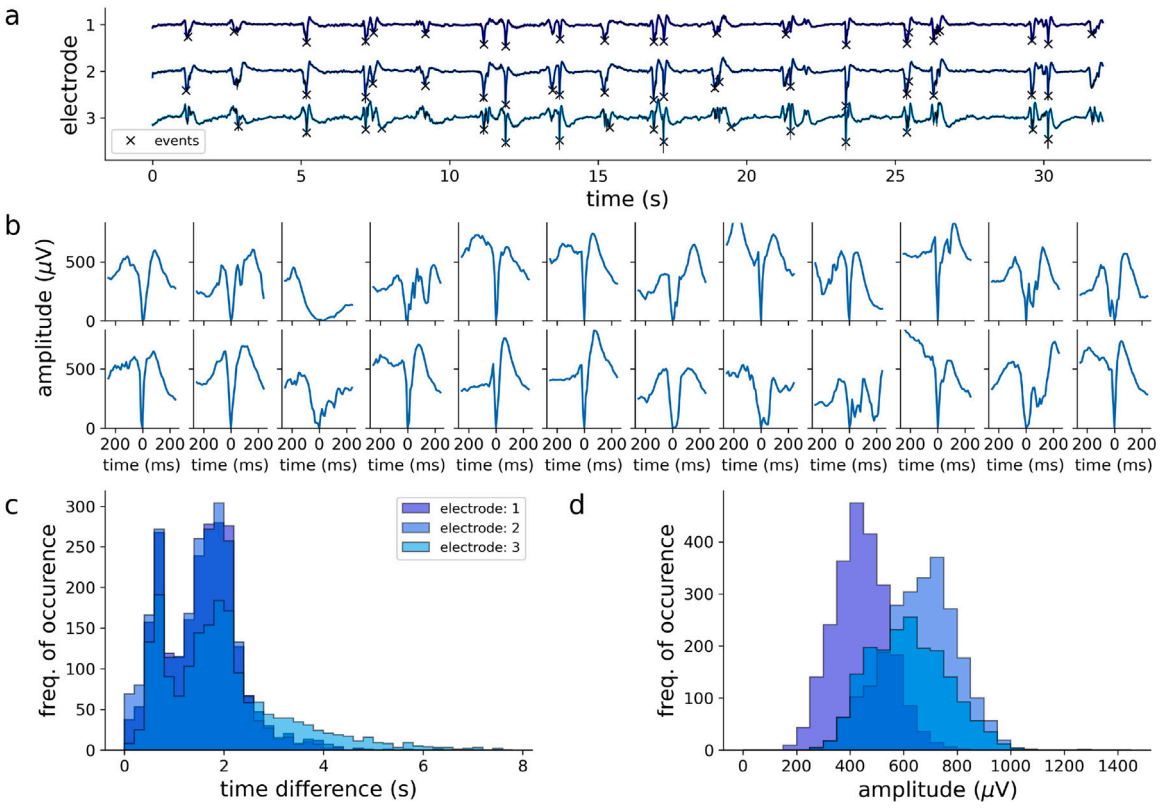

Fig. 2. LFP events in spontaneous activity. a: Spontaneous activity of 3 electrode channels. Black markers (X) show which events were detected as relevant LFP events. LFP events are detected via local minimum search combined with thresholding; b: Examples (n = 24) of detected LFP events; The shape of the events differs. c: Distribution of intervals between detected events; d: Peak-to-peak amplitude distribution of events.

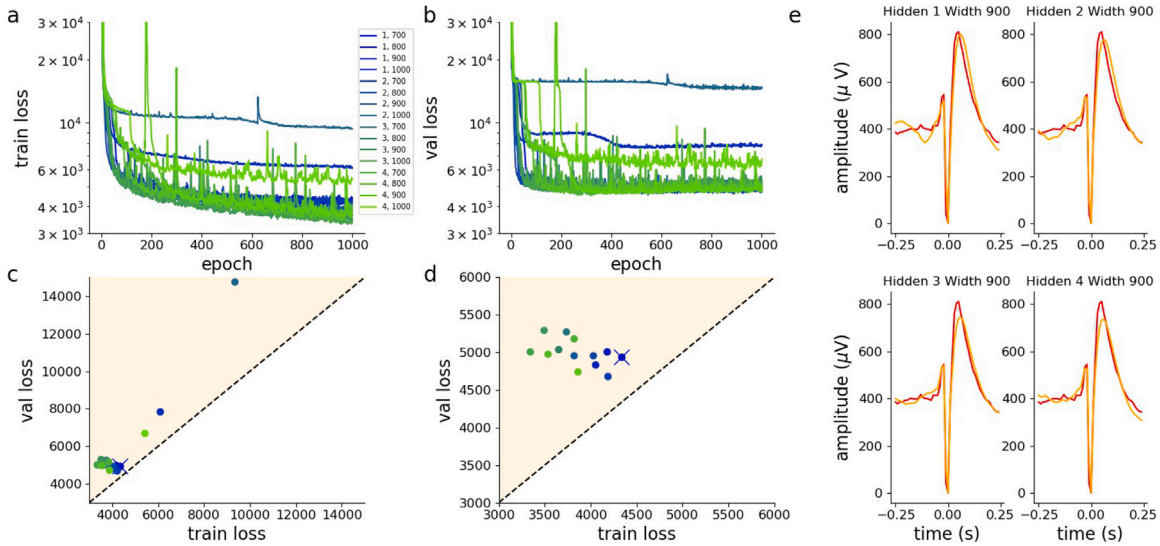

Fig. 3. Grid search to find auto-encoder hyper-parameters. To find a good parameter set for the auto-encoder we performed a grid search with a different number of hidden layers (1, 2, 3, 4; 1 means 1 hidden layer for encoder and 1 hidden unit for decoder, hidden layer means the dimensionality-expansion layer) and different layer widths of the hidden layers (700, 800, 900, 1000) a: Loss function (mean squared error) as a function of training epoch (learning curve) for training data set; b: Learning curve for the validation data set; c: Scatter plot showing validation loss as a function of the training loss. Values above the diagonal represent parameter combinations where the validation loss is higher than the training loss, which points to over-fitting (d: zoom of c, in d outliers are not shown). e: Exemplary LFP events (red) and reconstructions (orange) for exemplary parameter combinations. As the effect on the loss function of several and larger hidden layers was low and further layers cause over-fitting, the parameter combination (1 hidden layer and 700 neurons were chosen, see Table 1). Thus, the smaller network is assumed to have better generalization properties.

an asymmetric (e.g. log-normal distributed) way, whereas the peak-to-peak amplitudes are more Gaussian-like distributed. Furthermore, the inter-event-intervals show a bi-modal distribution with 2 peaks at different time-points (approx. 1 s and 2 s). Besides these basic measures of the LFP events such as amplitude or inter-event-interval duration, the LFP events contain much more information. For instance, the shape of the events can provide further information on how information is processed in the cortex.

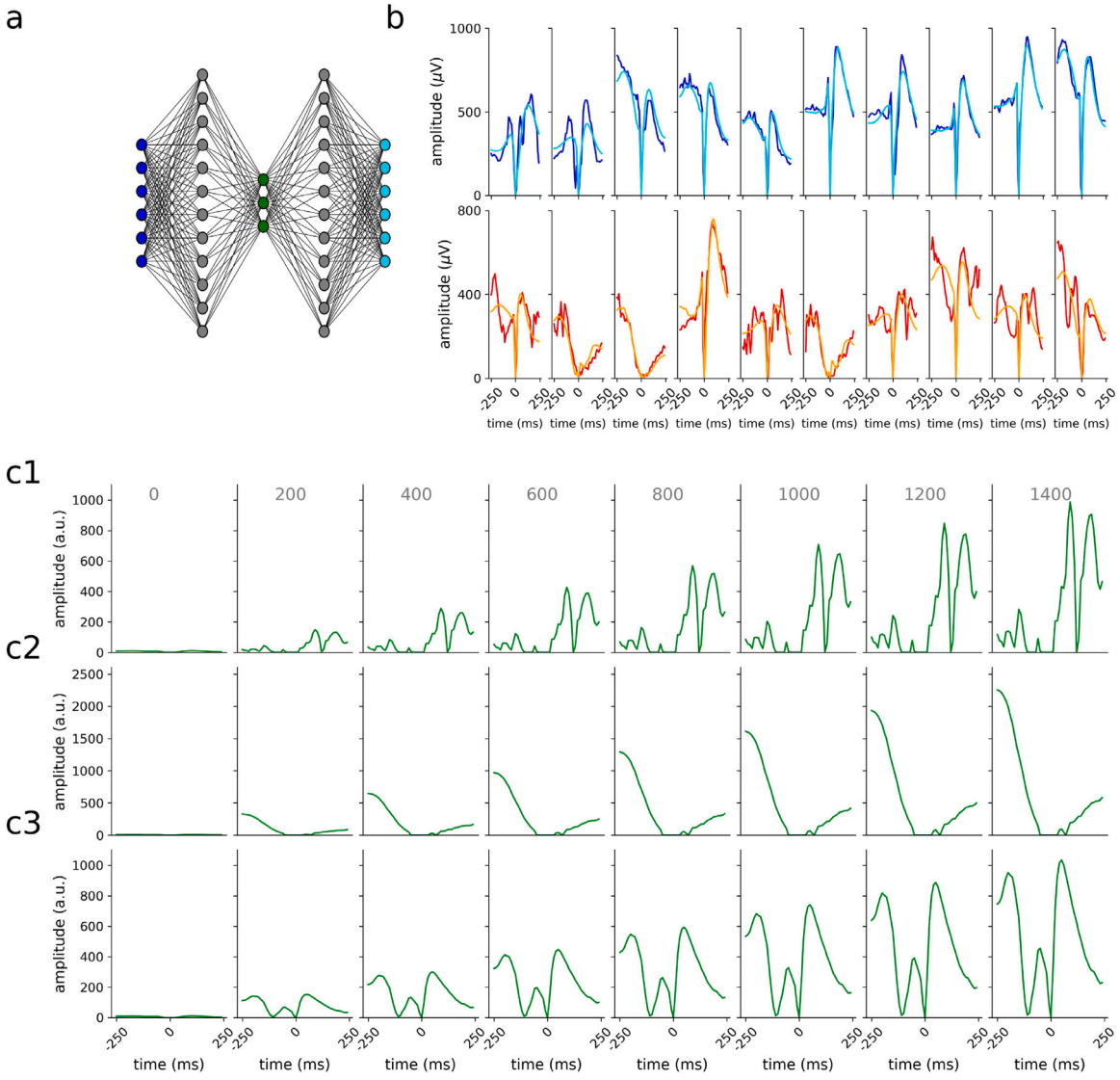

Fig. 4. Auto-Encoder for LFP events. a: Scheme of the used auto-encoder for data compression with input layer (dark blue), hidden layers (gray), bottleneck/encoding layer (green), and output layer (cyan). b1: Exemplary input training data (blue) and corresponding outputs/reconstructions (cyan). b2: Exemplary input test data (red) and corresponding outputs/reconstructions (orange). c1–c3: Output of the auto-encoder, for different unity vector activations in the bottleneck/encoding layer, (c1: (x,0,0), c2: (0,x,0), c3: (0,0,x) $x \in \{0, 200, 400, 800, 1000, 1200, 1400\}$). The resulting outputs (c1–c3) represent different degrees of expression of three fundamental complementary event shapes. Any concrete LFP event shape correspond to a weighted superposition of these three prototype shapes.

3.2. Calculation of embeddings (latent space encodings) of LFP events

In order to extract meaningful information from the LFP events (e.g. the shapes of LFP events) it is necessary to find an efficient representation of these shapes.

Therefore, we calculated so called feature space embeddings using an auto-encoder, i.e. an encoder–decoder-network. This type of artificial neural network has a special network architecture, where the input layer and the output layer are of the same dimension, i.e. consist of equal number of neurons. The network is trained in a self-supervised way, i.e. on reproducing the input in the output layer. Thus, the cost function is the mean-squared error between input and output (schematic drawing of an auto-encoder in Fig. 4a). A further important property of the auto-encoder is the so called bottleneck layer (green layer in Fig. 4a). This means that the intermediate layer is smaller, i.e. has fewer neurons, than the input and output layer. Hence, the auto-encoder has to compress (encode) the input, and subsequently to decompress (decode) the activity in the bottleneck layer in order to reconstruct the input as output. The activation patterns of the bottleneck layer are called latent space or feature space embeddings (encodings). The idea behind this processing principle is, that the network has to reduce the dimensionality of the input without loosing relevant information for subsequently reconstructing the input as output again (cyan neurons in 4a). The performance of a given auto-encoder can be assessed by comparing input and output (Fig. 4a). Over-fitting is prevented by adding dropout layers and the grade of over-fitting was determined by testing the trained network with an unknown test-data set (see Fig. 4b).

In order to find the best hyper-parameters for the auto-encoder, we performed a grid search with varying numbers of hidden layers and different widths for these layers. The results are summarized in Fig. 3. Despite minimal effects on the loss function from increasing the number and size of hidden layers, additional layers tended to cause overfitting. Consequently, we selected a configuration with one hidden layer and 700 neurons, suggesting that this smaller network may offer better generalization capabilities (see Fig. 3).

In General, auto-encoders are black boxes and it is not known which features of the input space are used to create the lower-dimensional latent space, i.e. what is the meaning of a particular dimension (neuron activity). Therefore, in order to make the encodings more interpretable,

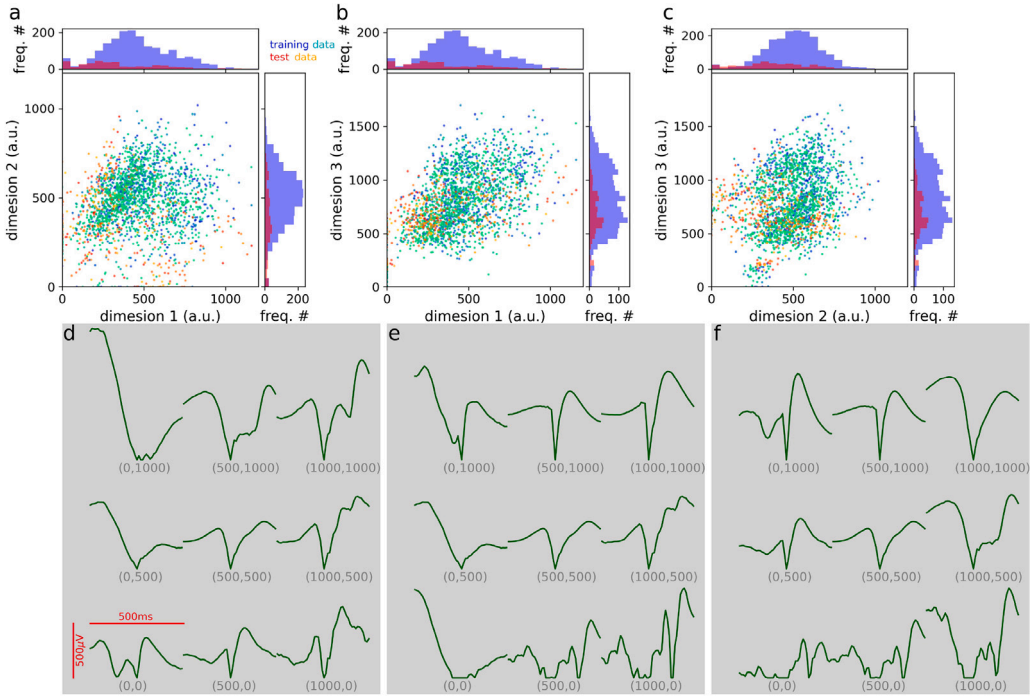

Fig. 5. Dimensionality reduction using the auto-encoder. a–c: 2-dimensional projections of the encoded 3-dimensional embeddings of LFP events for the training (blue–cyan) and the test data set (red–orange). Training and test data is spontaneous activity recorded for 1 h or 0.5 h respectively. The time of occurrence of the detected LFP events is color coded (blue to cyan for training data, red to yellow for test data). d–f: Histograms show the distribution of the encoded LFP events. The shapes of the LFP events for different encoding vectors (x,y,z-values $\in \{0, 500, 1000\}$). Note that, the plots do not show the exact reconstructed curve shapes since they are based on only 2 dimensions, whereas the embedding vector is 3-dimensional. For each column (a/d, b/e, c/f), we have set the respective missing third dimension to a constant medium range value of 500.

we systematically analyzed the emerging latent feature space embeddings. To this end, we directly activated the neurons in the bottleneck layer one by one without providing any input to the network. In particular, the activation vectors of the bottleneck layer were ((x,0,0), (0,x,0), or (0,0,x) with x$\in$ $\{0, 200, 400, 800, 1000, 1200, 1400\}$). These activation vectors were further propagated through the network to the output layer. The resulting outputs (Fig. 4c1–c3) indicate that each encoding neuron is specialized to certain LFP shapes. For example, neuron 2 seems to encode low-frequency LFP events, whereas neuron 3 is associated with high-amplitude W-shaped signals.

The dimensionality reduction through the auto-encoding can be used to analyze large electrophysiological data sets such as recordings of spontaneous activity for e.g. 1 hour (Fig. 5a–c). Note that, the encodings do neither depend on prior knowledge nor any labeled data. Actually, the auto-encoder only uses statistical features of the input such as the frequency of occurrence of certain LFP events. However, the encodings do not contain the full information of the underlying LFP event shapes, as certain information which is crucial to reconstruct the input is also stored in the decoding part of the neural network itself. Nevertheless, when certain clusters are identified, the shapes of this LFP events can be reconstructed using the decoder (examples of decoded inputs in Fig. 5d–f). The two processing steps: (1) finding certain clusters in the embeddings, and (2) identifying the underlying LFP event shapes, provide efficient data processing on the one hand, minimize the information loss and preserve biological interpretability.

As the auto-encoder is exclusively trained on statistical features, the trained network can be applied to any kind of LFP data. Thus, we tested the auto-encoder trained on spontaneous activity to find clusters in evoked activity. Therefore, we used the evoked activity shown in Fig. 1 to check, if the auto-encoder actually helps to extract meaningful information. We analyzed 4 different sound pressure levels (110 dB–80 dB) and 3 frequencies (500 Hz, 1000 Hz, 2000 Hz). We could show that indeed the LFP responses induced by different stimulus frequencies form clusters (Fig. 6a1–d1, for exemplary events and decoded LFP events see a–d 2–4). We quantified the quality of the clustering by calculating the generalized discrimination value (GDV) (Schilling et al., 2021b; Krauss et al., 2018a). We could show that the GDV is best for a stimulus intensity of 100 dB. Thus, too loud stimuli lead to unspecific activation due to recruiting of auditory nerve fibers. In contrast, too low stimulus intensities do not evoke clear responses. To put it in a nutshell, the auto-encoder trained on a different data set has proven to serve as universal tool to identify clusters of LFP events carrying the information on the stimulus frequency.

3.3. Estimation of information loss caused by auto-encoding using a classifier network

However, the dimensionality reduction by auto-encoding (Fig. 7a) obviously causes a loss of information which needs to be quantified. In a previous study, we already introduced a classifier network as a tool to objectively quantify the amount of meaningful information of certain data (Schilling et al., 2022). Training a classifier network (Fig. 7b) requires labeled data. We used a convolutional neural network which was trained on LFP responses induced by four different stimulus frequencies (500 Hz, 1000 Hz, 2000 Hz, 4000 Hz at 100 dB). To quantify the information loss due to auto-encoding/dimensionality reduction with respect to discriminability of different categories of LFP responses, two classifier networks were trained on stimulus frequencies. The original LFP events served as training input for the first classifier, whereas the second classifier was trained on the LFP events which were encoded and decoded by the auto-encoder. The evoked LFP events (raw and decoded) were split in 60% training LFP events and 40% validation LFP events, which were not used for training the classifier. Note that, the auto-encoder used in this experiment – used to encode and decode the evoked LFP events – was trained on data from a different animal (see Fig. 7c). This demonstrates that the auto-encoding method is universal and the auto-encoder do not have to be adjusted for each animal individually. This fact is emphasized by the comparison of unprocessed and reconstructed LFP events (see Fig. 7d).

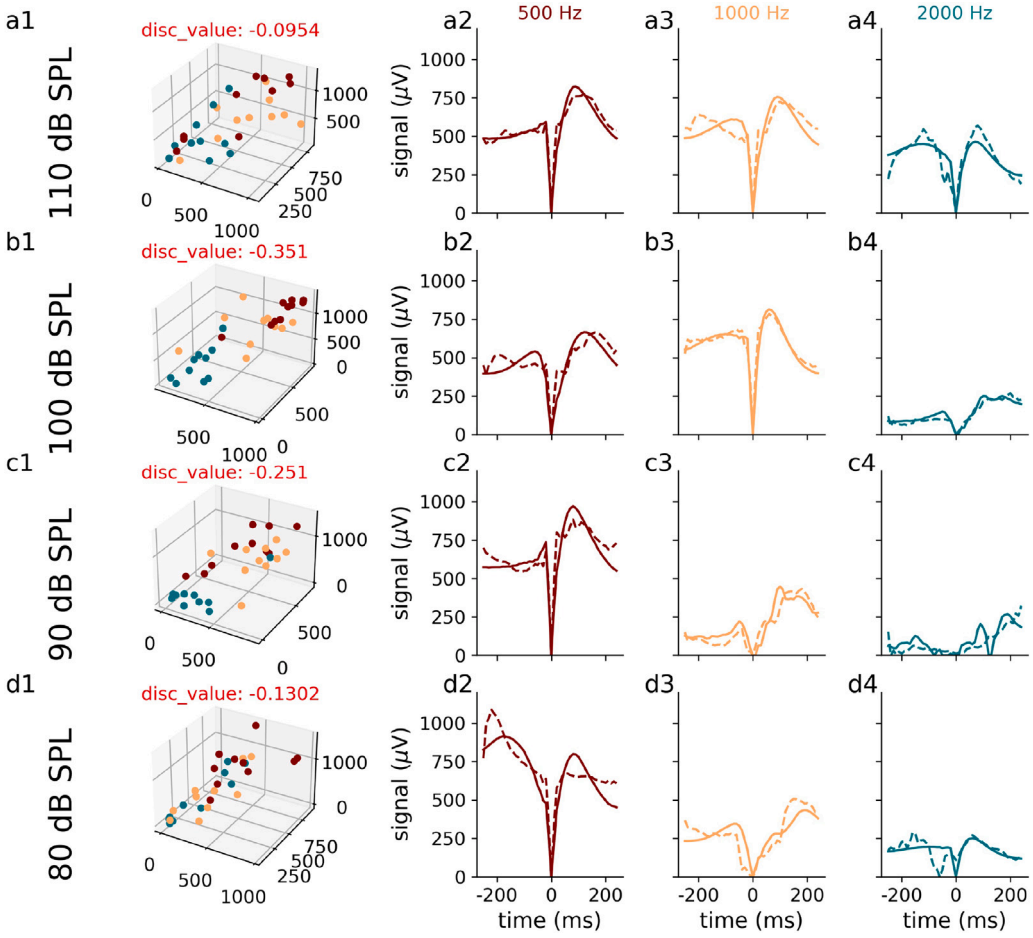

Fig. 6. Auto-Encoded LFPs from pure-tone stimulation. Column 1: LFP event shapes in 3-dimensional embedding space. Markers represent encoded LFP responses induced by pure tone stimuli (500 Hz, 1000 Hz, 2000 Hz, best frequency at 500 Hz, compare Fig. 1 electrode 3) for different stimulus intensities, i.e. sound pressure levels (a: 110 dB SPL, b: 100 dB SPL, c: 90 dB SPL, d: 80 dB SPL). The best separability, i.e. lowest generalized discrimination value (GDV, shown in red above plots) of the events can be observed for a optimum stimulus intensity of 100 dB. Louder stimuli cause further recruitment and the specificity to the stimuli is reduced, whereas lower stimulus intensities evoke a decreased signal intensity. Columns 2–4: Examples of input and corresponding reconstructed output shapes for the LFP events shown in a1–d1.

We could show that the classification accuracy (classes correspond to different stimulus frequencies) is slightly reduced for the reconstructed input compared to the unprocessed original input. Nevertheless, there is no large difference between validation accuracies (Fig. 7f). Furthermore, the auto-encoded input data lead to less over-fitting, as the neural network is not trained on specific features or artifacts unique to certain trials (see Fig. 7f). These features get lost during the process of auto-encoding since they are not relevant, i.e. carry no meaningful information, for reconstructing the input from the latent space embedding. The tendency of the auto-encoder to remember specific details rather than generalize is evident when comparing the training and validation accuracies of the classifiers. The classifier trained on raw LFP events shows a significant discrepancy between these accuracies, suggesting that it relies on non-stimulus-related information, likely artifacts, in the training data-set, and is consequently a special type of over-fitting. Conversely, the reduced discrepancy in the classifier trained on reconstructed data suggests effective noise removal by the auto-encoder, an advantage of auto-encoders already described e.g. by Hwaidi and Chen (2021).

To evaluate whether auto-encoding produces more meaningful embeddings of Local Field Potential (LFP) events compared to standard algorithms, we conducted a comparative analysis using an auto-encoder and two types of principal component analysis (PCA). The results are summarized in Fig. 8. In particular, we compared auto-encoder embeddings, PCA embeddings for each loudness level individually, and PCA embeddings for all loudness levels combined. Notably, the PCA was performed on the evoked dataset, whereas the auto-encoder was trained on spontaneous LFP events only. Despite this difference, the Generalized Discrimination Value (GDV) scores were lower (indicating better performance) for the auto-encoder embeddings. This demonstrates that the auto-encoder not only extracts significant and universal features of LFP events but also outperforms PCA in this context (see Fig. 8)

Furthermore, our analysis provides evidence that the information loss due to auto-encoding is acceptable and that the auto-encoder is a valid tool to reduce the dimensionality of the data. Up to now, the auto-encoding procedure was only applied to single trials of single channel data. In a next step, we show that the encoded data can be used to draw conclusions about information processing in the brain.

3.4. The relationship of spontaneous and evoked LFP events

Our research provides additional insight into the realm of neural response possibilities within the auditory cortex. Building upon Luczak et al.'s 2009 findings on population coding, we delve deeper into the subspace dynamics of LFP event embeddings. We illustrate that the embeddings of LFP events, when stimulated by pure tones, occupy a specific subspace within a broader state space. This broader state space is defined by the embeddings of spontaneous LFP events, as depicted in Fig. 9. Our findings underscore the brain's capacity to sample from a spectrum of potential stimulus-evoked event shapes during periods of spontaneous activity, suggesting a preconfigured

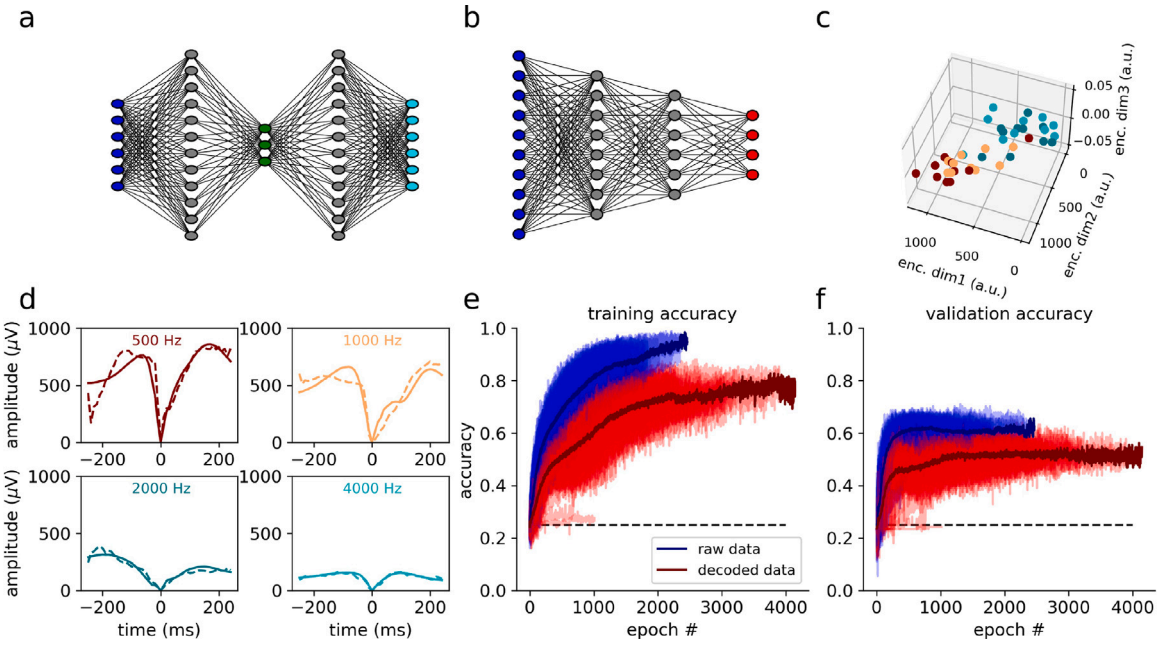

Fig. 7. Measurement of the meaningful information of auto-encoded events using a classifier network. a, b: Schematic representation of the two artificial neural networks. Note that, depicted network architectures are sketches, for detailed network parameters see Table 1. a: Auto-encoder network (compare Fig. 4a). b: The classifier network is designed to identify the frequency of the stimulus (500 Hz, 1000 Hz, 2000 Hz, 4000 Hz) that elicited the LFP event, based on the event's shape inputted into the network. The network was trained on 60% of evoked LFPs. The other 40% were used as validation data. This is true for the classifier trained on raw LFP events as well as reconstructed LFP events c: Reconstructed embeddings of one test data set (10 trials with 4 stimulation frequencies at a stimulation intensity of 100 dB SPL). d: Exemplary LFP responses for different stimulus frequencies (dashed line: measured signal, solid line: reconstructed signal from auto-encoder embeddings). e, f: Training accuracy (e) and validation accuracy (f) of the trained classifier network (dashed line: chance level, dark red/dark blue: average accuracy, red/blue: learning curves for 50 repetitions).

neural response template. Luczak and colleagues, in their 2009 study, analyzed the multi-channel spiking activity of several neurons through multi-electrode arrays, revealing similar phenomena in the context of population coding. Our study parallels and extends these findings by examining the cortical activity patterns through the lens of single-channel LFP event shapes. By doing so, we not only confirm their observations but also provide a novel perspective on how the brain's spontaneous activity mirrors its response to external stimuli. This analysis bridges the gap between spontaneous and stimulus-evoked neural dynamics, offering a more comprehensive understanding of cortical processing and its implications for sensory perception and brain function.

So far, we restricted our analyses on LFP events from a single recording channel. In the following, we show that we can even derive information about the temporal dynamics of information flow from LFP events by taking into account two spatially separated recording channels.

3.5. Auto-encoding, clustering and information flow

Therefore, we detected spontaneously occurring LFP events in a certain reference channel and analyzed the activity in a spatially separated, neighboring channel (example for events in reference channel in Fig. 10a blue curve and neighboring channel green).

The sign of the time difference between corresponding LFP events in two different channels indicates the direction of the information flow between the two recording sites. To quantify the time difference (latency) between corresponding events, the event-wise cross-correlation function between the two channels of interested was calculated (see Fig. 10b black curve). For each pair of events in the two channels, the lag-time which leads to the maximum cross-correlation coefficient corresponds to the respective time difference between the occurrence of the same event at the two different channels. It turns out that there is no general lag-time/latency value that fits to all pairs of events. Instead, spontaneous activity is characterized by a distribution of latencies as shown in Fig. 10c. This means that simply calculating the cross-correlation function over the complete LFP time series or spike trains of two recording channels as is frequently done as an objective function of the synchrony between two channels is insufficient. For instance, an increased synchrony could either be caused by a more balanced distribution of negative and positive time delays (lag-times) between corresponding LFP events, or by smaller absolute values of the time delays. This information gets lost by simply calculating cross-correlation functions for the entire time series of both channels. Furthermore, it turns out that the resulting cross-correlation coefficients are also broadly distributed between 0 and 1 (Fig. 10d), whereas there seems to be no clear dependence between resulting lag-time and maximum correlation coefficient. In Fig. 10e, a scatter plot of all pairs of cross-correlation coefficients and corresponding lag-times is shown. This indicates that the value of the cross-correlation coefficient is no reliable measure for the synchrony between two channels.

We took advantage of the previously calculated latent space embeddings of LFP event shapes to assess, if the time delays of LFP events between two channels correspond to the shape of the respective LFP event. Remarkably, the event shape embeddings, and consequently the event shapes, cluster according to the direction of information flow, i.e. the sign of the lag-time between two channels (Fig. 10f). Finally, we decoded the embeddings again to visualize the corresponding LFP event shapes of the channel of interest. We find that LFP event shapes characterized by larger amplitudes correspond to negative (blue) time delays, whereas broader shapes with smaller amplitudes correspond to positive (red) time delays, and hence indicate information flow in the opposite direction. Thus, it is possible to estimate the direction of information flow in the cortex by analyzing only a single recording channel. These findings indicate that it might be possible to assess changes in information flow direction, when the system is distorted e.g. by damages along the sensory pathway such as hearing loss caused by a noise trauma.

Indeed, noise trauma leads to changes of the information flow as seen in Fig. 11. Depending on the electrode position, the embedding

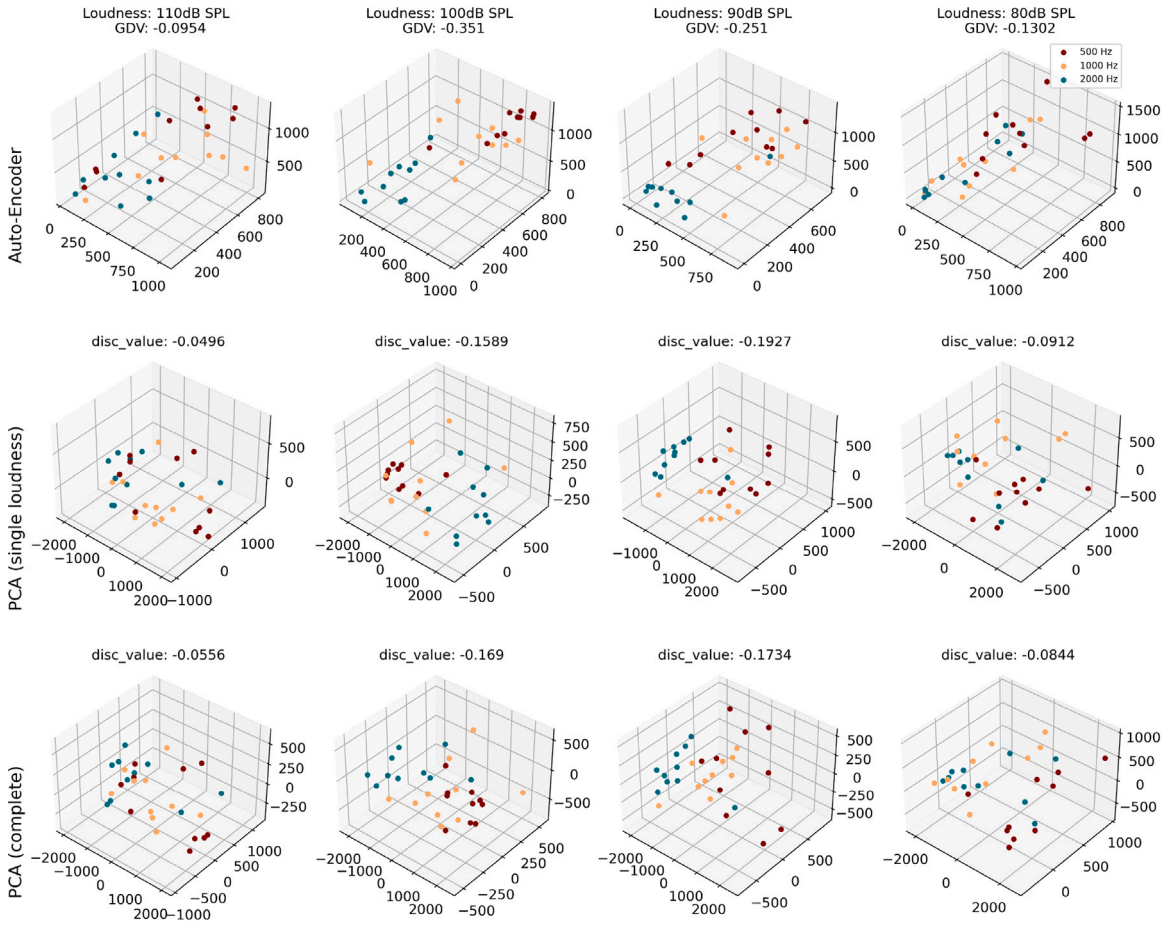

Fig. 8. Comparison Principal component analysis (PCA) and auto-encoding. To test, if auto-encoding produces more meaningful embeddings (encodings) of the LFP events than standard algorithms, we compared the embeddings of the evoked activity LFP events (shown in Fig. 6) created via the auto-encoder (first line) and two PCAs (second line: PCA for each loudness individually, third line: for all loudness levels together). The PCA was calculated on the basis of the evoked data set (shown in the figure). In contrast to that, the auto-encoder was trained on spontaneous LFP events only. However, the GDV values are lower (better) for the auto-encoder embeddings, although the PCA embeddings calculated on the basis of the here shown data set. This demonstrates that the auto-encoder extracts important and universal features of the LFP-events, which are related to auditory processing in the brain.

clusters changed when the silence condition (11a,b 1–2) is replaced by a 115 dB, 2 kHz auditory pure-tone trauma (11a,b 3–4).

3.6. Application of the method to human intra-cranial EEG (iEEG) data

The fact that the information flow can be determined by analyzing single channel LFP event shapes could be interesting for analyzing human intra-cranial EEG (iEEG) data. We applied the auto-encoding procedure on iEEG data recorded in the auditory cortex of a human epilepsy patient. Indeed it is possible to use the local-minimum search algorithm on the recorded data to identify LFP events (see 12a,b, events are marked by x). Besides spontaneous activity also evoked activity was recorded: currents between 1 mA and 15 mA were applied to certain channels of the recording device. We distinguish between stimulation of channels, which are not our recording channels (see red markers in Fig. 12a) and currents which cause artifacts (green markers Fig. 12a). Note that, in principle, events labeled as spontaneous events might nevertheless be evoked events since we cannot fully exclude that there were no sounds or other auditory stimuli in the room during recording. Thus, for this analysis evoked events are defined as events which are evoked by an intra-cranially applied stimulation current. We find that auto-encoding is again a valuable technique to extract meaningful data from the iEEG data.

First, we demonstrate that the auto-encoder network produces valid reconstructions as output compared to the input data (see Fig. 12c,d). Furthermore, evoked events lead to different LFP shapes than spontaneous LFP events. Even though, no distinct clusters in the embedding space could be observed, the median coordinate values of the three different conditions lead to three different reconstructed prototypical LFP event shapes (see Fig. 12f).

The inter-channel cross-correlation analysis (Fig. 13a,b) indicates that LFP event shape are correlated with lag-times between the channels. Thus, we find again a bi-modal lag-time distribution (Fig. 13c). Furthermore, a correlation between LFP event shape and lag-time can be observed, at least for lag-times with an absolute value larger than 10 ms, (Fig. 13d). Thus, the reconstruction of the average (prototypical) embeddings for lag-times larger than 10 ms and smaller than −10 ms (see Fig. 13e) indicate that at least the amplitude of the LFP events with negative lag times are increased (see also corresponding animal data in Figs. 10g and 11a2, b4).

4. Discussion

In the present study, we established an analytical pipeline specifically designed to identify, extract and characterize events within continuous local field potential (LFP) recordings. This pipeline was applied to data from an animal model (Meriones unguiculatus) and intracranial EEG (iEEG) recordings from a single human epilepsy patient.. We applied a local minimum search algorithm in combination with a thresholding procedure to identify significant LFP events. In a next step,

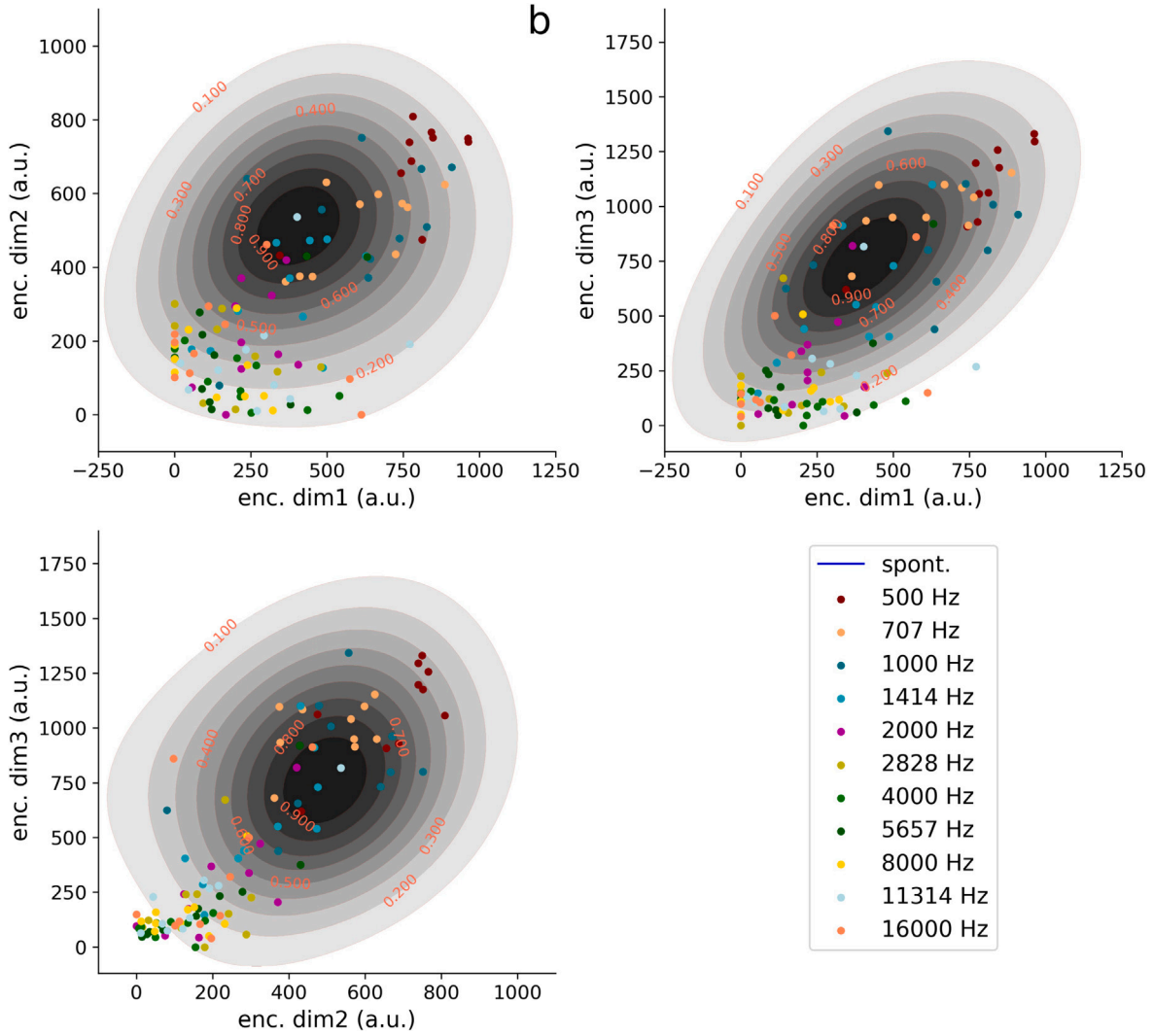

Fig. 9. Spontaneous LFP events outline the realm of possible evoked LFP events. Shown are three 2D projections of the 3D embedding space (a: dim 1 and dim 2, b: dim 1 and dim 3, c: dim 2 and dim 3). Markers represent auto-encoder embeddings of LFP events evoked by auditory stimuli of different frequencies (500 Hz–16 kHz, half octave steps, 100 dB SPL, 10 repetitions for each frequency). Spontaneous activity is shown as contour plots (black) representing the kernel density estimation of spontaneous LFP events.

the dimensionality of the LFP event shapes is reduced using an auto-encoder network. In its bottleneck layer, the auto-encoder provides a low-dimensional representation (embedding) of the input data, which conserves relevant information to reconstruct the original shapes again as output. These embeddings are used to visualize the data and to identify potential stimulus-related clusters. Note that, the clusters result from the properties of the electrophysiological recordings and are not due to any pre-defined labels. The embeddings were further used to show that the shape of the LFP events is correlated with the direction of the information flow between different recording sites/channels (see Fig. 10). In our example, sharp high-amplitude LFP events indicate that the source of the LFP event is located at the respective channel, whereas broad low-amplitude shapes indicate that the source is at the other channel. This means that LFP event shapes can be used to identify the location of sub-cortical input. It is important to note that the local field potential shapes are the most sharp in the input layer, specifically layer IV, of the auditory cortex. This layer is significant as it is the primary recipient of sensory inputs from the thalamus, a sub-cortical structure in the brain Schilling et al. (2023a). However, auto-encoder based dimensionality reduction leads to worse interpretability of the underlying LFP event shapes, as the low-dimensional representations are highly abstract and auto-encoders, as deep learning in general, suffer from the so called black box problem (Voosen, 2017). Therefore, we calculated prototypical embeddings by averaging over all embeddings belonging to a certain stimulus conditions. Subsequently, we reconstructed the corresponding typical LFP shapes. These prototypical LFP event shapes could be used to make further assessments on cortical information processing.

For instance, it turns out that the brain's preparation for sound is much more dynamic than previously thought. In particular, we have made a remarkable discovery by mapping the auditory cortex's response to stimuli, revealing a pre-configured neural landscape. We found that LFP event embeddings evoked by pure tones reside within a larger state space, which is typically occupied by spontaneous LFP event embeddings. This indicates that the brain's spontaneous activity already contains the spectrum of responses to potential stimuli. Luczak et al.'s 2009 study provided the foundation for this discovery (Luczak et al., 2009). They analyzed the multi-channel spiking activity of neurons using multi-electrode arrays, uncovering the brain's intrinsic coding mechanisms in response to external stimuli. However, our study extends this understanding by focusing on the shape of single-channel LFP events, revealing a more nuanced and comprehensive view of how the brain's inherent activity patterns mirror its response to the external world. This work not only confirms the earlier findings but also significantly expands our knowledge of cortical processing, offering new insights into sensory perception and neural dynamics. The fact

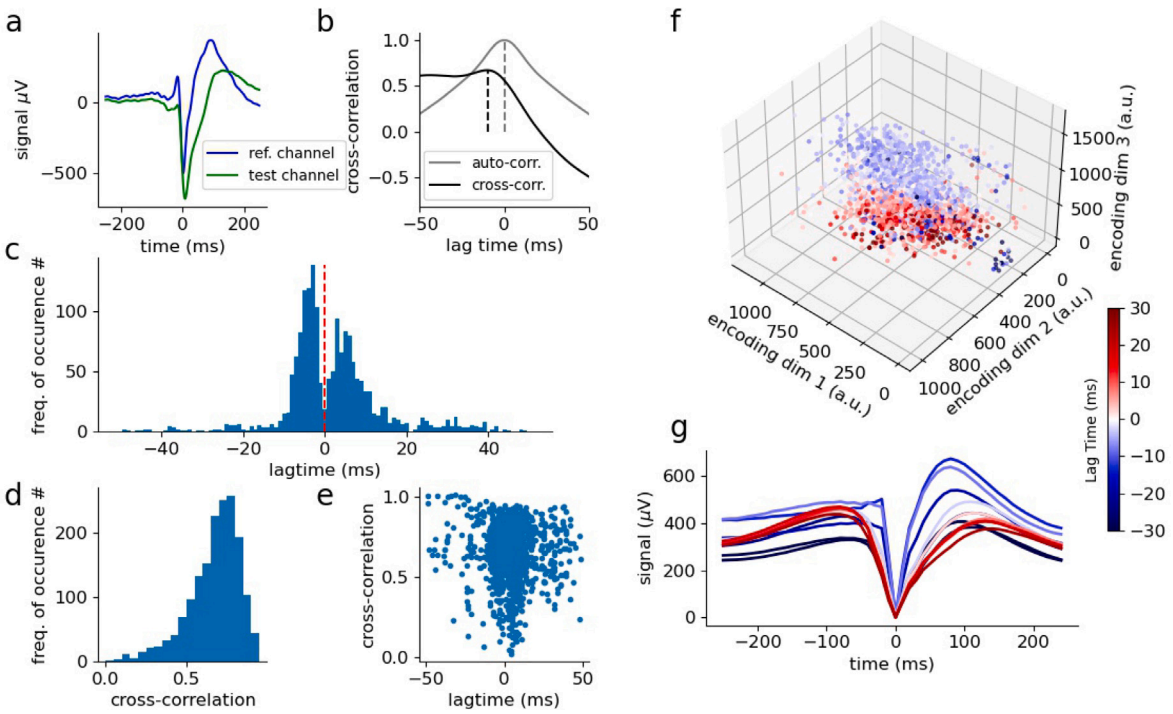

Fig. 10. Event-wise cross-correlation of spontaneous LFP events. a: LFP events in reference (blue) and in a neighboring channel (green). b: Auto-correlation of reference-channel (gray) and cross-correlation between channels (black). c: Histogram of detected lag-times. d: Histogram of maximum cross-correlation coefficients. e: Scatter plot of lag-times and corresponding cross-correlation coefficients. f: Encoding of all events of reference channel. Colors represent different lag times between reference channel and neighboring channel. g: Reconstructed event shapes from embeddings shown in f. Again, colors represent different lag-times as shown in f (blue: negative lag-times, red: positive lag-times). The sign of the lag-time indicates direction of information flow. Thus, blue represents input from thalamus and red represents input from other cortical areas. The LFP event shape embeddings cluster according to lag-times. Sharp asymmetric curves are related to thalamic input (blue curves g).

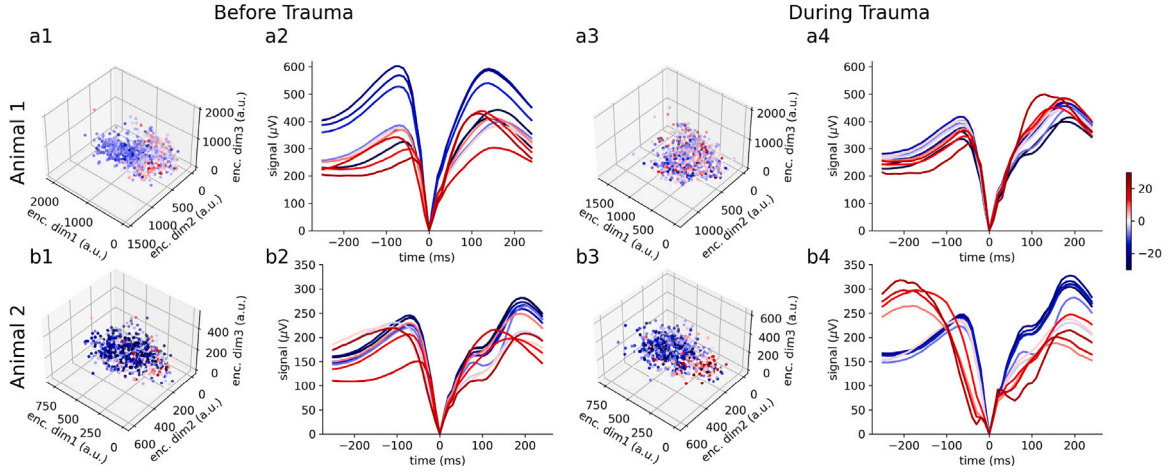

Fig. 11. Event-wise cross-correlation of spontaneous LFP events before and during noise trauma. a1: LFP event shape embeddings from an exemplary animal before noise trauma. a2: Reconstructed event shapes of embeddings from a1. a3–a4: Embeddings (a3) and corresponding reconstructions (a4) of LFP event shapes of the same animal shown in a1 and a2 but during application of an auditory noise trauma (2 kHz pure tone, 115 dB SPL, 60 min); b1–b4: Same as a1–a4 for a second animal.

that the summed activity, represented by the exact shape of the LFP events, is well suited to reproduce the results obtained by analyzing the spiking activity from multiple electrodes, shows that a precise analysis of these LFP events can open up great opportunities for cognitive and experimental neuroscience.

In our study, we leveraged the capabilities of autoencoders to analyze and interpret local field potential (LFP) data within the cerebral cortex. This approach not only allowed us to reduce the dimensionality of our high-density electrophysiological datasets but also enabled us to uncover meaningful patterns and dynamics within the neural signals. We opted for autoencoders due to their enhanced capability in capturing complex, non-linear relationships within the data, a critical aspect given the multifaceted nature of LFPs. In contrast, other methods like principle component analysis (PCA) are inherently limited to linear dimensionality reduction (Van Der Maaten et al., 2009; Abdi and Williams, 2010), which can oversimplify the neural data by failing to account for the nonlinear interactions among LFP components.

There are also other nonlinear techniques such as kernel PCA which extends the capabilities of conventional PCA by applying a kernel function to project data into a higher-dimensional space, enabling the capture of complex, nonlinear relationships within the data (Schölkopf et al., 1997). This approach provides a more flexible framework compared to standard PCA, allowing for the identification of patterns that are not linearly separable. However, kernel PCA still relies on the

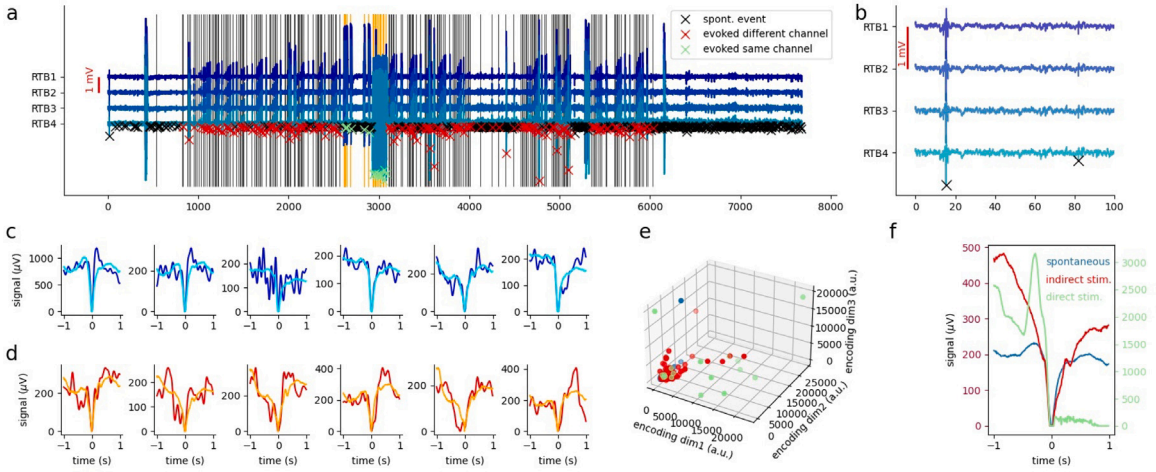

Fig. 12. Application of auto-encoding approach on human iEEG data. a: Intracranial recording (iEEG) in the auditory cortex of a human epilepsy patient (shown are four channels: RTB1-4). Markers (black, green, red) indicate LFP events identified by the local minimum search algorithm (see Methods). Red markers: Events evoked by stimulation currents induced in electrodes different from RTB1-4. Green markers:Events which were directly induced by current induction into the electrodes RTB1-4. Black markers: Spontaneous or auditory evoked events. b: Temporal zoom of time series shown in a. c, d: Examples of spontaneous LFP event shapes and corresponding reconstructions from auto-encoder embeddings of the training data set (blue: input, i.e. original recorded LFP events, cyan: reconstructed LFP events from embeddings) and the test data set (red: input LFP events, orange: reconstructions from embeddings). e: Embeddings of LFP events (blue: spontaneous, green: evoked by current in channels RTB1-4, red: evoked by current in different channel). f: Reconstructions calculated from median embeddings (from e) of the three different conditions. Note that, the reconstructions represent prototypical LFP shapes for the different conditions.

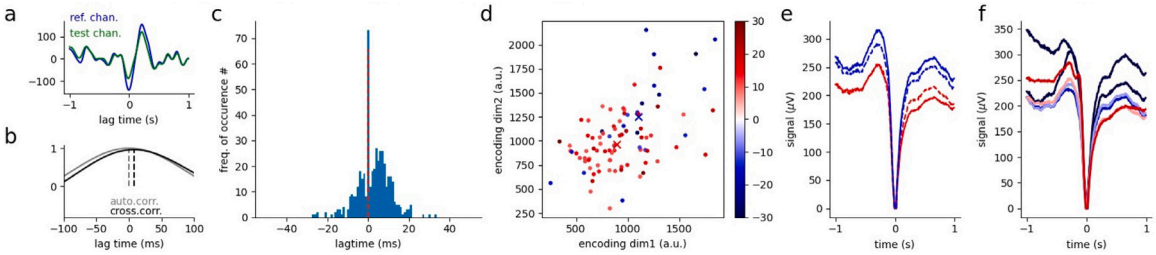

Fig. 13. iEEG cross-correlation analysis. a, b: Exemplary LFP events of channel RTB4 (green, reference channel) and channel RTB3 (blue, test channel) and corresponding cross correlation (b, black curve). The gray curve shows the auto-correlation of the reference channel. c: Distribution of the lag-times between the two channels (position of black peak). d: Embeddings of all spontaneous LFP events with lag-times smaller than −10 ms (shades of blue) or larger than 10 ms (shades of red). Markers (x) represent the average embeddings (blue cross: average of LFP events with latency smaller than −10 ms, red: >10 ms). f: Reconstructions corresponding to the mean embedding coordinate values shown in e (solid line: mean of LFP events with lag-times $in[-100,-10]\cup[10,100]$, dashed line: weighted mean over all events). l: Systematic reconstructions for different lag times.

choice of an appropriate kernel function, which may not always be straightforward. In our study, we opted for an autoencoder due to its inherent ability to learn complex, nonlinear mappings directly from the data without the need for a predefined kernel, offering a more adaptable and potentially powerful alternative for extracting meaningful features from Local Field Potential (LFP) events.

Our results clearly demonstrate that autoencoders are superior in capturing more informative and discriminative features for our specific type of data, which is crucial for accurately interpreting the complex neural dynamics inherent in LFP studies. The advantages of autoencoders over PCA in our context are manifold. First, autoencoders can model the higher-order interactions between variables that are typically present in neuroscientific data, which PCA might miss due to its linear nature. Second, autoencoders provide a more flexible architecture to learn representations that can be more abstract and nuanced compared to the principal components derived through PCA. This is particularly useful in neuroscience where the data may have underlying structures not alignable along linear axes. Third, unlike PCA which only focuses on variance maximization, autoencoders are trained to reconstruct the input data, minimizing reconstruction error. This process potentially preserves more useful information, even from less variant data aspects, which are often significant in biological interpretations. Finally, autoencoders can be easily extended to include regularization terms, such as sparsity or denoising, which help in learning more robust features from noisy neural data—a common challenge in LFP analyses. Hence, autoencoders, by design, excel in extracting hierarchical features from data, making them particularly suitable for neuroscientific applications where data complexity is high. The lower-dimensional embeddings produced by the bottleneck layer of the autoencoder effectively capture the essential characteristics of the input data while discarding noise and redundant information. This capability is critical when working with LFPs, which are often obscured by various sources of variability and noise. Indeed, recent studies have challenged the belief that local field potentials (LFPs) are confined to very small areas, finding instead that LFPs combine local activity with signals from distant sites, spreading passively over a centimeter or more. This broader integration range, which contrasts with the more localized measurements from current source density (CSD) and multiunit activity (MUA), suggests that LFPs have a more complex and extensive spatial influence than previously thought (Kajikawa and Schroeder, 2011; Lindén et al., 2011). The features learned by our autoencoder provide insights into the underlying network dynamics in several key ways. By compressing the LFP data into a compact representation, the autoencoder highlights the most salient features of the neural signals. These features often correspond to fundamental aspects of neuronal activity, such as synaptic inputs or network oscillations, that are crucial for understanding cortical processing. Furthermore, the process of decoding the embeddings back into the original data space allows us to interpret the types of neural activity patterns that dominate the recorded signals. For instance,

specific shapes and temporal patterns within the LFPs can be associated with particular cognitive processes or sensory inputs. Finally, the embeddings can reveal the direction and flow of information across the cortex by highlighting temporal relationships and synchrony between different regions. This is particularly evident in the way the embeddings cluster, suggesting pathways through which information travels during various cognitive states.

The physiological implications of our findings are profound. The autoencoder's ability to distill complex data into understandable and meaningful patterns provides a window into the brain's functional architecture. The patterns we identify are likely reflective of underlying connectivity and circuitry in the brain, offering clues about how different brain regions interact during various tasks or in different states. Additionally, changes in the embeddings over time or across different conditions can inform us about neural plasticity and how the brain adapts to new information or recovers from injury. For clinical applications such as epilepsy, where understanding the locus and network dynamics of seizures is crucial, our approach can pinpoint critical areas and network patterns that are involved in seizure genesis.

In summary, the benefit of using auto-encoders for processing LFP data can be divided into three major points. First, the dimensionality reduction allows for visualizing highly complex data sets. The fact that the embeddings can be reverse-engineered to prototypical LFP event shapes increases interpretability. Furthermore, there is another significant advantage of this procedure. The auto-encoder can be used to de-noise the data: as the auto-encoders are exclusively trained on statistical features of the LFP events, unique artifacts do not play a significant role. Thus, the encoding-decoding procedure can be used for artifact suppression (for some existing approaches see Hardcastle et al. (2019), Nishio et al. (2017), Li et al. (2021), De Coster et al. (2022)). A further central finding of the study is, that lag-times of LFP events measured between two channels are broadly distributed and this distribution is bi-modal, i.e. with two maxima for positive and negative lag-times, respectively (see Fig. 10c and Fig. 12i). Our data indicates that the different LFP event shapes correspond to different information processing mechanisms. Since the analysis of neural synchrony is an important target to investigate auditory processing in the cortex, and is of particular meaning for different phantom perceptions such as tinnitus theories (Tass and Popovych, 2012; Eggermont and Tass, 2015), this finding could be a starting point for further investigations. Indeed in most studies, synchrony between LFP streams from different channels is quantified by calculating the cross-correlation function between the entire signal streams, analyzing lag-times and the area under the cross-correlation curve (Eggermont et al., 2011). However, our findings that lag-times correspond to different LFP event shapes and that lag-time histograms show two maxima indicate that standard synchrony analyses fail to provide the full picture. Applying the here presented novel approach to further investigate the functional plasticity after hearing loss, which is hypothesized to be the cause of tinnitus (Schilling et al., 2023c; Krauss et al., 2019a; Schaette and McAlpine, 2011; Schilling et al., 2021d; Krauss et al., 2016; Schilling et al., 2020, 2021a, 2023b; Schilling and Krauss, 2022), might lead to a deeper understanding of the underlying processes.

In the following we discuss the data-related as well as methodological limitations of the study. A notable limitation of our study is the use of data from a single epilepsy patient. Obtaining intracranial EEG (iEEG) data is inherently difficult and requires the rare occurrence of epilepsy patients with specific brain region involvement. Nevertheless, the data from this single patient effectively served as a proof of principle, demonstrating the viability of our methods for analyzing such complex electro-physiological data. Additionally, the use of auto-encoder networks to analyze LFP events presents technical challenges that need to be considered when analyzing and interpreting the results. As we use the decoder part of the auto-encoder to generate reconstructions from average embeddings, we create novel LFP shapes and thus the decoder part of the network works as a generative model (Ullanat, 2020). Overfitting is an important problem for all machine learning applications (Ying, 2019; Gerum et al., 2020) and especially for generative models. When the auto-encoder network has too many trainable parameters negative side effects can occur. For instance, the network could store the whole information about different LFP shapes within the large weight matrix (Sun et al., 2016). Therefore, the weight matrix would serve as some kind of list or look-up table for different LFP shapes, and the embedding layer would just learn random labels (indices) for the content of this list. This unwanted case would cause the effect that the embeddings alone do not contain any useful information about the LFP shape, because it would be possible to train further auto-encoder with arbitrary permutations of the embeddings, which perform equally good. However, in order to allow an interpretation of clusters in embedding space, neighborhood relations between different LFP shapes should be conserved in the embedding layer (Hu et al., 2021). Thus, in our study we used shallow-networks to reduce the number of parameters, and added drop-out layers to prevent the neural network from overfitting. We could show that the encoder does not simply add labels to the different LFP-shapes, because the self-organized emerging clusters actually correspond to different stimulus conditions (see Fig. 6). As the neural network was not trained on these LFP event shapes yet from another data set and the clusters automatically emerge we could show that embeddings are not just random representations of the different LFP event shapes. Nevertheless, in a follow-up study, as a more sophisticated approach, variational auto-encoders (Kingma et al., 2019) could be used instead, which are further optimized to lead to better embeddings and thus to more interpretable decodings (Doersch, 2016; Girin et al., 2020).

Summing up, following the trend of integrating artificial intelligence and neuroscience (Marblestone et al., 2016; De Schutter, 2018; Richards et al., 2019; Tanaka et al., 2019; Krauss et al., 2019b; Krauss and Maier, 2020; Saxe et al., 2021; Schilling et al., 2022; Maier et al., 2022; Gerum et al., 2023; Stoll et al., 2023), machine learning provides valuable tools to extract information from electrophysiological data (Vogt, 2018; Storrs and Kriegeskorte, 2019; Mathis and Mathis, 2020). As described above in most studies the data is averaged over many measurement trials to increase the signal to noise-ratio. However, the frequently performed averaging procedure erases any correlates of information processing taking place during recording of the ongoing continuous signal stream. Potentially, it is possible to translate that stream of voltage fluctuations into signs, which could be interpreted by humans using approaches from deep learning based natural language processing such as machine translation, or even from *animal linguistics*, where e.g. killer whale calls are identified, segmented, extracted and classified from ongoing continuous sound streams according to recurring feature patterns (Bergler et al., 2019; Schröter et al., 2019; Bergler et al., 2020, 2021). By that, we are convinced that our approach might further push the progress in neuroscience in order to extract meaningful information from continuous electrophysiological data streams.

Data and code availability statement

The complete data and analysis programs will be made available upon reasonable request.

CRediT authorship contribution statement

Achim Schilling: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Richard Gerum:** Writing – original draft, Visualization, Validation, Software, Methodology, Formal analysis. **Claudia Boehm:** Data curation. **Jwan Rasheed:** Data curation. **Claus Metzner:** Writing – original draft, Validation. **Andreas Maier:** Writing – original draft, Validation. **Caroline Reindl:** Writing – original draft, Methodology, Data curation. **Hajo Hamer:** Writing – original draft, Resources, Methodology,

Data curation. **Patrick Krauss:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

The authors declare no competing interests.

Data availability

Data will be made available on request.

Acknowledgments

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation): grants KR 5148/2-1 (project number 436456810), KR 5148/3-1 (project number 510395418), KR 5148/5-1 (project number 542747151), and GRK 2839 (project number 468527017) to PK, and grant SCHI 1482/3-1 (project number 451810794) to AS. Furthermore, the research leading to these results has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (ERC Grant No. 810316 to AM).

References

Abdi, Hervé, Williams, Lynne J., 2010. Principal component analysis. Wiley Interdiscip. Rev.: Comput. Statist. 2 (4), 433–459.

Başar, E., Rosen, B., Başar-Eroglu, C., Greitschus, F., 1987. The associations between 40 hz-eeg and the middle latency response of the auditory evoked potential. Int. J. Neurosci. 33 (1–2), 103–117.

Bergler, Christian, Schmitt, Manuel, Maier, Andreas, Smeele, Simeon, Barth, Volker, Nöth, Elmar, 2020. Orca-clean: A deep denoising toolkit for killer whale communication. In: INTERSPEECH. pp. 1136–1140.

Bergler, Christian, Schmitt, Manuel, Maier, Andreas K., Symonds, Helena, Spong, Paul, Ness, Steven R., Tzanetakis, George, Nöth, Elmar, 2021. Orca-slang: An automatic multi-stage semi-supervised deep learning framework for large-scale killer whale call type identification. In: Interspeech. pp. 2396–2400.

Bergler, Christian, Schröter, Hendrik, Cheng, Rachael Xi, Barth, Volker, Weber, Michael, Nöth, Elmar, Hofer, Heribert, Maier, Andreas, 2019. Orca-spot: An automatic killer whale sound detection toolkit using deep learning. Sci. Rep. 9 (1), 1–17.

2021. BlackrockNeurotech. https://github.com/BlackrockNeurotech/Python-Utilities.

Bollmann, Saskia, Barth, Markus, 2021. New acquisition techniques and their prospects for the achievable resolution of fmri. Prog. Neurobiol. 207, 101936.

Bourlard, Hervé, Kabil, SelenHande, 2022. Autoencoders reloaded. Biol. Cybernet. 116 (4), 389–406.

Bukhtiyarova, Olga, Chauvette, Sylvain, Seigneur, Josée, Timofeev, Igor, 2022. Brain states in freely behaving marmosets. Sleep.

Buzsáki, György, Anastassiou, Costas A., Koch, Christof, 2012. The origin of extracellular fields and currents—eeg, ecog, lfp and spikes. Nat. Rev. Neurosci. 13 (6), 407–420.

Chollet, Francois, 2018. Deep Learning Mit Python und Keras: Das Praxis-Handbuch vom Entwickler der Keras-Bibliothek. MITP-Verlags GmbH & Co. KG.

Constantinou, Maria, Cogno, Soledad Gonzalo, Elijah, Daniel H., Kropff, Emilio, Gigg, John, Samengo, Inés, Montemurro, Marcelo A., 2016. Bursting neurons in the hippocampal formation encode features of lfp rhythms. Front. Comput. Neurosci. 10, 133.

Cunningham, John P., Byron, M. Yu, 2014. Dimensionality reduction for large-scale neural recordings. Nat. Neurosci. 17 (11), 1500–1509.

Dasgupta, Sanjoy, Stevens, Charles F., Navlakha, Saket, 2017. A neural algorithm for a fundamental computing problem. Science 358 (6364), 793–796.

De Coster, T., Kudryashova, N., Derevyanko, G., De Vries, A.A.F., Pijnappels, D.A., Panfilov, A.V., 2022. Identification of electrical rotational activity in noisy cardiac tissue recordings using a deep neural network. Europace 24 (Supplement_1), euac053–620.

De Schutter, Erik, 2018. Deep learning and computational neuroscience.

Doersch, Carl, 2016. Tutorial on variational autoencoders. arXiv preprint arXiv:1606.05908.

Eggermont, Jos J., Munguia, Raymundo, Pienkowski, Martin, Shaw, Greg, 2011. Comparison of lfp-based and spike-based spectro-temporal receptive fields and cross-correlation in cat primary auditory cortex. PLoS One 6 (5), e20046.

Eggermont, Jos J., Tass, Peter A., 2015. Maladaptive neural synchrony in tinnitus: origin and restoration. Front. Neurol. 6, 29.

Gajraj, R.J., Doi, M., Mantzaridis, H., Kenny, G.N., 1998. Analysis of the eeg bispectrum, auditory evoked potentials and the eeg power spectrum during repeated transitions from consciousness to unconsciousness. Br. J. Anaesth. 80 (1), 46–52.

Garibyan, Armine, Schilling, Achim, Boehm, Claudia, Zankl, Alexandra, Krauss, Patrick, 2022. Neural correlates of linguistic collocations during continuous speech perception. bioRxiv.

Gerum, Richard, 2019. Pylustrator: code generation for reproducible figures for publication. arXiv preprint arXiv:1910.00279.

Gerum, Richard C., Erpenbeck, André, Krauss, Patrick, Schilling, Achim, 2020. Sparsity through evolutionary pruning prevents neuronal networks from overfitting. Neural Netw. 128, 305–312.

Gerum, Richard, Erpenbeck, André, Krauss, Patrick, Schilling, Achim, 2023. Leaky-integrate-and-fire neuron-like long-short-term-memory units as model system in computational biology. In: 2023 International Joint Conference on Neural Networks. IJCNN, IEEE, pp. 1–9.

Girin, Laurent, Leglaive, Simon, Bie, Xiaoyu, Diard, Julien, Hueber, Thomas, Alameda-Pineda, Xavier, 2020. Dynamical variational autoencoders: A comprehensive review. arXiv preprint arXiv:2008.12595.

Golshan, Hosein M., Hebb, Adam O., Hanrahan, Sara J., Nedrud, Joshua, Mahoor, Mohammad H., 2016. A multiple kernel learning approach for human behavioral task classification using stn-lfp signal. In: 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. EMBC, IEEE, pp. 1030–1033.

Hagen, Espen, Dahmen, David, Stavrinou, Maria L., Lindén, Henrik, Tetzlaff, Tom, Van Albada, Sacha J., Grün, Sonja, Diesmann, Markus, Einevoll, Gaute T., 2016. Hybrid scheme for modeling local field potentials from point-neuron networks. Cerebral Cortex 1–36.

Hardcastle, Thomas J., Lee, Susannah, Wernisch, Lorenz, Fortier-Poisson, Pascal, Shunmugam, Sudha, Hewage, Kalon, Edwards, Tris, Armitage, Oliver, Hewage, Emil, 2019. Coordinate-vae: Unsupervised clustering and de-noising of peripheral nervous system data.

Hu, Youpeng, Li, Xunkai, Wang, Yujie, Wu, Yixuan, Zhao, Yining, Yan, Chenggang, Yin, Jian, Gao, Yue, 2021. Adaptive hypergraph auto-encoder for relational data clustering. IEEE Trans. Knowl. Data Eng..

Hunter, John D., 2007. Matplotlib: A 2d graphics environment. Comput. Sci. Eng. 9 (03), 90–95.

Hwaidi, Jamal F., Chen, Thomas M., 2021. A noise removal approach from eeg recordings based on variational autoencoders. In: 2021 13th International Conference on Computer and Automation Engineering. ICCAE, IEEE, pp. 19–23.

Jackson, Andrew, Hall, Thomas M., 2016. Decoding local field potentials for neural interfaces. IEEE Trans. Neural Syst. Rehabil. Eng. 25 (10), 1705–1714.

Kaiser, David A., 2007. What is quantitative eeg? J. Neurother. 10 (4), 37–52.

Kajikawa, Yoshinao, Schroeder, Charles E., 2011. How local is the local field potential? Neuron 72 (5), 847–858.

Keshtkaran, Mohammad Reza, Yang, Zhi, 2017. Noise-robust unsupervised spike sorting based on discriminative subspace learning with outlier handling. J. Neural Eng. 14 (3), 036003.

Kingma, Diederik P., Welling, Max, et al., 2019. An introduction to variational autoencoders. Found. Trends® Mach. Learn. 12 (4), 307–392.

Koelbl, Nikola, Schilling, Achim, Krauss, Patrick, 2023. Adaptive ica for speech eeg artifact removal. In: 2023 5th International Conference on Bio-Engineering for Smart Technologies. BioSMART, IEEE, pp. 1–4.

Kovac, Stjepana, Vakharia, Vejay N., Scott, Catherine, Diehl, Beate, 2017. Invasive epilepsy surgery evaluation. Seizure 44, 125–136.

Kraskov, Alexander, Quiroga, Rodrigo Quian, Reddy, Leila, Fried, Itzhak, Koch, Christof, 2007. Local field potentials and spikes in the human medial temporal lobe are selective to image category. J. Cognit. Neurosci. 19 (3), 479–492.

Krauss, Patrick, Maier, Andreas, 2020. Will we ever have conscious machines? Front. Comput. Neurosci. 116.

Krauss, Patrick, Metzner, Claus, Joshi, Nidhi, Schulze, Holger, Traxdorf, Maximilian, Maier, Andreas, Schilling, Achim, 2021. Analysis and visualization of sleep stages based on deep neural networks. Neurobiol. Sleep Circadian Rhythms 10, 100064.

Krauss, Patrick, Metzner, Claus, Lange, Janina, Lang, Nadine, Fabry, Ben, 2012. Parameter-free binarization and skeletonization of fiber networks from confocal image stacks. PLoS One 7 (5), e36575.

Krauss, Patrick, Metzner, Claus, Schilling, Achim, Tziridis, Konstantin, Traxdorf, Maximilian, Wollbrink, Andreas, Rampp, Stefan, Pantev, Christo, Schulze, Holger, 2018a. A statistical method for analyzing and comparing spatiotemporal cortical activation patterns. Sci. Rep. 8 (1), 1–9.

Krauss, Patrick, Schilling, Achim, Bauer, Judith, Tziridis, Konstantin, Metzner, Claus, Schulze, Holger, Traxdorf, Maximilian, 2018b. Analysis of multichannel eeg patterns during human sleep: a novel approach. Front. Hum. Neurosci. 12, 121.

Krauss, P., Schilling, A., Tziridis, K., Schulze, H., 2019a. Models of tinnitus development: From cochlea to cortex. HNO 67 (3), 172–177.

Krauss, Patrick, Schuster, Marc, Dietrich, Verena, Schilling, Achim, Schulze, Holger, Metzner, Claus, 2019b. Weight statistics controls dynamics in recurrent neural networks. PLoS One 14 (4), e0214541.

Krauss, Patrick, Tziridis, Konstantin, Metzner, Claus, Schilling, Achim, Hoppe, Ulrich, Schulze, Holger, 2016. Stochastic resonance controlled upregulation of internal noise after hearing loss as a putative cause of tinnitus-related neuronal hyperactivity. Front. Neurosci. 10, 597.

Kreiman, Gabriel, Hung, Chou P., Kraskov, Alexander, Quiroga, Rodrigo Quian, Poggio, Tomaso, DiCarlo, James J., 2006. Object selectivity of local field potentials and spikes in the macaque inferior temporal cortex. Neuron 49 (3), 433–445.

Kriegeskorte, Nikolaus, Kievit, Rogier A., 2013. Representational geometry: integrating cognition, computation, and the brain. Trends Cognit. Sci. 17 (8), 401–412.

Li, Jing, Yan, Jiaqing, Liu, Xianzeng, Ouyang, Gaoxiang, 2014. Using permutation entropy to measure the changes in eeg signals during absence seizures. Entropy 16 (6), 3049–3061.

Li, Xinyang, Zhang, Guoxun, Wu, Jiamin, Zhang, Yuanlong, Zhao, Zhifeng, Lin, Xing, Qiao, Hui, Xie, Hao, Wang, Haoqian, Fang, Lu, et al., 2021. Reinforcing neuron extraction and spike inference in calcium imaging using deep self-supervised denoising. Nat. Methods 18 (11), 1395–1400.

Lindén, Henrik, Tetzlaff, Tom, Potjans, Tobias C., Pettersen, Klas H., Grün, Sonja, Diesmann, Markus, Einevoll, Gaute T., 2011. Modeling the spatial reach of the lfp. Neuron 72 (5), 859–872.

Logothetis, Nikos K., 2003. The underpinnings of the bold functional magnetic resonance imaging signal. J. Neurosci. 23 (10), 3963–3971.

Luczak, Artur, Barthó, Peter, Harris, Kenneth D., 2009. Spontaneous events outline the realm of possible sensory responses in neocortical populations. Neuron 62 (3), 413–425.

Mackevicius, Emily L., Bahle, Andrew H., Williams, Alex H., Gu, Shijie, Denisenko, Natalia I., Goldman, Mark S., Fee, Michale S., 2019. Unsupervised discovery of temporal sequences in high-dimensional datasets, with applications to neuroscience. eLife 8, e38471.

Mahmud, Mufti, Cecchetto, Claudia, Vassanelli, Stefano, 2016. An automated method for characterization of evoked single-trial local field potentials recorded from rat barrel cortex under mechanical whisker stimulation. Cogn. Comput. 8 (5), 935–945.

Maier, Andreas, Köstler, Harald, Heisig, Marco, Krauss, Patrick, Yang, Seung Hee, 2022. Known operator learning and hybrid machine learning in medical imaging—a review of the past, the present, and the future. Prog. Biomed. Eng..

Marblestone, Adam H., Wayne, Greg, Kording, Konrad P., 2016. Toward an integration of deep learning and neuroscience. Front. Comput. Neurosci. 94.

Mathis, Mackenzie Weygandt, Mathis, Alexander, 2020. Deep learning tools for the measurement of animal behavior in neuroscience. Curr. Opin. Neurobiol. 60, 1–11.

Meier, Burkhard, 2019. Python GUI Programming Cookbook: Develop Functional and Responsive User Interfaces with Tkinter and PyQt5. Packt Publishing Ltd.

Metzner, Claus, Schilling, Achim, Traxdorf, Maximilian, Schulze, Holger, Krauss, Patrick, 2021. Sleep as a random walk: a super-statistical analysis of eeg data across sleep stages. Commun. Biol. 4 (1), 1–11.

Metzner, Claus, Schilling, Achim, Traxdorf, Maximilian, Schulze, Holger, Tziridis, Konstantin, Krauss, Patrick, 2023. Extracting continuous sleep depth from eeg data without machine learning. Neurobiol. Sleep Circadian Rhythms 14, 100097.

Mormann, Florian, Kornblith, Simon, Cerf, Moran, Ison, Matias J., Kraskov, Alexander, Tran, Michelle, Knieling, Simeon, Quiroga, Rodrigo Quian, Koch, Christof, Fried, Itzhak, 2017. Scene-selective coding by single neurons in the human parahippocampal cortex. Proc. Natl. Acad. Sci. 114 (5), 1153–1158.

Newson, Jennifer J., Thiagarajan, Tara C., 2019. Eeg frequency bands in psychiatric disorders: a review of resting state studies. Front. Hum. Neurosci. 12, 521.

Nishio, Mizuho, Nagashima, Chihiro, Hirabayashi, Saori, Ohnishi, Akinori, Sasaki, Kaori, Sagawa, Tomoyuki, Hamada, Masayuki, Yamashita, Tatsuo, 2017. Convolutional auto-encoder for image denoising of ultra-low-dose ct. Heliyon 3 (8), e00393.

Nurse, Ewan, Mashford, Benjamin S., Yepes, Antonio Jimeno, Kiral-Kornek, Isabell, Harrer, Stefan, Freestone, Dean R., 2016. Decoding eeg and lfp signals using deep learning: heading truenorth. In: Proceedings of the ACM International Conference on Computing Frontiers. pp. 259–266.

Pang, Rich, Lansdell, Benjamin J., Fairhall, Adrienne L., 2016. Dimensionality reduction in neuroscience. Curr. Biol. 26 (14), R656–R660.

Pang, Bo, Nijkamp, Erik, Wu, Ying Nian, 2020. Deep learning with tensorflow: A review. J. Educ. Behav. Stat. 45 (2), 227–248.

Ran, Xuming, Zhang, Jie, Ye, Ziyuan, Wu, Haiyan, Xu, Qi, Zhou, Huihui, Liu, Quanying, 2021. Deep auto-encoder with neural response. arXiv preprint arXiv:2111.15309.

Richards, Blake A., Lillicrap, Timothy P., Beaudoin, Philippe, Bengio, Yoshua, Bogacz, Rafal, Christensen, Amelia, Clopath, Claudia, Costa, Rui Ponte, Berker, Archyde, Ganguli, Surya, et al., 2019. A deep learning framework for neuroscience. Nat. Neurosci. 22 (11), 1761–1770.

Saxe, Andrew, Nelli, Stephanie, Summerfield, Christopher, 2021. If deep learning is the answer, what is the question? Nat. Rev. Neurosci. 22 (1), 55–67.

Schaette, Roland, McAlpine, David, 2011. Tinnitus with a normal audiogram: physiological evidence for hidden hearing loss and computational model. J. Neurosci. 31 (38), 13452–13457.

Schilling, Achim, Choi, Byunghee, Parameshwarappa, Vinay, Norena, Arnaud J., 2023a. Offset responses in primary auditory cortex are enhanced after notched noise stimulation. J. Neurophysiol. 129 (5), 1114–1126.

Schilling, Achim, Gerum, Richard, Krauss, Patrick, Metzner, Claus, Tziridis, Konstantin, Schulze, Holger, 2019. Objective estimation of sensory thresholds based on neurophysiological parameters. Front. Neurosci. 13, 481.

Schilling, Achim, Gerum, Richard, Metzner, Claus, Maier, Andreas, Krauss, Patrick, 2022. Intrinsic noise improves speech recognition in a computational model of the auditory pathway. Front. Neurosci. 795.

Schilling, Achim, Gerum, Richard, Zankl, Alexandra, Metzner, Claus, Maier, Andreas, Krauss, Patrick, 2020. Intrinsic noise improves speech recognition in a computational model of the auditory pathway. bioRxiv.

Schilling, Achim, Krauss, Patrick, 2022. Tinnitus is associated with improved cognitive performance and speech perception–can stochastic resonance explain? Front. Aging Neurosci. 14, 1073149.

Schilling, Achim, Krauss, Patrick, Gerum, Richard, Metzner, Claus, Tziridis, Konstantin, Schulze, Holger, 2017. A new statistical approach for the evaluation of gap-prepulse inhibition of the acoustic startle reflex (gpias) for tinnitus assessment. Front. Behav. Neurosci. 11, 198.

Schilling, A., Krauss, P., Hannemann, R., Schulze, H., Tziridis, K., 2021a. Reduktion der tinnituslautstärke: Pilotstudie zur abschwächung von tonalem tinnitus mit schwellennahem, individuell spektral optimiertem rauschen. HNO 69 (11), 891.

Schilling, Achim, Maier, Andreas, Gerum, Richard, Metzner, Claus, Krauss, Patrick, 2021b. Quantifying the separability of data classes in neural networks. Neural Netw. 139, 278–293.

Schilling, Achim, Schaette, Roland, Sedley, William, Gerum, Richard Carl, Maier, Andreas K., Krauss, Patrick, 2023b. Auditory perception and phantom perception in brains, minds and machines. Front. Neurosci. 17, 1293552.

Schilling, Achim, Sedley, William, Gerum, Richard, Metzner, Claus, Tziridis, Konstantin, Maier, Andreas, Schulze, Holger, Zeng, Fan-Gang, Friston, Karl J., Krauss, Patrick, 2023c. Predictive coding and stochastic resonance as fundamental principles of auditory phantom perception. Brain awad255.

Schilling, Achim, Tomasello, Rosario, Henningsen-Schomers, Malte R., Zankl, Alexandra, Surendra, Kishore, Haller, Martin, Karl, Valerie, Uhrig, Peter, Maier, Andreas, Krauss, Patrick, 2021. Analysis of continuous neuronal activity evoked by natural speech with computational corpus linguistics methods. Lang. Cognit. Neurosci. 36 (2), 167–186.

Schilling, Achim, Tziridis, Konstantin, Schulze, Holger, Krauss, Patrick, 2021d. The stochastic resonance model of auditory perception: A unified explanation of tinnitus development, zwicker tone illusion, and residual inhibition. Prog. Brain Res. 262, 139–157.

Schölkopf, Bernhard, Smola, Alexander, Müller, Klaus-Robert, 1997. Kernel principal component analysis. In: International Conference on Artificial Neural Networks. Springer, pp. 583–588.

Schröter, Hendrik, Nöth, Elmar, Maier, Andreas, Cheng, Rachael, Barth, Volker, Bergler, Christian, 2019. Segmentation, classification, and visualization of orca calls using deep learning. In: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP, IEEE, pp. 8231–8235.

Schüller, Alina, Schilling, Achim, Krauss, Patrick, Rampp, Stefan, Reichenbach, Tobias, 2023. Attentional modulation of the cortical contribution to the frequency-following response evoked by continuous speech. J. Neurosci. 43 (44), 7429–7440.

Schüller, Alina, Schilling, Achim, Krauss, Patrick, Reichenbach, Tobias, 2024. The early subcortical response at the fundamental frequency of speech is temporally separated from later cortical contributions. J. Cogn. Neurosci. 36 (3), 475–491.

Staresina, Bernhard P., Bergmann, Til Ole, Bonnefond, Mathilde, Van Der Meij, Roemer, Jensen, Ole, Deuker, Lorena, Elger, Christian E., Axmacher, Nikolai, Fell, Juergen, 2015. Hierarchical nesting of slow oscillations, spindles and ripples in the human hippocampus during sleep. Nat. Neurosci. 18 (11), 1679–1686.

Stoll, Andreas, Maier, Andreas, Krauss, Patrick, Gerum, Richard, Schilling, Achim, 2023. Coincidence detection and integration behavior in spiking neural networks. Cogn. Neurodyn. 1–13.

Storrs, Katherine R., Kriegeskorte, Nikolaus, 2019. Deep learning for cognitive neuroscience. arXiv preprint arXiv:1903.01458.

Sun, Wenjun, Shao, Siyu, Zhao, Rui, Yan, Ruqiang, Zhang, Xingwu, Chen, Xuefeng, 2016. A sparse auto-encoder-based deep neural network approach for induction motor faults classification. Measurement 89, 171–178.

Surendra, Kishore, Schilling, Achim, Stoewer, Paul, Maier, Andreas, Krauss, Patrick, 2023. Word class representations spontaneously emerge in a deep neural network trained on next word prediction. arXiv preprint arXiv:2302.07588.

Tanaka, Hidenori, Nayebi, Aran, Maheswaranathan, Niru, McIntosh, Lane, Baccus, Stephen, Ganguli, Surya, 2019. From deep learning to mechanistic understanding in neuroscience: the structure of retinal prediction. Adv. Neural Inf. Process. Syst. 32.

Tass, Peter A., Popovych, Oleksandr V., 2012. Unlearning tinnitus-related cerebral synchrony with acoustic coordinated reset stimulation: theoretical concept and modelling. Biol. Cybernet. 106 (1), 27–36.

Tonner, P.H., Bein, B., 2006. Classic electroencephalographic parameters: median frequency, spectral edge frequency etc. Best Pract. Res. Clin. Anaesthesiol. 20 (1), 147–159.

Ullanat, Varun, 2020. Variational autoencoder as a generative tool to produce de-novo lead compounds for biological targets. In: 2020 14th International Conference on Innovations in Information Technology. IIT, IEEE, pp. 102–107.

Van Der Maaten, Laurens, Postma, Eric O., van den Herik, H. Jaap, et al., 2009. Dimensionality reduction: A comparative review. J. Mach. Learn. Res. 10 (66–71), 13.

Van Der Walt, Stefan, Chris Colbert, S., Varoquaux, Gael, 2011. The numpy array: a structure for efficient numerical computation. Comput. Sci. Eng. 13 (2), 22–30.

Virtanen, Pauli, Gommers, Ralf, Oliphant, Travis E., Haberland, Matt, Reddy, Tyler, Cournapeau, David, Burovski, Evgeni, Peterson, Pearu, Weckesser, Warren, Bright, Jonathan, et al., 2020. Scipy 1.0: fundamental algorithms for scientific computing in python. Nat. Methods 17 (3), 261–272.

Vogt, Nina, 2018. Machine learning in neuroscience. Nat. Methods 15 (1), 33–33.

Voosen, Paul, 2017. The ai detectives.

Wang, Zhisong, Maier, Alexander, Leopold, David A., Logothetis, Nikos K., Liang, Hualou, 2007. Single-trial evoked potential estimation using wavelets. Comput. Biol. Med. 37 (4), 463–473.

Wang, Yasi, Yao, Hongxun, Zhao, Sicheng, 2016. Auto-encoder based dimensionality reduction. Neurocomputing 184, 232–242.

Waterstraat, Gunnar, Körber, Rainer, Storm, Jan-Hendrik, Curio, Gabriel, 2021. Noninvasive neuromagnetic single-trial analysis of human neocortical population spikes. Proc. Natl. Acad. Sci. 118 (11).

Yang, Zijin, Schilling, Achim, Maier, Andreas, Krauss, Patrick, 2021. Neural networks with fixed binary random projections improve accuracy in classifying noisy data. In: Bildverarbeitung FÜr Die Medizin 2021. Springer, pp. 211–216.

Ying, Xue, 2019. An overview of overfitting and its solutions. In: Journal of Physics: Conference Series. Vol. 1168, IOP Publishing, 022022.

Yoo, Jae-Chern, Han, Tae Hee, 2009. Fast normalized cross-correlation. Circuits Syst. Signal Process. 28 (6), 819–843.

Zhou, Ding, Wei, Xue-Xin, 2020. Learning identifiable and interpretable latent models of high-dimensional neural activity using pi-vae. Adv. Neural Inf. Process. Syst. 33, 7234–7247.