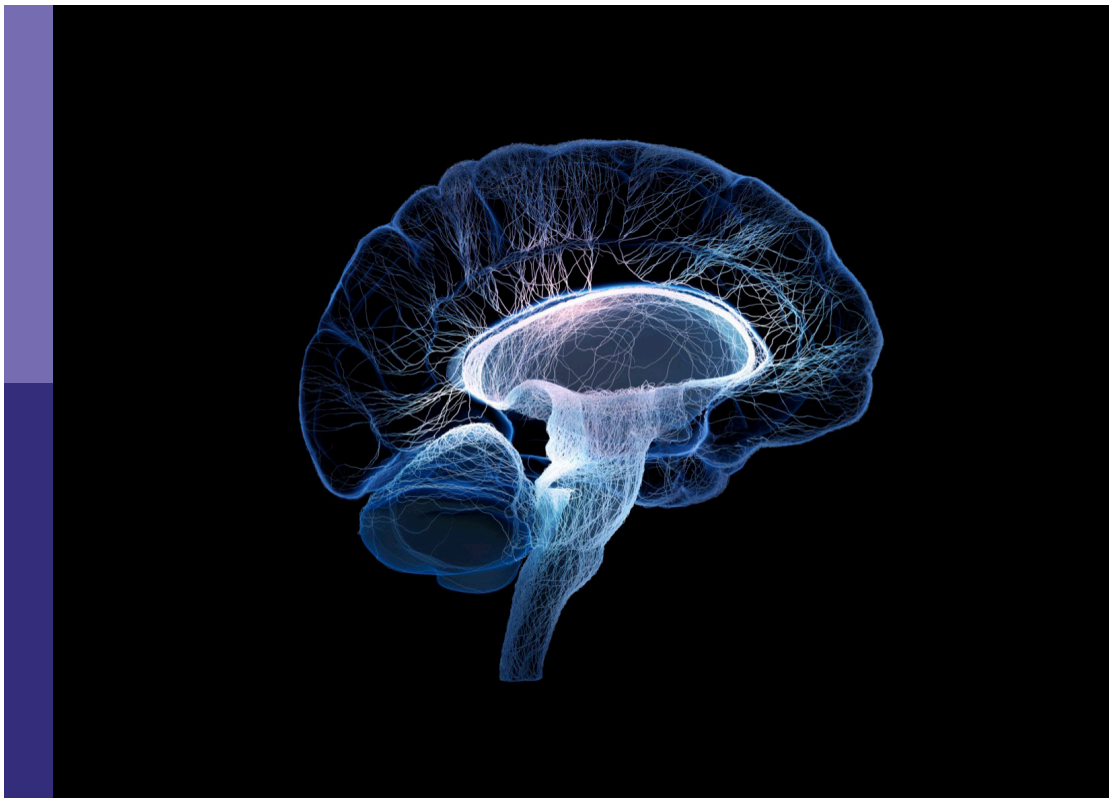
frontiers | Research Topics

Auditory perception and phantom perception in brains, minds and machines

Edited by
Achim Schilling, Patrick Krauss, Andreas K. Maier, Roland Schaette, William Sedley and Richard Carl Gerum

Published in
Frontiers in Neuroscience
Frontiers in Human Neuroscience
Frontiers in Psychology



ISSN 1664-8714
ISBN 978-2-8325-3756-5
DOI 10.3389/978-2-8325-3756-5

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Auditory perception and phantom perception in brains, minds and machines

Topic editors

Achim Schilling — University Hospital Erlangen, Germany
Patrick Krauss — University of Erlangen Nuremberg, Germany
Andreas K. Maier — University of Erlangen Nuremberg, Germany
Roland Schaette — University College London, United Kingdom
William Sedley — Newcastle University, United Kingdom
Richard Carl Gerum — York University, Canada

Citation

Schilling, A., Krauss, P., Maier, A. K., Schaette, R., Sedley, W., Gerum, R. C., eds. (2023). *Auditory perception and phantom perception in brains, minds and machines*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-3756-5

Table of contents

frontiers | Frontiers in Neuroscience

TYPE Editorial
PUBLISHED 06 October 2023
DOI 10.3389/fnins.2023.1293552

OPEN ACCESS

EDITED AND REVIEWED BY
Robert J. Zatorre,
McGill University, Canada

*CORRESPONDENCE
Achim Schilling
achim.schilling@fau.de

RECEIVED 13 September 2023
ACCEPTED 25 September 2023
PUBLISHED 06 October 2023

CITATION
Schilling A, Schaette R, Sedley W, Gerum RC, Maier A and Krauss P (2023) Editorial: Auditory perception and phantom perception in brains, minds and machines.
Front. Neurosci. 17:1293552.
doi: 10.3389/fnins.2023.1293552

Editorial: Auditory perception and phantom perception in brains, minds and machines

Achim Schilling[1,2]*, Roland Schaette[3], William Sedley[4], Richard Carl Gerum[5], Andreas Maier[6] and Patrick Krauss[1,2]

[1]Neuroscience Lab, University Hospital Erlangen, Erlangen, Germany, [2]Cognitive Computational Neuroscience Group, University Erlangen-Nürnberg, Erlangen, Germany, [3]UCL Ear Institute, University College London, London, United Kingdom, [4]Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom, [5]Department of Physics and Astronomy, York University, Toronto, ON, Canada, [6]Pattern Recognition Lab, University Erlangen-Nürnberg, Erlangen, Germany

KEYWORDS

tinnitus, deep neural networks, artificial intelligence, auditory neuroscience, auditory phantom perception, cognitive computational neuroscience, neuroimaging, hearing loss

Editorial on the Research Topic
Auditory perception and phantom perception in brains, minds and machines

Background and aim

The idea of combining artificial intelligence, in particular deep neural networks, and computational modeling with neuroscience and cognitive sciences has recently gained popularity leading to a novel research philosophy (Kietzmann et al., 2017; Kriegeskorte and Douglas, 2018; Naselaris et al., 2018). Regarding the auditory system, this interdisciplinary approach and close cooperation with computational sciences, would open up many interesting opportunities for auditory neuroscience. In analogy to lesion studies, phantom perceptions may serve as a vehicle to understand the fundamental processing principles underlying auditory perception (Schilling et al., 2023). The prime example of an auditory phantom perception is subjective tinnitus, i.e., the perception of a sound without any physical sound source involved (Krauss et al., 2016). Impairments of the auditory system due to cochlear damage, in particular synaptopathy (Tziridis et al., 2021), can lead to subjective tinnitus. Approximately 12% of the population are affected by tinnitus and 2% suffer heavily from that phantom perception, which causes severe side-effects ranging from concentration difficulties up to depression and suicide (Cederroth et al., 2020). A recent study estimated that in Germany alone the annual socio-economic costs of tinnitus are approximately 20 billion euros, which is in the same order of magnitude as the socio-economic costs of diabetes (Tziridis et al., 2022). Unfortunately, the underlying mechanisms of tinnitus are not yet fully understood and thus the development of accurate therapy approaches is difficult. The fact that tinnitus is associated with several mal-adaptations of the brain in different brain regions ranging from the brainstem up to the auditory cortex, makes it difficult to identify the exact underlying mechanisms (see e.g., Schilling et al., 2023). However, novel tools in neuroscience as well as recent progress in artificial intelligence and computational science provide novel approaches to unravel the mysteries of impaired auditory processing in the brain.

The aim of this Research Topic was to bring together researchers from different scientific fields such as neuroscience, artificial intelligence, psychology, medicine, computational science, and cognitive science to generate a trans-disciplinary view on auditory phantom perception and to spread novel research and therapy approaches with researchers from all fields. In this Research Topic, 19 publications were submitted, from which 13 (68.4%) were accepted for publication, resulting in a rejection rate of 31.6%.

Main

The 13 publications were grouped into four different sub-topics: (1) Mechanistic Models of (impaired) Auditory Perception and Phantom Perception and the influence of Cross-Modality, (2) The Cerebellar Forward Model, Prediction Errors, Auditory Cognition and Schizophrenia, (3) Artificial Intelligence as a Tool and a Model to Understand the Brain, (4) Artificial Intelligence to Validate and Improve Treatment of Disorders of the Auditory System.

Mechanistic models of (impaired) auditory perception and phantom perception and the influence of cross-modality

In this paragraph, we summarize studies, which deal with the effects of cross-modal connections, attentional fluctuations, as well as circadian rhythms on the auditory system and the other way around (Yakunina and Nam; Wang et al.; Grimm et al.; Kondo et al.). Furthermore, different mechanistic tinnitus models such as central noise in central gain are discussed amongst others in the light of correlations of tinnitus and decreased cognitive decline in certain tinnitus populations (Chen F. et al.; Hamza and Zeng).

In their study, Wang et al. analyzed the different neural correlates of Idiopathic Sudden Sensorineural Hearing Loss (ISSNHL, 30 dB hearing loss at several frequencies) with and without vertigo. The authors applied resting state functional magnetic resonance imaging and analyzed the regional homogeneity (ReHo, a measure of neural synchrony in certain brain regions) and the functional connectivity between brain regions. The main findings were: that in ISSNHL with additional vertigo the ReHo was decreased especially in the ipsilateral auditory cortex. Furthermore, the functional connectivity within the inferior parietal gyrus—a multi-sensory area processing vestibular information—was increased. The authors speculate that the dysfunction of the inferior parietal cortex affects the auditory cortex and worsens auditory processing.

Kondo et al. investigated the attention mechanisms in the auditory system. In their study, the authors applied functional magnetic resonance imaging (fMRI) to investigate which brain dynamics lead to attentional fluctuations. Thus, the authors measured the reaction time and the BOLD (blood oxygenation level dependent) responses during an auditory gradual onset, continuous performance task (gradCPT), where subjects have to quickly distinguish between a male (go trial) and female (no go trial) narrator. The authors applied energy landscape analysis and could show that task-related activation patterns cluster in different attractors. Indeed, it is assumed that attention plays a critical role also in the manifestation and modulation of tinnitus (Roberts et al., 2013). Thus, the methods and findings of this study could be a further step to unravel the neural origin of tinnitus.

The focus of the following four studies lies on tinnitus mechanisms, cross-modal influence on tinnitus and the connection of tinnitus and cognitive decline.

Yakunina and Nam discuss in their study the applicability, safety, and effectiveness of transcutanious and standard invasive vagus nerve (tVNS respectively VNS) stimulation for the treatment of tinnitus. The authors state that although the VNS technique paired with sensory stimuli drives plasticity in the cortex, the exact neural mechanisms regarding neuromodulation through (t)VNS remain elusive. The authors are clearly critical of (t)VNS stimulation in combination with auditory stimuli as tinnitus therapy for several reasons. In particular, the authors state that the entire field of combined VNS and auditory stimulation is based on the work of a single research group and was not independently confirmed by author researchers. However, recently the positive effect of combined somato-sensory and auditory stimulation on tinnitus has been confirmed in a double-blind study with a sample size of 326 tinnitus patients (Conlon et al., 2020). Yakunina and Nam further criticize that the combined VNS and auditory stimulation technique is exclusively based on the view that the main neural correlate of tinnitus is a reorganization of the tonotopy of the auditory cortex. Indeed, this view has been refuted several times in the meantime (see e.g., Koops et al., 2020). Hence, the study of Yakunina and Nam illustrates that a mechanistic theory behind auditory phantom perception is needed to develop perfect-fit therapy approaches.

The meta-analysis of Chen F. et al. is a further study dealing with mechanistic tinnitus models, and especially the connection of tinnitus and hyperacusis. The authors statistically evaluated and summarized the effect of acoustic trauma on auditory brainstem responses and discuss their evaluations in the light of synaptopathy at the inner hair cells, hyperacusis and tinnitus. The authors report a strong evidence for reduced wave I amplitudes—indicating spiral ganglion activity—in tinnitus patients compared to control groups. Furthermore, the authors state that wave I amplitude decrease is a proxy for synapse loss. However, for wave V the results are contradictory. In the literature, a tinnitus related increase as well as decrease of wave V is reported. Thus, the authors refer to the publication of Knipper et al. (2020), and speculate that the increased wave V amplitudes might be related to hyperacusis and could be explained with the central gain model that postulates an increased sensitivity of neurons through homeostatic plasticity, whereas central noise as the potential cause of tinnitus [would not lead to increased wave V amplitudes (Zeng, 2013, 2020; Schilling et al., 2023)].

In their study on circadian sensitivity of noise trauma-induced hearing loss and tinnitus in Mon- golian gerbils, Grimm et al. discovered a connection of tinnitus and decreased hearing thresholds [which has already been proposed by Krauss et al. (2016)]. In particular, the authors showed that in gerbils the highest hearing loss in the more affected ear occurs when the noise trauma is applied at 5 p.m., when the male gerbils have their activity minimum. This is a counter-intuitive finding, as mice and other

rodent species are more sensitive at night, where their activity level is maximal. Furthermore, the authors report that the correlation of hearing loss and effect size of the behavioral signs of tinnitus in rodents (compare e.g., Schilling et al., 2017) is significant, i.e., that tinnitus is correlated with better hearing. This finding is in line with the stochastic resonance model of auditory phantom perception (Krauss et al., 2016; Schilling et al., 2021).

Hamza and Zeng investigated the connection of higher brain functions and tinnitus in elderly people. They report that in a cohort of non-hispanic people aged between 60 and 69 years with a hearing loss of at least 25 dB, the presence of a tinnitus perception was associated with improved cognitive performance. The described effect is surprising in a sense that usually hearing loss is associated with impaired cognitive function, and hence tinnitus as a frequent comorbid condition of hearing loss was often suspected to also lead to impaired cognition. Earlier studies also described a negative effect of tinnitus on cognitive performance. However, the authors criticize that these studies did not control for confounding and interactive factors such as age. The authors speculate that tinnitus may have a benefit in a sense that tinnitus patients have less speech perception difficulties compared to patients suffering from hearing loss alone without tinnitus. This hypothesis is indeed in line with several recent studies (Schilling et al., 2021, 2022; Schilling and Krauss, 2022).

The cerebellar forward model, prediction errors, auditory cognition and schizophrenia

The second section of this Research Topic focuses on schizophrenia and additional auditory verbal hallucinations ("hearing voices") and consists of two studies that provide further insight into impaired multimodal integration and neural mechanisms in the cerebellum (Chen J. et al.; He et al.).

In their study, Chen J. et al. used resting state functional magnetic resonance imaging (fMRI) and arterial spin labeling to investigate cerebral blood flow and functional connectivity strength alterations in schizophrenia patients with verbal hallucinations compared to patients with- out verbal hallucinations and a healthy control group. "Hearing voices" or verbal hallucinations are characteristic symptom in schizophrenia. The authors report an increased CBF/FCS ratio (cerebral blood flow functional connectivity) in auditory perception and language processing areas of the cerebral cortex (left superior and middle temporal gyri) in the auditory verbal hallucinations group compared to the healthy control group. Further alterations are found in cerebellar structures. The authors speculate that the failure of the forward model of the cerebellum, which describes the cerebellum as a structure calculating discrepancies between sensory input and predictions (efferent copies) (Manto et al., 2012), may be a potential cause of auditory verbal hallucinations.

A further study on schizophrenia, where He et al. used the rubber hand illusion—i.e. the misinterpretation of a rubber hand as own hand induced by synchronous visual and tactile stimulation—to investigate differences in multi-sensory integration in schizophrenia patients with and without auditory verbal hallucinations. As schizophrenia patients are in many cases not able to identify the actual sources of their sensory perception, all schizophrenia patients (with and without auditory verbal hallucinations) showed an enhanced and earlier rubber hand illusion compared to a healthy control group. Surprisingly, patients with auditory verbal hallucinations experienced weaker illusions than the group with no auditory verbal illusions. However, when stimuli were presented asynchronously only in the group without auditory verbal hallucinations the rubber hand illusion decreased, which indicates that the impairment in temporal processing is increased in schizophrenia patients with auditory verbal hallucinations. The authors conclude that patients with auditory verbal hallucinations have multi-sensory processing dysfunctions and internal timing deficits.

Artificial intelligence as a tool and a model to understand the brain

In the third paragraph of this editorial, two studies using modern AI approaches to extract information from neuroimaging data (fNIRS and EEG), in order to get a better understanding of (impaired) brain mechanism, are summarized (Yoo et al.; Li et al.).

Yoo et al. used Long-Short-Term-Memory (LSTM) networks to classify neural activity patterns in the auditory cortex, which were evoked by six different auditory stimuli. To measure the neural activity patterns the near infrared spectroscopy (fNIRS) was used, a technique based on near infrared light penetrating the head and thus used to measure oxygenation changes. The authors report that the LSTM networks achieve a better classification accuracy compared to a support vector machine (SVM), when the data of all 18 participants was used for training, whereas the SVM worked better, when each participant was regarded individually. Thus, the authors speculated that the poorer performance of the LSTM network was a result of too few training examples. However, the authors emphasized that for the LSTM networks no hand-crafted features were needed.

In contrast to the previously described study, Li et al. analyzed EEG data with machine learning techniques. In particular, the authors investigate how to apply machine learning techniques to EEG-connectivity features to estimate, whether the tinnitus is right/left-lateralized or binaurally perceived. The authors analyzed four different connectivity features and two time-frequency domain features, and applied four different machine learning techniques (two different support vector machines (SVM), a multi-layer perceptron (MLP), and a convolutional neural network) to these features. They conclude that bilateral tinnitus was characterized as altered connectivity in both auditory cortices, whereas left-lateralized tinnitus affected only the contra-lateral auditory cortex. Right-lateralized tinnitus, however, affects connectivity of the auditory cortices in both hemispheres.

Furthermore, the authors report tinnitus classification accuracies above 99 % for a SVM and MLP network. Potentially, these models could be used as objective method to determine tinnitus locations.

Artificial intelligence to validate and improve treatment of disorders of the auditory system

In the fourth and last paragraph of this editorial we summarize studies, which have used AI approaches to validate and improve treatments of disorders of the auditory system.

Yoon investigates how the recognition of certain consonants can be improved by identifying and amplifying certain frequency and time ranges (target frequency and time ranges) of the presented consonants. The main finding of the study is that for conversationally produced (in contrast to artificial) consonants the removal of mutually disturbing frequency ranges (conflicting frequencies) has a negative effect on consonant recognition, whereas the amplification of the target frequency and time ranges is an efficient way to improve consonant recognition in healthy listeners. The author states that the findings presented in this study serve as the basis to develop an artificial intelligence (AI) based, individual fitting scheme for patients, who wear a hearing aid and a cochlear implant. This idea is in line with the call of Lesica et al. (2021) to exploit the novel opportunities of AI to improve the treatment of hearing impairments.

One further study analyzing medical treatment success using machine learning, is the study by Puga et al. where they analyze the differences in patient population properties at two German tinnitus centers (University Hospital Regensburg and Tinnitus Center-Charit'e Berlin). The authors report that the age distribution of patients in both tinnitus centers are significantly different from general German population and also slightly differ in both centers. Furthermore, the population of patients in Regensburg contains more males than the German population and the Berlin data set. The tinnitus questionnaire score is more similar in Berlin and Regensburg in the female population. The authors used machine learning techniques (linear regression, lasso, ridge, support vector regressor) trained on the Berlin data set to predict the treatment outcome in the Regensburg data set. In accordance with the finding that tinnitus questionnaire scores of females are more similar across tinnitus centers it was also easier to predict the treatment outcome in the female population than in the male population. The authors conclude that in future these analyses should be extended to other research centers, order to learn more about inter-and intra-center differences.

A further study combines AI algorithms with mobile health apps. Shahania et al. used data of 21 persons acquired with a mobile health app to investigate how to make better predictions on tinnitus perception and the effects on the patients. A contextual multi-armed bandit algorithm was used to answer the question, if it is advantageous to involve the data from all patients in specific predictions (global model) or if it is better to regard each patient separately or small groups of similar patients. The authors summarize that entity-centric models (each user regarded separately) are preferred, and stated that there is no guaranty that the multi-armed bandit converges to an optimal solution, which could be caused by the relatively small sample size.

Concluding remarks

The 13 studies published in the Research Topic "*Auditory perception and phantom perception in brains, minds and machines*" demonstrate that an inter-disciplinary view on the auditory system is necessary to unravel mechanisms underlying healthy and impaired processing in this sensory system. The heterogeneity of the studies shows that widening the view to research on other sensory modalities such as the somato-sensory system, or even to other disciplines as AI and computational sciences can lead to significant progress in these research strands.

The fast growing field of AI and high-performance computing in combination with innovative trans-disciplinary ideas have the potential to further boost auditory neuroscience and audiology. Or as the famous neuroscientist Gershman (2023) states: "Happily, algorithms optimized for solving engineering problems frequently turn out to be successful models of brain function."

Furthermore, the novel algorithms help to analyze and thus to transform neuro-imaging data into scientific knowledge on the one hand, and furthermore help to validate and improve therapy approaches for disorders of the auditory system such as tinnitus on the other hand. We hope that the interdisciplinary approach of this Research Topic inspires scientists of different fields to start co-operations in order to unravel the functions and mechanisms of the human brain.

Author contributions

AS: Conceptualization, Funding acquisition, Project administration, Writing—original draft, Writing—review and editing. RS: Conceptualization, Project administration, Writing—review and editing. WS: Conceptualization, Project administration, Writing—review and editing. RG: Conceptualization, Project administration, Writing—review and editing. AM: Conceptualization, Project administration, Writing—review and editing. PK: Conceptualization, Writing—original draft, Writing—review and editing, Funding acquisition, Project administration.

Funding

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), GRK 2839 (project number 468527017) to PK, and grant SCHI 1482/3-1 (project number 451810794) to AS.

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.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

References

Cederroth, C. R., Lugo, A., Edvall, N. K., Lazar, A., Lopez-Escamez, J. A., Bulla, J., et al. (2020). Association between hyperacusis and tinnitus. *J. Clin. Med.* 9, 2412. doi: 10.3390/jcm9082412

Conlon, B., Langguth, B., Hamilton, C., Hughes, S., Meade, E., Connor, C. O., et al. (2020). Bimodal neuromodulation combining sound and tongue stimulation reduces tinnitus symptoms in a large randomized clinical study. *Sci. Transl. Med.* 12, eabb2830. doi: 10.1126/scitranslmed.abb2830

Gershman, S. J. (2023). What have we learned about artificial intelligence from studying the brain? Available online at: https://gershmanlab.com/pubs/NeuroAI_critique.pdf

Kietzmann, T. C., McClure, P., and Kriegeskorte, N. (2017). Deep neural networks in computational neuroscience. *BioRxiv* 4, 133504. doi: 10.1101/133504

Knipper, M., Van Dijk, P., Schulze, H., Mazurek, B., Krauss, P., Scheper, V., et al. (2020). The neural bases of tinnitus: lessons from deafness and cochlear implants. *J. Neurosci.* 40, 7190–7202. doi: 10.1523/JNEUROSCI.1314-19.2020

Koops, E. A., Renken, R. J., Lanting, C. P., and van Dijk, P. (2020). Cortical tonotopic map changes in humans are larger in hearing loss than in additional tinnitus. *J. Neurosci.* 40, 3178–3185. doi: 10.1523/JNEUROSCI.2083-19.2020

Krauss, P., Tziridis, K., Metzner, C., Schilling, A., Hoppe, U., Schulze, H., 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. doi: 10.3389/fnins.2016.00597

Kriegeskorte, N., and Douglas, P. K. (2018). Cognitive computational neuroscience. *Nat. Neurosci.* 21, 1148–1160. doi: 10.1038/s41593-018-0210-5

Lesica, N. A., Mehta, N., Manjaly, J. G., Deng, L., Wilson, B. S., Zeng, F. G., et al. (2021). Harnessing the power of artificial intelligence to transform hearing healthcare and research. *Nat. Mach. Int.* 3, 840–849. doi: 10.1038/s42256-021-00394-z

Manto, M., Bower, J. M., Conforto, A. B., Delgado-García, J. M., Da Guarda, M., Gerwig, S. N. F., et al. (2012). Consensus paper: roles of the cerebellum in motor control—the diversity of ideas on cerebellar involvement in movement. *The Cereb.* 11, 457–487. doi: 10.1007/s12311-011-0331-9

Naselaris, T., Bassett, D. S., Fletcher, A. K., Kording, K., Kriegeskorte, N., Nienborg, H., et al. (2018). Cognitive computational neuroscience: a new conference for an emerging discipline. *Trends Cognit. Sci.* 22, 365–367. doi: 10.1016/j.tics.2018.02.008

Roberts, L. E., Husain, F. T., and Eggermont, J. J. (2013). Role of attention in the generation and modulation of tinnitus. *Neurosci. Biobehav. Rev.* 37, 1754–1773. doi: 10.1016/j.neubiorev.2013.07.007

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. *Front. Neurosci.* 16, 908330. doi: 10.3389/fnins.2022.908330

Schilling, A., and Krauss, P. (2022). Tinnitus is associated with improved cognitive performance and speech perception–can stochastic resonance explain? *Front. Aging Neurosci.* 14, 1073149. doi: 10.3389/fnagi.2022.1073149

Schilling, A., Krauss, P., Gerum, R., Metzner, C., Tziridis, K., Schulze, H., et al. (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. doi: 10.3389/fnbeh.2017.00198

Schilling, A., Sedley, W., Gerum, R., Metzner, C., Tziridis, K., Maier, A., et al. (2023). Predictive coding and stochastic resonance as fundamental principles of auditory phantom perception. *Brain* 28, awad255. doi: 10.1093/brain/awad255

Schilling, A., Tziridis, K., Schulze, H., and Krauss, P. (2021). 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. doi: 10.1016/bs.pbr.2021.01.025

Tziridis, K., Forster, J., Buchheidt-Dörfler, I., Krauss, P., Schilling, A., Wendler, O., et al. (2021). Tinnitus development is associated with synaptopathy of inner hair cells in mongolian gerbils. *Eur. J. Neurosci.* 54, 4768–4780. doi: 10.1111/ejn.15334

Tziridis, K., Friedrich, J., Brüeggemann, P., Mazurek, B., and Schulze, H. (2022). Estimation of tinnitus-related socioeconomic costs in Germany. *Int. J. Environ. Res. Pub. Health* 19, 10455. doi: 10.3390/ijerph191610455

Zeng, F. G. (2013). An active loudness model suggesting tinnitus as increased central noise and hyperacusis as increased nonlinear gain. *Hearing Res.* 295, 172–179. doi: 10.1016/j.heares.2012.05.009

Zeng, F. G. (2020). Tinnitus and hyperacusis: central noise, gain and variance. *Curr. Opin. Physiol.* 18, 123–129. doi: 10.1016/j.cophys.2020.10.009

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 28 April 2021
doi: 10.3389/fnhum.2021.636191

Decoding Multiple Sound-Categories in the Auditory Cortex by Neural Networks: An fNIRS Study

So-Hyeon Yoo[1], *Hendrik Santosa*[2], *Chang-Seok Kim*[3] *and Keum-Shik Hong*[1]*

[1]School of Mechanical Engineering, Pusan National University, Busan, South Korea, [2]Department of Radiology, University of Pittsburgh, Pittsburgh, PA, United States, [3]Department of Cogno-Mechatronics Engineering, Pusan National University, Busan, South Korea

OPEN ACCESS

Edited by:
Andreas K. Maier,
University of Erlangen Nuremberg, Germany

Reviewed by:
Evangelos Paraskevopoulos,
Aristotle University of Thessaloniki, Greece
Giancarlo Valente,
Maastricht University, Netherlands

***Correspondence:**
Keum-Shik Hong
kshong@pusan.ac.kr

Specialty section:
This article was submitted to Sensory Neuroscience, a section of the journal Frontiers in Human Neuroscience

Received: 01 December 2020
Accepted: 31 March 2021
Published: 28 April 2021

Citation:
Yoo S-H, Santosa H, Kim C-S and Hong K-S (2021) Decoding Multiple Sound-Categories in the Auditory Cortex by Neural Networks: An fNIRS Study. Front. Hum. Neurosci. 15:636191. doi: 10.3389/fnhum.2021.636191

This study aims to decode the hemodynamic responses (HRs) evoked by multiple sound-categories using functional near-infrared spectroscopy (fNIRS). The six different sounds were given as stimuli (English, non-English, annoying, nature, music, and gunshot). The oxy-hemoglobin (HbO) concentration changes are measured in both hemispheres of the auditory cortex while 18 healthy subjects listen to 10-s blocks of six sound-categories. Long short-term memory (LSTM) networks were used as a classifier. The classification accuracy was 20.38 ± 4.63% with six class classification. Though LSTM networks' performance was a little higher than chance levels, it is noteworthy that we could classify the data subject-wise without feature selections.

Keywords: functional near-infrared spectroscopy (fNIRS), long short-term memories (LSTMs), auditory cortex, decoding, deep learning

INTRODUCTION

Recognizing sound is one of the important senses in everyday life. People are always exposed to a variety of sounds, and they can know what sound it is without being conscious. This ability allows people to avoid various dangers and facilitates communication with others. The auditory stimulus that enters through the outer ear is transmitted to the auditory cortex through the auditory nerve. It is clear that the temporal cortex is activated differently by different sounds. Neural responses in the auditory cortices have been studied using diverse modalities like electroencephalography (EEG; Wong et al., 2007; Hill and Scholkopf, 2012; Liu et al., 2015), magnetoencephalography (MEG; Hyvarinen et al., 2015), eletrocorticogram (ECoG; Pasley et al., 2012; Herff et al., 2015), functional magnetic resonance imaging (fMRI; Wong et al., 2008; Gao et al., 2015; Zhang et al., 2016), functional near-infrared spectroscopy (fNIRS; Plichta et al., 2011; Kovelman et al., 2012; Dewey and Hartley, 2015), and multimodal imaging (i.e., concurrent fNIRS, fMRI, and/or MEG; Kovelman et al., 2015; Corsi et al., 2019) to identify this process. In these studies, the complexities of brain responses evoked by the perception of sounds have been investigated to improve the quality of life.

Griffiths and Warren (2004) argued that analyzing auditory objects in the two-dimensional space (frequency and time) rather than one-dimensional space (frequency or time) is more meaningful, and thus acoustic experiences that produce two-dimensional images need to be investigated. But, they did not provide specific sound categories in their work. Theunissen and Elie (2014) showed that natural sounds facilitate the characterizations of the stimulus-response functions for neurons than white noise or simple synthetic sounds. Salvari et al. (2019) demonstrated significant activation and interconnection differences between natural sounds and human-made object sounds (music and artificial sounds) in the prefrontal areas using MEG. However, there were no significant differences between music and artificial sounds. Liu et al. (2019) demonstrated that predictions of tuning properties of putative feature-selective neurons match data from the marmoset's primary auditory cortex. Also, they showed that the exact algorithm of marmoset's call classification could successfully be applied in call classification in other species.

Identifying the sound that a person hears using a brain-computer interface (BCI) enables us to know what the person is hearing. The more diverse sounds a BCI device can discern, the more variant conditions are identified. For those who have lost vision, sound may be an alternative tool to communicate with other people in a non-contact way. If it is possible to classify more sounds, we can increase the control commands for an external device. Zhang et al. (2015) have researched decoding brain activation from multiple sound categories in the human temporal cortex. Seven different sound categories (English, non-English, vocal, Animal, mechanical, music, and nature) were used for classification in their fMRI work. They reported sound-category-selective brain maps showing distributed patterns of brain activity in the superior temporal gyrus and the middle temporal gyrus. However, analyses of such responses were hampered by the machine noise produced during fMRI experiments (Scarff et al., 2004; Fuchino et al., 2006).

fNIRS is a non-invasive brain imaging method that uses near-infrared light (700–900 nm) to penetrate the head and records oxygenation changes in the cerebral blood flow. fNIRS is a promising method for analyzing sound and speech processing. Compared to fMRI, fNIRS measurement is not noisy, and such measurements can be made in an environment more conducive to infant studies. Owing to these advantages, fNIRS shows significant potential for real-time brain monitoring while the subject is moving. According to fNIRS analyses, newborns consistently exhibit a strong hemodynamic response to universally preferred syllables, which suggests that the early acquisition and perception of language can be detected using categorical linguistic sounds (Gomez et al., 2014). The applications of this technology have the potential to provide feedback for speech therapy or in the tuning of hearing aid devices (e.g., cochlear implants) at an early stage of development based on brain recordings (Mushtaq et al., 2020). Several groups have demonstrated fNIRS use for measuring brain responses in deaf children with cochlear implants (Sevy et al., 2010; Pollonini et al., 2014).

In fNIRS applications, classification has been used in lie detection (Bhutta et al., 2015), drowsiness detection (Khan and Hong, 2015), mental workload detection (Herff et al., 2014), brain disease (Yoo et al., 2020), and the fNIRS-EEG-based hybrid BCI (Yuan et al., 2019; Lin et al., 2020). In fMRI applications, classification has also been used to decode the brain responses evoked by sight (Kohler et al., 2013; Smith, 2013) and sound (Staeren et al., 2009; Zhang et al., 2015). Lotte et al. (2007) and Pereira et al. (2009) reviewed the classification algorithms for EEG and fMRI data, respectively.

Recently, numerous studies have focused on improving classification accuracy by applying deep learning technology to brain signal classification, in addition to artificial neural networks (ANNs; Badai et al., 2020; Flynn et al., 2020). Convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are representative forms of ANNs. The CNN is robust in processing large image data sets (Lee et al., 2020). It has been widely implemented in brain signal processing, including fMRI (Erturk et al., 2020), deep brain stimulation (Kakusa et al., 2020), and EEG (Lun et al., 2020). Besides, CNNs have been used to diagnose brain diseases in the fNIRS domain (Xu et al., 2019a; Yang et al., 2019, 2020). RNNs are capable of predicting and classifying sequential data. They have been widely applied in robotics for various purposes and systems such as obstacle avoidance control (Xu et al., 2019b; Zheng et al., 2019; Zhao et al., 2020), self-organizing robot control (Smith et al., 2020), collision-free compliance control (Zhou et al., 2019), dynamic neural robots (Tekulve et al., 2019), and self-driving system (Chen et al., 2019). Recently, RNNs have achieved impressive results in detecting seizures (Sirpal et al., 2019), brain injuries (Ieong et al., 2019), and pain (Hu et al., 2019b), as well as in discriminating attention-deficit hyperactivity disorder (Dubreuil-Vall et al., 2020).

Long short-term memory (LSTM) is a type of RNN incorporating a progressive model (Hochreiter and Schmidhuber, 1997). Compared to RNN, LSTM networks possess a "gate" to reduce the vanishing gradient problem and allow the algorithm to more precisely control the information that needs to be retained in memory and the information that must be removed. LSTM is also considered superior to RNNs when handling large sequences of data. Additionally, compared to CNN, it exhibits better performance in classifying highly dynamic nonlinear time-series data such as EEG data (Tsiouris et al., 2018; Li et al., 2020).

This study aims to develop a communication method for the completely paralyzed with no vision. We identify the sound that a person hears by measuring task-evoked hemodynamic responses from the auditory cortex: In our early work (Hong and Santosa, 2016), four sound categories were classified. When remotely communicating with people without vision, visual or motor cortex-based BCIs may not be applicable. Sound will be a vital tool to communicate. In this article, we increased the number of sound categories from four to six. The more diverse sounds are classified, the more variant conditions are identified. Eventually, we can diversify the control commands to operate an external device. Sound-based BCI using audio stimuli is promising because we can use such audio signals in our daily

lives (i.e., a passive BCI is possible). In this article, auditory-evoked HRs are measured using fNIRS, and subsequently, LSTM is applied to analyze fNIRS' ability to distinguish individual sounds out of six classes.

MATERIALS AND METHODS

Subjects

A total of 18 subjects participated in the experiment (age: 26.89 ± 3.49 years; seven females, two left-handed). All subjects had normal hearing and no history of any neurological disorder. All subjects were informed about the nature and purpose of the respective experiments before obtaining their written consent. For the experiment, each subject lay down on a bed. All subjects were asked to remain relaxed, close their eyes, and avoid significant body movements during the experiment. The subjects were asked to listen attentively to various audio stimuli and guess the category of each stimulus. After the experiment, all participants were asked to explain verbally whether they could precisely distinguish what they heard. The fNIRS experimentation was done on healthy subjects and the entire experimental procedure was carried out in accordance with the Declaration of Helsinki and guidelines approved by the Ethics Committee of the Institutional Review Board of Pusan National University.

Audio Stimuli

The audio stimuli consisted of six different sound categories selected from a popular website (http://www.youtube.com). As shown in **Table 1**, the first and second categories entailed speech in English and other languages (non-English). The subjects were Indonesian, Korean, Chinese, Vietnamese, and Pakistani. Each participant had a common recognition of English but failed to recognize the other languages. The third and fourth categories were annoying sounds and nature sounds. The fifth category was a segment of classical music (Canon in D by Pachelbel). The sixth category was gunshot sounds at a frequency of 1 Hz. Each category consisted of six different sounds (except the gunshots, which had the same repeated sound). Each subject was exposed to 36 trials (i.e., six sound categories × six trials). The audio stimuli were presented in a pseudo-randomized order. Each stimulus consisted of 10 s of the sound followed by 20 s of silence.

Additionally, pre- and post-trials of classical music were added (to avoid sudden hearing), neither of which was included in the data processing. Accordingly, the entire fNIRS recording lasted for approximately 19 min. All audio stimuli were digitally mixed using the Adobe Audition software (MP3-format file: 16-bit quantification, 44.1 kHz sampling, stereo channel) and normalized to the same intensity level. Active noise-cancellation earbuds (Sony MDR-NC100D) were utilized for acoustic stimulation of all subjects with the same sound-level setting. After each fNIRS recording session, all subjects reported that they could accurately distinguish the sound among the sound categories for all trials.

fNIRS Measurements

Figure 1 shows the continuous-wave fNIRS system's optode configuration (DYNOT: DYnamic Near-infrared Optical Tomography; NIRx Medical Technologies, Brooklyn, NY, USA) for bilateral imaging of the auditory cortex in both hemispheres. The emitter–detector distance was 23 mm, while the sampling rate was set to 1.81 Hz at two wavelengths (760 and 830 nm). The optode configuration consisted of 3 × 5 arrays (eight emitters and seven detectors) with 22 channels for each hemisphere. The two 22-channel sets were placed on the scalp, covering the left (Chs. 1–22) and right (Chs. 23–44) temporal lobes. According to the International 10-20 System, Chs. 16 and 38 were placed at T2 and T4, respectively (Santosa et al., 2014). In the left hemisphere, both Broca's area and Wernicke's area were covered by this configuration. Finally, the lights in the room were switched off to minimize signal contamination from ambient light sources during the experiments.

Preprocessing

The optical data of two wavelengths were converted into relative oxy-hemoglobin (HbO) and deoxy-hemoglobin (HbR) concentration changes using the modified Beer-Lambert law (Hiraoka et al., 1993) using MATLAB^{TM} (2020b, MathWorks, USA). Owing to the uniform emitter-detector distance, constant values of differential path-length factors were used for all channels (i.e., d = 7.15 for λ = 760 nm and d = 5.98 for λ = 830 nm). In previous studies, HbO data were activated significantly higher than HbR for given stimuli. Therefore, only HbO data were processed in this study. The HbO data were filtered to remove physiological and artificial noises using the fifth-order Butterworth band-pass filter with cutoff frequencies of 0.01 Hz and 0.1 Hz. The filtered data was chopped for each trial.

TABLE 1 | Audio categories (M: male, F: female).

Trial	Human vocal hearing		Nonvocal hearing			
	English	Non-English	Annoying sound	Nature sound	Music	Gunshot
1	M	Russian (F)	Baby cry	River	Canon in D	10 times
2	F	German (F)	Car alarm	Forest (day time)	Canon in D	10 times
3	M	French (F)	Police siren	Rain	Canon in D	10 times
4	MF*	Bulgarian (MF*)	Horror sound	Jungle	Canon in D	10 times
5	F	Italian (MF)	Male scream	Ocean waves	Canon in D	10 times
6	F	Japanese (F)	Nuclear alarm siren	Waterfall	Canon in D	10 times

**MF denotes male–female conversation.*

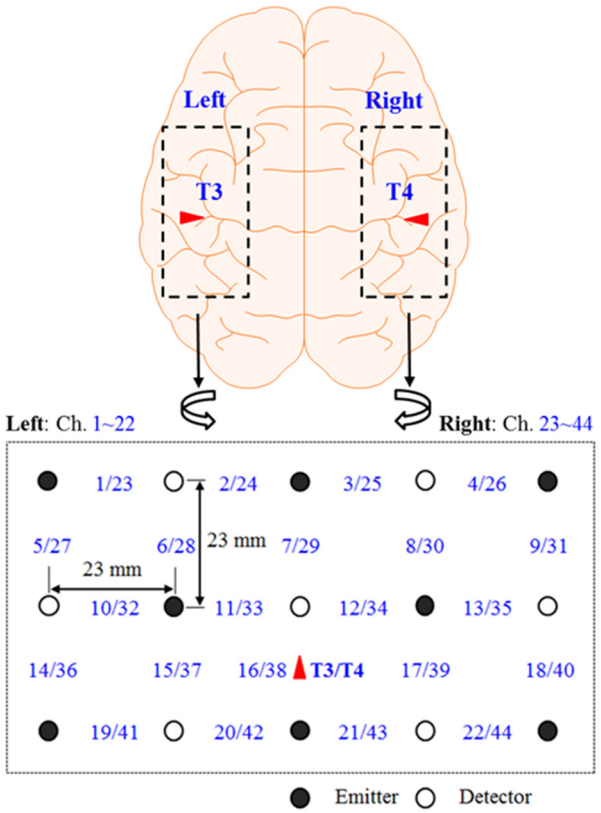


FIGURE 1 | Optode configuration: the numbers represent the measurement channels, where Chs. 16 and 38 coincide with T3 and T4 locations in the International 10–20 system (Hong and Santosa, 2016).

Feature Extraction for Support Vector Machine

The mean, slope, kurtosis, and skewness values of HbO signals were used as support vector machine (SVM) features. SVM classification was performed twice: One for "within-subject" and the other for "across-subject." Within-subject classification is a standard classification method for the fNIRS study. Considering the total number of trials for one subject was 36, 6-fold cross-validation was performed for each subject. For the across-subject classification, we used the entire data set. In this case, the total number of trials was 648 (i.e., multiplication of the number of subjects and the number of trials. Ten-fold and leave-one-out cross-validation were performed for the across-subject classification. Training and testing sets were divided randomly by MATLAB^{TM} function *cvpartition* for cross-validation. The same data partitions were used for SVM and LSTM.

LSTM

A recurrent neural network (RNN) is a type of artificial neural network wherein hidden nodes are connected with directional edges as a directed cycle. It is well known as an effective tool to process sequential data such as voice and handwriting. The RNN has the following structure (Hochreiter and Schmidhuber, 1997):

$$y_t = W_{hy}\, h_t + b_y, \tag{1}$$

$$h_t = \tanh\left(W_{hh}\, h_{t-1} + W_{xh}\, x_t + b_h\right), \tag{2}$$

where y_t indicates the output of the present state; subscript t is the discrete time step; W_{hy}, W_{hh}, and W_{xh} are the parameters from layer to layer; b_y is the bias of the output y; h_t is the hidden state vector; x_t is the input vector; and b_h is the bias of the hidden state vector h.

The LSTM is a special kind of recurrent neural network, compensating for the vanishing gradient problem. It has a structure of cell-states in the hidden state of RNN. The basic formulas for LSTM are as follows.

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + b_f), \tag{3}$$

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + b_o), \tag{4}$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + b_o), \tag{5}$$

$$g_t = \tanh(W_{xg}x_t + W_{hg}h_{t-1} + b_g), \tag{6}$$

$$c_t = f_t \circ c_{t-1} + i_t \circ g_t, \tag{7}$$

$$h_t = o_t \circ \tanh(c_t), \tag{8}$$

where f_t is the activation vector of forgetting gate to forget past information; i_t is the activation vector of input gate to memorize the current information; o_t is the activation vector of output gate; g_t is the activation vector of the cell input; c_t is the cell state vector; W_{xf}, W_{hf}, W_{xi}, W_{hi}, W_{xo}, W_{ho}, W_{xg}, and W_{hg} are the weight matrices of the input and recurrent connections; b_f, b_i, b_o, and b_g are the parametric bias vectors; and $\circ$ is the Hadamard product. In the LSTM networks, cell state and hidden state are calculated recursively.

In this article, LSTM networks are applied in two ways, like the two cases (within-subject, across-subject) in the SVM classification. **Figure 2** represents the LSTM networks used in this article. First, for within-subject classification, a bi-LSTM layer of eight hidden layers was used with two maximum epochs and three mini-batch sizes (Kang et al., 2020). Second, the bi-LSTM layer of 16 hidden layers was used for across-subject classification with three maximum epochs and three mini-batch sizes, see **Table 2**. The number of hidden layers, maximum epoch, and mini-batch size were selected to avoid overfitting (Sualeh and Kim, 2019). Additionally, a bi-LSTM layer of 256 hidden layers was examined for across-subject classification (to compare with 16 hidden layers). Also, 6- and 10-fold and leave-one-out cross-validations were performed in the same way as SVM.

RESULTS

In the experiment, a total of 18 subjects listened to six repetitions of each of the six categories of sound stimuli (36 total trials). The six categories were English (E), non-English (nE), nature sounds (NS), music (M), annoying sounds (AS), and gunshot (GS). The within-subject classification accuracies were 21.35 ± 6.71% for SVM and 19.14 ± 9.16% for LSTM, respectively; see **Figure 3**.

When the cross-validations of SVM and LSTM were performed separately, the accuracies of the 10-fold across-subject

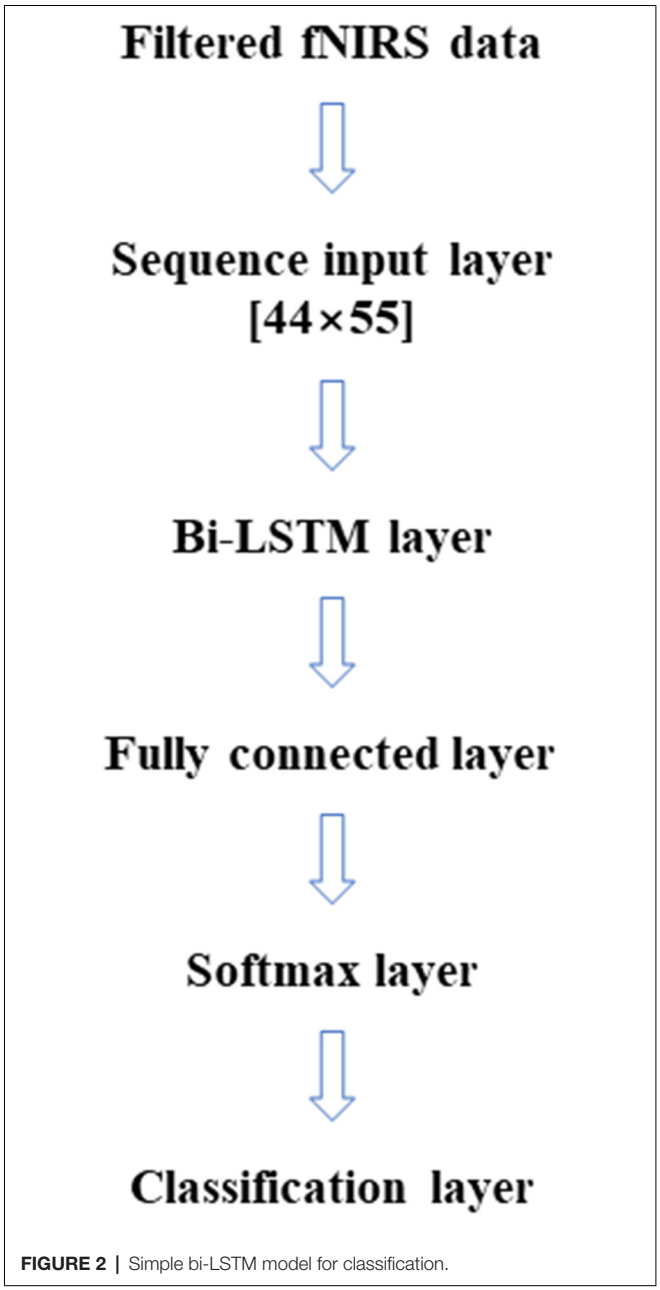


FIGURE 2 | Simple bi-LSTM model for classification.

TABLE 2 | A single bi-LSTM network.

bi-LSTM structure			
	Across-subject		Within-subject
	10-fold	Leave-one-out	
Input size	44 × 55		44 × 55
Training data set	584	612	30
Testing data set	64	36	6
The number of hidden layers	16		8

classification with 16 hidden layers were 16.83 ± 3.90% for SVM and 20.38 ± 4.63% for LSTM, respectively. **Figure 4** shows the confusion matrices for training and testing for the 10-fold across-subject classification. The hypergeometric *p*-values were calculated using the confusion matrix in **Figure 4B**. The *p*-values were 0.3745 (E), 0.3123 (nE), 0.0232 (NS), 0.0946 (M), 0.0129 (AS), and 0.3701 (GS). For a fair comparison between SVM and LSTM, we repeated the 10-fold cross-validation using the same data partitioning using the same code. In this case, the results were 15.73 ± 3.00% for SVM and 21.44 ± 4.57% for LSTM. Henceforth, we could not find a significant difference between the two cases regarding the data partitioning method.

Using the same data partitioning for SVM and LSTM, we also performed the leave-one-out across-subject classification. The results were 16.83 ± 3.90% for SVM and 20.52 ± 6.15% for LSTM. **Figure 5** shows the confusion matrices for training and testing in the leave-one-out case. The hypergeometric *p*-values were calculated using the confusion matrix in **Figure 5B**. The *p*-values were 0.4274 (E), 0.5488 (nE), 0.0129 (NS), 0.0036 (M), 0.1078 (AS), and 0.3769 (GS). The low *p*-values indicate that the classifier could successfully classify the sounds.

LSTM showed better accuracies in the across-subject classification but worse accuracies in the within-subject classification: This result was somewhat unexpected in comparison to the four-sound case (Hong and Santosa, 2016). It suggests that, in the six categories case, the subjects heard too many sound-categories, and they had difficulty in distinguishing them. Overall, when there are many hidden layers in the classifier, training becomes better than when there are few, but overfitting to the training data occurs. It is noted that the across-subject classification accuracies of LSTM with 256 hidden layers were 99.9% for training and 23.15% for testing, which is an overfitting case.

DISCUSSIONS

The previous studies in the literature have shown that various sound categories were processed differently in the brain. Staeren et al. (2009) showed that different sound categories evoked significant BOLD responses in a large expanse of the auditory cortex, including bilaterally the Heschl's gyrus, the superior temporal gyrus, and the upper bank of the superior temporal sulcus. Zhang et al. (2015) revealed that sound category-selective brain maps demonstrated distributed brain activity patterns in the superior temporal gyrus and the middle temporal gyrus. Plichta et al. (2011) reported that pleasant and unpleasant sounds increased auditory cortex activation than neutral sounds in their fNIRS research. Our results showed that nature sounds, music, and annoying sounds were classified better than other categories. Nature sounds and music are considered pleasant sounds, and annoying sounds are unpleasant sounds (Plichta et al., 2011). Classifying emotionally-neutral sounds (i.e., E, nE, and GS) "individually" from other categories is considered difficult.

Deep learning algorithms have been developed to increase classification accuracy and stability (Shan et al., 2019; Park and Jung, 2020; Sung et al., 2020). RNNs have been developed to improve their performance likewise; memristor-based RNNs (Yang et al., 2021), chaotic delayed RNNs with unknown parameters and stochastic noise (Yan et al., 2019), reformed recurrent Hermite polynomial neural network (Lin and Ting,

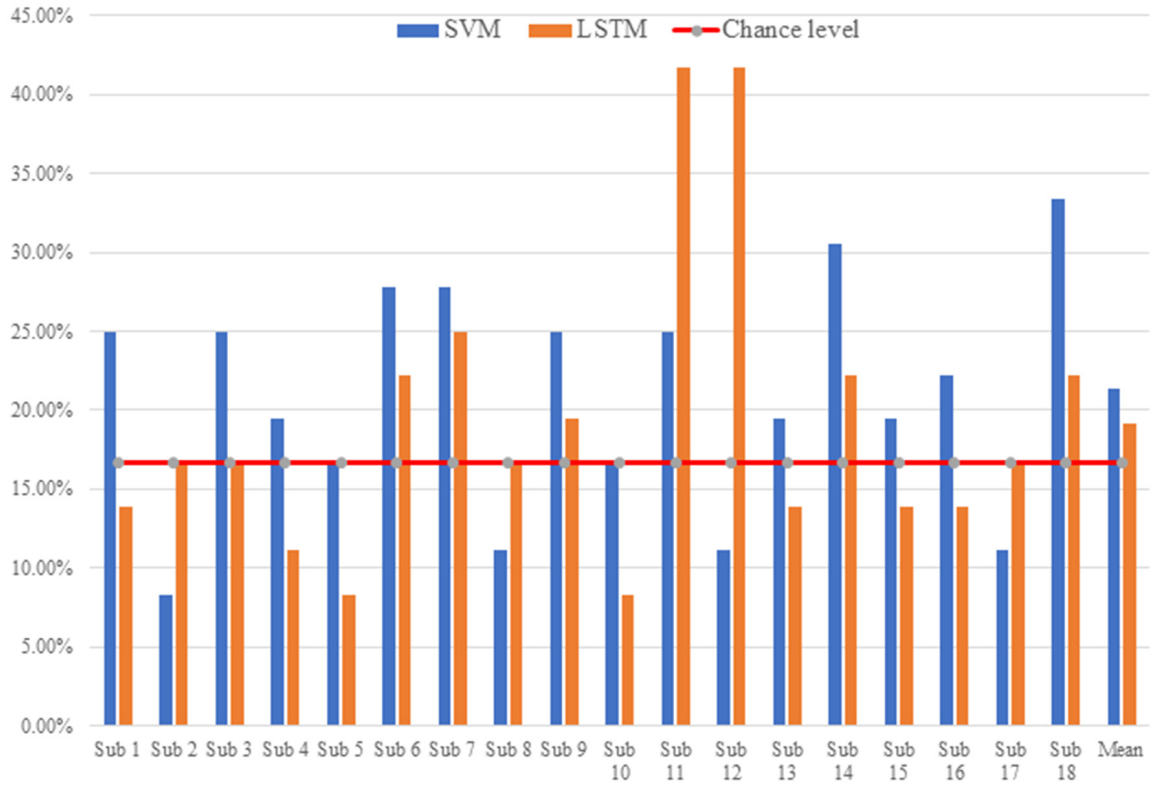


FIGURE 3 | Within-subject classification accuracies (SVM vs. LSTM).

2019). The developed RNNs have been applied in various brain imaging techniques (Hu et al., 2019a; Plakias and Boutalis, 2019). Wang et al. (2019) achieved 98.50% of classification accuracy by convolutional RNNs for individual recognition based on resting-state fMRI data. Qiao et al. (2019) proposed the application of bi-LSTM to decoding visual stimuli based on fMRI images from the visual cortex. The classification accuracies were 60.83 ± 1.17% and 42.50 ± 0.74% for each subject in five categories. The number of training samples and validation samples were 1,750 and 120 for each subject. Compared with the existing research, a limitation of this research is the small size of the data set.

The conventional classification technique is the process of distinguishing data from a set of categories based on a training data set on the category classes of which are known (Klein and Kranczioch, 2019; Pinti et al., 2019). The individual observations are analyzed into a set of features selected and executed by a classifier (Khan et al., 2014). In a more detailed process, a classifier is a function that takes the values of various features. For example, the mean, slope, skewness, and kurtosis values of HbO and HbR signals from individual trials can be used as the feature set (Tai and Chau, 2009). In our previous research (Hong and Santosa, 2016), decoding four-class sounds categories using fNIRS showed the 46.17 ± 6.25% (left) and 40.28 ± 6.00% (right) accuracies using LDA, while showing 38.35 ± 5.39% (left) and 36.99 ± 4.23% (right) using SVM. In the previous study, the classification was performed with the following steps: filtering, selecting region of interest, feature extraction from the region of interest, and classification. In the current study, to compare with the proposed method, the conventional classification technique was applied with the following steps: filtering, feature extraction from all channels, and classification. For LSTM networks, only filtering has been applied before classification.

The LSTM network may indirectly extract unstructured features from the data, and the network's weighting factors are optimized during the training session. The network can be trained only after simple filtering. The results showed that SVM is better for within-subject classification and LSTM is better for across-subject classification. It seems that fNIRS data involve different physiological data per subject, but this physiological difference is not removed with simple filtering. Also, the sample sizes of within-subject classification were 30 for training and six for validation. The sample sizes of across-subject classification were 584 for training and 64 for validation. Additionally, there are no significant differences between 10-fold and leave-one-out validations. The dataset of 10-fold validation might use the same subject's data in either training or testing. The size of

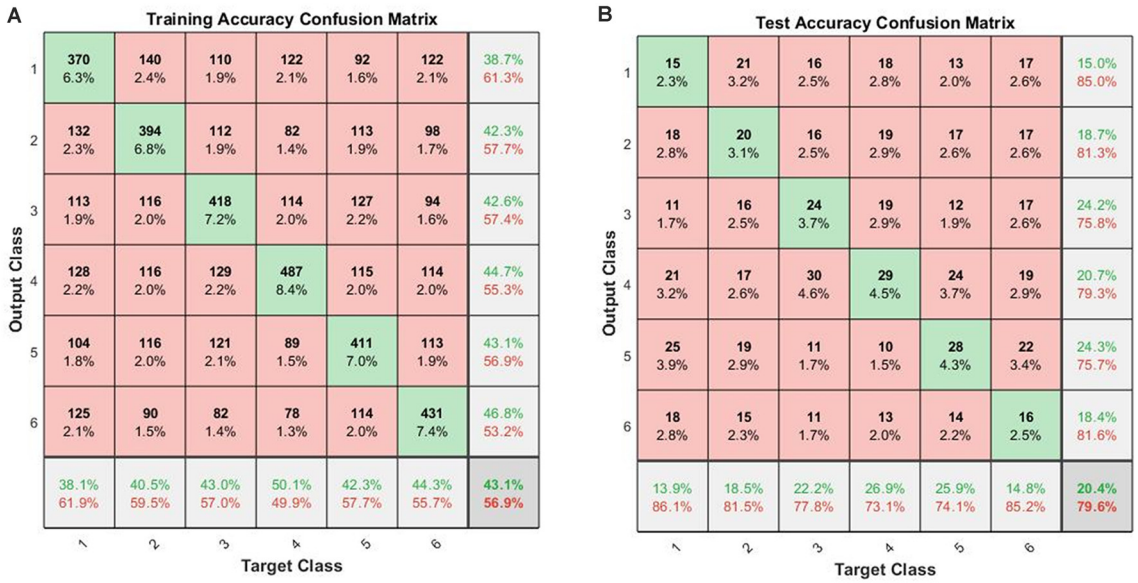


FIGURE 4 | Across-subject classification accuracies (LSTM): **(A)** confusion matrix for training, **(B)** confusion matrix for testing (Class 1: English; Class 2: non-English; Class 3: Nature sound; Class 4: Music; Class 5: Annoying sound; Class 6: Gunshot).

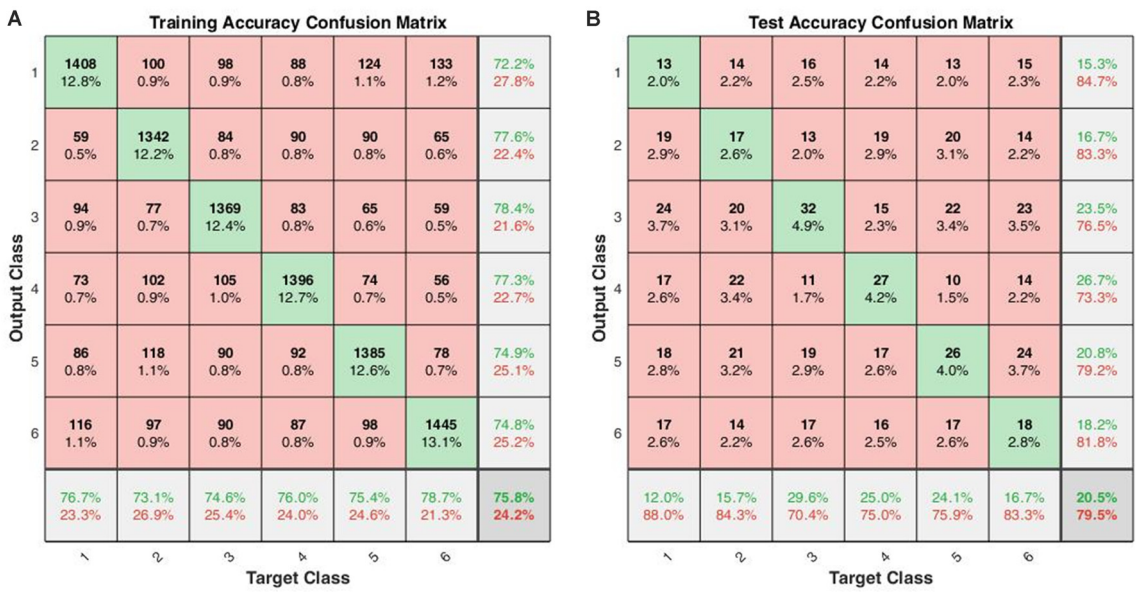


FIGURE 5 | Across-subject classification accuracies (leave-one-out validation of the LSTM with 16 hidden layers): **(A)** confusion matrix for training, **(B)** confusion matrix for testing (Class 1: English; Class 2: non-English; Class 3: Nature sound; Class 4: Music; Class 5: Annoying sound; Class 6: Gunshot).

the training data set was bigger in the leave-one-out validation. According to this result, if the sample size increases, the LSTM network would show better performance than the conventional method (SVM). Also, if we have enough training data, it would be enough for ignoring the subjects' physiological differences in the LSTM network. The LSTM network with 256 hidden layers showed slightly better performance than others, but the network overfitted with the training data set rapidly. Simplifying the data classification time can contribute to the commercialization of future diagnostics using fNIRS or BCI technology, given that it can reduce the classification time. Although the results in this study are not outstanding, it is worthwhile to show the potential of deep learning-based fNIRS signal classification technology.

CONCLUSION

This article aimed to identify hearing sounds using the HRs from the auditory cortices. The proposed audio-signal-

based BCI is to be used for completely paralyzed people, for whom visual or motor cortex-based BCI may not be suitable. In this study, we used fNIRS signals evoked by audio-stimuli from multiple sound-categories. Compared with the conventional method, the LSTM-based approach could decode the brain activities without heavy pre-processing of the data, such as regression, feature selection, and feature extraction. Though the LSTM network's performance was a little higher than the chance level, it is noteworthy that we could classify the six sounds virtually without defining the region of interest and feature extraction. The approach using audio stimuli is promising for a passive-type BCI using ordinary sounds in our daily lives. This study has a limitation on the number of data, which needs to be improved in the future.

DATA AVAILABILITY STATEMENT

The datasets generated in this article are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of the Institutional Review Board of Pusan National University. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

S-HY conducted the literature review and wrote the first draft of the manuscript. HS obtained the experimental data and initiated the work. C-SK participated in the revision of the manuscript. K-SH conceived the idea, corrected the manuscript, and finalized the work. All the authors have approved the final manuscript.

FUNDING

This work was supported by the National Research Foundation (NRF) of Korea under the auspices of the Ministry of Science and ICT, Korea (grant no. NRF-2020R1A2B5B03096000).

REFERENCES

Badai, J., Bu, Q., and Zhang, L. (2020). Review of artificial intelligence applications and algorithms for brain organoid research. *Interdiscip. Sci.* 12, 383–394. doi: 10.1007/s12539-020-00386-4

Bhutta, M. R., Hong, M. J., Kim, Y. H., and Hong, K. S. (2015). Single-trial lie detection using a combined fNIRS-polygraph system. *Front. Psychol.* 6:709. doi: 10.3389/fpsyg.2015.00709

Chen, J. N., Chen, J. Y., Zhang, R. M., and Hu, X. B. (2019). Toward a brain-inspired system: deep recurrent reinforcement learning for a simulated self-driving agent. *Front. Neurorobot.* 13:40. doi: 10.3389/fnbot.2019.00040

Corsi, M. C., Chavez, M., Schwartz, D., Hugueville, L., Khambhati, A. N., Bassett, D. S., et al. (2019). Integrating EEG and MEG signals to improve motor imagery classification in brain-computer interface. *Int. J. Neural Syst.* 29:1850014. doi: 10.1142/S0129065718500144

Dewey, R. S., and Hartley, D. E. H. (2015). Cortical cross-modal plasticity following deafness measured using functional near-infrared spectroscopy. *Hear. Res.* 325, 55–63. doi: 10.1016/j.heares.2015.03.007

Dubreuil-Vall, L., Ruffini, G., and Camprodon, J. A. (2020). Deep learning convolutional neural networks discriminate adult ADHD from healthy individuals on the basis of event-related spectral EEG. *Front. Neurosci.* 14:251. doi: 10.3389/fnins.2020.00251

Erturk, M. A., Panken, E., Conroy, M. J., Edmonson, J., Kramer, J., Chatterton, J., et al. (2020). Predicting *in vivo* MRI gradient-field induced voltage levels on implanted deep brain stimulation systems using neural networks. *Front. Hum. Neurosci.* 14:34. doi: 10.3389/fnhum.2020.00034

Flynn, M., Effraimidis, D., Angelopoulou, A., Kapetanios, E., Williams, D., Hemanth, J., et al. (2020). Assessing the effectiveness of automated emotion recognition in adults and children for clinical investigation. *Front. Hum. Neurosci.* 14:70. doi: 10.3389/fnhum.2020.00070

Fuchino, Y., Sato, H., Maki, A., Yamamoto, Y., Katura, T., Obata, A., et al. (2006). Effect of fMRI acoustic noise on sensorimotor activation examined using optical topography. *NeuroImage* 32, 771–777. doi: 10.1016/j.neuroimage.2006.04.197

Gao, P. P., Zhang, J. W., Fan, S. J., Sanes, D. H., and Wu, E. X. (2015). Auditory midbrain processing is differentially modulated by auditory and visual cortices: an auditory fMRI study. *NeuroImage* 123, 22–32. doi: 10.1016/j.neuroimage.2015.08.040

Gomez, D. M., Berent, I., Benavides-Varela, S., Bion, R. A. H., Cattarossi, L., Nespor, M., et al. (2014). Language universals at birth. *Proc. Natl. Acad. Sci. U S A* 111, 5837–5841. doi: 10.1073/pnas.1318261111

Griffiths, T., and Warren, J. (2004). What is an auditory object? *Nat. Rev. Neurosci.* 5, 887–892. doi: 10.1038/nrn1538

Herff, C., Heger, D., de Pesters, A., Telaar, D., Brunner, P., Schalk, G., et al. (2015). Brain-to-text: decoding spoken phrases from phone representations in the brain. *Front. Neurosci.* 9:217. doi: 10.3389/fnins.2015.00217

Herff, C., Heger, D., Fortmann, O., Hennrich, J., Putze, F., and Schultz, T. (2014). Mental workload during n-back task-quantified in the prefrontal cortex using fNIRS. *Front. Hum. Neurosci.* 7:935. doi: 10.3389/fnhum.2013.00935

Hill, N. J., and Scholkopf, B. (2012). An online brain-computer interface based on shifting attention to concurrent streams of auditory stimuli. *J. Neural Eng.* 9:026011. doi: 10.1088/1741-2560/9/2/026011

Hiraoka, M., Firbank, M., Essenpreis, M., Cope, M., Arridge, S. R., Vanderzee, P., et al. (1993). A monte-carlo investigation of optical pathlength in inhomogeneous tissue and its application to near-infrared spectroscopy. *Phys. Med. Biol.* 38, 1859–1876. doi: 10.1088/0031-9155/38/12/011

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

Hong, K. S., and Santosa, H. (2016). Decoding four different sound-categories in the auditory cortex using functional near-infrared spectroscopy. *Hear. Res.* 333, 157–166. doi: 10.1016/j.heares.2016.01.009

Hu, R. H., Huang, Q. J., Wang, H., He, J., and Chang, S. (2019a). Monitor-based spiking recurrent network for the representation of complex dynamic patterns. *Int. J. Neural Syst.* 29:1950006. doi: 10.1142/S0129065719500060

Hu, X. S., Nascimento, T. D., Bender, M. C., Hall, T., Petty, S., O'Malley, S., et al. (2019b). Feasibility of a real-time clinical augmented reality and artificial intelligence framework for pain detection and localization from the brain. *J. Med. Internet Res.* 21:e13594. doi: 10.2196/13594

Hyvarinen, P., Yrttiaho, S., Lehtimaki, J., Ilmoniemi, R. J., Makitie, A., Ylikoski, J., et al. (2015). Transcutaneous vagus nerve stimulation modulates tinnitus-related beta- and gamma-band activity. *Ear Hear.* 36, E76–E85. doi: 10.1097/AUD.0000000000000123

Ieong, H. F. H., Gao, F., and Yuan, Z. (2019). Machine learning: assessing neurovascular signals in the prefrontal cortex with non-invasive bimodal electro-optical neuroimaging in opiate addiction. *Sci. Rep.* 9:18262. doi: 10.1038/s41598-019-54316-6

Kakusa, B., Saluja, S., Dadey, D. Y. A., Barbosa, D. A. N., Gattas, S., Miller, K. J., et al. (2020). Electrophysiology and structural connectivity of the posterior hypothalamic region: much to learn from a rare indication of deep brain stimulation. *Front. Hum. Neurosci.* 14:164. doi: 10.3389/fnhum.2020.00164

Kang, H., Yang, S., Huang, J., and Oh, J. (2020). Time series prediction of wastewater flow rate by bidirectional LSTM deep learning. *Int. J. Control Autom. Syst.* 18, 3023–3030. doi: 10.1007/s12555-019-0984-6

Khan, M. J., and Hong, K. S. (2015). Passive BCI based on drowsiness detection: an fNIRS study. *Biomed. Opt. Express* 6, 4063–4078. doi: 10.1364/BOE.6.004063

Khan, M. J., Hong, M. J. Y., and Hong, K. S. (2014). Decoding of four movement directions using hybrid NIRS-EEG brain-computer interface. *Front. Hum. Neurosci.* 8:244. doi: 10.3389/fnhum.2014.00244

Klein, F., and Kranczioch, C. (2019). Signal processing in fNIRS: a case for the removal of systemic activity for single trial data. *Front. Hum. Neurosci.* 13:331. doi: 10.3389/fnhum.2019.00331

Kohler, P. J., Fogelson, S. V., Reavis, E. A., Meng, M., Guntupalli, J. S., Hanke, M., et al. (2013). Pattern classification precedes region-average hemodynamic response in early visual cortex. *NeuroImage* 78, 249–260. doi: 10.1016/j.neuroimage.2013.04.019

Kovelman, I., Mascho, K., Millott, L., Mastic, A., Moiseff, B., and Shalinsky, M. H. (2012). At the rhythm of language: brain bases of language-related frequency perception in children. *NeuroImage* 60, 673–682. doi: 10.1016/j.neuroimage.2011.12.066

Kovelman, I., Wagley, N., Hay, J. S. F., Ugolini, M., Bowyer, S. M., Lajiness-O'Neill, R., et al. (2015). Multimodal imaging of temporal processing in typical and atypical language development. *Ann. N. Y. Acad. Sci.*. 1337, 7–15. doi: 10.1111/nyas.12688

Lee, S. J., Choi, H., and Hwang, S. S. (2020). Real-time depth estimation using recurrent CNN with sparse depth cues for SLAM system. *Int. J. Control Autom. Syst.* 18, 206–216. doi: 10.1007/s12555-019-0350-8

Li, R. L., Wu, Q., Liu, J., Li, C., and Zhao, Q. B. (2020). Monitoring depth of anesthesia based on hybrid features and recurrent neural network. *Front. Neurosci.* 14:26. doi: 10.3389/fnins.2020.00026

Lin, C. T., King, J. T., Chuang, C. H., Ding, W. P., Chuang, W. Y., Liao, L. D., et al. (2020). Exploring the brain responses to driving fatigue through simultaneous EEG and fNIRS measurements. *Int. J. Neural Syst.* 30:1950018. doi: 10.1142/S0129065719500187

Lin, C. H., and Ting, J. C. (2019). Novel nonlinear backstepping control of synchronous reluctance motor drive system for position tracking of periodic reference inputs with torque ripple consideration. *Int. J. Control Autom. Syst.* 17, 1–17. doi: 10.1007/s12555-017-0703-0

Liu, F., Maggu, A. R., Lau, J. C. Y., and Wong, P. C. M. (2015). Brainstem encoding of speech and musical stimuli in congenital amusia: evidence from Cantonese speakers. *Front. Hum. Neurosci.* 8:1029. doi: 10.3389/fnhum.2014.01029

Liu, S. T., Montes-Lourido, P., Wang, X., and Sadagopan, S. (2019). Optimal features for auditory categorization. *Nat. Commun.* 10:1302. doi: 10.1038/s41467-019-09115-y

Lotte, F., Congedo, M., Lecuyer, A., Lamarche, F., and Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain-computer interfaces. *J. Neural Eng.* 4, R1–R13. doi: 10.1088/1741-2560/4/2/R01

Lun, X. M., Yu, Z. L., Chen, T., Wang, F., and Hou, Y. M. (2020). A simplified CNN classification method for MI-EEG *via* the electrode pairs signals. *Front. Hum. Neurosci.* 14:338. doi: 10.3389/fnhum.2020.00338

Mushtaq, F., Wiggins, I. M., Kitterick, P. T., Anderson, C. A., and Hartley, D. E. H. (2020). The benefit of cross-modal reorganization on speech perception in pediatric cochlear implant recipients revealed using functional near-infrared spectroscopy. *Front. Hum. Neurosci.* 14:308. doi: 10.3389/fnhum.2020.00308

Park, J., and Jung, D. J. (2020). Deep convolutional neural network architectures for tonal frequency identification in a lofargram. *Int. J. Control Autom. Syst.* 19, 1103–1112. doi: 10.1007/s12555-019-1014-4

Pasley, B. N., David, S. V., Mesgarani, N., Flinker, A., Shamma, S. A., Crone, N. E., et al. (2012). Reconstructing speech from human auditory cortex. *PLoS Biol.* 10:e1001251. doi: 10.1371/journal.pbio.1001251

Pereira, F., Mitchell, T., and Botvinick, M. (2009). Machine learning classifiers and fMRI: a tutorial overview. *NeuroImage* 45, S199–S209. doi: 10.1016/j.neuroimage.2008.11.007

Pinti, P., Scholkmann, F., Hamilton, A., Burgess, P., and Tachtsidis, I. (2019). Current status and issues regarding pre-processing of fNIRS neuroimaging data: an investigation of diverse signal filtering methods within a general linear model framework. *Front. Hum. Neurosci.* 12:505. doi: 10.3389/fnhum.2018.00505

Plakias, S., and Boutalis, Y. S. (2019). Lyapunov theory-based fusion neural networks for the identification of dynamic nonlinear systems. *Int. J. Neural Syst.* 29:1950015. doi: 10.1142/S0129065719500151

Plichta, M. M., Gerdes, A. B. M., Alpers, G. W., Harnisch, W., Brill, S., Wieser, M. J., et al. (2011). Auditory cortex activation is modulated by emotion: a functional near-infrared spectroscopy (fNIRS) study. *NeuroImage* 55, 1200–1207. doi: 10.1016/j.neuroimage.2011.01.011

Pollonini, L., Olds, C., Abaya, H., Bortfeld, H., Beauchamp, M. S., and Oghalai, J. S. (2014). Auditory cortex activation to natural speech and simulated cochlear implant speech measured with functional near-infrared spectroscopy. *Hear. Res.* 309, 84–93. doi: 10.1016/j.heares.2013.11.007

Qiao, K., Chen, J., Wang, L. Y., Zhang, C., Zeng, L., Tong, L., et al. (2019). Category decoding of visual stimuli from human brain activity using a bidirectional recurrent neural network to simulate bidirectional information flows in human visual cortices. *Front. Neurosci.* 13:692. doi: 10.3389/fnins.2019.00692

Salvari, V., Paraskevopoulos, E., Chalas, N., Muller, K., Wollbrink, A., Dobel, C., et al. (2019). Auditory categorization of man-made sounds versus natural sounds by means of MEG functional brain connectivity. *Front. Hum. Neurosci.* 13:1052. doi: 10.3389/fnins.2019.01052

Santosa, H., Hong, M. J., and Hong, K. S. (2014). Lateralization of music processing auditory cortex: an fNIRS study. *Front. Behav. Neurosci.* 8:418. doi: 10.3389/fnbeh.2014.00418

Scarff, C. J., Dort, J. C., Eggermont, J. J., and Goodyear, B. G. (2004). The effect of MR scanner noise on auditory cortex activity using fMRI. *Hum. Brain Mapp.* 22, 341–349. doi: 10.1002/hbm.20043

Sevy, A. B. G., Bortfeld, H., Huppert, T. J., Beauchamp, M. S., Tonini, R. E., and Oghalai, J. S. (2010). Neuroimaging with near-infrared spectroscopy demonstrates speech-evoked activity in the auditory cortex of deaf children following cochlear implantation. *Hear. Res.* 270, 39–47. doi: 10.1016/j.heares.2010.09.010

Shan, C. H., Guo, X. R., and Ou, J. (2019). Deep leaky single-peaked triangle neural networks. *Int. J. Control Autom. Syst.* 17, 2693–2701. doi: 10.1007/s12555-018-0796-0

Sirpal, P., Kassab, A., Pouliot, P., Nguyen, D. K., and Lesage, F. (2019). fNIRS improves seizure detection in multimodal EEG-fNIRS recordings. *J. Biomed. Opt.* 24, 1–9. doi: 10.1117/1.JBO.24.5.051408

Smith, K. (2013). Reading minds. *Nature* 502, 428–430. doi: 10.1038/502428a

Smith, S. C., Dharmadi, R., Imrie, C., Si, B. L., and Herrmann, J. M. (2020). The DIAMOND model: deep recurrent neural networks for self-organizing robot control. *Front. Neurorobot.* 14:62. doi: 10.3389/fnbot.2020.00062

Staeren, N., Renvall, H., De Martino, F., Goebel, R., and Formisano, E. (2009). Sound categories are represented as distributed patterns in the human auditory cortex. *Curr. Biol.* 19, 498–502. doi: 10.1016/j.cub.2009.01.066

Sualeh, M., and Kim, G. W. (2019). Simultaneous localization and mapping in the epoch of semantics: a survey. *Int. J. Control Autom. Syst.* 17, 729–742. doi: 10.1007/s12555-018-0130-x

Sung, H. J., Park, M. K., and Choi, J. W. (2020). Automatic grader for flatfishes using machine vision. *Int. J. Control Autom. Syst.* 18, 3073–3082. doi: 10.1007/s12555-020-0007-7

Tai, K., and Chau, T. (2009). Single-trial classification of NIRS signals during emotional induction tasks: towards a corporeal machine interface. *J. Neuroeng. Rehabil.* 6:39. doi: 10.1186/1743-0003-6-39

Tekulve, J., Fois, A., Sandamirskaya, Y., and Schoner, G. (2019). Autonomous sequence generation for a neural dynamic robot: scene perception, serial order and object-oriented movement. *Front. Neurorobot.* 13:95. doi: 10.3389/fnbot.2019.00095

Theunissen, F., and Elie, J. (2014). Neural processing of natural sounds. *Nat. Rev. Neurosci.* 15, 355–366. doi: 10.1038/nrn3731

Tsiouris, K. M., Pezoulas, V. C., Zervakis, M., Konitsiotis, S., Koutsouris, D. D., and Fotiadis, D. I. (2018). A long short-term memory deep learning network for the prediction of epileptic seizures using EEG signals. *Comput. Biol. Med.* 99, 24–37. doi: 10.1016/j.compbiomed.2018.05.019

Wang, L. B., Li, K. M., Chen, X., and Hu, X. P. P. (2019). Application of convolutional recurrent neural network for individual recognition based on resting state fMRI data. *Front. Neurosci.* 13:434. doi: 10.3389/fnins.2019.00434

Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., and Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat. Neurosci.* 10, 420–422. doi: 10.1038/nn1872

Wong, P. C. M., Uppunda, A. K., Parrish, T. B., and Dhar, S. (2008). Cortical mechanisms of speech perception in noise. *J. Speech Lang. Hear. Res.* 51, 1026–1041. doi: 10.1044/1092-4388(2008/075)

Xu, L. Y., Geng, X. L., He, X. Y., Li, J., and Yu, J. (2019a). Prediction in autism by deep learning short-time spontaneous hemodynamic fluctuations. *Front. Neurosci.* 13:1120. doi: 10.3389/fnins.2019.01120

Xu, Z. H., Zhou, X. F., and Li, S. (2019b). Deep recurrent neural networks based obstacle avoidance control for redundant manipulators. *Front. Neurorobot.* 13:47. doi: 10.3389/fnbot.2019.00047

Yan, Z. L., Liu, Y. M., Huang, X., Zhou, J. P., and Shen, H. (2019). Mixed script capital H-infinity and script capital L-2—script capital L-infinity anti-synchronization control for chaotic delayed recurrent neural networks. *Int. J. Control Autom. Syst.* 17, 3158–3169. doi: 10.1007/s12555-019-0263-6

Yang, C., Liu, Y. C., Li, F. M., and Li, Y. F. (2021). Finite-time synchronization of a class of coupled memristor-based recurrent neural networks: static state control and dynamic control approach. *Int. J. Control Autom. Syst.* 19, 426–438. doi: 10.1007/s12555-019-0616-1

Yang, D., Hong, K.-S., Yoo, S.-H., and Kim, C.-S. (2019). Evaluation of neural degeneration biomarkers in the prefrontal cortex for early identification of patients with mild cognitive impairment: an fNIRS study. *Front. Hum. Neurosci.* 13:317. doi: 10.3389/fnhum.2019.00317

Yang, D., Huang, R., Yoo, S.-H., Shin, M.-J., Yoon, J. A., Shin, Y.-I., et al. (2020). Detection of mild cognitive impairment using convolutional neural network: temporal-feature maps of functional near-infrared spectroscopy. *Front. Aging Neurosci.* 12:141. doi: 10.3389/fnagi.2020.00141

Yoo, S.-H., Woo, S.-W., Shin, M.-J., Yoon, J. A., Shin, Y.-I., and Hong, K.-S. (2020). Diagnosis of mild cognitive impairment using cognitive tasks: a functional near-infrared spectroscopy study. *Curr. Alzeimer Res.* 17, 1145–1160. doi: 10.2174/1567205018666210212154941

Yuan, Z., Zhang, X., and Ding, M. Z. (2019). Editorial: techniques advances and clinical applications in fused EEG-fNIRS. *Front. Hum. Neurosci.* 13:408. doi: 10.3389/fnhum.2019.00408

Zhang, C. C., Pugh, K. R., Mencl, W. E., Molfese, P. J., Frost, S. J., Magnuson, J. S., et al. (2016). Functionally integrated neural processing of linguistic and talker information: an event-related fMRI and ERP study. *NeuroImage* 124, 536–549. doi: 10.1016/j.neuroimage.2015.08.064

Zhang, F. Q., Wang, J. P., Kim, J., Parrish, T., and Wong, P. C. M. (2015). Decoding multiple sound categories in the human temporal cortex using high resolution fMRI. *PLoS One* 10:e0117303. doi: 10.1371/journal.pone.0117303

Zhao, W. F., Li, X. X., Chen, X., Su, X., and Tang, G. R. (2020). Bi-criteria acceleration level obstacle avoidance of redundant manipulator. *Front. Neurorobot.* 14:54. doi: 10.3389/fnbot.2020.00054

Zheng, W., Wang, H. B., Zhang, Z. M., Li, N., and Yin, P. H. (2019). Multi-layer feed-forward neural network deep learning control with hybrid position and virtual-force algorithm for mobile robot obstacle avoidance. *Int. J. Control Autom. Syst.* 17, 1007–1018. doi: 10.1007/s12555-018-0140-8

Zhou, X. F., Xu, Z. H., and Lie, S. (2019). Collision-free compliance control for redundant manipulators: an optimization case. *Front. Neurorobot.* 13:50. doi: 10.3389/fnbot.2019.00050

Conflict of Interest: The authors declare that they have no conflict of interest. This research was conducted in the absence of any commercial or financial relationship that could be construed as a potential conflict of interest.

frontiers in Neuroscience

REVIEW
published: 28 May 2021
doi: 10.3389/fnins.2021.680590

Direct and Transcutaneous Vagus Nerve Stimulation for Treatment of Tinnitus: A Scoping Review

Natalia Yakunina[1] *and Eui-Cheol Nam*[2]*

[1] Institute of Medical Science, School of Medicine, Kangwon National University, Chuncheon, South Korea, [2] Department of Otolaryngology, School of Medicine, Kangwon National University, Chuncheon, South Korea

Recent animal research has shown that vagus nerve stimulation (VNS) paired with sound stimuli can induce neural plasticity in the auditory cortex in a controlled manner. VNS paired with tones excluding the tinnitus frequency eliminated physiological and behavioral characteristics of tinnitus in noise-exposed rats. Several clinical trials followed and explored the effectiveness of VNS paired with sound stimuli for alleviating tinnitus in human subjects. Transcutaneous VNS (tVNS) has received increasing attention as a non-invasive alternative approach to tinnitus treatment. Several studies have also explored tVNS alone (not paired with sound stimuli) as a potential therapy for tinnitus. In this review, we discuss existing knowledge about direct and tVNS in terms of applicability, safety, and effectiveness in diminishing tinnitus symptoms in human subjects. This review includes all existing clinical and neuroimaging studies of tVNS alone or paired with acoustic stimulation in tinnitus patients and outlines the present limitations that must be overcome to maximize the potential of (t)VNS as a therapy for tinnitus.

Keywords: tinnitus, vagus nerve stimulation, transcutaneous vagus nerve stimulation, neuromodulation, auricular branch of vagus nerve

OPEN ACCESS

Edited by:
William Sedley,
Newcastle University, United Kingdom

Reviewed by:
Phillip Evan Gander,
The University of Iowa, United States
Rich Tyler,
The University of Iowa, United States

***Correspondence:**
Eui-Cheol Nam
birdynec@kangwon.ac.kr

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

Received: *15 March 2021*
Accepted: *05 May 2021*
Published: *28 May 2021*

Citation:
Yakunina N and Nam E-C (2021) Direct and Transcutaneous Vagus Nerve Stimulation for Treatment of Tinnitus: A Scoping Review. Front. Neurosci. 15:680590. doi: 10.3389/fnins.2021.680590

INTRODUCTION

Tinnitus is the perception of a phantom auditory sensation in the absence of an external sound source. It is one of the most prevalent auditory disorders, affecting 10–15% of the population, sometimes severely impairing quality of life (Davis and Rafaie, 2000; Baguley et al., 2013). In this paper we will discuss only subjective idiopathic tinnitus. The psychological model of tinnitus suggests that the overall annoyance of the tinnitus is a result of the tinnitus characteristics and the psychological make-up of each individual patient (Tyler et al., 1992). Treatments can be focused on reducing the tinnitus sensation (e.g., pharmacological) or on reducing the reactions to the tinnitus (e.g., psychological and cognitive training). A variety of treatments exist, including acoustic stimulation-based and brain stimulation therapies, but most tinnitus cases remain refractory to treatment.

One of the existing tinnitus models describes the tinnitus-generation mechanism as "maladaptive plastic re-organization of the auditory cortex" and suggests that tinnitus may develop as a result of auditory deafferentation related to peripheral hearing loss (Muhlnickel et al., 1998;

Norena and Eggermont, 2003; Shore et al., 2016). Reduced output from the affected cochlear region could induce loss of lateral inhibition from the damaged frequency areas, which may lead to elevated neural synchrony and hyperexcitability of the central auditory system (Eggermont and Roberts, 2004).

Vagus nerve stimulation (VNS) via surgical implantation is an FDA-approved procedure for the treatment of epilepsy and depression that is believed to trigger the release of neuromodulators in the brain (Schachter and Saper, 1998; Groves and Brown, 2005; Milby et al., 2008). Presentation of a tone together with a neuromodulator release, i.e., targeted neuromodulation, could increase the proportion of auditory cortical neurons that respond to the paired tone (Bakin and Weinberger, 1996; Kilgard and Merzenich, 1998; Martins and Froemke, 2015). In 2011, Engineer et al. (2011) reversed tinnitus-related cortical plastic maladaptation in a targeted and controlled way by applying VNS combined with acoustic stimulation. They demonstrated that VNS paired with interleaved multiple tones spanning the hearing range but excluding the tinnitus frequency eliminated the behavioral and neural manifestations of tinnitus in noise-exposed rats (Engineer et al., 2011). A few human studies that followed suggested that VNS paired with tones stripped of the tinnitus frequency might improve tinnitus-related symptoms (De Ridder et al., 2014a, 2015; Tyler et al., 2017).

Transcutaneous VNS (tVNS), applied at either the auricular or cervical branch of the vagus nerve, has been adopted as a non-invasive alternative to VNS (Yap et al., 2020). Multiple neuroimaging studies have confirmed that tVNS activates the same brain networks and pathways as those by direct VNS (Kraus et al., 2007, 2013; Frangos et al., 2015; Yakunina et al., 2017). A few clinical trials have explored the safety of tVNS alone (Kreuzer et al., 2012, 2014; Ylikoski et al., 2017, 2020), and neuroimaging studies have mapped its neuromodulatory effects (Lehtimaki et al., 2013; Hyvarinen et al., 2015; Yakunina et al., 2018) on the brain. The effectiveness of tVNS combined with acoustic stimulation has been explored by only two clinical trials to date (Lehtimaki et al., 2013; Shim et al., 2015).

Tailor-made notched music training (TMNMT) has been developed to reverse maladaptive plastic changes in the auditory cortex by enhancing lateral inhibition (Pantev et al., 2004). TMNMT employs music with an octave (or half octave) band centered on the patient's tinnitus, with the tinnitus frequency filtered out (notched) to augment lateral inhibition to the notched region (Okamoto et al., 2010; Pantev et al., 2012). Thus, TMNMT could be applied instead of the multiple tones stripped of the tinnitus frequency (paired with VNS) used in the 2011 study by Engineer et al. (2011) to increase the ratio of non-tinnitus frequency neurons and suppress cortical overrepresentation of the tinnitus frequency.

A recent systematic review reviewed nine studies on (t)VNS effect in reducing tinnitus symptoms (Stegeman et al., 2021). The review concluded that due to methodological limitations and low reporting quality of the studies, the effect of VNS on tinnitus remains unclear. In this review, in addition to discussing the applicability, safety, and effectiveness of direct and tVNS for treating tinnitus, we also review general knowledge about the mechanism, stimulation parameters, anatomical basis, and electrode placement in tVNS as well as relevant neuroimaging experiments. We then consider in detail how tVNS is currently applied for tinnitus treatment, all existing clinical and neuroimaging studies on tVNS alone or paired with sound stimuli in tinnitus patients, and its effectiveness. We discuss the quality of the existing theoretical basis for VNS for tinnitus. Finally, we outline fundamental gaps that must be overcome to maximize the efficacy of tVNS as a part of tinnitus therapy and discuss possible future directions for facilitating tNVS therapy.

VNS: TARGETING PLASTICITY

The cervical branches of the vagus nerve are mainly afferent sensory fibers that synapse in the nucleus tractus solitarius (NTS) and then project to the noradrenergic locus coeruleus (LC) in the brainstem and the cholinergic nucleus basalis (Leslie et al., 1982; Berthoud and Neuhuber, 2000) in the basal forebrain. They are involved in the release of neuromodulators such as acetylcholine, norepinephrine, serotonin, and brain-derived neurotrophic factor (Detari et al., 1983; Hassert et al., 2004; Follesa et al., 2007; Manta et al., 2009; Raedt et al., 2011) and subsequently influence the limbic, reticular, and autonomic centers of the brain (Sumal et al., 1983; Berntson et al., 1998; Berthoud and Neuhuber, 2000; Henry, 2002; Lulic et al., 2009; Manta et al., 2009). These neuromodulators seem to play key roles in promoting plastic changes (Bear and Singer, 1986; Kirkwood et al., 1999; Bramham and Messaoudi, 2005; Seol et al., 2007) (see Hays et al., 2013 for a review), although the precise mechanism of VNS neuromodulation remains unknown.

VNS paired with sensory stimuli or an active task has been shown to drive reorganization in various parts of the cerebral cortex. Pairing VNS with movement reorganizes the motor cortex (Porter et al., 2012), and VNS paired with physical rehabilitation improves recovery of motor function after a stroke (Khodaparast et al., 2013; Dawson et al., 2016; Pruitt et al., 2016; Kimberley et al., 2018). VNS paired with sensory stimuli restores sensory function after neurological injury (Meyers et al., 2019; Darrow et al., 2020a,b). Pairing extinction training with VNS reduces conditioned fear by modulating plasticity in the pathway from the ventromedial prefrontal cortex to the amygdala (Peña et al., 2014; Childs et al., 2015; Noble et al., 2019).

VNS paired with tones could induce reorganization of the tonotopic map in the auditory cortex, enhancing cortical responses to the paired tone and altering the fields across the entire auditory pathway, increasing the percentage of each field that responds to the paired tone frequency (Engineer et al., 2011, 2015, 2017; Shetake et al., 2012; Buell et al., 2019; Adcock et al., 2020).

tVNS: A NON-INVASIVE ALTERNATIVE

Anatomical Basis for tVNS

tVNS can be performed on the cervical branch of the vagus nerve in the neck. Cervical tVNS has been used to treat various disorders such as migraines, cluster headaches, and asthma

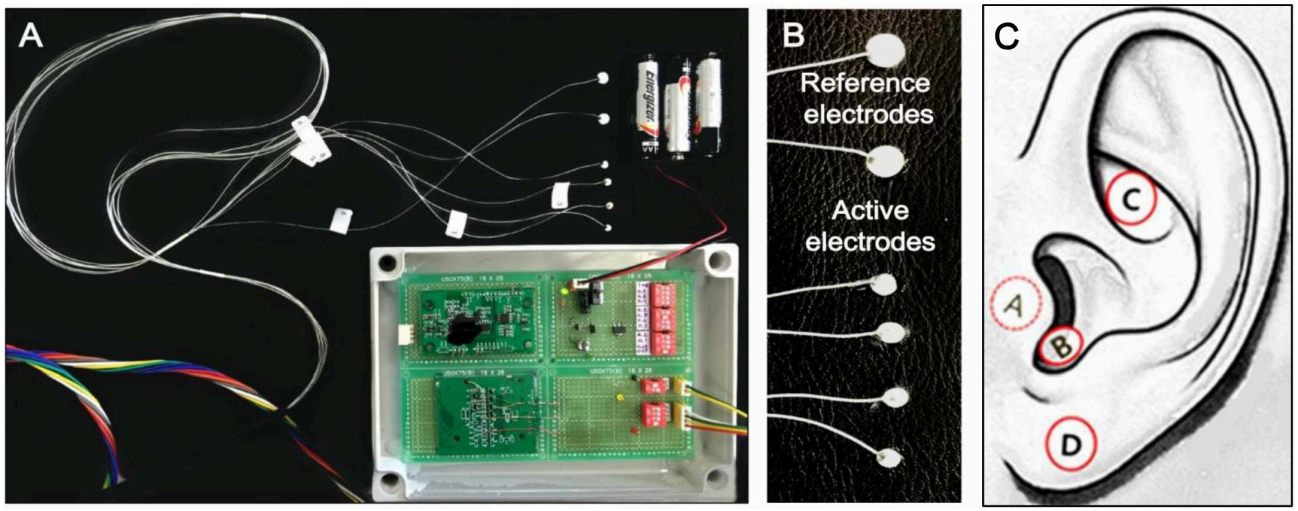


FIGURE 1 | (A) A custom-built MRI-compatible tVNS stimulator used in the tVNS fMRI studies by Yakunina et al., 2017, 2018 **(B)** Six silver electrodes (four active, two reference electrodes). **(C)** tVNS stimulation locations: inner surface of the tragus (A), inferoposterior wall of the external acoustic meatus (B), cymba conchae (C), and earlobe (sham; D).

(Yuan and Silberstein, 2016; Frangos and Komisaruk, 2017). In tinnitus patients, tVNS has been exclusively applied in the ear on the auricular branch of the vagus nerve (ABVN); hence, tVNS refers to auricular tVNS in this paper.

tVNS is usually performed in the left ear to avoid cardiac complications, as efferent vagal fibers to the heart are predominantly located on the right side (Nemeroff et al., 2006; Ogbonnaya and Kaliaperumal, 2013). The ABVN originates from the superior ganglion of the vagus nerve and innervates the external acoustic meatus and auricle (Tekdemir et al., 1998; Kiyokawa et al., 2014). A widely cited cadaver study on the AVBN conducted by Peuker and Filler (2002) concluded that the ABVN innervates most prominently the antihelix, tragus, and cymba conchae, but other studies have suggested denser innervation on the posterior wall of the external acoustic meatus (Kiyokawa et al., 2014; Watanabe et al., 2016). However, a number of functional magnetic resonance (fMRI) studies have shown a practical preference for the inner tragus (Dietrich et al., 2008; Kraus et al., 2013; Yakunina et al., 2017; Badran et al., 2018) and cymba concha (Frangos et al., 2015; Yakunina et al., 2017; Wang et al., 2018) over the ear canal's posterior wall as sites for stimulation of the ABVN. **Figure 1** demonstrates the tVNS locations used in a previous fMRI study (Yakunina et al., 2017). A recent extensive review concluded that the current literature lacks a clear consensus on the location most densely innervated by the ABVN, but given the existing research, it is reasonable to assume that the cymba concha and inner tragus are suitable locations for vagal modulation (Butt et al., 2020). Between the two locations, tVNS was suggested to activate the vagal pathway slightly more effectively at the concha than at the tragus (Yakunina et al., 2017). Locating and fixing an electrode might be structurally easier in the tragus (using a clip-on or insert-type electrode) than in the concha. However, a fMRI study showed that the maximal tolerable intensity of electrical stimulation was significantly higher in the conchae than in the tragus, so the degree of brain (de)activation was stronger with concha stimulation (Yakunina et al., 2018).

Optimal tVNS Parameters

A variety of pulse frequencies, widths, and intensities and shapes of electrical currents have been evaluated (see Yap et al., 2020 for a review), but no consensus on the optimal values has been achieved. In fMRI studies of normal human subjects using a pulse width of 250 μs and stimulation frequency of 25 Hz, tragus stimulation activated the left LC (Dietrich et al., 2008), and concha stimulation activated the LC and NTS (Frangos et al., 2015; **Table 1**). Using 500 μs at 25 Hz, stimulation of both locations activated the LC and NTS (Yakunina et al., 2017). In a subsequent study of tinnitus patients using the same tVNS parameters (25 Hz and 500 μs), Yakunina et al. (2018) replicated fMRI activation of the NTS and LC. However, a later study using the same parameters demonstrated similar cortical effects but not activation of the brainstem nuclei (Badran et al., 2018). That study set the stimulation intensity as 200% of the sensory threshold, whereas the Frangos et al. (2015) and Yakunina et al. (2017) used a current intensity just a step below the pain threshold (**Table 1**). Therefore, based on neuroimaging experiences so far, the following tVNS parameters could be applied to activate the classical vagal pathway: stimulation frequency of 25 Hz, pulse width of 250 or 500 μs, and tolerable maximal intensity.

To our knowledge, only one non-tinnitus study paired tVNS with sensory stimuli (Llanos et al., 2020). tVNS paired with speech stimuli robustly enhanced speech category learning and retention of correct stimulus-response associations. The authors used pulse width of 150 μs, frequency of 25 Hz, and amplitudes below each participant's perceptual level (1.24–1.67 mA). However, the observed behavioral changes immediately followed

TABLE 1 | Parameters used in the tVNS fMRI studies.

Study	Subjects	Location	Pulse width (μs)	Intensity level	LC/NTS activation
Dietrich et al. (2008)	Normal	Tragus	250	4–8 mA	Left LC
Frangos et al. (2015)	Normal	Concha	250	Just below pain threshold. 0.3–0.8 mA 0.43 ± 0.14	LC, NTS
Yakunina et al. (2017)	Normal	Tragus, concha	500	Just below pain threshold. 0.2–1.8 mA 0.77 ± 60.42 tragus 0.91 ± 60.47 concha	LC, NTS
Yakunina et al. (2018)	Tinnitus	Tragus, concha	500	Just below pain threshold. 0.1–1.8 mA 0.71 ± 0.43 tragus 0.80 ± 0.47 concha	LC, NTS
Badran et al. (2018)	Normal	Tragus	500	200% of sensory threshold. Sensory threshold: 1.57 ± 0.48 mA	None

All studies used a stimulation frequency of 25 Hz.

the single 25-min tVNS session and thus did not seem to be due to long-term plastic changes in sensory representation of the stimuli, but most likely resulted from processes related to the adjustment of the functional mapping between representations of stimulus signals and examined stimuli categories (Llanos et al., 2020). Polak et al. (2009) measured vagus somatosensory evoked potentials (VSEP) in healthy participants as a measure of vagus brainstem nuclei activity, and concluded that 8 mA without pain perception was the optimal tVNS intensity to maximize VSEP. However, it should be considered that the amount of stimulation actually delivered to the tissue depends on the electrode material and tissue impedance (and thus electrode location), and therefore stating the exact amplitude as optimal may not be proper (Yap et al., 2020). Furthermore, although individual tolerances may widely vary, 8 mA still appears well above pain threshold of most people (**Table 1**). Polak et al. (2009) also acknowledged that VSEP amplitudes are directly correlated to stimulation intensity, and thus our previous conclusion that a maximal intensity without feeling of pain may be optimal for tVNS appears valid. Altogether, further studies that would pair tVNS with sensory stimuli are needed to establish more optimized parameters for improving the efficiency of the tVNS method.

It is sensible to consider how tVNS compares to VNS in terms of its effectiveness for other disorders. A recent systematic review of tVNS in epilepsy reported the overall mean seizure reduction of approximately 42% in the treatment group, with about 43.4% of patients being responders, which was similar to the results of direct VNS with 50.6% of responders and mean seizure reduction of 44.6% (Wu et al., 2020). In depression, several systematic reviews concluded that the existing evidence for VNS efficacy in depression is not of sufficient quality to make clear conclusions (Lv et al., 2019). Non-randomized and not controlled studies report 42–53% response rate to VNS in depression (Nahas et al., 2005; Bajbouj et al., 2010), while a tVNS study reported similar response rate of 27–80% depending on the treatment duration (Rong et al., 2016). Therefore, tVNS appears to be similar to VNS in terms of its effectiveness for other disorders, and both treatments have about 50% rate of success.

NEUROIMAGING STUDIES: VNS AND tVNS EFFECTS ON BRAIN ACTIVITY

The effects of VNS have been studied using various neuroimaging methods such as single-photon emission computed tomography (SPECT), positron emission tomography (PET), and fMRI (Chae et al., 2003). Results using all methods suggested that VNS induces immediate as well as lasting changes in the thalamus, cerebellum, orbitofrontal cortex, limbic system, hypothalamus, and medulla (Bohning et al., 2001; Lomarev et al., 2002; Narayanan et al., 2002; Chae et al., 2003). It is noteworthy that the majority of the structures VNS affects are subcortical, while the paired VNS approach targets sensory cortex by pairing VNS-triggered release of neuromodulators with a sensory stimulus.

One of the most consistent findings in studies of acute VNS effects is diminished activity and reduced cerebral blood flow in the limbic system, namely the amygdala, hippocampus, cingulate cortex, ventral anterior cingulum, and parahippocampal gyrus (Zobel et al., 2005; Vonck et al., 2008). There is no universally accepted list of structures that comprise the limbic system, but the above areas are believed to constitute its core (Mega et al., 1997; Rajmohan and Mohandas, 2007).

Brain activation under tVNS in normal subjects has been explored by several fMRI studies (Kraus et al., 2007, 2013; Dietrich et al., 2008; Frangos et al., 2015; Yakunina et al., 2017; Badran et al., 2018; Sclocco et al., 2019), allowing human subjects to avoid radiation hazards from imaging using such methods as SPECT and PET, although the cables connected to stimulation electrodes (when they are placed in a circular pattern and generate an electrical current) can burn the skin in the contact area. These studies used different stimulation parameters and demonstrated diverse results other than deactivation of the limbic system (the amygdala, hippocampus, and parahippocampal gyri) (Kraus et al., 2007, 2013; Frangos et al., 2015; Yakunina et al., 2017). Additionally, activation in the thalamus, cerebellum, insula, and frontal gyrus (Kraus et al., 2007, 2013; Dietrich et al., 2008; Frangos et al., 2015; Yakunina et al., 2017; Badran et al., 2018) has been reported frequently. Activation of the LC and NTS in the brainstem is also often considered robust evidence of vagal activation (Dietrich et al., 2008; Frangos et al., 2015; Yakunina et al., 2017). **Figure 2** demonstrates activation of the LC and NTS following tVNS at the tragus and concha, but not the infero-posterial ear canal or earlobe (Yakunina et al., 2017). A recent ultrahigh-field 7T fMRI study explored the brainstem response to tVNS delivered during exhalation and found activation in the NTS, LC, and raphe nuclei (Sclocco et al., 2019). The authors suggested that exhalation-gated tVNS enhances NTS targeting.

A noteworthy difference between brain activation induced by tVNS and that by VNS is deactivation of the auditory cortices

TABLE 2 | Summary of neuroimaging studies on (t)VNS in patients with tinnitus.

Study	Method	Stimulation	Study group	Control group	Application time	Stimulation parameters	Results	Safety/side effects
Lehtimaki et al. (2013)	MEG	tVNS (tragus)	tVNS-on, tVNS-off, + tone matched with Ftinn $n = 8$	–	One session tVNS-on tVNS-off; Duration NS	∼0.8 mA (above sensory threshold) 25 Hz	tVNS decreased the amplitude of the auditory N1m response.	Possible acute adverse effects were not monitored during MEG
Hyvarinen et al. (2015)	MEG	tVNS (tragus)	tVNS-on, tVNS-off, + tone matched with Ftinn $n = 7$	tVNS in normal subjects (earlobe), +1 kHz tone $n = 8$	One session 6 min active 6 min sham	∼ 0.5 mA (just above sensory threshold) 25 Hz 500 μs	tVNS modulates the synchrony of tone-evoked tinnitus-related brain activity, especially at the beta and gamma bands.	NS
Yakunina et al. (2018)	fMRI	tVNS (tragus, concha)	tVNS $n = 36$	tVNS in normal subjects (earlobe) $n = 37$	Two sessions per location 30 s on 30 s off 5 min per session	0.3–3.0 mA 0.71 ± 0.43 tragus, 0.80 ± 0.47 concha (just below pain threshold) 25 Hz 500 μs	tVNS deactivated multiple brain areas related to tinnitus generation and related distress. Both locations activated the vagal pathway in the brainstem.	No adverse effects

EEG, electroencephalography; fMRI, functional magnetic resonance imaging; Ftinn, tinnitus frequency; MEG, magnetoencephalography; NS, not stated.

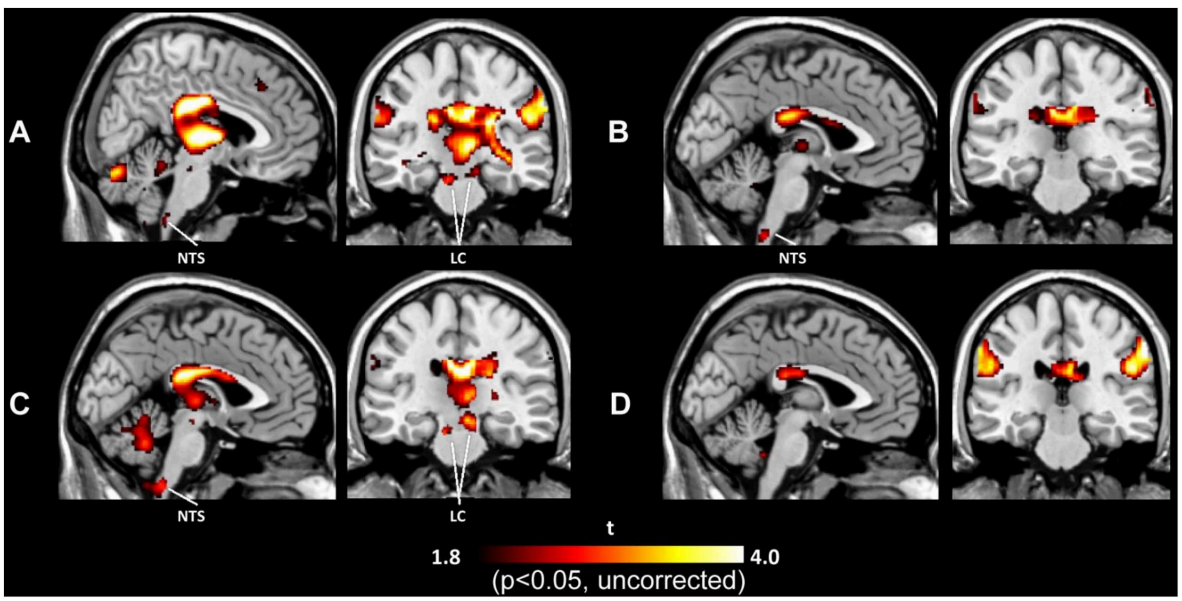


FIGURE 2 | Activation maps induced by tVNS in the tragus **(A)**, inferoposterior wall of the external acoustic meatus **(B)**, concha **(C)**, and earlobe **(D)**. tVNS at the tragus and concha activated the locus coeruleus (LC) and nucleus of the solitary tract (NTS), while tVNS at the ear canal and earlobe did not activate either brain center.

(superior and middle temporal gyri) following tVNS but not VNS (Kraus et al., 2007; Yakunina et al., 2017). Deactivation of these temporal cortices might reflect multisensory integration of the somatosensory and auditory systems. The medullary somatosensory nuclei receive sensory inputs from cranial and cervical nerves including the trigeminal, facial, glossopharyngeal, vagal, C1, and C2 nerves and then project to the dorsal cochlear nucleus in the auditory system (Young et al., 1995). This auditory–somatosensory connection seems to be involved more with ABVN than with the other branches of the vagus nerve. Electrical stimulation of the earlobe (innervated mostly by the cervical nerve) induces similar auditory cortex deactivation (Yakunina et al., 2017), possibly via the same multisensory integration. Therefore, the earlobe may not be a good sham location for tVNS studies. Earlobe stimulation has been used in cranial electrotherapy stimulation, which is FDA approved for the treatment of insomnia, depression, and anxiety and induces deactivation in several

brain regions that overlap with those deactivated by tVNS (Yakunina et al., 2017).

(t)VNS EFFECTS IN PATIENTS WITH TINNITUS

Neuroimaging Studies of tVNS in Tinnitus Patients

Lehtimaki et al. (2013) performed magnetoencephalography (MEG) while stimulating tinnitus patients with tVNS; a pure tone matched to the tinnitus frequency was presented continuously under both tVNS-on and tVNS-off conditions (**Table 2**). The amplitude of auditory N1m responses to the tone was reduced following the application of tVNS. This was a pilot study performed on only eight subjects; no control group was used. Another MEG study used the same stimulation setup but included a control group of normal hearing subjects who underwent sham tVNS at the earlobe (Hyvarinen et al., 2015). Tinnitus patients had higher beta- and gamma-band synchrony compared with the control group at baseline. tVNS at the tragus induced a reduction in beta and gamma synchrony in accordance with tinnitus severity. Sham tVNS had only a weak effect on the normalized spectrum at frontal alpha and beta and no effect on the measures of synchrony. The amount of gamma-band synchronization in the human auditory cortex was correlated with the subjective loudness of tinnitus (Van Der Loo et al., 2009). The authors concluded that tVNS was successful in modulating tinnitus-related beta- and gamma-band activity and thus could have potential in the treatment of tinnitus. The study was done on a small sample size (7–8 subjects).

The only extant tVNS fMRI study in tinnitus patients found that tVNS deactivated the auditory and limbic areas (**Figure 3**; Yakunina et al., 2018). Numerous neuroimaging studies supported Jastreboff's neurophysiological model, which suggests an abnormally strong connection between the auditory and limbic systems in tinnitus patients (Jastreboff, 1990; Chen et al., 2017). Various other non-auditory brain areas associated with tinnitus were also deactivated by tVNS, such as the cingulate cortex, precuneus, and frontal gyrus.

It should be noted that both controlled studies used the earlobe for sham stimulation which, as discussed above, may not be a good location for sham tVNS.

Together, these findings indicate that tVNS suppresses tinnitus-related brain networks, reduces auditory N1m responses, and reduces the level of beta- and gamma-band synchrony, which is elevated in individuals with tinnitus and is correlated with tinnitus loudness. Therefore, neuroimaging results suggest that tVNS affects brain activity in a way that potentially reduces the generation of tinnitus and tinnitus-related distress.

Clinical Studies of (t)VNS Paired With Sound Stimuli in Tinnitus Patients

Since the 2011 study by Engineer et al. (2011) on VNS paired with tones in a rat model, numerous studies on direct and tVNS with and without paired sound stimuli in tinnitus patients have been published. Among them, seven studies explored tVNS paired with sound stimuli (**Table 3**), of which six explored the efficacy of paired tVNS for reducing tinnitus symptoms, and the remaining study explored the effects of paired VNS on voice and hearing. Five and two of the seven studies evaluated VNS and tVNS, respectively.

All five VNS studies paired VNS with pure tones that excluded the tinnitus frequency (De Ridder et al., 2014a, 2015; Tyler et al., 2017; Vanneste et al., 2017; Kochilas et al., 2020). The study durations ranged from 3 to 12 weeks. VNS parameters were identical in all studies. In the only randomized double-blind trial (Tyler et al., 2017), after 6 weeks of VNS paired with tones, the scores on the Tinnitus Handicap Inventory (THI) were significantly improved in the active group compared with controls who received VNS only. Improvement was seen in 50% of the participants in the paired VNS group compared with only 28% of controls. However, the sample size was small (16 patients in the VNS group). The other controlled study used partial data from a previous study exploring the effect of paired VNS on voice and showed that paired VNS does not have a negative effect on voice in tinnitus patients (Kochilas et al., 2020). One of the most common side effects of VNS is its possible effect on voice; it can reduce vocal cord motion on the implantation side with secondary supraglottic muscle tension, causing voice changes or hoarseness (Al Omari et al., 2017).

A case report of a patient with chronic tinnitus unresponsive to various previous therapies (De Ridder et al., 2015) showed improvement in tinnitus-related symptoms that lasted for 2 months after treatment with VNS paired with tones without tinnitus frequency.

Other VNS studies were uncontrolled and recruited too few subjects. Four of ten subjects exhibited clinically meaningful improvement in their tinnitus [THI and minimum masking level (MML)] (De Ridder et al., 2014a) that lasted more than 2 months after therapy. They claimed that the patients on medications that might interfere with VNS-released neuromodulators showed no improvement in their tinnitus. The same group retrospectively analyzed the results of EEG done before and immediately after 1–3 months of VNS paired with tones (Vanneste et al., 2017). The study group was a subset of patients from two previous VNS + tones trials (De Ridder et al., 2014b; Tyler et al., 2017). VNS–tone pairing reduced gamma-band activity in the left auditory cortex. The reduction in gamma-band activity was correlated with the degree of loudness reduction. Paired VNS also reduced the phase coherence between the auditory cortex and areas associated with tinnitus distress, including the cingulate cortex and the parahippocampus. The authors argued that these results support the hypothesis that VNS paired with sound stimuli could direct therapeutic neural plasticity.

Two uncontrolled clinical trials investigated the feasibility of tVNS paired with sound stimuli (TMNMT) for tinnitus treatment. Ten sessions of paired tVNS at the tragus in 10 subjects showed significant reductions in the mean THI and Mini Tinnitus Questionnaire (mini-TQ) scores, and subjective loudness and annoyance were also decreased by more than

TABLE 3 | Summary of the existing studies on (t)VNS paired with sound stimuli in patients with tinnitus.

Study	Stimulation	Study group	Control group	Application time	Stimulation parameters	Explored variables	Results	Safety/side effects
De Ridder et al. (2014a)	VNS	VNS + tones without Ftinn *n* = 10	–	20 days 2.5 h/day 2-month follow-up	0.8 mA 100 μs 30 Hz Every 30 s	THI MML	Average THI decreased by 11.78% (28.17 in the no-drug group). Clinically significant improvements in 4 of 10.	Post-surgery complications in two patients (redness; vocal cord hypomobility; infection)
De Ridder et al. (2015)	VNS	VNS + tones without Ftinn *n* = 1 (case report)	Tones *n* = 1 (same subject)	20 days 5 d/week 2.5 h/day	0.8 mA 100 μs 30 Hz Every 30 s	THI TRQ THQ TAQ MML	THI reduced by 48% TRQ reduced by 68% Control: both THI and TRQ increased.	No adverse effects
Tyler et al. (2017)	VNS	VNS + tones without Ftinn *n* = 16	VNS *n* = 14	6 weeks 7 d/week 2.5 h/day	0.8 mA 100 μs 30 Hz 0.5-s train delivered every 30 s	THI	Active group had significantly improved THI, while control group did not. THI improved in 50% of active group vs. 28% of controls. VNS + tones is more effective for subgroup of tinnitus patients that do not have hissing or blast-induced tinnitus.	Mild, well-tolerated adverse effects; similar profile to VNS in epilepsy
Vanneste et al. (2017)	VNS	VNS + tones without Ftinn *n* = 18	–	4–12 weeks 2.5 h/day	0.8 mA 100 μs 30 Hz 0.5-s train delivered every 30 s	EEG THI VAS loudness	Reduced gamma band activity in the left auditory cortex and reduced phase coherence between the auditory cortex and areas associated with tinnitus distress. Significant reduction of THI and VAS.	NS
Kochilas et al. (2020)	VNS	VNS + tones without Ftinn *n* = 7	VNS and tones separately	6 weeks 2.5 h/day	0.8 mA 100 μs 30 Hz 0.5-s train delivered every 30 s	Voice characteristics; PTA; MML; tinnitus pitch; monosyllabic word recognition	VNS + tones have no adverse impact on voice. Slight PTA threshold increases were observed at 2, 3, and 4 kHz.	Minimal or no adverse effects
Lehtimaki et al. (2013)	tVNS (tragus)	tVNS + TMNM *n* = 10	–	10 days 7 sessions, each session 45–60 min	∼0.8 mA (above sensory threshold) 25 Hz	WHO-5-point VAS THI Mini-TQ	Increased WHO well-being scores (56 → 76); THI, mean VAS loudness and annoyance, mini-TQ reduced significantly.	No adverse effects observed
Shim et al. (2015)	tVNS (concha)	tVNS + TMNM *n* = 30	–	10 sessions, each session 30 min, 1–4 days between sessions	1–10 mA (highest possible) 200 μs 25 Hz	GI index VAS loudness Tinnitus awareness score THI	GI score improved in 50% of patients. Mean loudness and awareness improved significantly THI improved, but not significantly (41.5 → 35.4).	No adverse effects reported

Ftinn, tinnitus frequency; GI, global improvement; Mini-TQ, Mini Tinnitus Questionnaire; MML, Minimum masking level; NS, not stated; PTA, pure tone average; TAQ, Tinnitus Activities Questionnaire; THI, Tinnitus Handicap Inventory; THQ, Tinnitus Handicap Questionnaire; TMNM, tailor-made notched music; TRQ, Tinnitus Reaction Questionnaire; VAS, visual analog scale; WHO-5, World Health Organization-Five Well-Being Index.

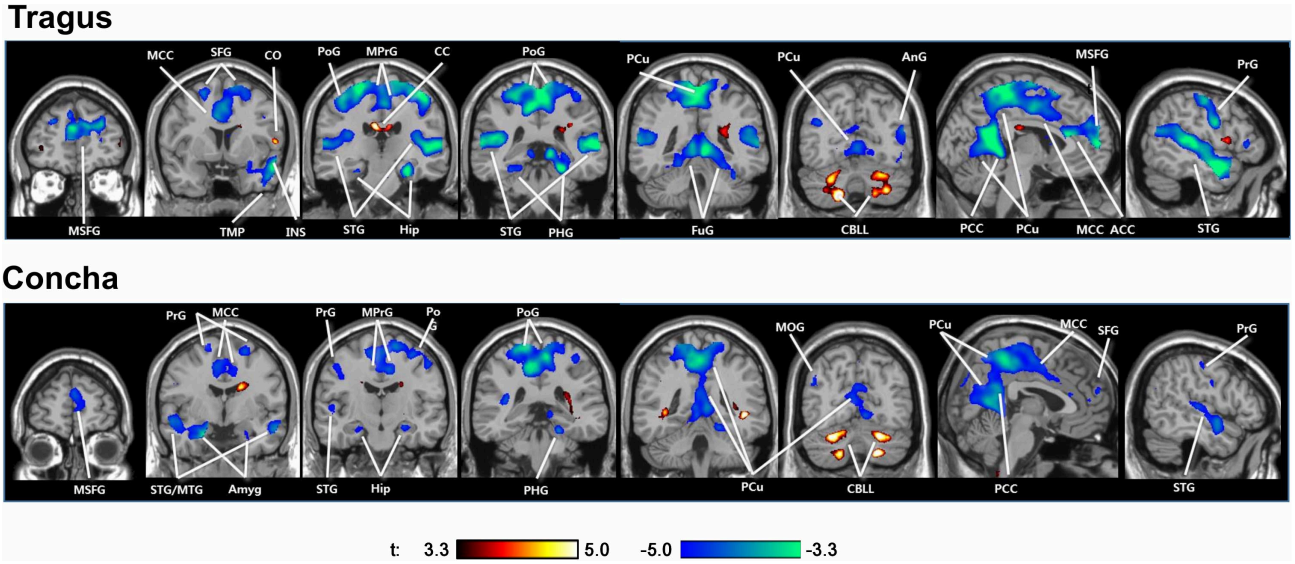


FIGURE 3 | Activations (red) and deactivations (blue) induced by tVNS at the tragus and cymba conchae in tinnitus patients. tVNS resulted in deactivation of the auditory and limbic systems, as well as a number of other cortical areas. ACC/MCC/PCC, anterior/middle/posterior cingulate cortex; Amyg, amygdala; AnG, angular gyrus; CC, corpus callosum; CBLL, cerebellum; FuG, fusiform gyrus; Hip, hippocampus; LiG, lingual gyrus; MOG, middle orbital gyrus; MTG/STG, middle/superior temporal gyrus; PCu, precuneus; PoG/PrG, postcentral/precentral gyrus; SFG, superior frontal gyrus; TMP, temporal pole.

20 points (Lehtimaki et al., 2013). The other study applied 10 sessions of paired tVNS at the concha in 30 subjects; the mean subjective loudness and tinnitus awareness score (but not THI) were significantly improved after treatment (Shim et al., 2015).

Overall, direct or tVNS paired with sound stimuli seems to produce a positive therapeutic effect on reducing tinnitus symptoms. However, small sample sizes, the absence of controls, and the lack of long-term follow-up seriously limit the reliability of these studies. Furthermore, given the wide heterogeneity of tinnitus pathophysiology and symptoms, well-organized systematic studies are needed to establish the effects of paired tVNS, particularly in different subgroups of tinnitus patients.

No adverse effects have been observed in tVNS studies, but VNS studies have often reported some adverse effects related to the invasiveness of the procedure (**Table 3**). VNS is reportedly a well-tolerated and relatively simple surgical procedure, but it is costly and carries the risk of side effects such as cough, hoarseness, voice alteration, and paresthesia (Ben-Menachem, 2001). Both direct and tVNS present cardiac-related risks, such as bradycardia and cardiac asystole (Ben-Menachem, 2001). Therefore, patients with cardiac disorders and implanted devices such as pacemakers should be screened out, and heart rate should be monitored during the stimulation periods.

Clinical Studies on (t)VNS Alone in Tinnitus Patients

Several studies have investigated the effects of stand-alone tVNS in tinnitus patients (**Table 4**). The cardiac safety of long-term tVNS in tinnitus patients (Kreuzer et al., 2012) was demonstrated by ECG in subjects without cardiac disease. In another study, tVNS alone induced no clinically relevant improvements in tinnitus-related complaints even after 24 weeks of treatment (Kreuzer et al., 2014). Suk et al. (2018) applied four sessions of 4-min tVNS at each of three stimulation sites for 2 weeks. At 4 weeks after the end of treatment, tinnitus-related visual analog scale (VAS) scores were improved by at least 50% in 33–45% of the patients, and THI and Beck Depression Inventory scores also improved significantly. The authors reported no significant differences in tinnitus relief according to tVNS intensity; a stepwise increase in intensity up to the tolerable maximum level did not produce significant differences in the treatment outcome. However, in addition to the absence of a control group, the short treatment duration (4-min session at each location for 2 weeks) was different from other tVNS protocols, which performed treatment for several hours each day for several weeks (**Table 3**). Wichova et al. (2018) retrospectively evaluated epilepsy patients with tinnitus who received VNS as a treatment for their epilepsy. Phone inquiries regarding changes in the loudness of tinnitus (VAS) showed that 16 of 20 patients had at least one quieter moment. Furthermore, the difference between pre- and postoperative loudness was statistically significant. However, the results of that study were difficult to interpret reliably because of the retrospective nature, fact that the preoperative VAS was recorded postoperatively, inconsistent VNS settings among patients, absence of a control group, small sample size, and unknown treatment period, among other factors.

Ylikoski et al. (2017, 2020) evaluated acute tVNS effects on the autonomic nervous system in tinnitus patients by measuring heart rate variability before and after a tVNS session. They concluded that tVNS induces a shift in autonomic

TABLE 4 | Summary of the existing studies on (t)VNS alone (not paired with sound stimuli) in patients with tinnitus.

Study	Stimulation	Study purpose	Sample size	Application time	Stimulation parameters	Primary variable	Results	Safety/side effects
Kreuzer et al. (2014)	tVNS (site NS)	Feasibility, safety, efficacy for tinnitus treatment	*n* = 24 in phase 1, *n* = 26 in phase 2	Phase 1: 30 s on/180 s off 6 h/day 45.5 ± 21.0 days Phase 2: 30 s on/30 s off 4 h/day 24 weeks	0.1–10 mA (just above sensory threshold) 25 Hz	TQ THI TBF-12 BDI VAS ECG	No clinically relevant improvement of tinnitus complaints. tVNS considered safe in patients with no history of cardiac disease.	Twitching and pressure at electrode placement site. No significant changes in cognitive testing.
Suk et al. (2018)	tVNS (cavum, concha, tragus)	Clinical significance of tVNS intensity in tinnitus	*n* = 24	4 min per site sequentially; 4 sessions over 2 weeks	Intensity increased by 1 mA every 5 s until maximum without pain 30 Hz 200 μs	THI BDI VAS	VAS scores improved at least 50% in 33.3–62.5% of patients THI and BDI were reduced significantly. No effect of stimulus intensity on treatment outcome	No adverse effects
Wichova et al. (2018)	VNS	Effect on perception of tinnitus in epilepsy patients	*n* = 20	NS	NS	VAS loudness	Post-operative VAS loudness decreased significantly. At least a one-point decrease in VAS loudness in 80% of patients	Transient post-operative hoarseness in one patient
Kreuzer et al. (2012)	tVNS (site NS)	Cardiac safety in tinnitus	*n* = 24	30 s on/180 s off 5.15 ± 1.80 h/day 24.0 ± 19.3 days	3.2 ± 2.5 (0.1–10) mA (just above sensory threshold) 25 Hz	ECG	tVNS is feasible in tinnitus patients without signs of long-term worsening of tinnitus complaints. In subjects with no known pre-existing cardiac pathology no arrhythmic effects of tVNS were seen.	Two adverse cardiac events (one severe) registered, but considered very unlikely to be tVNS-induced
Ylikoski et al. (2017)	tVNS (site NS)	Acute effects on autonomic nervous system imbalance	*n* = 97	One session 15–60 min	0.1–130 mA (above sensory threshold) 25 Hz 250 μs	HRV	tVNS can induce a shift from sympathetic preponderance toward parasympathetic predominance	No major morbidity, no cardiac or circulatory effects during active tVNS
Ylikoski et al. (2020)	tVNS (tragus)	Clinical features, psychophysiological characteristics, and HRV	tVNS *n* = 171 tVNS + TRT *n* = 78	One session 15–60 min for HRV assessment/ 60–90 min/day 5 d/week 1 year	0.3–3.0 mA (above sensory threshold) 25 Hz 250 μs	HRV TQ	tVNS improved parasympathetic function, most efficiently in patients with a low starting HRV level. tVNS + TRT alleviated tinnitus stress and handicap in 76% of patients	No cardiac adverse effects in 250 tVNS patients

None of the studies had a control group.

BDI, Beck Depression Inventory; ECG, electroencephalography; Ftinn, tinnitus frequency; HRV, heart rate variability; NS, not stated; TBF-12, Tinnitus Impairment Questionnaire; THI, Tinnitus Handicap Inventory; TRT, tinnitus retraining therapy; TQ, Tinnitus Questionnaire; VAS, visual analog scale.

nervous system functioning from sympathetic preponderance toward parasympathetic predominance, thus reducing the stress-related imbalance in the autonomic nervous system. A consistent improvement in heart rate variability, which is considered a useful marker of mental stress, was observed in 90% of the patients (Ylikoski et al., 2017, 2020). The authors argued that tVNS can be a helpful therapeutic tool in reducing tinnitus-related mental stress. Their second study reported significant decreases in loudness and tinnitus-related annoyance in 78 tVNS-treated patients; however, all patients simultaneously received tinnitus retraining therapy (TRT), a form of instructive counseling combined with sound therapy using white noise. Thus, the tVNS effect could not be evaluated separately from the counseling effect of the tinnitus retraining therapy.

Engineer's method for tinnitus treatment using VNS, based on the tonotopic model of tinnitus, involves paired sound stimuli; it is the sound stimuli that initiates rearrangement of the auditory cortex to eliminate the origin of tinnitus percept (Engineer et al., 2011). VNS enforces this process through action of several neuromodulators, thus promoting neural plasticity. Therefore, (t)VNS alone is not expected to be effective as tinnitus therapy following the original "targeted plasticity" method. Nevertheless, a number of studies attempted to demonstrate that tVNS alone without any paired stimuli can be successful in relieving tinnitus symptoms. The main reasoning behind these attempts is that tVNS modulates the auditory and limbic areas, as shown by neuroimaging studies in normal subjects and tinnitus patients. Additional arguments include VNS positive effect on habituation, its antidepressant mode of action and influence on vegetative nervous system, and clinical data on electrical auricular stimulation that might have involved unintentional vagal stimulation (Kreuzer et al., 2014). tVNS also reduces imbalance of the autonomic nervous system related to tinnitus-induced stress (Ylikoski et al., 2020).

Nevertheless, no strong or clear evidence that tVNS alone reduces tinnitus-related symptoms is currently available. However, this does not mean that tVNS has no potential as a tinnitus-relief therapy but rather points to the need for more reliable studies to provide robust evidence of its ability to relieve tinnitus.

DISCUSSION

VNS paired with sound stimuli as tinnitus treatment was first proposed by Engineer et al. (2011). The study was performed on rats; tinnitus was induced by exposing rats to loud octave-band noise centered on a certain frequency. Inability of noise-exposed rats to detect silent gaps in narrow-band noise centered on the assumed tinnitus frequency was taken as a behavioral correlate of tinnitus; the gap impairment was eliminated after rats were continuously treated with VSN paired with tones excluding the frequency of tinnitus. To the best of our knowledge, all existing studies replicating these results are from the same group of authors (Shetake et al., 2012; Engineer et al., 2015, 2017; Buell et al., 2019; Adcock et al., 2020). No replication of this work by an independent research team exists. This is the first and most glaring gap in existing literature on the subject; the entire field of VNS paired with sound stimuli as a therapy for tinnitus is based on work by a single research team. The first and most urgent need is to replicate and verify these results.

Furthermore, the method is based on the tonotopic model of tinnitus which claims that auditory cortical reorganization is a primary cause for tinnitus. This model has multiple flaws and has been strongly criticized. Among arguments against it is the fact that tinnitus can develop without preceding extensive hearing loss which is generally required for cortical reorganization; no human study demonstrated the evidence for cortical reorganization in tinnitus patients; tinnitus often occurs instantaneously within a few seconds following noise exposure which is not enough time to generate tonotopical plasticity in the cortex (see Knipper et al., 2013 for a review). Additionally, overwhelming majority of studies exploring the relationship between the tinnitus pitch and audiometric edge frequency found no connection between the two variables (see Yakunina and Nam, 2021 for a review). Hence, while we cannot exclude the possibility that cortical map reorganization may be the reason for developing tinnitus in some individuals, it is by far nor necessary neither always existing condition for tinnitus.

Despite the fact that VNS is normally a well-tolerated and relatively simple surgical procedure, it nevertheless is invasive. tVNS is a non-invasive and thus cheaper, safer and easier-to-implement alternative. A new invasive experimental therapy should generally not be used in human subjects until there is a strong effectiveness evidence base. Therefore, (paired) tVNS rather than VNS for tinnitus therapy should receive primary focus in future research, particularly given its own challenges such as finding a proper site for active and sham stimulation as discussed above. Before VNS is considered for tinnitus treatment, the effectiveness of tVNS for tinnitus should be firmly established.

As of now, there is no evidence showing effectiveness of either treatment in relieving tinnitus as all existing studies are of poor quality, which is confirmed by a recent systematic review (Stegeman et al., 2021). Future studies should follow the Consolidated Standards of Reporting Trials (CONSORT) recommendations, which provide guidelines for designing and reporting randomized control trials that include double blinding and a suitable sample size (Schulz et al., 2010). The calculation of the sample size is one of the most important steps in designing a randomized controlled trial; it majorly influences the statistical reliability of the findings. And yet, this step is most frequently omitted in existing clinical studies. At the same time, individual results should not be neglected and are encouraged to be reported, since there are many different subtypes of tinnitus, and the reaction to treatment can vary widely depending on individual tinnitus and psychological characteristics (Tyler et al., 2007, 2008). Primary and secondary variables should be carefully chosen based on the existing considerations for the design of tinnitus clinical trials (Tyler et al., 2006, 2007).

CONCLUSION

Direct VNS or tVNS paired with sound stimuli excluding the individual's tinnitus frequency appears to have potential as a treatment method for alleviating tinnitus symptoms; however, no reliable study exists on this topic as yet. All existing studies have major flaws such as the absence of a control group, small sample size, and lack of randomization and blinding. Similarly, there is no reliable evidence to date showing that (t)VNS alone without paired sound stimuli is effective for tinnitus treatment. Higher-quality research is needed for both paired and unpaired tVNS to compensate for the flaws of existing studies and address the gaps in the current knowledge on the subject, such as proper stimulation parameters, longer follow-up periods, and the most responsive tinnitus populations.

AUTHOR CONTRIBUTIONS

NY wrote the manuscript. E-CN curated and supervised the study, and edited the manuscript. Both authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by the Bio and Medical Technology Development Program of the National Research Foundation funded by the Ministry of Science and Information and Communications Technologies of the Korean Government (Grant No. 2016M3A9F1941022).

REFERENCES

Adcock, K. S., Chandler, C., Buell, E. P., Solorzano, B. R., Loerwald, K. W., Borland, M. S., et al. (2020). Vagus nerve stimulation paired with tones restores auditory processing in a rat model of Rett syndrome. *Brain Stimul.* 13, 1494–1503. doi: 10.1016/j.brs.2020.08.006

Al Omari, A. I., Alzoubi, F. Q., Alsalem, M. M., Aburahma, S. K., Mardini, D. T., and Castellanos, P. F. (2017). The vagal nerve stimulation outcome, and laryngeal effect: otolaryngologists roles and perspective. *Am. J. Otolaryngol.* 38, 408–413. doi: 10.1016/j.amjoto.2017.03.011

Badran, B. W., Dowdle, L. T., Mithoefer, O. J., LaBate, N. T., Coatsworth, J., Brown, J. C., et al. (2018). Neurophysiologic effects of transcutaneous auricular vagus nerve stimulation (taVNS) via electrical stimulation of the tragus: a concurrent taVNS/fMRI study and review. *Brain Stimul.* 11, 492–500. doi: 10.1016/j.brs.2017.12.009

Baguley, D., McFerran, D., and Hall, D. (2013). Tinnitus. *Lancet* 382, 1600–1607.

Bajbouj, M., Merkl, A., Schlaepfer, T. E., Frick, C., Zobel, A., Maier, W., et al. (2010). Two-year outcome of vagus nerve stimulation in treatment-resistant depression. *J. Clin. Psychopharmacol.* 30, 273–281.

Bakin, J. S., and Weinberger, N. M. (1996). Induction of a physiological memory in the cerebral cortex by stimulation of the nucleus basalis. *Proc. Natl. Acad. Sci. U S A* 93, 11219–11224. doi: 10.1073/pnas.93.20.11219

Bear, M. F., and Singer, W. (1986). Modulation of visual cortical plasticity by acetylcholine and noradrenaline. *Nature* 320, 172–176. doi: 10.1038/320172a0

Ben-Menachem, E. (2001). Vagus nerve stimulation, side effects, and long-term safety. *J. Clin. Neurophys.* 18, 415–418. doi: 10.1097/00004691-200109000-00005

Berntson, G. G., Sarter, M., and Cacioppo, J. T. (1998). Anxiety and cardiovascular reactivity: the basal forebrain cholinergic link. *Behav. Brain Res.* 94, 225–248. doi: 10.1016/s0166-4328(98)00041-2

Berthoud, H., and Neuhuber, W. L. (2000). Functional and chemical anatomy of the afferent vagal system. *Autonom. Neurosci.* 85, 1–17. doi: 10.1016/s1566-0702(00)00215-0

Bohning, D. E., Lomarev, M. P., Denslow, S., Nahas, Z., Shastri, A., and George, M. S. (2001). Feasibility of vagus nerve stimulation-synchronized blood oxygenation level-dependent functional MRI. *Invest. Radiol.* 36, 470–479. doi: 10.1097/00004424-200108000-00006

Bramham, C. R., and Messaoudi, E. B. D. N. F. (2005). function in adult synaptic plasticity: the synaptic consolidation hypothesis. *Prog. Neurobiol.* 76, 99–125. doi: 10.1016/j.pneurobio.2005.06.003

Buell, E. P., Borland, M. S., Loerwald, K. W., Chandler, C., Hays, S. A., Engineer, C. T., et al. (2019). Vagus nerve stimulation rate and duration determine whether sensory pairing produces neural plasticity. *Neuroscience* 406, 290–299. doi: 10.1016/j.neuroscience.2019.03.019

Butt, M. F., Albusoda, A., Farmer, A. D., and Aziz, Q. (2020). The anatomical basis for transcutaneous auricular vagus nerve stimulation. *J. Anat.* 236, 588–611. doi: 10.1111/joa.13122

Chae, J., Nahas, Z., Lomarev, M., Denslow, S., Lorberbaum, J. P., Bohning, D. E., et al. (2003). review of functional neuroimaging studies of vagus nerve stimulation (VNS). *J. Psychiatr. Res.* 37, 443–455. doi: 10.1016/s0022-3956(03)00074-8

Chen, Y. C., Xia, W., Chen, H., Feng, Y., Xu, J. J., Gu, J. P., et al. (2017). Tinnitus distress is linked to enhanced resting-state functional connectivity from the limbic system to the auditory cortex. *Hum. Brain Mapp.* 38, 2384–2397. doi: 10.1002/hbm.23525

Childs, J. E., Alvarez-Dieppa, A. C., McIntyre, C. K., and Kroener, S. (2015). Vagus nerve stimulation as a tool to induce plasticity in pathways relevant for extinction learning. *JoVE* 102:e53032.

Darrow, M. J., Mian, T. M., Torres, M., Haider, Z., Danaphongse, T., Rennaker, Jr RL, et al. (2020a). Restoration of somatosensory function by pairing vagus nerve stimulation with tactile rehabilitation. *Ann. Neurol.* 194–205. doi: 10.1002/ana.25664

Darrow, M. J., Mian, T. M., Torres, M., Haider, Z., Danaphongse, T., Seyedahmadi, A., et al. (2020b). The tactile experience paired with vagus nerve stimulation determines the degree of sensory recovery after chronic nerve damage. *Behav. Brain Res.* 396:112910. doi: 10.1016/j.bbr.2020.112910

Davis, A., and Rafaie, E. A. (2000). Epidemiology of tinnitus. *Tinnitus Handbook* 1:23.

Dawson, J., Pierce, D., Dixit, A., Kimberley, T. J., Robertson, M., Tarver, B., et al. (2016). Safety, feasibility, and efficacy of vagus nerve stimulation paired with upper-limb rehabilitation after ischemic stroke. *Stroke* 47, 143–150. doi: 10.1161/strokeaha.115.010477

De Ridder, D., Kilgard, M., Engineer, N., and Vanneste, S. (2015). Placebo-controlled vagus nerve stimulation paired with tones in a patient with refractory tinnitus: a case report. *Otol. Neurotol.* 36, 575–580. doi: 10.1097/mao.0000000000000704

De Ridder, D., Vanneste, S., Engineer, N. D., and Kilgard, M. P. (2014a). Safety and efficacy of vagus nerve stimulation paired with tones for the treatment of tinnitus: a case series. *Neuromodulation* 17, 170–179. doi: 10.1111/ner.12127

De Ridder, D., Vanneste, S., Weisz, N., Londero, A., Schlee, W., Elgoyhen, A. B., et al. (2014b). An integrative model of auditory phantom perception: tinnitus as a unified percept of interacting separable subnetworks. *Neurosci. Biobehav. Rev.* 44, 16–32. doi: 10.1016/j.neubiorev.2013.03.021

Detari, L., Juhasz, G., and Kukorelli, T. (1983). Effect of stimulation of vagal and radial nerves on neuronal activity in the basal forebrain area of anaesthetized cats. *Acta Physiol. Hung.* 61, 147–154.

Dietrich, S., Smith, J., Scherzinger, C., Hofmann-Preiss, K., Freitag, T., Eisenkolb, A., et al. (2008). novel transcutaneous vagus nerve stimulation leads to brainstem and cerebral activations measured by functional MRI. *Biomed. Tech.* 53, 104–111. doi: 10.1515/bmt.2008.022

Eggermont, J. J., and Roberts, L. E. (2004). The neuroscience of tinnitus. *Trends Neurosci.* 27, 676–682.

Engineer, C. T., Engineer, N. D., Riley, J. R., Seale, J. D., and Kilgard, M. P. (2015). Pairing speech sounds with vagus nerve stimulation drives stimulus-specific cortical plasticity. *Brain Stimul.* 8, 637–644. doi: 10.1016/j.brs.2015.01.408

Engineer, C. T., Shetake, J. A., Engineer, N. D., Vrana, W. A., Wolf, J. T., and Kilgard, M. P. (2017). Temporal plasticity in auditory cortex improves neural discrimination of speech sounds. *Brain stimul.* 10, 543–552. doi: 10.1016/j.brs.2017.01.007

Engineer, N. D., Riley, J. R., Seale, J. D., Vrana, W. A., Shetake, J. A., Sudanagunta, S. P., et al. (2011). Reversing pathological neural activity using targeted plasticity. *Nature* 470, 101–104. doi: 10.1038/nature09656

Follesa, P., Biggio, F., Gorini, G., Caria, S., Talani, G., Dazzi, L., et al. (2007). Vagus nerve stimulation increases norepinephrine concentration and the gene expression of BDNF and bFGF in the rat brain. *Brain Res.* 1179, 28–34. doi: 10.1016/j.brainres.2007.08.045

Frangos, E., and Komisaruk, B. R. (2017). Access to vagal projections via cutaneous electrical stimulation of the neck: fMRI evidence in healthy humans. *Brain Stimul.* 10, 19–27. doi: 10.1016/j.brs.2016.10.008

Frangos, E., Ellrich, J., and Komisaruk, B. R. (2015). Non-invasive Access to the Vagus Nerve Central Projections via Electrical Stimulation of the External Ear: fMRI Evidence in Humans. *Brain Stimul.* 8, 624–636. doi: 10.1016/j.brs.2014.11.018

Groves, D. A., and Brown, V. J. (2005). Vagal nerve stimulation: a review of its applications and potential mechanisms that mediate its clinical effects. *Neurosci. Biobehav. Rev.* 29, 493–500. doi: 10.1016/j.neubiorev.2005.01.004

Hassert, D., Miyashita, T., and Williams, C. (2004). The effects of peripheral vagal nerve stimulation at a memory-modulating intensity on norepinephrine output in the basolateral amygdala. *Behav. Neurosci.* 118:79. doi: 10.1037/0735-7044.118.1.79

Hays, S. A., Rennaker, R. L., and Kilgard, M. P. (2013). Targeting plasticity with vagus nerve stimulation to treat neurological disease. *Prog. Brain Res.* 207, 275–299. doi: 10.1016/b978-0-444-63327-9.00010-2

Henry, T. R. (2002). Therapeutic mechanisms of vagus nerve stimulation. *Neurology* 59(6 Suppl. 4), S3–S14.

Hyvarinen, P., Yrttiaho, S., Lehtimaki, J., Ilmoniemi, R. J., Makitie, A., Ylikoski, J., et al. (2015). Transcutaneous vagus nerve stimulation modulates tinnitus-related beta- and gamma-band activity. *Ear. Hear.* 36, e76–e85.

Jastreboff, P. J. (1990). Phantom auditory perception (tinnitus): mechanisms of generation and perception. *Neurosci. Res.* 8, 221–254. doi: 10.1016/0168-0102(90)90031-9

Khodaparast, N., Hays, S. A., Sloan, A. M., Hulsey, D. R., Ruiz, A., Pantoja, M., et al. (2013). Vagus nerve stimulation during rehabilitative training improves forelimb strength following ischemic stroke. *Neurobiol. Dis.* 60, 80–88. doi: 10.1016/j.nbd.2013.08.002

Kilgard, M. P., and Merzenich, M. M. (1998). Cortical map reorganization enabled by nucleus basalis activity. *Science* 279, 1714–1718. doi: 10.1126/science.279.5357.1714

Kimberley, T. J., Pierce, D., Prudente, C. N., Francisco, G. E., Yozbatiran, N., Smith, P., et al. (2018). Vagus nerve stimulation paired with upper limb rehabilitation after chronic stroke: a blinded randomized pilot study. *Stroke* 49, 2789–2792. doi: 10.1161/strokeaha.118.022279

Kirkwood, A., Rozas, C., Kirkwood, J., Perez, F., and Bear, M. F. (1999). Modulation of long-term synaptic depression in visual cortex by acetylcholine and norepinephrine. *J. Neurosci.* 19, 1599–1609. doi: 10.1523/jneurosci.19-05-01599.1999

Kiyokawa, J., Yamaguchi, K., Okada, R., Maehara, T., and Akita, K. (2014). Origin, course and distribution of the nerves to the posterosuperior wall of the external acoustic meatus. *Anat. Sci. Int.* 89, 238–245. doi: 10.1007/s12565-014-0231-4

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. *Prog. Neurobiol.* 111, 17–33. doi: 10.1016/j.pneurobio.2013.08.002

Kochilas, H. L., Cacace, A. T., Arnold, A., Seidman, M. D., and Tarver, W. B. (2020). Vagus nerve stimulation paired with tones for tinnitus suppression: Effects on voice and hearing. *Laryngoscope Investigat. Otolaryngol.* 5, 286–296. doi: 10.1002/lio2.364

Kraus, T., Hosl, K., Kiess, O., Schanze, A., Kornhuber, J., and Forster, C. B. O. L. D. (2007). fMRI deactivation of limbic and temporal brain structures and mood enhancing effect by transcutaneous vagus nerve stimulation. *J. Neural. Transm.* 114, 1485–1493. doi: 10.1007/s00702-007-0755-z

Kraus, T., Kiess, O., Hosl, K., Terekhin, P., Kornhuber, J., and Forster, C. C. N. S. B. O. L. D. (2013). fMRI effects of sham-controlled transcutaneous electrical nerve stimulation in the left outer auditory canal - a pilot study. *Brain Stimul.* 6, 798–804. doi: 10.1016/j.brs.2013.01.011

Kreuzer, P. M., Landgrebe, M., Husser, O., Resch, M., Schecklmann, M., Geisreiter, F., et al. (2012). Transcutaneous vagus nerve stimulation: retrospective assessment of cardiac safety in a pilot study. *Front. Psychiatr.* 3:70. doi: 10.3389/fpsyt.2012.00070

Kreuzer, P. M., Landgrebe, M., Resch, M., Husser, O., Schecklmann, M., Geisreiter, F., et al. (2014). Feasibility, safety and efficacy of transcutaneous vagus nerve stimulation in chronic tinnitus: an open pilot study. *Brain Stimul.* 7, 740–747. doi: 10.1016/j.brs.2014.05.003

Lehtimaki, J., Hyvarinen, P., Ylikoski, M., Bergholm, M., Makela, J. P., Aarnisalo, A., et al. (2013). Transcutaneous vagus nerve stimulation in tinnitus: a pilot study. *Acta Otolaryngol.* 133, 378–382. doi: 10.3109/00016489.2012.750736

Leslie, R., Gwyn, D., and Hopkins, D. (1982). The central distribution of the cervical vagus nerve and gastric afferent and efferent projections in the rat. *Brain Res. Bull.* 8, 37–43. doi: 10.1016/0361-9230(82)90025-9

Llanos, F., McHaney, J. R., Schuerman, W. L., Han, G. Y., Leonard, M. K., and Chandrasekaran, B. (2020). Non-invasive peripheral nerve stimulation selectively enhances speech category learning in adults. *Npj Sci. Learn.* 5, 1–11.

Lomarev, M., Denslow, S., Nahas, Z., Chae, J. H., George, M. S., and Bohning, D. E. (2002). Vagus nerve stimulation (VNS) synchronized BOLD fMRI suggests that VNS in depressed adults has frequency/dose dependent effects. *J. Psychiatr. Res.* 36, 219–227. doi: 10.1016/s0022-3956(02)00013-4

Lulic, D., Ahmadian, A., Baaj, A. A., Benbadis, S. R., and Vale, F. L. (2009). Vagus nerve stimulation. *Neurosurg. Focus* 27:E5.

Lv, H., Zhao, Y., Chen, J., Wang, D., and Chen, H. (2019). Vagus nerve stimulation for depression: a systematic review. *Front. Psychol.* 10:64. doi: 10.3389/fpsyg.2019.00064

Manta, S., Dong, J., Debonnel, G., and Blier, P. (2009). Enhancement of the function of rat serotonin and norepinephrine neurons by sustained vagus nerve stimulation. *J. Psychiatr. Neurosci.* 34, 272–280.

Martins, A. R. O., and Froemke, R. C. (2015). Coordinated forms of noradrenergic plasticity in the locus coeruleus and primary auditory cortex. *Nat. Neurosci.* 18, 1483–1492. doi: 10.1038/nn.4090

Mega, M. S., Cummings, J. L., Salloway, S., and Malloy, P. (1997). The limbic system: an anatomic, phylogenetic, and clinical perspective. *J. Neuropsychiatr. Clin. Neurosci.* 9, 315–330. doi: 10.1176/jnp.9.3.315

Meyers, E. C., Kasliwal, N., Solorzano, B. R., Lai, E., Bendale, G., Berry, A., et al. (2019). Enhancing plasticity in central networks improves motor and sensory recovery after nerve damage. *Nat. Comm.* 10, 1–14. doi: 10.1155/2020/9484298

Milby, A. H., Halpern, C. H., and Baltuch, G. H. (2008). Vagus nerve stimulation for epilepsy and depression. *Neurotherapeutics* 5, 75–85.

Muhlnickel, W., Elbert, T., Taub, E., and Flor, H. (1998). Reorganization of auditory cortex in tinnitus. *Proc. Natl. Acad. Sci. U S A* 95, 10340–10343. doi: 10.1073/pnas.95.17.10340

Nahas, Z., Marangell, L. B., Husain, M. M., Rush, A. J., Sackeim, H. A., Lisanby, S. H., et al. (2005). Two-year outcome of vagus nerve stimulation (VNS) for treatment of major depressive episodes. *J. Clin. Psychiatr.* 66, 1097–1104. doi: 10.4088/jcp.v66n0902

Narayanan, J. T., Watts, R., Haddad, N., Labar, D. R., Li, P. M., and Filippi, C. G. (2002). Cerebral activation during vagus nerve stimulation: a functional MR study. *Epilepsia* 43, 1509–1514. doi: 10.1046/j.1528-1157.2002.16102.x

Nemeroff, C. B., Mayberg, H. S., Krahl, S. E., McNamara, J., Frazer, A., Henry, T. R., et al. (2006). therapy in treatment-resistant depression: clinical evidence and putative neurobiological mechanisms. *Neuropsychopharmacology* 31, 1345–1355. doi: 10.1038/sj.npp.1301082

Noble, L. J., Souza, R. R., and McIntyre, C. K. (2019). Vagus nerve stimulation as a tool for enhancing extinction in exposure-based therapies. *Psychopharmacology* 236, 355–367. doi: 10.1007/s00213-018-4994-5

Norena, A. J., and Eggermont, J. J. (2003). Changes in spontaneous neural activity immediately after an acoustic trauma: implications for neural correlates of tinnitus. *Hear Res.* 183, 137–153. doi: 10.1016/s0378-5955(03)00225-9

Ogbonnaya, S., and Kaliaperumal, C. (2013). Vagal nerve stimulator: Evolving trends. *J. Nat. Sci. Biol. Med.* 4, 8–13. doi: 10.4103/0976-9668.107254

Okamoto, H., Stracke, H., Stoll, W., and Pantev, C. (2010). Listening to tailor-made notched music reduces tinnitus loudness and tinnitus-related auditory cortex activity. *Proc. Natl. Acad. Sci. U S A* 107, 1207–1210. doi: 10.1073/pnas.0911268107

Pantev, C., Okamoto, H., and Teismann, H. (2012). Music-induced cortical plasticity and lateral inhibition in the human auditory cortex as foundations for tonal tinnitus treatment. *Front. Syst. Neurosci.* 6:50. doi: 10.3389/fnsys.2012.00050

Pantev, C., Okamoto, H., Ross, B., Stoll, W., Ciurlia−Guy, E., Kakigi, R., et al. (2004). Lateral inhibition and habituation of the human auditory cortex. *Eur. J. Neurosci.* 19, 2337–2344. doi: 10.1111/j.0953-816x.2004.03296.x

Peña, D. F., Childs, J. E., Willett, S., Vital, A., McIntyre, C. K., and Kroener, S. (2014). Vagus nerve stimulation enhances extinction of conditioned fear and modulates plasticity in the pathway from the ventromedial prefrontal cortex to the amygdala. *Front. Behav. Neurosci.* 8:327. doi: 10.3389/fnbeh.2014.00327

Peuker, E. T., and Filler, T. J. (2002). The nerve supply of the human auricle. *Clin. Anat.* 15, 35–37. doi: 10.1002/ca.1089

Polak, T., Markulin, F., Ehlis, A., Langer, J. B., Ringel, T. M., and Fallgatter, A. J. (2009). Far field potentials from brain stem after transcutaneous vagus nerve stimulation: optimization of stimulation and recording parameters. *J. Neural. Transm.* 116, 1237–1242. doi: 10.1007/s00702-009-0282-1

Porter, B. A., Khodaparast, N., Fayyaz, T., Cheung, R. J., Ahmed, S. S., Vrana, W. A., et al. (2012). Repeatedly pairing vagus nerve stimulation with a movement reorganizes primary motor cortex. *Cereb. Cortex* 22, 2365–2374. doi: 10.1093/cercor/bhr316

Pruitt, D. T., Schmid, A. N., Kim, L. J., Abe, C. M., Trieu, J. L., Choua, C., et al. (2016). Vagus nerve stimulation delivered with motor training enhances recovery of function after traumatic brain injury. *J. Neurotrauma.* 33, 871–879. doi: 10.1089/neu.2015.3972

Raedt, R., Clinckers, R., Mollet, L., Vonck, K., El Tahry, R., Wyckhuys, T., et al. (2011). Increased hippocampal noradrenaline is a biomarker for efficacy of vagus nerve stimulation in a limbic seizure model. *J. Neurochem.* 117, 461–469. doi: 10.1111/j.1471-4159.2011.07214.x

Rajmohan, V., and Mohandas, E. (2007). The limbic system. *Indian J. Psychiatr.* 49, 132–139.

Rong, P., Liu, J., Wang, L., Liu, R., Fang, J., Zhao, J., et al. (2016). Effect of transcutaneous auricular vagus nerve stimulation on major depressive disorder: a nonrandomized controlled pilot study. *J. Affect Disord.* 195, 172–179. doi: 10.1016/j.jad.2016.02.031

Schachter, S. C., and Saper, C. B. (1998). Vagus nerve stimulation. *Epilepsia* 39, 677–686.

Schulz, K. F., Altman, D. G., and Moher, D. C. O. N. S. O. R. T. (2010). 2010 statement: updated guidelines for reporting parallel group randomised trials. *Trials* 11, 1–8.

Sclocco, R., Garcia, R. G., Kettner, N. W., Isenburg, K., Fisher, H. P., Hubbard, C. S., et al. (2019). The influence of respiration on brainstem and cardiovagal response to auricular vagus nerve stimulation: A multimodal ultrahigh-field (7T) fMRI study. *Brain Stimul.* 12, 911–921. doi: 10.1016/j.brs.2019.02.003

Seol, G. H., Ziburkus, J., Huang, S., Song, L., Kim, I. T., Takamiya, K., et al. (2007). Neuromodulators control the polarity of spike-timing-dependent synaptic plasticity. *Neuron* 55, 919–929. doi: 10.1016/j.neuron.2007.08.013

Shetake, J. A., Engineer, N. D., Vrana, W. A., Wolf, J. T., and Kilgard, M. P. (2012). Pairing tone trains with vagus nerve stimulation induces temporal plasticity in auditory cortex. *Exp. Neurol.* 233, 342–349. doi: 10.1016/j.expneurol.2011.10.026

Shim, H. J., Kwak, M. Y., An, Y. H., Kim, D. H., Kim, Y. J., and Kim, H. J. (2015). Feasibility and Safety of Transcutaneous Vagus Nerve Stimulation Paired with Notched Music Therapy for the Treatment of Chronic Tinnitus. *J. Audiol. Otol.* 19, 159–167. doi: 10.7874/jao.2015.19.3.159

Shore, S. E., Roberts, L. E., and Langguth, B. (2016). Maladaptive plasticity in tinnitus–triggers, mechanisms and treatment. *Nat. Rev. Neurol.* 12, 150–160. doi: 10.1038/nrneurol.2016.12

Stegeman, I., Velde, H., Robe, P., Stokroos, R., and Smit, A. (2021). Tinnitus treatment by vagus nerve stimulation: A systematic review. *PLoS One* 16:e0247221. doi: 10.1371/journal.pone.0247221

Suk, W. C., Kim, S. J., Chang, D. S., and Lee, H. Y. (2018). Characteristics of stimulus intensity in transcutaneous vagus nerve stimulation for chronic tinnitus. *J. Internat. Adv. Otol.* 14:267.

Sumal, K. K., Blessing, W., Joh, T. H., Reis, D. J., and Pickel, V. M. (1983). Synaptic interaction of vagal afferents and catecholaminergic neurons in the rat nucleus tractus solitarius. *Brain Res.* 277, 31–40. doi: 10.1016/0006-8993(83)90904-6

Tekdemir, I., Aslan, A., and Elhan, A. A. (1998). clinico-anatomic study of the auricular branch of the vagus nerve and Arnold's ear-cough reflex. *Surg. Radiol. Anat.* 20, 253–257. doi: 10.1007/s00276-998-0253-5

Tyler, R. S., Aran, J., and Dauman, R. (1992). Recent advances in tinnitus. *Am. J. Audiol.* 1, 36–44. doi: 10.1044/1059-0889.0104.36

Tyler, R. S., Noble, W., and Coelho, C. (2006). Considerations for the design of clinical trials for tinnitus. *Acta Otolaryngol.* 126, 44–49. doi: 10.1080/03655230600895424

Tyler, R. S., Oleson, J., Noble, W., Coelho, C., and Ji, H. (2007). Clinical trials for tinnitus: study populations, designs, measurement variables, and data analysis. *Prog. Brain Res.* 166, 499–509. doi: 10.1016/s0079-6123(07)66048-8

Tyler, R., Cacace, A., Stocking, C., Tarver, B., Engineer, N., Martin, J., et al. (2017). Vagus Nerve Stimulation Paired with Tones for the Treatment of Tinnitus: A Prospective Randomized Double-blind Controlled Pilot Study in Humans. *Sci. Rep.* 7:11960.

Tyler, R., Coelho, C., Tao, P., Ji, H., Noble, W., Gehringer, A., et al. (2008). Identifying tinnitus subgroups with cluster analysis. *Age* 1:3.

Van Der Loo, E., Gais, S., Congedo, M., Vanneste, S., Plazier, M., Menovsky, T., et al. (2009). Tinnitus intensity dependent gamma oscillations of the contralateral auditory cortex. *PLoS One* 4:e7396. doi: 10.1371/journal.pone.0007396

Vanneste, S., Martin, J., Rennaker, R. L., and Kilgard, M. P. (2017). Pairing sound with vagus nerve stimulation modulates cortical synchrony and phase coherence in tinnitus: An exploratory retrospective study. *Sci. Rep.* 7, 1–11.

Vonck, K., De Herdt, V., Bosman, T., Dedeurwaerdere, S., Van Laere, K., and Boon, P. (2008). Thalamic and limbic involvement in the mechanism of action of vagus nerve stimulation, a SPECT study. *Seizure* 17, 699–706. doi: 10.1016/j.seizure.2008.05.001

Wang, Z., Fang, J., Liu, J., Rong, P., Jorgenson, K., Park, J., et al. (2018). Frequency-dependent functional connectivity of the nucleus accumbens during continuous transcutaneous vagus nerve stimulation in major depressive disorder. *J. Psychiatr. Res.* 102, 123–131. doi: 10.1016/j.jpsychires.2017.12.018

Watanabe, K., Tubbs, R. S., Satoh, S., Zomorodi, A. R., Liedtke, W., Labidi, M., et al. (2016). Isolated deep ear canal pain: possible role of auricular branch of vagus nerve—case illustrations with cadaveric correlation. *World Neurosurg.* 96, 293–301. doi: 10.1016/j.wneu.2016.08.102

Wichova, H., Alvi, S. A., Shew, M., Lin, J., Sale, K., Larsen, C., et al. (2018). Tinnitus perception in patients after vagal nerve stimulator implantation for epilepsy. *Am. J. Otolaryngol.* 39, 599–602. doi: 10.1016/j.amjoto.2018.07.009

Wu, K., Wang, Z., Zhang, Y., Yao, J., and Zhang, Z. (2020). Transcutaneous vagus nerve stimulation for the treatment of drug-resistant epilepsy: a meta-analysis and systematic review. *ANZ J Surg.* 90, 467–471. doi: 10.1111/ans.15681

Yakunina, N., and Nam, E. (2021). Does the tinnitus pitch correlate with the frequency of hearing loss? *Acta Otolaryngol.* 141, 163–170. doi: 10.1080/00016489.2020.1837394

Yakunina, N., Kim, S. S., and Nam, E. B. O. L. D. (2018). fMRI effects of transcutaneous vagus nerve stimulation in patients with chronic tinnitus. *PLoS One* 13:e0207281. doi: 10.1371/journal.pone.0207281

Yakunina, N., Kim, S. S., and Nam, E. C. (2017). Optimization of Transcutaneous Vagus Nerve Stimulation Using Functional MRI. *Neuromodulation* 20, 290–300. doi: 10.1111/ner.12541

Yap, J. Y., Keatch, C., Lambert, E., Woods, W., Stoddart, P. R., and Kameneva, T. (2020). Critical review of transcutaneous vagus nerve stimulation: Challenges for translation to clinical practice. *Front. Neurosci.* 2020:14. doi: 10.3389/fnins.2020.00284

Ylikoski, J., Lehtimaki, J., Pirvola, U., Makitie, A., Aarnisalo, A., Hyvarinen, P., et al. (2017). Non-invasive vagus nerve stimulation reduces sympathetic preponderance in patients with tinnitus. *Acta Otolaryngol.* 137, 426–431. doi: 10.1080/00016489.2016.1269197

Ylikoski, J., Markkanen, M., Pirvola, U., Lehtimäki, J. A., Ylikoski, M., Jing, Z., et al. (2020). Stress and Tinnitus; Transcutaneous Auricular Vagal Nerve Stimulation Attenuates Tinnitus-Triggered Stress Reaction. *Front. Psychol.* 11:2442. doi: 10.3389/fpsyg.2020.570196

Young, E. D, Nelken, I., and Conley, R. A. (1995). Somatosensory effects on neurons in dorsal cochlear nucleus. *J. Neurophysiol.* 73, 743–765. doi: 10.1152/jn.1995.73.2.743

Yuan, H., and Silberstein, S. D. (2016). Vagus Nerve and Vagus Nerve Stimulation, a Comprehensive Review: Part II. *Headache* 56, 259–266. doi: 10.1111/head.12650

Zobel, A., Joe, A., Freymann, N., Clusmann, H., Schramm, J., Reinhardt, M., et al. (2005). Changes in regional cerebral blood flow by therapeutic vagus nerve stimulation in depression: an exploratory approach. *Psychiatr. Res.* 139, 165–179. doi: 10.1016/j.pscychresns.2005.02.010

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.

frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 27 September 2021
doi: 10.3389/fnhum.2021.719254

Functional Magnetic Resonance Imaging Reveals Early Connectivity Changes in the Auditory and Vestibular Cortices in Idiopathic Sudden Sensorineural Hearing Loss With Vertigo: A Pilot Study

Qiuxia Wang[1†], Qingguo Chen[2†], Ping Liu[1], Jing Zhang[1], Liangqiang Zhou[2] and Liyan Peng[2*]*

[1] Department of Radiology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [2] Department of Otorhinolaryngology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

The underlying pathophysiology of idiopathic sudden sensorineural hearing loss (ISSNHL) with vertigo has yet to be identified. The aims of the current study were (1) to elucidate whether there are functional changes of the intrinsic brain activity in the auditory and vestibular cortices of the ISSNHL patients with vertigo using resting-state functional magnetic resonance imaging (rs-fMRI) and (2) whether the connectivity alterations are related to the clinical performance associated with ISSNHL with vertigo. Twelve ISSNHL patients with vertigo, eleven ISSNHL patients without vertigo and eleven healthy subjects were enrolled in this study. Rs-fMRI data of auditory and vestibular cortices was extracted and regional homogeneity (ReHo) and seed-based functional connectivity (FC) were evaluated; the chi-square test, the ANOVA and the Bonferroni multiple comparison tests were performed. Significantly decreased ReHo in the ipsilateral auditory cortex, as well as increased FC between the inferior parietal gyrus and the auditory cortex were found in the ISSNHL with vertigo groups. These findings contribute to a characterization of early plastic changes in ISSNHL patients with vertigo and cultivate new insights for the etiology research.

Keywords: neural plasticity, idiopathic sudden sensorineural hearing loss, resting-state functional magnetic resonance imaging, regional homogeneity, vertigo

OPEN ACCESS

Edited by:
Achim Schilling, UMR 7260 Neurosciences Sensorielles et Cognitives, France

Reviewed by:
Joel Alan Goebel, Washington University School of Medicine in St. Louis, United States
Norma De Oliveira Penido, Federal University of São Paulo, Brazil
Elouise Koops, University Medical Center Groningen, Netherlands

***Correspondence:**
Liyan Peng
liyanpeng2050@163.com
Liangqiang Zhou
andqiang@163.com

[†] These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Sensory Neuroscience, a section of the journal Frontiers in Human Neuroscience

Received: *02 June 2021*
Accepted: *01 September 2021*
Published: *27 September 2021*

Citation:
Wang Q, Chen Q, Liu P, Zhang J, Zhou L and Peng L (2021) Functional Magnetic Resonance Imaging Reveals Early Connectivity Changes in the Auditory and Vestibular Cortices in Idiopathic Sudden Sensorineural Hearing Loss With Vertigo: A Pilot Study. Front. Hum. Neurosci. 15:719254. doi: 10.3389/fnhum.2021.719254

INTRODUCTION

Idiopathic sudden sensorineural hearing loss (ISSNHL) is defined as a sensorineural hearing loss of at least 30 dB for three or more contiguous audiometric frequencies. Idiopathic sudden sensorineural hearing loss typically develops over 72 h and affects 10, 20 and 300 out of every 100,000 people in China, the United States and Germany every year (Fang et al., 1999; Michel, 2011;

Abbreviations: BA, Brodmann area; BOLD, blood oxygen level-dependent; cVEMP, cervical vestibular-evoked myogenic potentials; DHI, dizziness handicap inventory; FC, functional connectivity; ISSNHL, idiopathic sudden sensorineural hearing loss; MATLAB, matrix laboratory platform; MNI, Montréal neurological institute; ReHo, regional homogeneity; ROI, region of interest; rs-fMRI, resting-state functional magnetic resonance imaging; AC, auditory cortex.

Stachler et al., 2012). Furthermore, vertigo develops in approximately 20% to 60% of the patients with ISSNHL (Moskowitz et al., 1984; Park et al., 2001; Rauch, 2008; Pogson et al., 2016; Chang et al., 2018). Vertigo indicated a poor prognosis for hearing recovery as the incidence of severe or profound hearing loss in ISSNHL patients with vertigo was increased (Shaia and Sheehy, 1976; Wang et al., 2009; Kim et al., 2018; Zhou et al., 2018). Chang et al. concluded that sudden hearing loss with vertigo portended greater stroke risk than sudden hearing loss or vertigo alone (Chang et al., 2018). Moreover, the burden of ISSNHL with vertigo on the patient is considerable since the impact of both the hearing loss and vertigo cannot be adequately diminished as there is a lack of effective treatments. Therefore it's necessary to identify the underlying pathophysiology of ISSNHL with vertigo which represents a unique clinical entity dissimilar to ISSNHL without vertigo (Rauch, 2018).

Brain structural alterations have been reported in patients with auditory impairments such as unilateral hearing loss, tinnitus or deafness (Fan et al., 2015). Resting-state functional magnetic resonance imaging (rs-fMRI) has enabled the mapping of brain activity based on the blood oxygen level-dependent (BOLD) signal in ISSNHL patients. In particular, the regional homogeneity (ReHo) metric reflects the consistency of neuronal activity in a local brain region by measuring the similarity of the BOLD signal fluctuation between adjacent voxels. Historically, the ReHo metric has successfully identified biomarkers of various neurological diseases such as Alzheimer's Disease and depression (Liu et al., 2008; Yao et al., 2009; Guo et al., 2011), and is thought to provide a pure measure of time-resolved brain connectivity patterns.

We speculated that the auditory and vestibular cortices have functional connectivity in the ISSNHL patients with vertigo. The aim of this study was to elucidate whether there are changes to the ReHo signals in the central auditory and vestibular cortices of the ISSNHL patients with vertigo and how the auditory-vestibular cortex is integrated. Another purpose was to determine whether these connectivity variations related to clinical performance changes occur during the onset of the disease.

MATERIALS AND METHODS

Participants

This study was approved by the Institutional Review Board of the Ethics Committee of the Huazhong University of Science and Technology. Informed consent was obtained by each subject before participating in this study. Patients were enrolled between January 2017 and December 2018.

Twelve right-handed, previously untreated patients with acute unilateral ISSNHL with vertigo participated in this study. The pure tone audiometry and dizziness handicap inventory (DHI) scores were listed in **Table 1**. All patients met the following inclusion criteria: (1) suffering from ISSNHL for the first time; (2) unknown cause of hearing loss; (3) the level of hearing loss was at least 30 dB in at least three contiguous frequencies with no air-bone gap; (4) a history of vertiginous episodes near the onset of hearing loss; (5) the interval between the onset of vertigo and the MRI examination was $\leq$7 days; (6) absence of other neurological signs existed; (7) CT and MRI were performed to ensure normal ear structure and no brain lesions. The exclusion criteria included: (1) vertigo caused by benign paroxysmal positional vertigo, Meniere's disease or acute vestibular neuritis; (2) fluctuating hearing loss; (3) inflammation of the external or middle ear; (4) a history of ear surgery; (5) spatial claustrophobia. The concomitant symptom of vertigo was defined as episodic rotational vertigo which occurred one day before/after hearing loss and lasted for several hours to several days. The onset was not related to head position and vertigo attacks did not recur after recovery.

Eleven right-handed, age, gender, and education matched patients with acute unilateral ISSNHL without vertigo were included (**Table 1**). The patients met the same inclusion criteria as the ISSNHL with vertigo group except the items 4 and 5.

Eleven age, gender, and education matched healthy people with normal hearing and negative otoscopic findings were included as the control group in the study. The control group had no history of auricular or neurological diseases.

Hearing and Vestibular Testing

Pure tone audiometry testing (CONERA OB922 Audiometer, Madsen, Denmark) was performed for all participants. The pure-tone hearing thresholds at 250, 500, 1000, 2000, 4000, and 8000 Hz were recorded at the beginning and the end of the therapy. The outcome assessment was performed according to the American clinical practice guidelines.

TABLE 1 | Demographic and clinical characteristics.

	Loss	Vertigo	F/X^2	P
Number (n)	11	12		
Sex (n)				
Female	7	8	0.023	0.879†
Male	4	4		
Age (year, mean ± SD)	41.27 ± 14.16	44.42 ± 10.57	2.146	0.550
Ear affected				
Left	6	6	0.048	0.827
Right	5	6		
Pretreatment PTA (mean ± SD)	91.2 ± 18.98 dB	97.03 ± 15.95 dB	0.877	0.362
Postreatment PTA (mean ± SD)	56.29 ± 23.61 dB	74.38 ± 22.30 dB		
DHI (mean ± SD)	-	62.67 ± 10.77		
PTA gain (n)				
No recovery	2	7	3.884	0.049‡
Partial recovery	8	5		
Complete recovery	1	0		

n means number. Loss: ISSNHL without vertigo; Vertigo: ISSNHL with vertigo; PTA, pure tone audiometry; DHI, dizziness handicap inventory. $p < 0.05$ was considered statistically significant.

†P value was derived from Fisher's Exact test. ‡ The comparison was between "No recovery" group and "Partial+Complete" recovery group and P value was derived from Fisher's Exact test.

Acoustic reflex measurements were performed by an acoustic impedance audiometer (Impedance Audiometer, Itera, Madsen, Denmark). Ipsilateral and contralateral stapes reflexes were examined at 500, 1000, 2000, and 4000 Hz. For each of these four frequencies, an acoustic stimulus of 80 dB HL was presented, and an additional 10 dB was used until a reflex curve was detected. To avoid acoustic trauma, a maximum acoustic stimulus of 110 dB was applied. The reflex curves were recorded and plotted and the latencies of the reflexes were calculated at the intersection of the baseline with the rising edge of the reflex curves.

The symptom of vertigo was assessed using the Chinese version of the dizziness handicap inventory (DHI) (Fang et al., 1999). Peripheral vestibular excitability was tested using videonystagmography with bithermal caloric irrigation (ICS CHARTR 200, Otometrics, Germany) and saccular function was assessed via the cervical vestibular-evoked myogenic potential (cVEMP) using a Medelec Synergy unit (ICS CHARTR diagnostic systems MOU-90, Otometrics, German). In all the ISSNHL patients with vertigo, the initial nystagmus examination was performed at the first visit and daily during the acute stage. Spontaneous nystagmus was checked for in an upright-seated position. Nystagmus examination lasted for 2 min.

Resting-State Functional Magnetic Resonance Imaging Data Acquisition

fMRI data was collected using a 3.0 T MRI scanner (GE Medical Systems, Milwaukee, WI) equipped with a 32-channel head coil in the department of Radiology. The head of the subject was fixed in a head coil using rubber pads and both ears were plugged. Patients were instructed to close eyes during the functional scans. Anatomical imaging included a high-resolution three-dimensional sagittal magnetization-prepared rapid acquisition gradient echo T1-weighted sequence with the following parameters: repetition time (TR) = 5000 ms, echo time (TE) = 2960 ms, flip angle = 12 °, field of view (FOV) = 256 × 256mm^2, matrix size = 256 × 256, slice thickness = 1 mm, no slice gap, voxel size = 1.0 × 1.0 × 1.0 mm, and slice number = 184. The resting-state functional images were acquired using a single-short gradient-echo echo-planar imaging sequence parallel to the anterior commissure-posterior commissure plane with the following parameters: TR = 2000 ms, TE = 35 ms, flip angle = 90 °, FOV = 224 × 224 mm^2, matrix size = 64 × 64, slice thickness = 3.5 mm, no slice gap, voxel size = 3.5 × 3.5 × 3.5 mm, and slice number = 40.

Resting-State Functional Magnetic Resonance Imaging Data Processing

The fMRI data were processed with SPM8[1] and Data Processing & Analysis of Brain Imaging (DPABI) software.[2] Preprocessing included: (a) discarding of the first 10 time points, (b) slice timing correction, (c) realignment and motion correction (framewise displacement, FD) (Power et al., 2013), (d) co-registration of the individual anatomical and the realigned functional volumes, (e) spatial normalization into Montreal Neurological Institute (MNI152) space through Diffeomorphic Anatomical Registration Through Exponentiated Lie Algebra (DARTEL) (Ashburner, 2018), (f) spatial smoothing with 6 mm full width half maximum (FWHM) Gaussian (this step was only performed for functional connectivity, and the ReHo was smoothed lastly), (g) reduction of confounding factors via linear regression, including the signals from the white matter and cerebrospinal fluid, and linear and quadratic trends, (h) temporal filtering (0.01–0.1 Hz) of the time series, and finally (i) motion scrubbing (Power et al., 2012; Yan et al., 2013) with a threshold of 0.5. According to the realignment parameters of fMRI run head motion, subjects were excluded from the analysis if they showed motion more than 2.0mm maximum displacement in any of the x, y, or z directions or more than 2.5° of angular motion.

[1]https://www.fil.ion.ucl.ac.uk/spm

[2]http://rfmri.org/dpabi

Regional Homogeneity and Seed-Based Functional Connectivity

ReHo is one of the frequently used methods to analyze image data of brain activities. An increase in ReHo means an increase in the neuronal synchrony in a specific brain region. In our experiment, the individual ReHo maps were computed using the DPABI, with the Kendall's coefficient of concordance (KCC) algorithm and local neighborhood of 26 voxels. The ReHo maps were smoothed with a 6 mm Gaussian kernel.

A voxel-wise FC analysis of the ROI was used. The two ROI seeds were selected from the clusters that were statistically significant in the ANOVA analysis of the ReHo at the bilateral auditory regions. The Pearson correlation coefficient was obtained between all brain voxels and seed time series and was then transformed using the Fisher Z transformation to ensure a normally distributed dataset.

Statistical Analyses

The Chi-square and Fisher's exact tests were performed to analyse the clinical data of the ISSNHL with vertigo (Vertigo group), without vertigo groups (Loss group) and those with normal hearing (Normal group) using SPSS 22.0 software (SPSS Inc., Chicago, Illinois, United States). The results were considered significant at a threshold of $p < 0.05$.

The DPABI (edition 4.3_200401) was used for the statistical analyses. An analysis of variance (ANOVA) was used to analyze voxel-wise whole brain inter-group differences. The resultant map was corrected using a cluster-level AlphaSim algorithm (voxel $p < 0.001$ and cluster $p < 0.05$) under effective smoothing kernel estimation. This correction is equivalent to a voxel level of $p < 0.001$ and a minimum cluster size of >54 voxels. The brain regions displaying significantly different ReHo values were used to create a mask for further Bonferroni multiple comparison tests (post-hoc analyses).

The ReHo values of each group were extracted within the clusters presenting statistical significance in the ANOVA results. Correlation of the ReHo and DHI scores mentioned above was assessed using Person's correlations. The result was considered significant at $p < 0.05$.

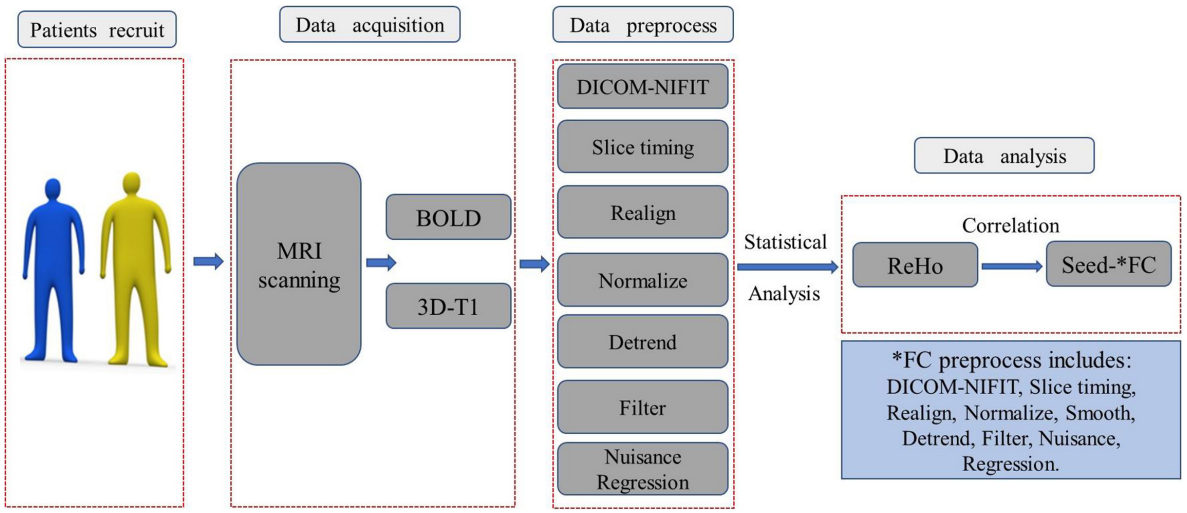


FIGURE 1 | The image acquisition and statistical analysis process were summarized and depicted in this flowchart. BOLD, blood oxygen level-dependent; ReHo, regional homogeneity; DICOM-NIFIT, three-dimensional fMRI data is generally in DICOM format or NIFIT format.

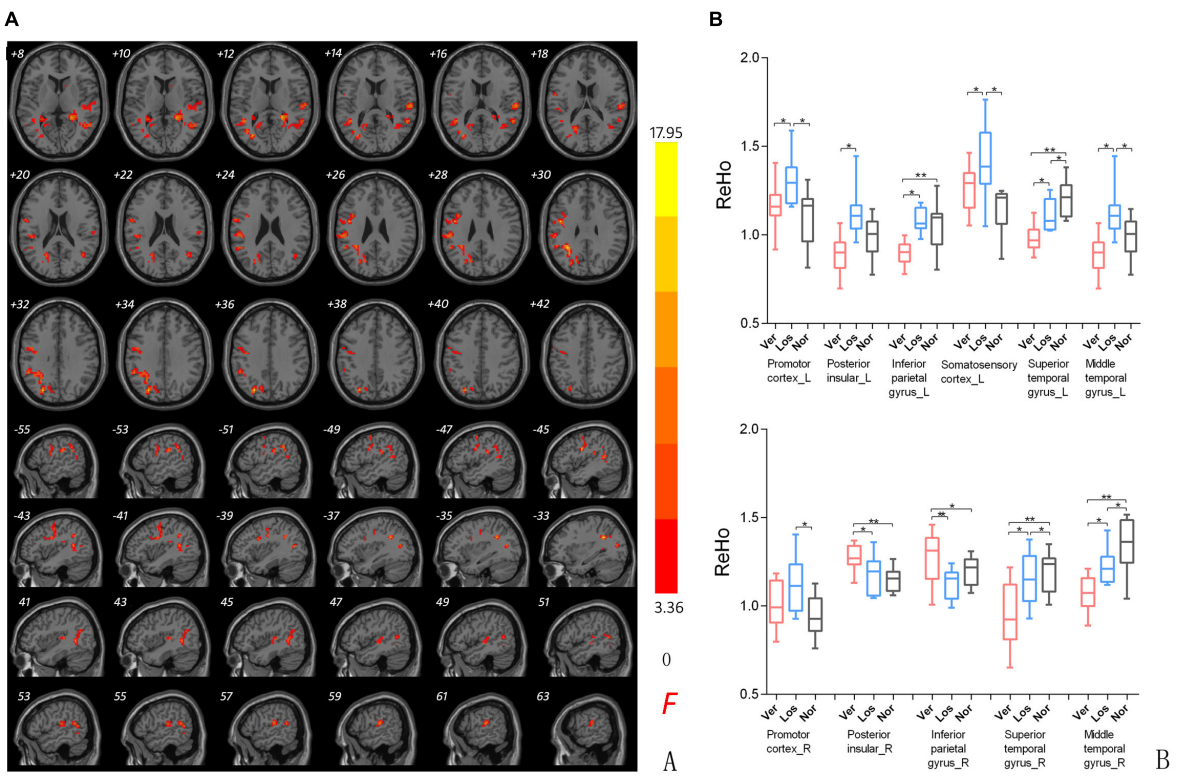


FIGURE 2 | (A) ReHo comparison among groups. ReHo signal differences among ISSNHL with vertigo (Ver), ISSNHL without vertigo (Los), and normal hearing (Nor) groups. Slice coordinates according to Montreal Neurological Institute space are shown in the left corner of the slices, indicating the Z-axis in axial orientation and the X-axis in sagittal slices. Red denotes the regions that differed among the three groups. Differences were considered significant at a threshold of $p < 0.001$, corrected *via* Alpahsim correction. **(B)** *Post hoc* comparison of clusters showing significant differences in the three groups pairwise analysis (Ver vs. Nor, Los vs. Nor, Ver vs. Los). $P^{*}/^{**} < 0.001$.

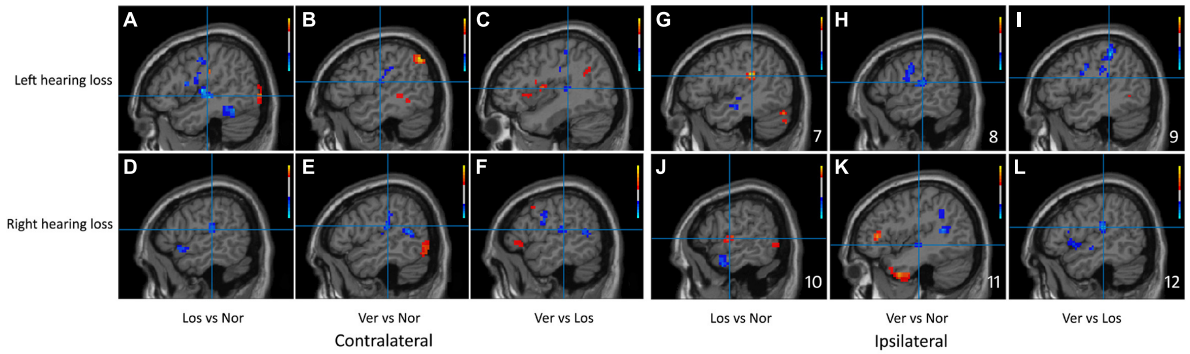


FIGURE 3 | Sagittal planes of the auditory cortex were exhibited in the ReHo maps of the left and right hearing loss patients separately. The clusters of significant difference in the Bonferroni multiple comparison tests correspond to regions listed in **Table 2**. Slices **(A–F)** showed contralateral brain changes and **(G–L)** showed ipsilateral changes. Crosshair demonstrated the peak value of the cluster and color bars were displayed in the right corner. Ver, ISSNHL with vertigo group; Los, ISSNHL without vertigo group, and Nor, normal hearing group.

TABLE 2 | Changes in ReHo signal of the auditory cortex by Vertigo vs. Loss vs. Normal hearing groups.

	Hemi sphere	Los VS Nor					Ver VS Nor					Ver VS Los				
		Voxel	T value	MNI coordinate			Voxel	T value	MNI coordinate			Voxel	T value	MNI coordinate		
				X	Y	Z			X	Y	Z			X	Y	Z
Left ear	L	41	3.472	–48	–33	22	53	–2.538	–50	–15	16	50	–4.101	–53	–25	25
	R	71	–3.394	48	–18	–1	46	–3.846	58	–23	13	35	–2.363	57	–22	13
Right ear	L	35	–2.420	–56	–20	19	72	–2.522	–44	–18	–4	28	–2.387	–45	–31	13
	R	63	2.342	60	–5	4	61	–3.252	53	–25	19	46	–4.962	50	–25	16

Voxel number, T-values were obtained from the statistical parametric mapping of the ReHo ($p < 0.001$). The MNI coordinates reflected the center of gravity of the cluster as found in the map. Los: ISSNHL without vertigo; Ver, ISSNHL with vertigo; Nor, Normal hearing group.

The image acquisition and statistical analysis process were summarized and depicted in the flowchart in **Figure 1**.

RESULTS

Demographic and Clinical Data

No significant difference in sex ($F = 0.023$, $p = 0.827$) or age ($X^2 = 2.146$, $p = 0.851$) was found among the three groups. There was a significant difference between the ISSNHL without vertigo group and with vertigo group in Pure Tone Audiometry (PTA) gain for prognosis ($X^2 = 3.884$, $P = 0.049$) (**Table 1**).

Comparison of Regional Homogeneity Among Groups

Significant differences in the ReHo values were found in both auditory and vestibular regions among the three groups, including the posterior insular, inferior parietal gyrus, superior and middle temporal regions (**Figure 2**); there were also differences in premotor and somatosensory cortices which are considered as vestibular related regions (Brandt et al., 1998; Frank and Greenlee, 2018). Specifically, when we focused on auditory cortex (AC) and separated the affected side, we found that the ReHo signals in the contralateral (contrary to the affected side) superior temporal gyrus were both decreased in the Vertigo and Loss groups, whereas the signals of ipsilateral (on the same side of the affected ear) AC was increased in the Loss group and decreased in the Vertigo group (**Figure 3** and **Table 2**).

Comparison of the Functional Connectivity Among Groups

In the participants with the left sided hearing loss, we identified altered FC of the left AC and cerebellum with the vestibular regions and related areas such as the inferior parietal gyrus, premotor areas, somatosensory cortex, angular gyrus, V2 and posterior insular cortex. A post-hoc analysis showed a statistically significant difference in the FC between the right inferior parietal gyrus, left and right premotor cortex, left and right somatosensory cortex for the Vertigo and Loss groups (**Figure 4A** and **Table 3**).

In the participants with the right sided hearing loss, we identified altered FC of the left AC and cerebellum with the premotor cortex, associative visual cortex, inferior parietal gyrus, posterior insular cortex, dorsolateral prefrontal cortex, posterior insular cortex and middle temporal gryus. A post-hoc analysis showed a statistically significant difference in the FC between the left and right inferior parietal gyrus, right middle temporal gryus and right premotor cortex for the Vertigo and Loss groups (**Figure 4B** and **Table 3**).

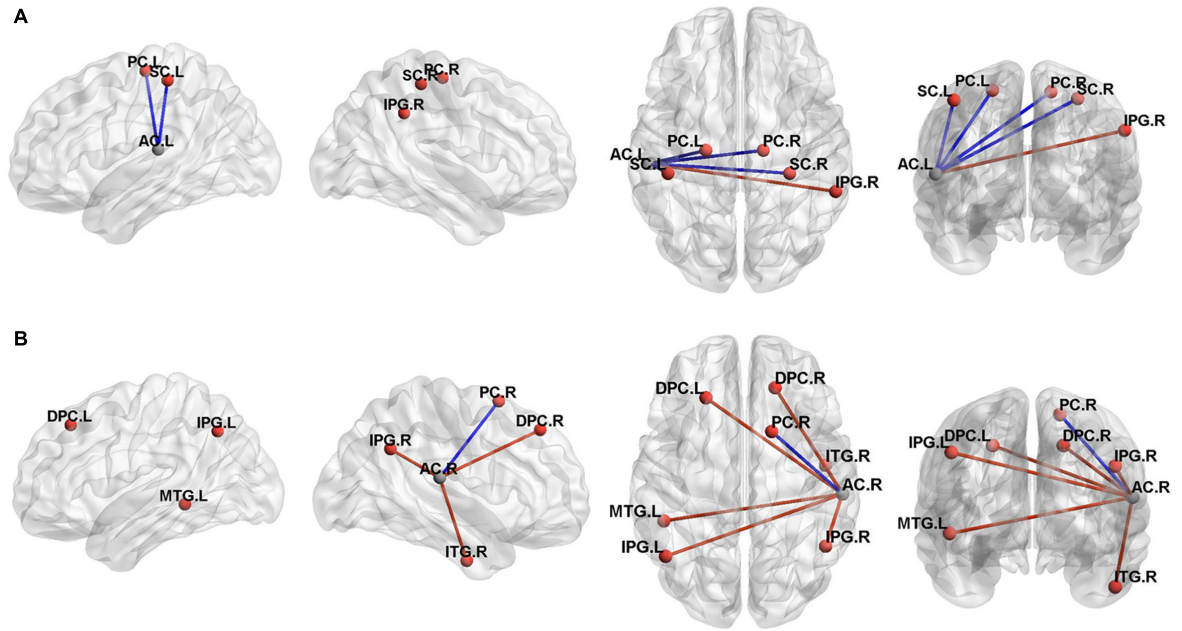


FIGURE 4 | Three-dimensional rendering of the functional connectivity results ($P < 0.001$, Alphasim correction). A represented the left auditory cortex functional connectivity results and B represented the right auditory cortex results. The clusters of significant difference in the two sample *T* tests correspond to regions listed in **Table 3**. MNI coordinate of the ROI seed (left auditory cortex, -55/-26/12; right auditory cortex, 58/-25/16) were shown in gray ball and the positive FC connectivity areas were shown in red ball. The blue line indicated decreased connectivity and the red line indicated increased connectivity. PC, Premotor Cortex;SC, Somatosensory Cortex;IPG, Inferior Parietal Gyrus;AC, auditory cortex;MTG, Middle Temporal Gyrus;ITG, Inferior Temporal Gyrus;DPC, Dorsolateral Prefrontal Cortex;L, left;R, Right.

Relationship Between the Dizziness Handicap Inventory Score and the Regional Homogeneity Value

There was a significant negative correlation between the DHI scores (**Table 4**) and the ReHo values in the right superior temporal gyrus of the Vertigo group ($r = -0.595$, $p = 0.031$), as well as a significant positive correlation in the right inferior parietal gyrus ($r = 0.834$, $p < 0.01$) and the left inferior parietal gyrus ($r = 0.579$, $p = 0.049$) (**Figure 5**).

DISCUSSION

To our best knowledge, this study firstly identified the altered activities in the auditory and vestibular cortices of the ISSNHL patients with vertigo by rs-fMRI. The major findings of this study demonstrated that: (1) the auditory and vestibular cortices both exhibited altered local activities; (2) In comparison to the normal hearing, the ReHo signals of the ipsilateral AC were increased in the ISSNHL without vertigo and decreased in the ISSNHL with vertigo; (3) There was a negative association between the DHI scores and the ReHo values in the inferior parietal gyrus. In sum, these results demonstrated that the ISSNHL patients with vertigo exhibited different intrinsic brain activity patterns compared to the ISSNHL patients without vertigo and healthy controls.

The Auditory and Vestibular Cortices Showed Functional Changes

The presence of vertigo in some but not all ISSNHL patients suggests that the brain areas involved in the additional dizziness may demonstrate a distinctive pattern of activity. From our results, we observed the ReHo signal changes in the auditory and vestibular cortices, such as the superior and middle temporal cortices, posterior insular cortex, inferior parietal gyrus (**Figure 2**). Furthermore, when comparing the ISSNHL without vertigo and the vestibular neuritis, the confirmed hypothesis was obtained: in the ISSNHL patients without vertigo, the activity of the contralateral AC was decreased, and the vestibular cortex was normal (Micarelli et al., 2017); in the vestibular neuritis patients (Bense et al., 2004), the activity of the right vestibular cortices was changed (increased or decreased) and the activity of the AC was normal; in the ISSNHL patients with vertigo, the activities of the AC and left vestibular cortices were decreased and the right parieto-insular vestibular cortex and inferior parietal cortex were increased (**Figure 1**).

The Activity of the Ipsilateral Auditory Cortex Was Decreased in Idiopathic Sudden Sensorineural Hearing Loss With Vertigo

The auditory projections from the inner ear to the AC cause contralateral activation of the brain in response to sound. Such

TABLE 3 | Brain regions with significant FC differences between ISSNHL with vertigo and without vertigo group.

Index	Brain region	ROI-1					ROI-2				
		Voxel	T value	MNI coordinate			Voxel	T value	NMI coordinate		
				X	Y	Z			X	Y	Z
1	Middle temporal gyrus (BA21_L)						115	2.678	–48	–41	–5
2	Somatosensory Cortex (BA3_L)	112	–4.962	–44	–32	56					
3	Somatosensory Cortex (BA3_R)	105	—3.312	30	–32	57					
4	Inferior parietal gyrus (BA40_L)						97	3.680	–47	–62	42
5	Inferior parietal gyrus (BA40_R)	106	3.526	58	–43	38	99	3.998	48	–56	34
6	Premotor cortex (BA6_R)	117	–3.254	14	–18	61	84	–4.874	16	12	64
7	premotor cortex (BA6_L)	101	–4.954	–21	–18	62					
8	Dorsolateral prefrontal cortex (BA9R)						116	4.095	18	38	46
9	Dorsolateral prefrontal cortex (BA9L)						90	6.655	–23	32	46
10	Inferior temporal gyrus (BA20R)						113	4.778	48	–8	–36

Voxel number, T-values were obtained from the statistical parametric mapping of the ReHo ($p < 0.001$).
ROI-1 (MNI coordinate: -55/-26/12) and ROI-2 (MNI coordinate: 58/-25/16) served as the seeds for the FC calculation.
L means left brain; R means right brain; BA means Brodmann area.

TABLE 4 | Summary of the characteristics of ISSNHL patients with vertigo.

Patient number	Age (years)	Sex (F/M)	Affected ear	Onset time(d)	PTA (dB)		ART Ipsi/con	Vestibular test			DHI
					before	after		Spont-N	BCT	cVEMP	
1	49	F	L	4	120	85	–/85	–	+	+	56
2	61	F	R	5	85	83.3	–/85	–	–	0	58
3	38	M	L	6	105	71.6	–/85	–	+	–	64
4	57	F	L	7	88.3	71.6	–/90	R	+	+	54
5	42	F	R	2	90	90	–/85	–	+	–	48
6	31	F	L	6	120	63	–/–	L	N	+	80
7	52	M	R	5	94	64	–/80	–	–	–	58
8	45	M	R	4	75	25	–/80	–	–	+	80
9	40	F	R	7	120	120	–/85	L	+	0	74
10	25	F	R	1	92	80	–/85	–	–	–	62
11	53	M	L	5	78	60	–/80	–	N	–	52
12	40	F	L	2	97	79	–/85	–	+	+	56

PTA, pure tone audiometry; ISSNHL, Idiopathic sudden sensorineural hearing loss; BCT, bithermal caloric test; cVEMP, cervical vestibular evoked myogenic potential; DHI, dizziness handicap inventory; "+", abnormal of waveforms or results; "–", normal waveforms or results; 0, no response; N, Test was not finished and terminated. ART, Acoustic Reflex Thresholds (ART) was a 3 frequency average (500, 1k, and 2k Hz) of the sensation level in dB HL; ipsi, ipsilesion; con, contralesion; "–", absent. Spon-N, spontaneous nystamus; "–", absent; R, right ear; L, left ear.

a cross-projection is useful in binaural facilitation of hearing and localization of sounds (Moore, 1991). However, when unilateral hearing is deprived, the corresponding AC activities may change in response.

In the contralateral AC of the unilateral ISSNHL patients (with or without vertigo, **Figure 2**), we observed a reduction of the ReHo values compared to the normal hearing. The underlying mechanism of this contralateral deactivation could be related to the misrepresentation of the sound intensity at the cortical level (Musiek et al., 2013). The perception of loudness or intensity at the cortical level is affected by the interaction between the excitatory and inhibitory neurons (Phillips et al., 1994). If deprived sound input activates fewer excitatory fibers, the activities of excitatory and inhibitory neurons can become unbalanced and, as a result, the neuronal response in one hemisphere will be reduced compared to the other (Fan et al., 2015). This finding is consistent with the previous studies in patients with chronic deafness or tinnitus reporting reduced blood flow in the AC (Lanting et al., 2009; Okuda et al., 2013; Micarelli et al., 2017). However, some recent studies reported no significant difference in the ReHo value between the ISSNHL patients and the healthy controls in any brain region. The various neuroimaging methods employed and the participants' heterogeneity may have contributed to the inconsistent results (Cai et al., 2020; Chen et al., 2020).

We observed a different effect in the activity of the ipsilateral AC when vertigo was present. The ipsilateral AC showed the restrained status after vertigo happened: the ReHo signals of the ipsilateral AC were increased in the ISSNHL group without vertigo, whereas they were

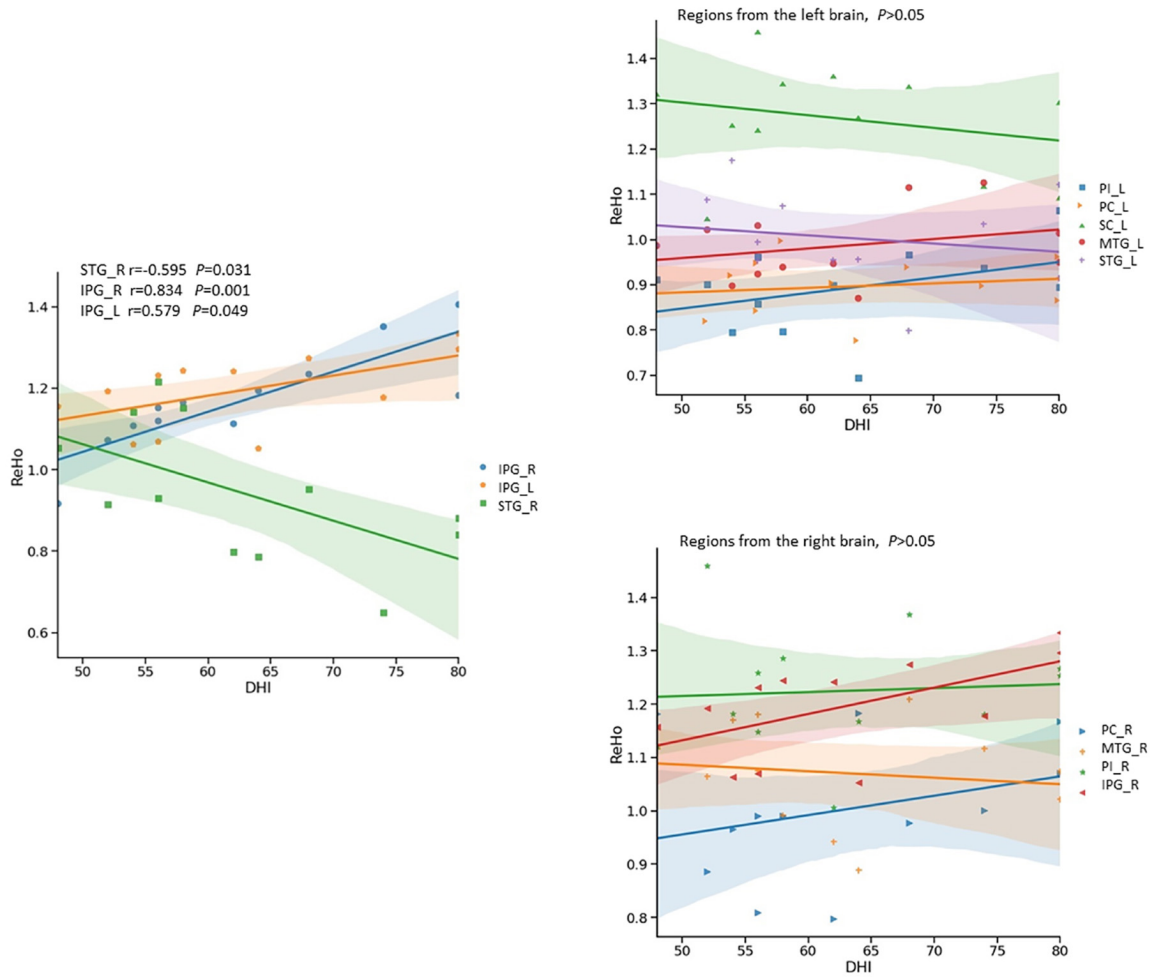


FIGURE 5 | Relationship between Dizziness Handicap Index (DHI) and ReHo signals in ISSNHL with vertigo by scatter plot. The DHI showed significantly negative correlation (r = -0.595, $p < 0.05$) with ReHo signals from the right superior temporal gyrus, and significantly positive correlations with ReHo signals in the left (r = 0.579, $p < 0.05$) and right inferior parietal gyri (r = 0.834, $p < 0.01$). PC, Premotor Cortex; SC, Somatosensory Cortex; IPG, Inferior Parietal Gyrus; STG, Superior Temporal Gyrus; MTG, Middle Temporal Gyrus; ITG, Inferior Temporal Gyrus; PI, Posterior Insular; L, left;R, Right.

decreased in the ISSNHL with vertigo (**Figure 3**). In two magnetoencephalography reports on ISSNHL, the activity of the ipsilateral AC was observed and elucidated. Morita and Li reported a stronger N100 evoked response in the ipsilateral compared to the contralateral hemisphere in response to ear stimulation during the acute phase of hearing loss (Li et al., 2006; Morita et al., 2007). The authors hypothesized that a cochlear lesion might induce a bilateral effect through retrocochlear crossing fibers that may influence the function of the auditory pathway associated with the ipsilateral healthy ear. This pattern of the reduced contralateral and increased ipsilateral AC activities in ISSNHL without vertigo may reflect the compensatory mechanism of the brain after unilateral hearing deprivation. However, how to explain the decreased activity of the ipsilateral AC in the ISSNHL patients with vertigo and what is the mechanism of this reduction? Did the activities of vestibular cortices lead to this reduction?

The Functional Connectivity of the Auditory-Vestibular Cortices in the Idiopathic Sudden Sensorineural Hearing Loss With Vertigo

Our research demonstrated that the inferior parietal gyrus and the AC displayed functional connectivity in the ISSNHL patients with vertigo (**Figures 4A,B**). It is known that the inferior parietal gyrus belongs to the multisensory area predominantly in the temporo-insular and temporo-parietal cortex of the human brain. These multisensory areas are involved in the processing of vestibular information and have been delineated during the last 7 years by functional imaging studies in humans (Bremmer et al., 2001; Fasold et al., 2002; Emri et al., 2003; Stephan et al., 2005).

In contrast to the integration of visual-vestibular perception, less is known about the neural mechanism that mediates the integration of vestibular and auditory processing. An early human functional study revealed a consistent focal activation

in the superior temporal region during vestibular activation (Friberg et al., 1985). A vestibular galvanic study demonstrated activations in the superior temporal gyrus and as well in the middle temporal gyrus (Bense et al., 2001). These results depicted a functional connectivity between the vestibular cortex and the auditory cortex in the human brain. Our findings also revealed that the inferior parietal cortex and the AC exhibited functional connectivity in the ISSNHL patients with vertigo (**Figures 4A,B**). A mechanism similar to the inhibitory visual-vestibular cortex interaction (Brandt and Dieterich, 1998; Della-Justina et al., 2015; Frank et al., 2020) may explain our experimental findings in the auditory-vestibular cortex of the patients. That means the dysfunction of the inferior parietal cortex perturbs the activity of the temporal cortex and worsens the auditory function. Confirmation of this theory and the mechanism or interaction must be examined in further experiments.

Correlation Analysis

A positive correlation was found between the DHI scores and the inferior parietal gyrus activity (**Figure 5**). In fact, activation of inferior parietal gyrus has been seen during caloric vestibular stimulation in fMRI and PET studies (Dieterich and Brandt, 2008; Helmchen et al., 2014). Brodmann Area 40 belongs to a multisensory area in the inferior parietal lobe, part of with strong projections the temporal lobe. Monkey studies revealed two vestibular areas in the parietal lobe, named the visual temporal sylvian area and area 7b (Faugier-Grimaud and Ventre, 1989; Guldin and Grusser, 1998). The location of these regions is comparable to parts of our activation area in inferior parietal gyrus. However, these findings do not imply that the observed fMRI changes are a contributing factor to the vertigo symptom. Our results are based on an observational study focusing on the central correlates of symptoms with a peripheral origin. The mechanisms involved in the observed changes to the central auditory areas, that occur in those that develop vertigo, need further investigation.

Limitations

Despite the promising results, there were certain limitations in our study. The small sample size of this research may cause type II statistical errors. Acute otoneurological patients, due to their very severe symptoms, often refused or delayed the MRI examination due to discomfort of lying still. This issue had made the recruitment of an adequate number of subjects difficult in functional neuroimaging studies, including ours (Micarelli et al., 2017).

Secondly, our study did not touch upon the pathogenesis of peripheral vestibular organs in this disease. The occurrence of vertigo was an integrated response of the whole vestibular pathway and it is becoming increasingly suggested that peripheral vestibular organs play a role in vertigo symptoms in the patients with sudden deafness. An overall vestibular pathway investigation should be carried out to fully uncover the pathogenesis of vertigo during sudden deafness.

CONCLUSION

Our research provides the first evidence for the altered activity in the AC and vestibular cortical areas of the ISSNHL patients with vertigo by rs-fMRI. The decreased activity of the ipsilateral AC in ISSNHL patients with vertigo was our main finding. We also found that ReHo value of the inferior parietal cortex was related to the DHI scores of the patient and this area demonstrated positive functional connectivity with the AC. The auditory and vestibular cortices demonstrate functional connectivity in the ISSNHL patients with vertigo. These findings contribute to the identification of neural plasticity in ISSNHL patients with vertigo and warrant further investigation to understand the mechanisms involved in the generation of vertigo.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Institutional Review Board of the Ethics Committee of the Huazhong University of Science and Technology. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

LP, QW, and LZ developed the research ideas. LP and LZ developed hypotheses, conducted analyses, interpreted data, and drafted the manuscript. PL coded, analyzed data, revised drafts of the manuscript, and designed the data collection protocol. PL and JZ revised drafts of the manuscript, contributed in discussing analyses, and provided critical revisions of the manuscript. All authors approved the submitted version.

FUNDING

This study was supported by grants from the National Natural Science Foundation of China (project numbers: 81700909).

ACKNOWLEDGMENTS

We are grateful to Gao Lie and Yang Huaguang for fMRI data processing.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnhum.2021.719254/full#supplementary-material

REFERENCES

Ashburner, J. (2018). A fast diffeomorphic image registration algorithm. *Neuroimage* 38, 95–113. doi: 10.1016/j.neuroimage.2007.07.007

Bense, S., Bartenstein, P., Lochmann, M., Schlindwein, P., Brandt, T., and Dieterich, M. (2004). Metabolic changes in vestibular and visual cortices in acute vestibular neuritis. *Ann. Neurol.* 56, 624–630. doi: 10.1002/ana.20244

Bense, S., Stephan, T., Yousry, T. A., Brandt, T., and Dieterich, M. (2001). Multisensory cortical signal increases and decreases during vestibular galvanic stimulation (fMRI). *J. Neurophysiol.* 85, 886–899. doi: 10.1152/jn.2001.85.2.886

Brandt, T., and Dieterich, M. (1998). The vestibular cortex. Its locations, functions, and disorders. *Ann. N. Y. Acad. Sci.* 871, 293–312. doi: 10.1111/j.1749-6632.1999.tb09193.x

Brandt, T., Bartenstein, P., Janek, A., and Dieterich, M. (1998). Reciprocal inhibitory visual–vestibular interaction. Visual motion stimulation deactivates the parieto-insular vestibular cortex. *Brain* 121 (Pt 9), 1749–1758. doi: 10.1093/brain/121.9.1749

Bremmer, F., Schlack, A., Shah, N. J., Zafiris, O., Kubischik, M., Hoffmann, K., et al. (2001). Polymodal motion processing in posterior parietal and premotor cortex: a human fMRI study strongly implies equivalencies between humans and monkeys. *Neuron* 29, 287–296. doi: 10.1016/s0896-6273(01)00198-2

Cai, Y., Xie, M., Su, Y., Tong, Z., Wu, X., Xu, W., et al. (2020). Aberrant functional and causal connectivity in acute tinnitus with sensorineural hearing loss. *Front. Neurosci.* 14:592. doi: 10.3389/fnins.2020.00592

Chang, T. P., Wang, Z., Winnick, A. A., Chuang, H. Y., Urrutia, V. C., Carey, J. P., et al. (2018). Sudden hearing loss with vertigo portends greater stroke risk than sudden hearing loss or vertigo alone. *J. Stroke. Cerebrovasc. Dis.* 27, 472–478. doi: 10.1016/j.jstrokecerebrovasdis.2017.09.033

Chen, J., Hu, B., Qin, P., Gao, W., Liu, C., Zi, D., et al. (2020). Altered brain activity and functional connectivity in unilateral sudden sensorineural hearing loss. *Neural Plast.* 2020:9460364.

Della-Justina, H. M., Gamba, H. R., Lukasova, K., Nucci-da-Silva, M. P., Winkler, A. M., and Amaro, E. Jr. (2015). Interaction of brain areas of visual and vestibular simultaneous activity with fMRI. *Exp. Brain Res.* 233, 237–252. doi: 10.1007/s00221-014-4107-6

Dieterich, M., and Brandt, T. (2008). Functional brain imaging of peripheral and central vestibular disorders. *Brain* 131(Pt 10), 2538–2552. doi: 10.1093/brain/awn042

Emri, M., Kisely, M., Lengyel, Z., Balkay, L., Marian, T., Miko, L., et al. (2003). Cortical projection of peripheral vestibular signaling. *J. Neurophysiol.* 89, 2639–2646. doi: 10.1152/jn.00599.2002

Fan, W., Zhang, W., Li, J., Zhao, X., Mella, G., Lei, P., et al. (2015). Altered contralateral auditory cortical morphology in unilateral sudden sensorineural hearing loss. *Otol. Neurotol.* 36, 1622–1627. doi: 10.1097/mao.0000000000000892

Fang, J. Q., Hao, Y. T., and Li, C. X. (1999). Reliability and validity for Chinese version of WHO quality of lifet scale. *Chin. Ment. Health. J.* 13, 203–206.

Fasold, O., von Brevern, M., Kuhberg, M., Ploner, C. J., Villringer, A., Lempert, T., et al. (2002). Human vestibular cortex as identified with caloric stimulation in functional magnetic resonance imaging. *NeuroImage* 17, 1384–1393. doi: 10.1006/nimg.2002.1241

Faugier-Grimaud, S., and Ventre, J. (1989). Anatomic connections of inferior parietal cortex (area 7) with subcortical structures related to vestibulo-ocular function in a monkey (*Macaca fascicularis*). *J. Comp. Neurol.* 280, 1–14. doi: 10.1002/cne.902800102

Frank, S. M., and Greenlee, M. W. (2018). The parieto-insular vestibular cortex in humans: more than a single area? *J. Neurophysiol.* 120, 1438–1450. doi: 10.1152/jn.00907.2017

Frank, S. M., Pawellek, M., Forster, L., Langguth, B., Schecklmann, M., and Greenlee, M. W. (2020). Attention networks in the parietooccipital cortex modulate activity of the human vestibular cortex during attentive visual processing. *J. Neurosci.* 40, 1110–1119. doi: 10.1523/jneurosci.1952-19.2019

Friberg, L., Olsen, T. S., Roland, P. E., Paulson, O. B., and Lassen, N. A. (1985). Focal increase of blood flow in the cerebral cortex ofman during vestibular stimulation. *Brain* 108, 609–623. doi: 10.1093/brain/108.3.609

Guldin, W. O., and Grusser, O. J. (1998). Is there a vestibular cortex? *Trends Neurosci.* 21, 254–259.

Guo, W. B., Sun, X. L., Liu, L., Xu, Q., Wu, R. R., Liu, Z. N., et al. (2011). Disrupted regional homogeneity in treatment-resistant depression: a resting-state fMRI study. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 35, 1297–1302. doi: 10.1016/j.pnpbp.2011.02.006

Helmchen, C., Ye, Z., Sprenger, A., and Munte, T. F. (2014). Changes in resting-state fMRI in vestibular neuritis. *Brain. Struct. Funct.* 219, 1889–1900. doi: 10.1007/s00429-013-0608-5

Kim, C. H., Choi, H. R., Choi, S., Lee, Y. S., and Shin, J. E. (2018). Patterns of nystagmus conversion in sudden sensorineural hearing loss with vertigo. *Medicine (Baltimore)* 97:e12982. doi: 10.1097/md.0000000000012982

Lanting, C. P., de Kleine, E., and van Dijk, P. (2009). Neural activity underlying tinnitus generation: results from PET and fMRI. *Hear Res.* 255, 1–13. doi: 10.1016/j.heares.2009.06.009

Li, L. P., Shiao, A. S., Chen, L. F., Niddam, D. M., Chang, S. Y., Lien, C. F., et al. (2006). Healthy-side dominance of middle- and long-latency neuromagnetic fields in idiopathic sudden sensorineural hearing loss. *Eur. J. Neurosci.* 24, 937–946. doi: 10.1111/j.1460-9568.2006.04961.x

Liu, Y., Wang, K., Yu, C., He, Y., Zhou, Y., Liang, M., et al. (2008). Regional homogeneity, functional connectivity and imaging markers of Alzheimer's disease: a review of resting-state fMRI studies. *Neuropsychologia* 46, 1648–1656. doi: 10.1016/j.neuropsychologia.2008.01.027

Micarelli, A., Chiaravalloti, A., Viziano, A., Danieli, R., Schillaci, O., and Alessandrini, M. (2017). Early cortical metabolic rearrangement related to clinical data in idiopathic sudden sensorineural hearing loss. *Hear Res.* 350, 91–99. doi: 10.1016/j.heares.2017.04.011

Michel, O. (2011). The revised version of the german guidelines "sudden idiopathic sensorineural hearing loss". *Laryngorhinootologie* 90, 290–293.

Moore, D. R. (1991). Anatomy and physiology of binaural hearing. *Audiology* 30, 125–134. doi: 10.3109/00206099109072878

Morita, T., Hiraumi, H., Fujiki, N., Naito, Y., Nagamine, T., Fukuyama, H., et al. (2007). A recovery from enhancement of activation in auditory cortex of patients with idiopathic sudden sensorineural hearing loss. *Neurosci. Res.* 58, 6–11.

Moskowitz, D., Lee, K. J., and Smith, H. W. (1984). Steroid use in idiopathic sudden sensorineural hearing loss. *Laryngoscope* 94(5 Pt 1), 664–666. doi: 10.1288/00005537-198405000-00016

Musiek, F., Guenette, L., and Fitzgerald, K. (2013). Lateralized auditory symptoms in central neuroaudiology disorder. *J. Am. Acad. Audiol.* 24, 556–563. doi: 10.3766/jaaa.24.7.4

Okuda, T., Nagamachi, S., Ushisako, Y., and Tono, T. (2013). Glucose metabolism in the primary auditory cortex of postlingually deaf patients: an FDG-PET study. *ORL J. Otorhinolaryngol. Relat. Spec.* 75, 342–349. doi: 10.1159/000357474

Park, H. M., Jung, S. W., and Rhee, C. K. (2001). Vestibular diagnosis as prognostic indicator in sudden hearing loss with vertigo. *Acta Otolaryngol. Suppl.* 545, 80–83. doi: 10.1080/oto.121.533.80.83

Phillips, D. P., Semple, M. N., Calford, M. B., and Kitzes, L. M. (1994). Level-dependent representation of stimulus frequency in cat primary auditory cortex. *Exp. Brain Res.* 102, 210–226.

Pogson, J. M., Taylor, R. L., Young, A. S., McGarvie, L. A., Flanagan, S., Halmagyi, G. M., et al. (2016). Vertigo with sudden hearing loss: audio-vestibular characteristics. *J. Neurol.* 263, 2086–2096. doi: 10.1007/s00415-016-8214-0

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., and Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage* 59, 2142–2154. doi: 10.1016/j.neuroimage.2011.10.018

Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., and Petersen, S. E. (2013). Steps toward optimizing motion artifact removal in functional connectivity

MRI; a reply to carp. *Neuroimage* 76, 439–441. doi: 10.1016/j.neuroimage.2012.03.017

Rauch, S. D. (2008). Clinical practice. Idiopathic sudden sensorineural hearing loss. *N. Engl. J. Med.* 359, 833–840.

Rauch, S. D. (2018). The clinical value of vertigo as a prognostic indicator of outcome in sudden sensorineural hearing loss. *JAMA Otolaryngol. Head Neck Surg.* 144, 684–685. doi: 10.1001/jamaoto.2018.0674

Shaia, F. T., and Sheehy, J. L. (1976). Sudden sensori-neural hearing impairment: a report of 1,220 cases. *Laryngoscope* 86, 389–398. doi: 10.1288/00005537-197603000-00008

Stachler, R. J., Chandrasekhar, S. S., Archer, S. M., Rosenfeld, R. M., Schwartz, S. R., Barrs, D. M., et al. (2012). Clinical practice guideline: sudden hearing loss. *Otolaryngol. Head Neck Surg.* 146(3 Suppl), S1–S35.

Stephan, T., Deutschländer, A., Nolte, A., Schneider, E., Wiesmann, M., Brandt, T., et al. (2005). Functional MRI of galvanic vestibular stimulation with alternating currents at different frequencies. *Neuro-Image* 26, 721–732. doi: 10.1016/j.neuroimage.2005.02.049

Wang, C. T., Huang, T. W., Kuo, S. W., and Cheng, P. W. (2009). Correlation between audiovestibular function tests and hearing outcomes in severe to profound sudden sensorineural hearing loss. *Ear Hear* 30, 110–114. doi: 10.1097/aud.0b013e318192655e

Yan, C. G., Craddock, R. C., He, Y., and Milham, M. P. (2013). Addressing head motion dependencies for small-world topologies in functional connectomics. *Front. Hum. Neurosci.* 7:910. doi: 10.3389/fnhum.2013.00910

Yao, Z., Wang, L., Lu, Q., Liu, H., and Teng, G. (2009). Regional homogeneity in depression and its relationship with separate depressive symptom clusters: a resting-state fMRI study. *J. Affect. Disord.* 115, 430–438. doi: 10.1016/j.jad.2008.10.013

Zhou, F., Zhu, M. C., Wang, M., Wang, H. T., Jiao, Y. L., Huang, L. F., et al. (2018). Clinical analysis of idiopathic sudden sensorineural hearing loss with vertigo and without vertigo. *Lin Chuan Er Bi Yan Hou Tou Jing Wai Ke Za Zhi* 32, 920–923.

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 28 October 2021
doi: 10.3389/fnins.2021.735950

Tinnitus Is Associated With Improved Cognitive Performance in Non-hispanic Elderly With Hearing Loss

Yasmeen Hamza and Fan-Gang Zeng**

Center for Hearing Research, Department of Anatomy and Neurobiology, Biomedical Engineering, Cognitive Sciences, Otolaryngology–Head and Neck Surgery, University of California, Irvine, Irvine, CA, United States

Because hearing loss is a high-risk factor for cognitive decline, tinnitus, a comorbid condition of hearing loss, is often presumed to impair cognition. The present cross-sectional study aimed to delineate the interaction of tinnitus and cognition in the elderly with and without hearing loss after adjusting for covariates in race, age, sex, education, pure tone average, hearing aids, and physical well-being. Participants included 643 adults (60–69 years old; 51.3% females) from the National Health and Nutrition Examination Survey (NHANES, 2011–2012), and 1,716 (60–69 years old; 60.4% females) from the Hispanic Community Health Study (HCHS, 2008–2011). Multivariable linear and binary logistic regression was used to assess the association between tinnitus and cognition in the two sub-cohorts of normal hearing (NHANES, n = 508; HCHS, n = 1264) and hearing loss (NHANES, n = 135; HCHS, n = 453). Cognitive performance was measured as a composite z-score from four cognitive tests: The Consortium to Establish a Registry for Alzheimer's Disease (CERAD)-word learning, CERAD-animal fluency, CERAD-word list recall, and the digit symbol substitution test (DSST) in NHANES, and a comparable Hispanic version of these four tests in HCHS. Multivariable linear regression revealed no association between tinnitus and cognition, except for the NHANES (non-Hispanic) participants with hearing loss, where the presence of tinnitus was associated with improved cognitive performance (Mean = 0.3; 95% CI, 0.1–0.5; p, 0.018). Using the 25th percentile score of the control (i.e., normal hearing and no tinnitus) as a threshold for poor cognitive performance, the absence of tinnitus increased the risk for poor cognitive performance (OR = 5.6, 95% CI, 1.9–17.2; p, 0.002). Sensitivity analysis found a positive correlation between tinnitus duration and cognitive performance in the NHANES cohort [$F(4,140)$, 2.6; p, 0.037]. The present study finds no evidence for the assumption that tinnitus impairs cognitive performance in the elderly. On the contrary, tinnitus is associated with improved cognitive performance in the non-Hispanic elderly with hearing loss. The present result suggests that race be considered as an important and relevant factor in the experimental design of tinnitus research. Future longitudinal and imaging studies are needed to validate the present findings and understand their mechanisms.

Keywords: tinnitus, cognition, elderly, Hispanic, hearing loss

OPEN ACCESS

Edited by:
William Sedley,
Newcastle University, United Kingdom

Reviewed by:
Derek James Hoare,
University of Nottingham,
United Kingdom
Joel I. Berger,
The University of Iowa, United States

****Correspondence:***
Yasmeen Hamza
yhamza@hs.uci.edu
orcid.org/0000-0003-3294-6629
Fan-Gang Zeng
fzeng@uci.edu
orcid.org/0000-0002-4325-2780

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

Received: *04 July 2021*
Accepted: *04 October 2021*
Published: *28 October 2021*

Citation:
Hamza Y and Zeng FG (2021) Tinnitus Is Associated With Improved Cognitive Performance in Non-hispanic Elderly With Hearing Loss. Front. Neurosci. 15:735950. doi: 10.3389/fnins.2021.735950

INTRODUCTION

Hearing loss is one of the most prevalent conditions in the elderly, including nearly half of those aged 65–84 years (Nash et al., 2011; Davis et al., 2016). Not only does hearing loss contribute to age-related cognitive decline (Lin, 2011; Lin et al., 2011a,b; Deal et al., 2017), but also is a leading modifiable risk factor that may prevent or delay 40% of dementia cases (Livingston et al., 2020). Tinnitus, or ringing in the ear, affects 15% of the general population (Nondahl et al., 2010; McCormack et al., 2016). Tinnitus is not only a common comorbid condition of hearing loss in the elderly, with about 80% overlap (Lockwood et al., 2002; Baguley et al., 2013) but also a concomitant symptom of dementia, with 52% overlap (Spiegel et al., 2018).

Because of this high comorbidity, tinnitus has been often assumed to impair cognition (Hallam et al., 2004; Savastano, 2008; Tegg-Quinn et al., 2016; Chu et al., 2020; Lee, 2020). For example, Tegg-Quinn et al. (2016) reviewed 18 relevant studies to show an association between tinnitus and some aspects of cognitive function, such as executive control of attention. In a meta-analysis study involving 38 records from 1,863 participants, Clarke et al. (2020) found that tinnitus is additionally associated with lower processing speed and poorer short-term memory. Based on a national population retrospective study, Chu et al. (2020) showed that tinnitus is an independent risk factor for subsequent Alzheimer's and Parkinson's disease, suggesting a role of tinnitus in age-related cognitive decline.

There are three knowledge gaps about the assumption that tinnitus impairs cognition in these previous studies. First, most studies did not control for potential interactive factors such as age, sex, race, hearing loss, education, anxiety, depression, and physical wellbeing, potentially confounding the role of tinnitus in cognition (Mohamad et al., 2016; Tegg-Quinn et al., 2016; Jafari et al., 2019). Second, if we accepted the aforementioned tinnitus-impairing cognition assumption, it remains unclear whether tinnitus inserts a general influence on the global cognitive function or affects specific domains of cognition such as attention, executive function, episodic or working memory (Albert et al., 2011; Veríssimo et al., 2021). Third, tinnitus is highly heterogeneous, with different tinnitus attributes and types potentially affecting different domains of cognitive function (Baguley et al., 2013).

Here we partially addressed these three knowledge gaps by studying the relation of tinnitus to cognition in a specific age group of elderly adults (60–69 years old) who participated in the National Health and Nutrition Examination Survey 2011–2012 (NHANES) and the Hispanic Community Health Study 2008–2011 (HCHS). First, we delineated the tinnitus role in cognition, among different hearing statuses, after controlling for age, sex, race, education, and physical wellbeing covariates. Second, we examined not only a global cognitive measure but also several separate cognitive domains. Third, in conditions where tinnitus affects cognition, we examined the relationship between specific tinnitus attributes and the affected cognitive function. Our overarching hypothesis was that tinnitus impairs cognitive function even if hearing loss and other covariates are accounted for.

MATERIALS AND METHODS

Participants and Cohorts

The current cross-sectional study includes cohorts from the NHANES and the HCHS.

The NHANES is a biannual United States-representative cross-sectional study. The 2011–2012 cycle entailing 9,756 individuals was chosen, as it was the only cycle that assessed more than one cognitive domain. **Figure 1A** shows the process of identifying the participants for the present study. A total of 9,113 participants were excluded, including 8,069 who did not participate in the cognitive tests because, within that cycle, cognitive tests were only administered to participants aged between 60 and 69 years, and 1,044 who had missing data in various tests. The resulting full sample included 643 participants who had completed an assessment on tinnitus, pure-tone audiometry, other covariates, and cognitive tests. To delineate the role of hearing loss in tinnitus effect on cognition, the full sample was divided into two sub-cohorts: 508 participants with normal hearing (blue left box) and 135 with hearing loss (red right box). Hearing loss was defined as a threshold greater than 25 dB HL based on the unaided better-ear pure-tone average (PTA) of 0.5, 1, 2, and 4 kHz [World Health Organization (WHO), 2008], as in previous studies assessing the association between hearing loss and cognition in the elderly (Lin, 2011; Lin et al., 2011a,b; Deal et al., 2015, 2017). Finally, to examine the interactive effects of hearing loss and tinnitus on cognition, the normal hearing and hearing loss sub-cohorts were divided into four groups with no tinnitus and tinnitus, respectively (the bottom row).

To confirm NHANES findings or to assess their generalizability to the Hispanic population, the HCHS cohort was often chosen (Cruickshanks et al., 2015; Golub et al., 2017, 2020). The HCHS is a multicenter United States community-based prospective study in Hispanic/Latino populations conducted between 2008 and 2011 entailing 16,415 individuals. **Figure 1B** shows the process of identifying the participants for the present study. A total of 14,699 participants were excluded, including 8,050 who did not participate in the cognitive tests, and 822 who had missing data in various tests. An additional 5,827 participants were excluded to match the NHANES cohort age range of 60–69 years. The resulting full sample included 1,716 participants who had completed an assessment on tinnitus, pure-tone audiometry, other covariates, and cognitive tests. The full sample was divided into two sub-cohorts: 1,263 participants with normal hearing (blue left box) and 453 with hearing loss (red right box). The same definitions in NHANES regarding hearing loss and tinnitus were used to divide the cohort into four groups (the bottom row).

Cognitive Performance

Cognitive performance was the primary outcome measure. Four tests were conducted to assess cognitive function. The results from the four cognitive tests were normalized by standard deviations from the full-sample mean, then averaged to yield a single cognitive z-score. For example, in the NHANES cohort, a z-score of zero is equivalent to the mean of the 643 participants,

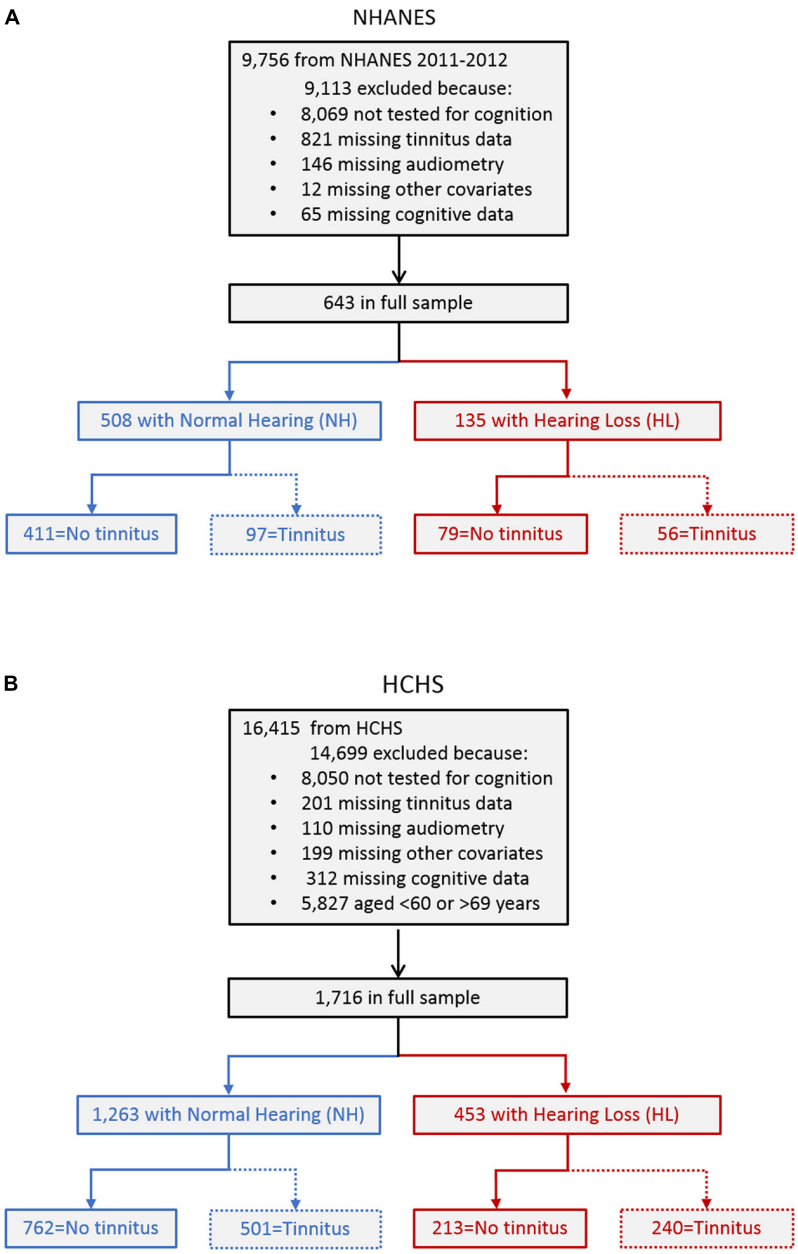


FIGURE 1 | Experimental design flowchart in the NHANES [panel **(A)**] and HCHS **(B)** database.

and a z-score of 1.0 indicates a value that is one standard deviation above the mean performance. A single cognitive score provides a better measure of global cognitive function than individual scores for clinical evaluation of patients with cognitive impairment, especially when individual test scores do not show concordance (Mistridis et al., 2015; Palumbo et al., 2020).

The four tests administered are as follows: To test immediate learning memory, the Consortium to Establish a Registry for Alzheimer's Disease (CERAD)-word learning test (Morris et al., 1989) in NHANES, and its HCHS-equivalent: the Spanish-English Verbal Learning Test (SEVLT: González et al., 2001) was used, in which the participant was presented with unrelated common words randomly in three trials and recalled them immediately (10 in CERAD and 15 in SEVLT). To test delayed learning memory, the CERAD-word list recall, and SEVLT-recall were used, in which the participant was asked to recall as many words as possible from the words learned in the first test after a short delay (CERAD) or a distracting word list (SEVLT). To test executive function, the CERAD-Animal Fluency test was used in NHANES, in which the participant named as many animals as possible in 1 min, and the Word Fluency Test (Spreen and Strauss, 1998) was used in HCHS, in which the participant named

as many words as possible that begin with certain letters in two trials of 1 min each. To test sustained attention, processing speed, and working memory, the Digit Symbol Substitution Test (Wechsler, 1997; Lezak et al., 2004) was used in both NHANES and HCHS. The participant was shown a key containing nine numbers and their paired symbols, then asked to, in 2 min in NHANES and 90 s in HCHS, transcribe as many symbols as possible corresponding to the numbers in a contiguous box that contains up to 133 symbols. High correlation was present between the individual tests in NHANES (r, 0.4–0.7; $p < 0.001$) and HCHS (r, 0.3–0.7; $p < 0.001$).

Tinnitus

Tinnitus was the primary exposure under test. It was encoded as a categorical variable, defined as whether the participant had experienced "Tinnitus" or "No Tinnitus" in the past year before examination. For those with tinnitus, three tinnitus-specific attributes were further examined in the NHANES cohort. Tinnitus duration was encoded as a five-level categorical variable, including less than 3 months, 3 months to a year, 1 to 4 years, 5 to 9 years, and ten or more years. Tinnitus severity was encoded as a five-level categorical variable, including no problem, a small problem, a moderate problem, a big problem, a very big problem. Tinnitus frequency was encoded as a five-level categorical variable, including less frequently than once a month, at least once a month, at least once a week, at least once a day, almost always. The reference category was set to the category of least magnitude. For example, for the tinnitus duration attribute, the shortest duration (less than 3 months) served as a reference, namely coded as 0, whereas the longest duration (ten or more years) was coded as 4. For the HCHS cohort, only data on tinnitus frequency was available and was encoded similar to the NHANES.

Covariates

Covariates included age, sex, education, physical well-being score, PTA in both NHANES and HCHS cohorts, and additionally race in the NHANES cohort. Hearing aid use was added as a covariate only in the hearing loss sub-cohorts of both the NHANES and HCHS. The selection of covariates was based on a statistically significant association with cognitive performance using univariable linear regression models in NHANES (see **Table 1**). Age was a continuous variable, ranging from 60 to 69 years old. Sex was a binary variable as either female or male. Education was coded as a four-level categorical variable, including less than ninth grade, high school, high school graduate or some college, and college degree or higher. The physical well-being score was a continuous variable, ranging from zero to five based on the presence of the following risk factors: coronary artery disease, hypertension, history of transient ischemic attack, impaired glucose tolerance, and diabetes (Golub et al., 2020). The PTA was a continuous variable, ranging from −7.5 to 100 dB HL. The race was coded as a binary variable as either Hispanic or non-Hispanic in the NHANES cohort. Additionally, in the hearing loss sub-cohorts of both the NHANES and HCHS, hearing aid use was coded as a binary variable as either worn hearing aids previously or not.

Statistical Analysis

Analysis was done for the NHANES and HCHS cohorts separately. First, descriptive analysis was performed on all variables, with continuous variables being described in terms of means (SDs) and range, while categorical variables were in frequencies and proportions (**Tables 2**, **3**). Raw cumulative frequency curves were generated as a function of cognitive z-score for the four groups to visualize the overall trends in the data (**Figures 2A,B**).

Second, to account for the significant covariate effect on cognition, multivariable linear regression models were used to obtain adjusted cognitive z-score, and to assess the association between tinnitus and cognition in the two sub-cohorts (normal hearing and hearing loss):

$$\text{Adjusted cognitive z} - \text{score} = \beta 0 + \beta 1 * \text{Age} + \beta 2 * \text{Sex} + \beta 3 * \text{Education} + \beta 4 * \text{Physical Well} - \text{being} + \beta 5 * \text{PTA} + \beta 6 * \text{Race} + \beta 7 * \text{Hearing Aid} + \beta 8 * \text{Tinnitus} + \varepsilon$$

The analysis utilized the cognitive z-score for global cognition and was replicated for each of the individual cognitive test z-scores. Regression $\beta 8$ coefficients and 95% confidence intervals (CI) were reported, which is the difference in adjusted z-score based on tinnitus status. Note that race was used only in the NHANES cohort, and hearing aid was used only in the hearing loss sub-cohort. Cumulative frequency curves were generated as a function of the adjusted cognitive z-score (**Figures 2C,D**).

Third, to test the consistency in the results between the NHANES and HCHS cohorts, the NHANES cohort was stratified into Hispanic and Non-Hispanic groups. Multivariable linear regression was conducted in the two groups to derive adjusted cognitive z-score and to assess the association between tinnitus

TABLE 1 | Covariate's association with cognition using univariable linear regression.

Covariate	Cognitive Z-score difference (95% CI)	p
Age	−0.03 (−0.1–0.0)	0.004*
Gender	0.4 (0.3–0.5)	< 0.001*
Educational level		
Less than ninth grade	Reference	Reference
High school (HS)	0.6 (0.4–0.8)	< 0.001*
HS graduate- some college	1.1 (0.9–1.3)	< 0.001*
College degree or higher	1.4 (1.2–1.6)	< 0.001*
Physical well-being score	−0.1 (−0.2 to −0.1)	< 0.001*
Psychological well-being score[a]	0.0 (0.02–0.01)	0.25
PTA	−0.01 (−0.02 to −0.01)	< 0.001*
Hearing aid	0.2 (−0.1−0.5)	0.22
Hearing aid in hearing loss	0.5 (0.1–0.8)	0.007*
Race	0.4 (0.2–0.5)	< 0.001*

[a] *The psychological well-being score is a depression screener score obtained by the patient health questionnaire (PHQ; Spitzer et al., 1999), that assesses the frequency of symptoms of depression over the past 2 weeks using nine items. * indicates a significant association between the covariate and cognition ($p < 0.05$).*

TABLE 2 | Participant characteristics and test scores of the NHANES sub-cohorts.

Variable	NHANES sub-cohorts			
	Normal hearing (508)		Hearing loss (135)	
	No tinnitus (411)	Tinnitus (97)	No tinnitus (79)	Tinnitus (56)
Age, mean (SD) (range), y	63.7 (2.7) (60–69)	63.7 (2.7) (60–69)	64.2 (2.7) (60–69)	64.5 (3.2) (60–69)
Gender				
Females, No. (%)	202 (49.1)	34 (35.1)	56 (70.9)	38 (67.9)
Males, No. (%)	209 (50.9)	63 (64.9)	23 (29.1)	18 (32.1)
Educational level, No. (%)				
Less than ninth grade	37 (9.0)	8 (8.2)	11 (13.9)	8 (14.3)
High school (HS)	142 (34.5)	35 (36.1)	40 (50.6)	19 (33.9)
HS graduate- some college	119 (29.0)	33 (34.0)	20 (25.3)	23 (41.1)
College degree or higher	113 (27.5)	21 (21.7)	8 (10.1)	6 (10.7)
Physical well-being score, mean (SD) (range)	1.1 (1.1) (0–4)	1.4 (1.2) (0–5)	1.1 (1.0) (0–4)	1.3 (1.3) (0–4)
PTA[a], mean (SD) (range), dB HL	13.6 (6.0) (−7.5–25)	14.9 (6.3) (1.3–25)	36.0 (10.1) (26.3–70)	35.6 (8.4) (26.3–60)
Hearing aids, No. (%)	N/A	N/A	12 (15.2)	12 (21.4)
DSST score, mean (SD) (range), z-score	0.0 (1.0) (−2.8–2.8)	0.2 (0.9) (−1.8–2.3)	−0.2 (0.9) (−2.3–1.7)	−0.1 (1.0) (−2.6–2.4)
CERAD learning score, mean (SD) (range), z-score	0.0 (1.0) (−4.6–2.3)	0.1 (0.9) (−2.2–2.1)	−0.4 (1.1) (−3.2–1.6)	−0.1 (0.9) (−2.4–2.1)
CERAD recall score, mean (SD) (range), z-score	0.0 (1.0) (−3.0–1.9)	0.1 (0.9) (−2.0–1.9)	−0.3 (1.0) (−3.0–1.4)	−0.1 (1.0) (−3.0–1.9)
Animal fluency Test score, mean (SD) (range), z-score	0.0 (1.0) (−2.6–4.0)	0.0 (1.0) (−1.7–3.1)	−0.2 (0.9) (−2.4–2.9)	0.1(1.0) (−2.1–2.4)
Cognitive performance, mean (SD) (range), z-score	0.0 (0.8) (−2.2–2.4)	0.1 (0.7) (−1.9–1.7)	−0.3 (0.8) (−1.7–1.4)	0.0 (0.8) (−1.9–2.0)

[a]*Pure-Tone Average threshold of four frequencies (0.5, 1, 2, and 4 KHz) in the better hearing ear.*

TABLE 3 | Participant characteristics and test scores of the HCHS sub-cohorts.

Variable	HCHS sub-cohorts			
	Normal hearing (1263)		Hearing loss (453)	
	No tinnitus (762)	Tinnitus (501)	No Tinnitus (213)	Tinnitus (240)
Age, mean (SD) (range), y	63.4 (2.7) (60–69)	63.3 (2.7) (60–69)	63.8 (2.8) (60–69)	64.2 (3.0) (60–69)
Gender				
Females, No. (%)	468 (61.4)	364 (72.7)	84 (39.4)	120 (50.0)
Males, No. (%)	294 (38.6)	137 (27.3)	129 (60.6)	120 (50.0)
Educational level, No. (%)				
Less than ninth grade	262 (34.4)	191 (38.1)	86 (40.4)	106 (44.2)
High school (HS)	210 (27.6)	139 (27.7)	72 (33.8)	83 (34.6)
HS graduate- some college	177 (23.2)	123 (24.6)	38 (17.8)	47 (19.6)
College degree or higher	113 (14.8)	48 (9.6)	17 (8.0)	4 (1.7)
Physical well-being score, mean (SD) (range)	1.7 (1.2) (0–5)	1.7 (1.2) (0–5)	1.8 (1.1) (0–5)	2.0 (1.1) (0–4)
PTA[a], mean (SD) (range), dB HL	15.5 (5.4) (0–25)	16.3 (5.4) (0–25)	34.1 (9.4) (26.3–83.8)	37.3 (12.7) (26.3–100)
Hearing aids, No. (%)	N/A	N/A	11 (5.2)	20 (8.4)
DSST score, mean (SD) (range), z-score	0.1 (1.0) (−2.4–3.4)	0.0 (1.0) (−2.4–3.2)	−0.1 (1.0) (−2.1–2.9)	−0.2 (0.9) (−2.3–2.9)
CERAD learning score, mean (SD) (range), z-score	0.1 (1.0) (−2.9–3.3)	0.1 (1.0) (−2.7–2.9)	−0.2 (1.0) (−2.5–2.5)	−0.2 (1.0) (−3.5–2.7)
CERAD recall score, mean (SD) (range), z-score	0.1 (0.9) (−2.7–2.6)	0.1 (1.0) (−2.7–2.3)	−0.1 (1.0) (−2.7–2.3)	−0.3 (1.0) (−2.7–2.3)
Animal fluency Test score, mean (SD) (range), z-score	0.1 (1.1) (−2.6–4.4)	0.0 (1.0) (−2.4–3.2)	−0.1 (1.0) (−2.3–2.7)	−0.1(0.9) (−2.4–2.7)
Cognitive performance, mean (SD) (range), z-score	0.1 (0.8) (−2.4–2.5)	0.1 (0.7) (−2.1–1.9)	−0.1 (0.7) (−2.2–2.0)	−0.2 (0.7) (−2.1–2.3)

[a]*Pure-tone average threshold of four frequencies (0.5, 1, 2, and 4 KHz) in the better hearing ear.*

and cognition. Cumulative frequency curves were generated as a function of the adjusted cognitive z-score (**Figures 2E,F**).

Fourth, to examine the between-group effect on cognitive performance, one-way ANOVAs were performed with the independent variable "Study Groups (4)" and dependent variable "adjusted z-score." *Post hoc* tests with Bonferroni corrections for multiple comparisons were conducted for pair-wise comparison between the four groups (**Figure 3**).

Fifth, to predict tinnitus probability based on cognitive performance, a multivariable binary logistic regression model

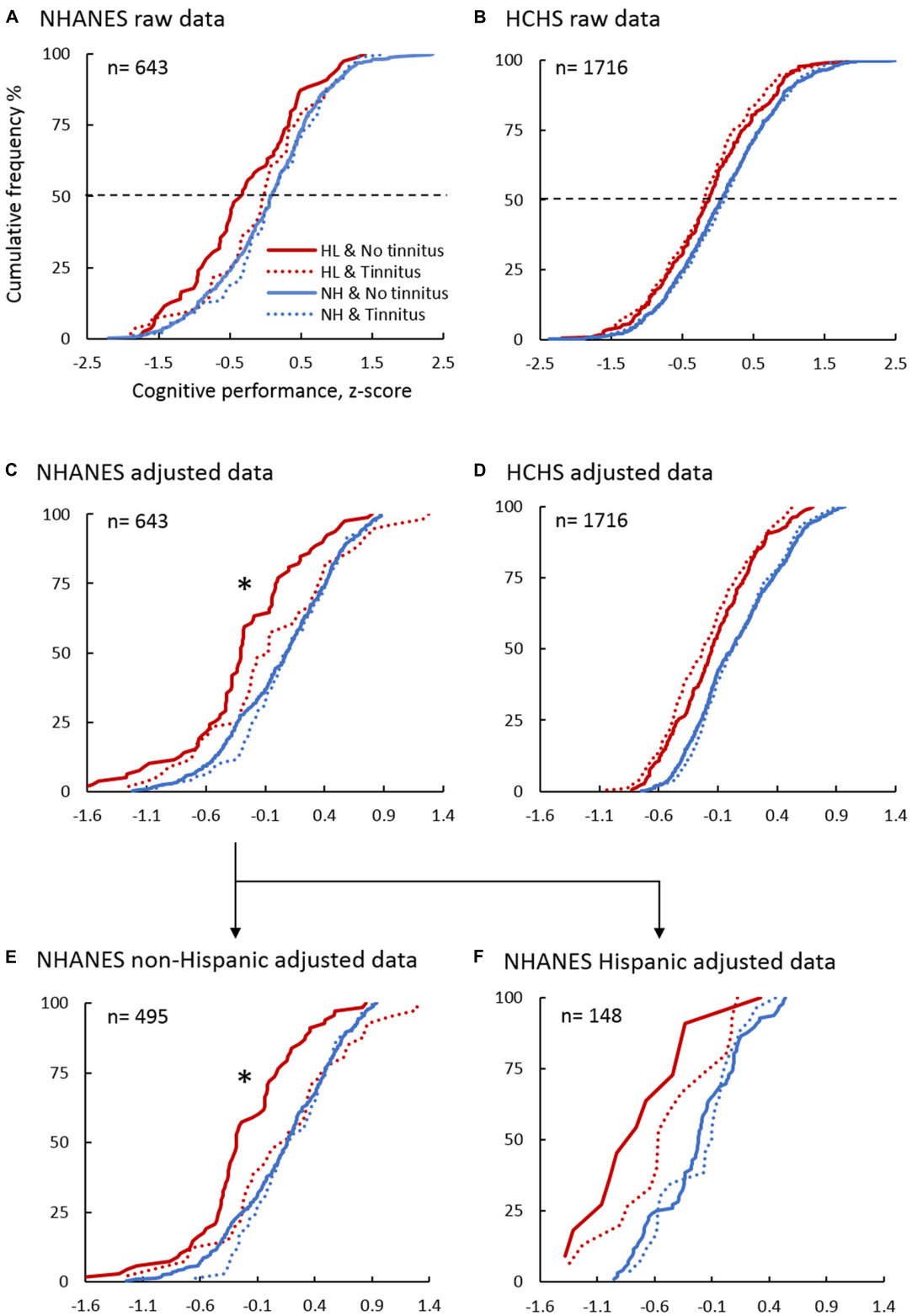


FIGURE 2 | Cumulative Frequency Distribution of the Cognitive Performance Z-score in the NHANES and HCHS Sub-cohorts. Panels show the raw data in NHANES **(A)** and HCHS **(B)**, the adjusted cognitive performance using multivariable linear regression in NHANES **(C)**, HCHS **(D)**, NHANES non-Hispanic **(E)**, and NHANES Hispanic **(F)** sub-cohorts. Blue lines denote normal hearing, and red lines are of hearing loss. Solid lines are of No tinnitus, and dashed lines are of tinnitus. *indicates a significant association between tinnitus and improved cognition in the hearing loss sub-cohort ($p < 0.05$).

was used with tinnitus being the binary outcome and cognitive z-score being the predictor under test (**Figure 4**):

$$\text{Probability (Tinnitus)} = \frac{1}{1 + e^{-(\beta 0 + \beta 1 * Age + \beta 2 * Sex + \beta 3 * Education + \beta 4 * Physical\ Well-being + \beta 5 * \text{PTA} + \beta 6 * \text{Race} + \beta 7 * \text{Hearing Aid} + \beta 8 * Cognitive\ performance + \varepsilon)}}$$

Odds ratio [exp(β8)] and 95% confidence intervals (CI) were reported, which are the change in the likelihood of tinnitus based on an increase in cognitive performance (per 1 z-score). The model was replicated using individual cognitive tests z-scores.

Sixth, to estimate the likelihood of poor cognitive performance based on tinnitus status, multivariable binary logistic regression models were used in the normal hearing and hearing loss sub-cohorts. The 25th percentile z-score of the normal hearing and no tinnitus group was calculated (i.e., z-score = −0.47) and used as a criterion to demarcate the outcome measure "poor cognitive performance," in accordance with the recommended criterion for the definition of mild cognitive impairment (American Psychiatric Association, 1980):

$$\text{Probability}\left(\text{Cognitive z} - \text{score} \leq -0.47\right) = \frac{1}{1 + e^{-(\beta 0 + \beta 1 * Age + \beta 2 * Sex + \beta 3 * Education + \beta 4 * \text{Physical Well-being} + \beta 5 * \text{PTA} + \beta 6 * \text{Race} + \beta 7 * \text{Hearing Aid} + \beta 8 * \text{Tinnitus} + \varepsilon)}}$$

Odds ratio [exp(β8)] and 95% confidence intervals (CI) were reported, which are the odds of poor cognitive performance of participants with no tinnitus relative to that of participants with tinnitus. The model was replicated for individual cognitive test z-scores. Note again that race was used only in the NHANES cohort, and hearing aid was used only in the hearing loss sub-cohorts.

Lastly, for those with tinnitus, sensitivity analysis was performed on the relationship between tinnitus-specific attributes and change in cognitive performance using multivariable linear regression models (**Figure 5**):

$$\text{Cognitive performance (z - score)} = \beta 0 + \beta 1 * \text{Age} + \beta 2 * \text{Sex} + \beta 3 * \text{Education} + \beta 4 * \text{Physical Well - being} + \beta 5 * \text{PTA} + \beta 6 * \text{Race} + \beta 7 * \text{Tinnitus attribute} + \varepsilon$$

For the NHANES cohort (n = 135), tinnitus attributes included tinnitus duration, severity, and frequency, whereas, for the HCHS cohort (n = 715), tinnitus attributes were confined to tinnitus frequency only. Regression β7 coefficients and 95% confidence intervals (CI) were reported, which are the adjusted difference in z-score based on the specified tinnitus attribute relative to the reference category. Note that race was used only in the NHANES cohort.

Data were analyzed using IBM SPSS software package version 26.0 (IBM_Corp, 2019). Significance was defined at the $p < 0.05$ level.

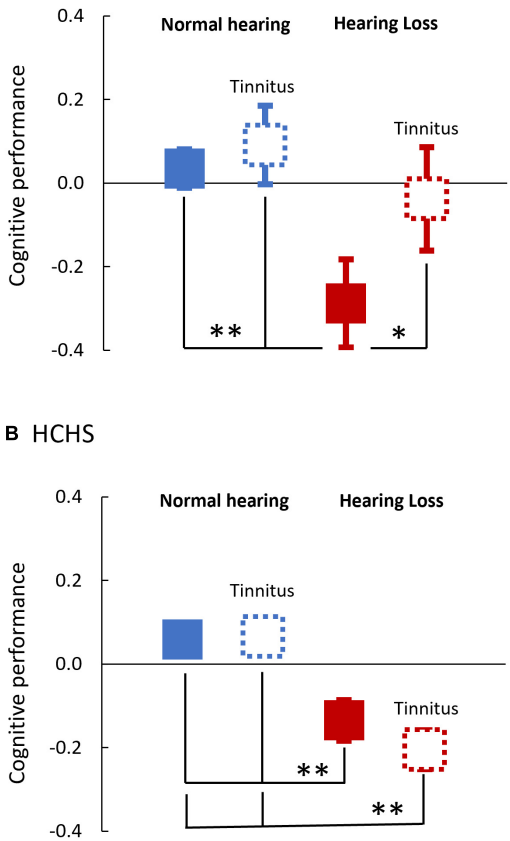


FIGURE 3 | Adjusted Mean Cognitive Performance of the Four Study Groups in both NHANES **(A)** and HCHS **(B)**. One-way ANOVAs with the *post hoc* Bonferroni test were used for between-groups comparison of the adjusted cognitive z-score. Boxes denote the adjusted mean and error bars denote the 95 CI%. Blue denotes normal hearing and red denotes hearing loss. Solid lines are of No tinnitus, and dashed lines are of tinnitus. *, **indicates a significant difference in the adjusted cognitive performance between groups (*$p < 0.05$; **$p < 0.001$).

RESULTS

Demographics and Descriptive Test Scores

Table 2 (NHANES) and **Table 3** (HCHS) show the demographic characteristics and test scores of the four groups stratified based on the hearing (normal hearing vs. hearing loss) and tinnitus (tinnitus vs. not tinnitus) status. In each of the NHANES and HCHS cohorts, there was little or no difference between the four stratified groups in the covariates (age, education, physical well-being score), except for gender. In the NHANES cohort, there were more females in the hearing loss group vs. more males in the normal hearing group. The HCHS cohort showed an opposite pattern.

The mean cognitive performance in the NHANES cohort was zero for the normal hearing and no tinnitus group, as well as

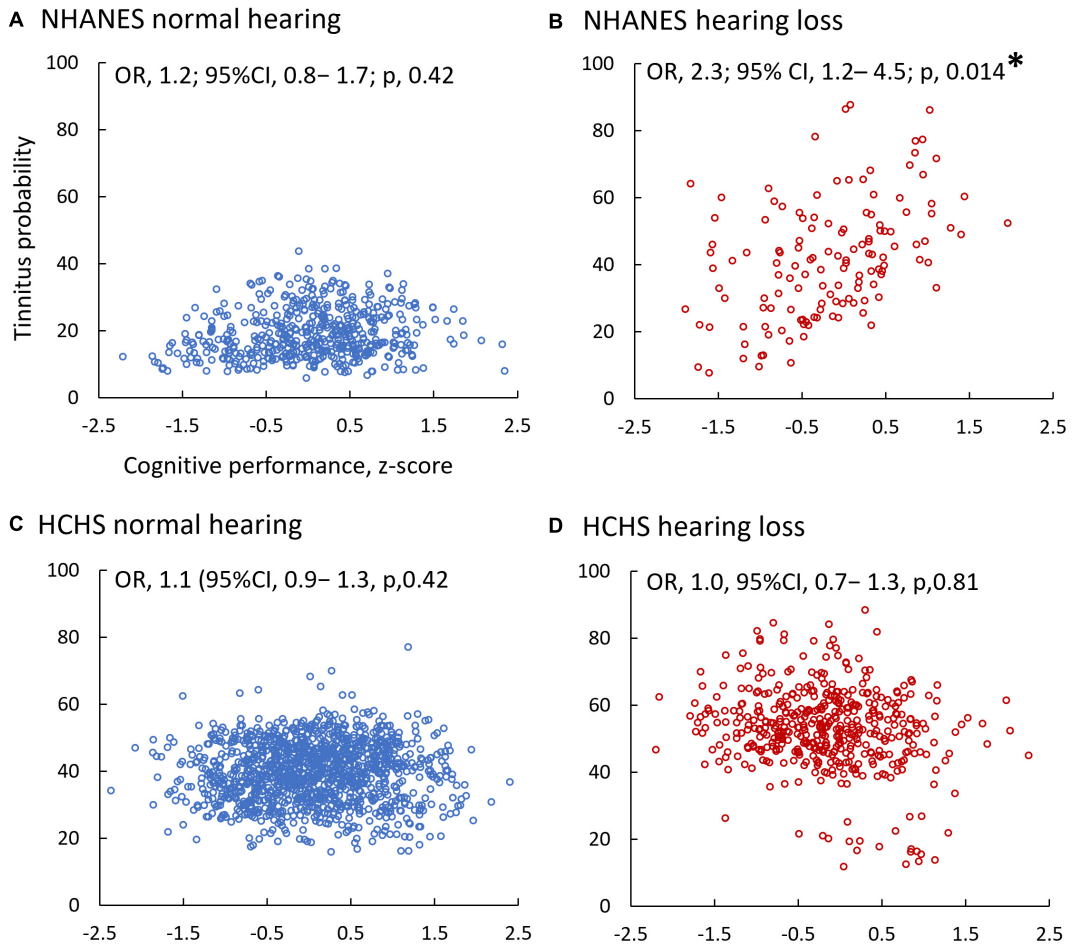


FIGURE 4 | Adjusted Tinnitus Probability as a Function of Cognitive Performance in the NHANES **(A,B)** and HCHS **(C,D)** Sub-Cohorts. Multivariable binary logistic regression models predicted the adjusted probability. Blue circles denote normal hearing **(A,C)**, and red circles are of hearing loss **(B,D)**. *indicates a significant association between tinnitus and cognition ($p < 0.05$).

the hearing loss and tinnitus group, 0.1 for the normal hearing and tinnitus group, but −0.3 for the hearing loss and no tinnitus group (bottom row in **Table 2**). In the HCHS cohort, regardless of tinnitus, the mean cognitive performance was 0.1 for the normal hearing group and −0.1 to −0.2 for the hearing loss group (bottom row in **Table 3**).

These trends in the mean performance can also be seen from the raw cumulative frequency distribution curves in **Figures 2A,B**. The x-axis value corresponding to the 50% cumulative frequency (dashed line) represents roughly the mean performance.

Tinnitus Correlates With Improved Cognition in Non-Hispanic Elderly With Hearing Loss

Figure 2C shows that while no association between tinnitus and cognition was observed in the normal hearing sub-cohort (β, 0.1; 95% CI, −0.1–0.2; *p*, 0.42), the hearing loss sub-cohort produced an unexpected result, namely, tinnitus was associated with better cognitive performance compared to no tinnitus (thick solid red line vs. dashed thin red line: β, 0.3; 95% CI, 0.05–0.5; *p*, 0.017). At the individual test level, only the CERAD word-learning test produced a significant association between tinnitus and improved cognitive performance for the hearing loss sub-cohort (β, 0.4; 95% CI, 0.02–0.7; *p*, 0.04). All the other individual tests showed no statistically significant association between tinnitus and cognition in the two sub-cohorts.

Figure 2D shows that within the HCHS cohort, tinnitus (dashed vs. solid line) was not associated with cognitive performance in the normal hearing (β, 0.0; 95% CI, 0.0–0.1; *p*, 0.42), nor the hearing loss (β, 0.0; 95% CI, −0.1–0.1; *p*, 0.83) sub-cohorts. The lack of a tinnitus effect on cognition in the HCHS cohort begs a question: Would the same result be obtained for the Hispanic sub-cohort in the NHANES database?

Figures 2E,F show the cumulative frequency distribution in the NHANES non-Hispanic ($n = 495$) and Hispanic ($n = 148$) sub-cohorts, respectively. In the non-Hispanic hearing loss sub-cohort ($n = 109$), tinnitus ($n = 41$) was significantly associated with better cognitive performance (β, 0.3; 95% CI, 0.1–0.5; *p*,

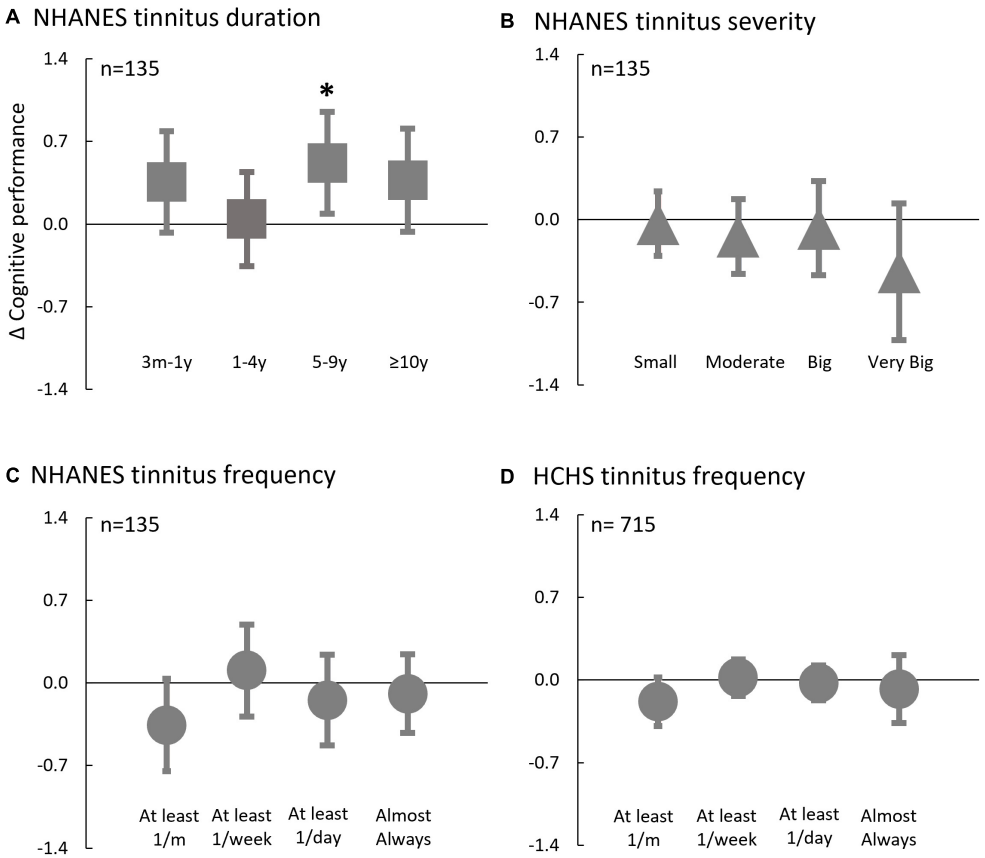


FIGURE 5 | Adjusted Difference in Cognitive Performance Associated with Tinnitus Attributes in NHANES **(A–C)** and HCHS **(D)**. Multivariable linear regression models were used in participants with tinnitus to predict the relative difference in cognitive performance to the reference groups based on tinnitus attributes. Reference groups: **(A)** Tinnitus Duration < 3 m; **(B)** Tinnitus Severity = No problem; **(C,D)** Tinnitus Frequency < 1/month; Error bars denote 95% CI. *indicates a significant association between tinnitus factor and cognition and the tinnitus-based category scoring statistically significantly different than the reference group ($p < 0.05$).

0.018), but this significant tinnitus effect disappeared, as expected from the HCHS result, in the NHANES Hispanic hearing loss sub-cohort (n = 26; β, 0.1; 95% CI, −0.5–0.8; p, 0.62). Normal hearing groups, Hispanic (n = 122) and non-Hispanic (n = 386), showed no association between tinnitus and cognition.

Figure 3 shows the adjusted cognitive performance of the four test groups. There was a significant main group effect on cognitive performance in both NHANES [F(3, 642), 11.8; $p < 0.001$, η2 = 0.1], and HCHS [F(3, 1714), 46.4; $p < 0.001$; η2 = 0.1]. *Post hoc* tests showed that for NHANES (**Figure 3A**), the hearing loss group and no tinnitus was the only group to score statistically significantly lower than the normal hearing groups ($p < 0.001$) and the group of hearing loss and tinnitus (p, 0.016). For the HCHS cohort (**Figure 3B**), both the hearing loss groups, with or without tinnitus, scored lower than the normal hearing groups ($p < 0.001$).

Tinnitus Decreases the Odds of Poor Cognitive Performance

Figure 4 shows the adjusted individual probability of tinnitus as a function of cognitive performance in the NHANES normal hearing (**Figure 4A**) and hearing loss (**Figure 4B**), and the HCHS normal hearing (**Figure 4C**) and hearing loss (**Figure 4D**) sub-cohorts. Three of the four sub-cohorts (**Figures 4A,C,D**) had an odds ratio close to 1 ($p > 0.05$), indicating no significant association between tinnitus and cognition. The exception was the NHANES hearing loss sub-cohort (**Figure 4B**), in which each unit increase in the z-score increased a participant's odds of having tinnitus by 2.3 times (OR, 2.3; 95% CI, 1.2–4.5; p, 0.014). At the individual test level, only the CERAD word-learning test showed a significant association, in which each unit increase in the z-score increased tinnitus odds by 1.6 times (OR, 1.6; 95% CI, 1.0–2.5; p, 0.03).

Using the 25th percentile threshold for poor cognitive performance, tinnitus status was not associated with poor cognitive performance in the HCHS normal hearing (OR, 0.9, 95% CI, 0.7–1.2; p, 0.56), hearing loss sub-cohorts (OR, 1.0, 95% CI, 0.7–1.2; p, 0.96), and the NHANES normal hearing sub-cohort (OR, 1.4; 95% CI, 0.7–2.6; p, 0.33). However, tinnitus status was a significant predictor in the NHANES hearing loss sub-cohort, in which a participant with tinnitus is 5.6 times less likely to have poor cognitive performance than a participant without tinnitus (OR, 5.6; 95% CI, 1.9–17.2; p, 0.002).

Tinnitus Duration Correlates With Cognitive Performance

Figure 5 shows the adjusted cognitive performance as a function of the tinnitus attributes in NHANES and HCHS. In the NHANES cohort, only tinnitus duration (**Figure 5A**) was associated with improved cognitive performance [$F(4,140)$, 2.6; p, 0.037]; relative to a participant with tinnitus duration less than 3 months, a participant with tinnitus duration of 5–9 years had better cognitive z-score by 0.5 (95% CI, 0.1–1.0; p, 0.02). Cognitive performance was not significantly associated with tinnitus severity [$F(4,136)$, 0.9; p, 0.47; **Figure 5B**] or frequency [$F(4,132)$, 1.5; p, 0.22; **Figure 5C**] in the NHANES, or with tinnitus frequency in the HCHS cohort [$F(4,703)$, 1.2; p, 0.31; **Figure 5D**].

DISCUSSION

Against the main hypothesis, we found no evidence that tinnitus is associated with poor cognition, and if anything, tinnitus is associated with improved cognitive performance in the NHANES (non-Hispanic) elderly population with hearing loss (**Figures 2C,E**, **4B**). Relative to participants without tinnitus in this population, tinnitus decreased the odds of having poor cognitive performance by 5.6 times. Tinnitus association with improved cognition was mainly with episodic memory (CERAD-word learning test), which is the main domain affected in mild cognitive impairment and Alzheimer's disease dementia (Albert et al., 2011). Sensitivity analysis showed that, among those with tinnitus, cognitive performance was improved with longer tinnitus duration.

Tinnitus Association With Better Cognition in the Elderly With Hearing Loss

The mechanisms responsible for hearing loss-associated cognitive decline are not well understood. This lack of understanding makes it difficult to delineate the mechanisms underlying tinnitus association with better cognition in age-related hearing loss (Whitson et al., 2018; Harris et al., 2019; Griffiths et al., 2020).

One possibility is tinnitus is a side effect of compensating for cortical changes in response to hearing loss that counteracts the increased central activities, and cortical tonotopic reorganization in response to hearing loss (Robertson and Irvine, 1989; Eggermont and Roberts, 2004). Koops et al. (2020a,b) showed that participants with hearing loss but no tinnitus, unlike those with tinnitus, had significantly different tonotopic maps, smaller gray matter volume, and thinner cortical surface both within and outside of the auditory pathway than controls. The presently observed association between longer tinnitus duration and better cognitive performance is consistent with the compensatory role of tinnitus in hearing loss-induced cortical plasticity. Another possibility is that tinnitus compensates for reduced auditory input due to hearing loss (Noreña, 2011; Zeng, 2020), which may, in turn, prevent auditory activity-related cognitive decline, for example, mnemonic memory (Gray et al., 2020; Veríssimo et al., 2021). This possibility is in line with the present finding that tinnitus improves episodic memory in individuals with hearing loss. A third possibility is that tinnitus is associated with less speech perception difficulties in those with hearing loss. While there is evidence for normal speech performance in individuals with tinnitus (Zeng et al., 2020), it is not known whether hearing loss individuals without tinnitus perform poorly in speech perception. If so speech perception difficulties may contribute to social disengagement that results in an accelerated cognitive decline (Bernabei et al., 2014). This third possibility would predict a global relationship between tinnitus and cognition.

Race Effect on Hearing Loss and Tinnitus

It is difficult to explain the race dependency of tinnitus compensation for hearing loss-associated cognitive decline. At present, few studies examined the race or ethnicity variations in hearing loss and tinnitus, producing inconsistent findings, especially related to the Hispanic community. For example, one report found that hearing loss is less common among Mexican Americans than non-Hispanic whites (Davanipour et al., 2000), another showed similar prevalence between the Hispanic and other populations (Cruickshanks et al., 2015). The present study showed a similar discrepancy that 21.0% of the NHANES and 26.4% of the HCHS participants had hearing loss, but 23.8% of the NHANES participants reported tinnitus, compared to 43.2% of the HCHS participants. Other studies have also reported that non-Hispanic whites have higher odds of frequent tinnitus compared with other racial/ethnic groups (Shargorodsky et al., 2010). Golub et al. (2020) found after adjusting for covariates there was no statistically significant association between PTA and cognition in the NHANES but a significant association in the HCHS. The inconsistent finding between the two databases can be partially accounted for by the present result that the hearing-impaired non-Hispanic population with tinnitus had better cognitive performance than that without tinnitus, which would reduce the overall effect of hearing loss on cognition in NHANES.

Limitations and Future Directions

There are two major limitations. First, this is a retrospective cross-sectional study with a narrow age range (60–69 years) and a relatively limited sample, where only an association rather than a causal relationship between tinnitus and cognition could be determined. Second, tinnitus characterization is based on a simple binary question on the absence or presence of tinnitus. Longitudinal data are needed to directly address whether and how tinnitus is associated with hearing loss-associated cognitive decline, and their interactions with age, sex, and race factors. Brain imaging and electrophysiological studies will likely shed light on mechanisms underlying both hearing loss and tinnitus, as well as their functional and structural relationship to dementia (Slade et al., 2020). Future studies need to investigate if tinnitus simulations in participants with hearing loss but no tinnitus might delay or even prevent dementia. Possibly, both actual and simulated tinnitus may counteract the neuroplastic changes in

response to auditory deprivation (Guitton, 2012). Additionally, future studies should include rehabilitation outcome measures that go beyond hearing and tinnitus assessment to include cognitive domains, particularly in the elderly (Naples et al., 2021). Finally, because the impact of hearing loss on cognition might be higher without tinnitus than with tinnitus, clinicians should pay special attention to individuals with hearing loss but no tinnitus to reduce the risk of cognitive decline.

Conclusion

The present study found that not only does tinnitus not aggregate hearing loss-related cognitive decline, but rather it is associated with better cognitive performance than those with hearing loss and no tinnitus, at least in the non-Hispanic elderly population. The present finding challenges the present assumption that tinnitus impairs cognitive function and provides interesting directions for future studies.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: NHANES Database from https://www.cdc.gov/nchs/nhanes/index.htm and HCHS-SOL Database from https://biolincc.nhlbi.nih.gov/home/.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

YH and F-GZ contributed to the concept and design, acquisition, analysis, and interpretation of data. YH contributed to the statistical analysis and drafted the manuscript. F-GZ supervised and critically revised the manuscript and obtained the funding. Both authors contributed to the article and approved the submitted version.

FUNDING

This study was supported in part by NIH NIDCD (3R01DC015587), NIA (3R01DC015587-Supplements) and the UC Irvine Institute for Memory Impairments and Neurological Disorders.

ACKNOWLEDGMENTS

Parts of the manuscript were presented at the UCI MIND 11th Annual Emerging Scientists Symposium, Irvine, CA, United States- February 2020. This manuscript was prepared using HCHSSOL Research Materials obtained from the NHLBI Biologic Specimen and Data Repository Information Coordinating Center and does not necessarily reflect the opinions or views of the HCHSSOL or the NHLBI. We thank the reviewers DH and JB for their helpful comments on the manuscript.

REFERENCES

Albert, M. S., DeKosky, S. T., Dickson, D., Dubois, B., Feldman, H. H., Fox, N. C., et al. (2011). The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement.* 7, 270–279.

American Psychiatric Association (1980). *Diagnostic and Statistical Manual of Mental Disorders*, Vol. 3. Washington, DC: American Psychiatric Association.

Baguley, D., McFerran, D., and Hall, D. (2013). Tinnitus. *Lancet* 382, 1600–1607. doi: 10.1016/S0140-6736(13)60142-7

Bernabei, R., Bonuccelli, U., Maggi, S., Marengoni, A., Martini, A., Memo, M., et al. (2014). Hearing loss and cognitive decline in older adults: questions and answers. *Aging Clin. Exp. Res.* 26, 567–573. doi: 10.1007/s40520-014-0266-3

Clarke, N. A., Henshaw, H., Akeroyd, M. A., Adams, B., and Hoare, D. J. (2020). Associations between subjective tinnitus and cognitive performance: Systematic review and meta-analyses. *Trends Hear.* 24:2331216520918416. doi: 10.1177/2331216520918416

Chu, H. T., Liang, C. S., Yeh, T. C., Hu, L. Y., Yang, A. C., Tsai, S. J., et al. (2020). Tinnitus and risk of Alzheimer's and Parkinson's disease: a retrospective nationwide population-based cohort study. *Sci. Rep.* 10:12134. doi: 10.1038/s41598-020-69243-0

Cruickshanks, K. J., Dhar, S., Dinces, E., Fifer, R. C., Gonzalez, F., Heiss, G., et al. (2015). Hearing impairment prevalence and associated risk factors in the Hispanic Community Health Study/Study of Latinos. *JAMA Otolaryngol. Head Neck Surg.* 141, 641–648. doi: 10.1001/jamaoto.2015.0889

Davanipour, Z., Lu, N. M., Lichtenstein, M., and Markides, K. S. (2000). Hearing problems in Mexican American elderly. *Otol. Neurotol.* 21, 168–172. doi: 10.1016/s0196-0709(00)80004-6

Davis, A., McMahon, C. M., Pichora-Fuller, K. M., Russ, S., Lin, F., Olusanya, B. O., et al. (2016). Aging and hearing health: the life-course approach. *Gerontol.* 56(Suppl. 2), S256–S267.

Deal, J. A., Betz, J., Yaffe, K., Harris, T., Purchase-Helzner, E., Satterfield, S., et al. (2017). Hearing impairment and incident dementia and cognitive decline in older adults: the health ABC study. *J. Gerontol. Ser. A* 72, 703–709.

Deal, J. A., Sharrett, A. R., Albert, M. S., Coresh, J., Mosley, T. H., Knopman, D., et al. (2015). Hearing impairment and cognitive decline: a pilot study conducted within the atherosclerosis risk in communities neurocognitive study. *Am. J. Epidemiol.* 181, 680–690. doi: 10.1093/aje/kwu333

Eggermont, J., and Roberts, L. (2004). The neuroscience of tinnitus. *Trends Neurosci.* 27, 676–682.

Golub, J. S., Brickman, A. M., Ciarleglio, A. J., Schupf, N., and Luchsinger, J. A. (2020). Association of subclinical hearing loss with cognitive performance. *JAMA Otolaryngol. Head Neck Surg.* 146, 57–67. doi: 10.1001/jamaoto.2019.3375

Golub, J. S., Luchsinger, J. A., Manly, J. J., Stern, Y., Mayeux, R., and Schupf, N. (2017). Observed hearing loss and incident dementia in a multiethnic cohort. *J. Am. Geriatr. Soc.* 65, 1691–1697. doi: 10.1111/jgs.14848

González, H. M., Mungas, D., Reed, B. R., Marshall, S., and Haan, M. N. (2001). A new verbal learning and memory test for English-and Spanish-speaking older people. *J. Int. Neuropsychol. Soc. JINS* 7:544. doi: 10.1017/s1355617701755026

Gray, D. T., Umapathy, L., De La Pena, N. M., Burke, S. N., Engle, J. R., Trouard, T. P., et al. (2020). Auditory processing deficits are selectively associated with

medial temporal lobe mnemonic function and white matter integrity in aging macaques. *Cereb. Cortex* 30, 2789–2803. doi: 10.1093/cercor/bhz275

Griffiths, T. D., Lad, M., Kumar, S., Holmes, E., McMurray, B., Maguire, E. A., et al. (2020). How can hearing loss cause dementia? *Neuron* 108, 401–412. doi: 10.1016/j.neuron.2020.08.003

Guitton, M. J. (2012). Tinnitus: pathology of synaptic plasticity at the cellular and system levels. *Front. Syst. Neurosci.* 6:12. doi: 10.3389/fnsys.2012.00012

Hallam, R. S., McKenna, L., and Shurlock, L. (2004). Tinnitus impairs cognitive efficiency. *Int. . Audiol.* 43, 218–226. doi: 10.1080/14992020400050030

Harris, M. S., Doerfer, K., and Moberly, A. C. (2019). Discussing age-related hearing loss and cognitive decline with patients. *JAMA Otolaryngol Head Neck Surg.* 145, 781–782. doi: 10.1001/jamaoto.2019.1667

IBM_Corp (2019). *IBM SPSS Statistics for Windows, Version 20.0*. Armonk, NY: IBM Corp.

Jafari, Z., Kolb, B. E., and Mohajerani, M. H. (2019). Age-related hearing loss and tinnitus, dementia risk, and auditory amplification outcomes. *Ageing Res. Rev.* 56:100963. doi: 10.1016/j.arr.2019.100963

Koops, E. A., de Kleine, E., and van Dijk, P. (2020a). Gray matter declines with age and hearing loss, but is partially maintained in tinnitus. *Sci. Rep.* 10:21801. doi: 10.1038/s41598-020-78571-0

Koops, E. A., Renken, R. J., Lanting, C. P., and van Dijk, P. (2020b). Cortical tonotopic map changes in humans are larger in hearing loss than in additional tinnitus. *J. Neurosci.* 40, 3178–3185. doi: 10.1523/jneurosci.2083-19.2020

Lee, H. Y. (2020). Beyond hearing loss: does tinnitus cause cognitive impairment? *Clin. Exp. Otorhinolaryngol.* 13, 2–3. doi: 10.21053/ceo.2019.01949

Lezak, M. D., Howieson, D. B., Loring, D. W., and Fischer, J. S. (2004). *Neuropsychological Assessment*. New York, NY: Oxford University Press.

Lin, F. R. (2011). Hearing loss and cognition among older adults in the United States. *J. Gerontol. A Biol. Sci. Med. Sci.* 66, 1131–1136. doi: 10.1093/gerona/glr115

Lin, F. R., Ferrucci, L., Metter, E. J., An, Y., Zonderman, A. B., and Resnick, S. M. (2011a). Hearing loss and cognition in the baltimore longitudinal study of aging. *Neuropsychology* 25:763. doi: 10.1037/a0024238

Lin, F. R., Metter, E. J., O'Brien, R. J., Resnick, S. M., Zonderman, A. B., and Ferrucci, L. (2011b). Hearing loss and incident dementia. *Arch. Neurol.* 68, 214–220.

Livingston, G., Huntley, J., Sommerlad, A., Ames, D., Ballard, C., Banerjee, S., et al. (2020). Dementia prevention, intervention, and care: 2020 report of the lancet commission. *Lancet* 396, 413–446. doi: 10.1016/s0140-6736(20)30367-6

Lockwood, A. H., Salvi, R. J., and Burkard, R. F. (2002). Tinnitus. *N. Engl. J. Med.* 347, 904–910. doi: 10.1056/NEJMra013395347/12/904

McCormack, A., Edmondson-Jones, M., Somerset, S., and Hall, D. (2016). A systematic review of the reporting of tinnitus prevalence and severity. *Hear. Res.* 337, 70–79. doi: 10.1016/j.heares.2016.05.009

Mistridis, P., Egli, S. C., Iverson, G. L., Berres, M., Willmes, K., Welsh-Bohmer, K. A., et al. (2015). Considering the base rates of low performance in cognitively healthy older adults improves the accuracy to identify neurocognitive impairment with the consortium to establish a registry for Alzheimer's disease-neuropsychological assessment battery (CERAD-NAB). *Eur. Arch. Psychiatry Clin. Neurosci.* 265, 407–417. doi: 10.1007/s00406-014-0571-z

Mohamad, N., Hoare, D. J., and Hall, D. A. (2016). The consequences of tinnitus and tinnitus severity on cognition: a review of the behavioural evidence. *Hear. Res.* 332, 199–209. doi: 10.1016/j.heares.2015.10.001

Morris, J., Heyman, A., Mohs, R., Hughes, J., van Belle, G., and Fillenbaum, G. (1989). The Consortium to establish a registry for Alzheimer's Disease (CERAD). Part 1. Clinical and neuropsychological assessment of Alzheimer's disease. *Neurology* 39, 1159–1165. doi: 10.1212/wnl.39.9.1159

Naples, J. G., Castellanos, I., and Moberly, A. C. (2021). Considerations for integrating cognitive testing into adult cochlear implant evaluations-foundations for the future. *JAMA Otolaryngol. Head. Neck Surg.* 147, 413–414. doi: 10.1001/jamaoto.2020.5487

Nash, S. D., Cruickshanks, K. J., Klein, R., Klein, B. E., Nieto, F. J., Huang, G. H., et al. (2011). The prevalence of hearing impairment and associated risk factors: the beaver dam offspring study. *Arch. Otolaryngol. Head. Neck Surg.* 137, 432–439. doi: 10.1001/archoto.2011.15

Nondahl, D. M., Cruickshanks, K. J., Wiley, T. L., Klein, B. E., Klein, R., Chappell, R., et al. (2010). The ten-year incidence of tinnitus among older adults. *Int. J. Audiol.* 49, 580–585. doi: 10.3109/14992021003753508

Noreña, A. J. (2011). An integrative model of tinnitus based on a central gain controlling neural sensitivity. *Neurosci. Biobehav. Rev.* 35, 1089–1109. doi: 10.1016/j.neubiorev.2010.11.003

Palumbo, R., Di Domenico, A., Piras, F., Bazzano, S., Zerilli, M., Lorico, F., et al. (2020). Measuring global functioning in older adults with cognitive impairments using the Rasch model. *BMC Geriatrics* 20:492. doi: 10.1186/s12877-020-01886-0

Robertson, D., and Irvine, D. R. (1989). Plasticity of frequency organization in auditory cortex of guinea pigs with partial unilateral deafness. *J. Comp. Neurol.* 282, 456–471. doi: 10.1002/cne.902820311

Savastano, M. (2008). Tinnitus with or without hearing loss: are its characteristics different? *Eur. Arch. OtoRhino Laryngol.* 265, 1295–1300. doi: 10.1007/s00405-008-0630-z

Shargorodsky, J., Curhan, G. C., and Farwell, W. R. (2010). Prevalence and characteristics of tinnitus among US adults. *Am. J. Med.* 123, 711–718. doi: 10.1016/j.amjmed.2010.02.015

Slade, K., Plack, C. J., and Nuttall, H. E. (2020). The effects of age-related hearing loss on the brain and cognitive function. *Trends Neurosci.* 43, 810–821. doi: 10.1016/j.tins.2020.07.005

Spiegel, R., Kalla, R., Mantokoudis, G., Maire, R., Mueller, H., Hoerr, R., et al. (2018). Ginkgo biloba extract EGb 761((R)) alleviates neurosensory symptoms in patients with dementia: a meta-analysis of treatment effects on tinnitus and dizziness in randomized, placebo-controlled trials. *Clin. Interv. Aging* 13, 1121–1127. doi: 10.2147/CIA.S157877

Spitzer, R. L., Kroenke, K., Williams, J. B., and Group PHQPCS (1999). Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. *JAMA* 282, 1737–1744.

Spreen, O., and Strauss, E. (1998). *A Compendium of Neuropsychological Tests*, 2nd Edn. New York, NY: Oxford University Press.

Tegg-Quinn, S., Bennett, R. J., Eikelboom, R. H., and Baguley, D. M. (2016). The impact of tinnitus upon cognition in adults: a systematic review. *Int. J. Audiol.* 55, 533–540. doi: 10.1080/14992027.2016.1185168

Veríssimo, J., Verhaeghen, P., Goldman, N., Weinstein, M., and Ullman, M. T. (2021). Evidence that ageing yields improvements as well as declines across attention and executive functions. *Nat. Hum. Behav.* 1–14. doi: 10.1038/s41562-021-01169-7

Wechsler, D. (1997). *WAIS Manual*, 3rd Edn. San Antonio, TX: Psychological Corporation.

Whitson, H. E., Cronin-Golomb, A., Cruickshanks, K. J., Gilmore, G. C., Owsley, C., Peelle, J. E., et al. (2018). American geriatrics society and national institute on aging bench-to-bedside conference: sensory impairment and cognitive decline in older adults. *J. Am. Geriatr. Soc.* 66, 2052–2058. doi: 10.1111/jgs.15506

World Health Organization (WHO) (2008). *Grades of Hearing Impairment*. Geneva: WHO.

Zeng, F. G. (2020). Tinnitus and hyperacusis: central noise, gain and variance. *Curr. Opin. Physiol.* 18, 23–129.

Zeng, F.-G., Richardson, M., and Turner, K. (2020). Tinnitus does not interfere with auditory and speech perception. *J. Neurosci.* 40, 6007–6017. doi: 10.1523/jneurosci.0396-20.2020

Conflict of Interest: F-GZ owns stock in Axonics, DiaNavi, Nurotron, Syntiant, Velox, and Xense.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Psychology

ORIGINAL RESEARCH
published: 15 November 2021
doi: 10.3389/fpsyg.2021.733100

Effect of the Target and Conflicting Frequency and Time Ranges on Consonant Enhancement in Normal-Hearing Listeners

*Yang-Soo Yoon**

Laboratory of Translational Auditory Research, Department of Communication Sciences and Disorders, Baylor University, Waco, TX, United States

In this paper, the effects of intensifying useful frequency and time regions (target frequency and time ranges) and the removal of detrimental frequency and time regions (conflicting frequency and time ranges) for consonant enhancement were determined. Thirteen normal-hearing (NH) listeners participated in two experiments. In the first experiment, the target and conflicting frequency and time ranges for each consonant were identified under a quiet, dichotic listening condition by analyzing consonant confusion matrices. The target frequency range was defined as the frequency range that provided the highest performance and was decreased 40% from the peak performance from both high-pass filtering (HPF) and low-pass filtering (LPF) schemes. The conflicting frequency range was defined as the frequency range that yielded the peak errors of the most confused consonants and was 20% less than the peak error from both filtering schemes. The target time range was defined as a consonant segment that provided the highest performance and was decreased 40% from that peak performance when the duration of the consonant was systematically truncated from the onset. The conflicting time ranges were defined on the coincided target time range because, if they temporarily coincide, the conflicting frequency ranges would be the most detrimental factor affecting the target frequency ranges. In the second experiment, consonant recognition was binaurally measured in noise under three signal processing conditions: unprocessed, intensified target ranges by a 6-dB gain (target), and combined intensified target and removed conflicting ranges (target-conflicting). The results showed that consonant recognition improved significantly with the target condition but greatly deteriorated with a target-conflicting condition. The target condition helped transmit voicing and manner cues while the target-conflicting condition limited the transmission of these cues. Confusion analyses showed that the effect of the signal processing on consonant improvement was consonant-specific: the unprocessed condition was the best for /da, pa, ma, sa/; the target condition was the best for /ga, fa, va, za, ʒa/; and the

OPEN ACCESS

Edited by:
Achim Schilling,
University Hospital Erlangen, Germany

Reviewed by:
Vijaya Kumar Narne,
All India Institute of Speech and Hearing (AIISH), India
Jennifer Lentz,
Indiana University Bloomington, United States

***Correspondence:**
Yang-Soo Yoon
yang-soo_yoon@baylor.edu

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Psychology

Received: *29 June 2021*
Accepted: *11 October 2021*
Published: *15 November 2021*

Citation:
Yoon YS (2021) Effect of the Target and Conflicting Frequency and Time Ranges on Consonant Enhancement in Normal-Hearing Listeners. Front. Psychol. 12:733100. doi: 10.3389/fpsyg.2021.733100

target-conflicting condition was the best for /na, ʃa/. Perception of /ba, ta, ka/ was independent of the signal processing. The results suggest that enhancing the target ranges is an efficient way to improve consonant recognition while the removal of conflicting ranges negatively impacts consonant recognition.

Keywords: spectral cues, temporal cues, articulation-index gram, conflicting ranges, target ranges, consonant recognition

INTRODUCTION

Consonant recognition depends on the listener's ability to discriminate details of spectral and temporal acoustic cues such as voicing, an onset of the noise burst, and spectral and temporal transitions (Miller and Nicely, 1955; Stevens and Klatt, 1974; Stevens and Blumstein, 1978; Blumstein and Stevens, 1979, 1980). Many classic studies used synthetic consonants, which require prior knowledge of the spectral and temporal cues that are critical for perception (Hughes and Halle, 1956; Heinz and Stevens, 1961; Blumstein et al., 1977; Stevens and Blumstein, 1978). Other studies used naturally produced consonants and identified the spectral and temporal cues for recognition (Soli, 1981; Baum and Blumstein, 1987; Behrens and Blumstein, 1988; Jongman et al., 2000). While the results obtained from these studies help characterize spectral and temporal cues and their variability, there are limited studies available that utilize these identified cues for an enhancement of consonant recognition.

To identify spectral and temporal cues for naturally produced consonants, Allen et al. collected consonant confusion matrices as a function of cutoff frequency for both low-pass filtering (LPF) and high-pass filtering (HPF) schemes, time truncation from an onset of consonants, and signal-to-noise ratio (SNR) in normal-hearing (NH) listeners (Phatak and Allen, 2007; Phatak et al., 2008; Li and Allen, 2011; Li et al., 2012)Phatak, Lovitt, and Allen. Through the analysis of confusion matrices, they were able to identify frequency and time ranges for each consonant, which resulted in a significant positive change in recognition and labeled them as "target frequency and time ranges." They also noticed specific frequency and time ranges that produced a significant negative change in recognition called "conflicting frequency and time ranges." In phonetics, a "conflicting" cue is an acoustic property that is phonetically inconsistent with another acoustic property in the same utterance (e.g., a stop consonant with a long voice onset time (VOT) but low fundamental frequency onset in the following vowel or *vice versa*). In this article, "conflicting" frequency and time ranges were defined as the ranges that generate more consonant confusion than enhancement.

To enhance consonant recognition with the target and conflicting frequency and time ranges, Allen et al. used a novel signal processing tool called the Articulation Index-Gram (AI-Gram) (Li and Allen, 2011; Li et al., 2012). The AI-Gram comprises the combined use of the articulation index model (French and Steinberg, 1947; Allen, 2005) and the linear peripheral cochlear model (Li et al., 2010). To determine whether consonant recognition was affected by the target and conflicting frequency and time ranges, they conducted two pilot studies with NH listeners. Two stop consonants (/ka, ga/) were tested in noise with a 6- and 12-dB gain on the target frequency and time ranges and complete removal of the conflicting frequency and time ranges (Li and Allen, 2011). In another study, they tested four stop consonants (/ta, ka, da, ga/) in noise with a 6-dB gain, 6-dB attenuation, complete removal of the target ranges, and an unprocessed control condition (Kapoor and Allen, 2012). The results obtained from these two studies indicated that the additional gain on the target ranges and the removal of conflicting ranges enhances the consonant perception by 10–70%. These findings led to our prediction that a greater enhancement for other consonants can be achieved if the target ranges are intensified while conflicting ranges are removed.

One challenging aspect of the two pilot studies conducted by Li and Allen (2011) and Kapoor and Allen (2012) is that their analyses were primarily based on the subjects' responses to consonant syllables produced by a few selective (good) talkers, which resulted in clearly spoken stimulus tokens. The use of the clearly spoken tokens may contribute to a higher benefit (min. 10% to max. 70%) of the AI-Gram processing in consonant recognition. It is known that clearly spoken consonants yield intelligibility advantages of 3–38% points relative to normal conversational speech for NH listeners in noise or reverberation (Helfer, 1997; Ferguson and Kewley-Port, 2002; Ferguson, 2004). Underlying the reasons for the intelligibility advantages of clearly spoken consonants include enhanced acoustic cues such as the following: overall longer durations, longer VOT for voiceless stops, and increased consonant–vowel amplitude ratio (CVR) for stops and some fricatives (Chen, 1980; Picheny et al., 1986). Thus, it is unclear whether the significant consonant enhancement reported in Allen's studies is due to AI-Gram processing on the target and conflicting cues or the combined effects with the use of the clearly spoken consonants produced by highly selective talkers. Another factor in Allen's pilot studies (Li and Allen, 2011; Kapoor and Allen, 2012) to further consider is that, for each consonant, the averaged target and conflicting ranges over the NH subjects were used. By using the averaged target and conflicting frequency and time ranges, the potential for an intersubject variation in the target and conflicting ranges is not accounted. An intersubject variation in the target and/or conflicting ranges might not be a major issue for NH listeners but should be a critical factor for listeners with different configurations and degrees of hearing loss (Revoile et al., 1982; Ferguson and Kewley-Port, 2002).

In summary, previous studies demonstrate some efforts to determine the effect of the AI-Gram processing on consonant enhancement for four stop consonants. However, a more comprehensive assessment that includes the control of important confounding variables (conversationally spoken consonants and

the use of tailored target and conflicting ranges) is warranted. In this study, using conversationally produced consonants by a single female talker and the target and conflicting frequency and time ranges identified from each NH subject, two experiments with NH listeners were conducted to determine how both the target and conflicting frequency and time ranges affected consonant enhancement when combined. The ideal approach for both experiments is to make direct comparisons between the results gained with conversationally and clearly spoken stimulus tokens, along with acoustical differences between the speech samples. Instead, extensive comparisons were made between the current study and Allen et al.'s studies (Li and Allen, 2011; Kapoor and Allen, 2012) in the "Discussion" section as the same AI-Gram was used to intensify and remove the target and conflicting ranges of the similar sets of consonants. In the first experiment, the target and conflicting frequency and time ranges for each consonant were identified on an individual listener basis. In the second experiment, the effects of the AI-Gram processing on the target and conflicting frequency and time ranges on consonant enhancement were determined.

EXPERIMENT 1: IDENTIFY TARGET AND CONFLICTING FREQUENCY AND TIME RANGES

Materials and Methods

Subject

Thirteen NH adults (seven women and six men; aged 19–43 with an average age of 28 years old) participated. All participants were native American English speakers and had thresholds better than a 20-dB hearing level at audiometric frequencies ranging from 0.25 to 8 kHz. All participants had interaural thresholds less than a 10-dB hearing level. All subjects provided informed consent, and all procedures were approved by the Texas Tech University Health Sciences Center Institution Review Board.

Stimuli

Stimuli included 14 frequently used consonants with the common vowel /a/ (/pa, ba, ta, da, ka, ga, ma, na, fa, va, sa, za, ʃa, ʒa/) in American English (Hayden, 1950). This study chose to test these consonants so that the results of this study could be directly compared to the results found in Allen's studies (Li and Allen, 2011; Kapoor and Allen, 2012). To obtain conversationally spoken stimuli, each consonant syllable was produced by a single female with three different speaking efforts: minimum, medium, and maximum. For the minimum speaking effort, the talker was instructed to speak as if she was speaking to one NH listener in a quiet room. For the medium speaking effort, she was instructed to speak as she would to an NH listener in an everyday conversation, and for the maximum speaking effort, she was instructed to speak as if she was talking to a person with a hearing loss (Ferguson and Morgan, 2018). There were 42 sound files (14 consonants × 3 speaking efforts). All sound files were resampled from their original recorded sampling rate of 22,050–44,100 Hz, which is the standard for most consumer audio, and then normalized to have the same long-term root mean square energy [65 dBA sound pressure level (SPL)].

To choose one "conversationally spoken" token per consonant syllable, five lab members (all female students with an average age of 24 ± 1.6 years) were asked to rate how clearly the sound was spoken. They had normal hearing, which was verified through a pure-tone audiometry test. Each of the 42 sound files was randomly presented five times in quiet *via* headphones, and each lab member was asked to rate "how clearly the sound was spoken" on a scale from 1 to 7: 1—lowest possible clarity, 2—very unclear, 3—somewhat unclear, 4—midway, 5—somewhat clear, 6—very clear, and 7—highest possible clarity (Ferguson and Morgan, 2018). The lab members made their selections by clicking on the desired category from the graphical user interface and then pressing the "next" key to continue. The lab members were instructed to use the whole 1–7 scale and to focus on how clearly the sound was spoken with reference to the sounds they heard that day instead of speech heard in the past. For each consonant syllable, 75 ratings were collected (i.e., 3 different speaking efforts × 5 repetitions × 5 lab members). The sound file with the median rating was chosen as a "conversationally spoken" syllable. It turned out that all conversationally spoken stimuli used for the experiments resulted in consonant phonemes recorded with medium speaking efforts. All other tokens were not used for the experiment. The acoustic analyses on the final conversationally spoken stimuli showed an average fundamental frequency of 228 Hz. Complete silent parts were manually removed, which were identified by looking at waveforms and spectrograms, from both the onset and offset of the consonant syllables. Each processed consonant syllable was presented five times in quiet to all the five lab members. The processed consonant syllables were accepted as stimuli if they were perceived at a level of 99% correct, averaged over five presentations. The average duration plus SD of consonants was 406.57 ± 102.61 ms. The duration of each consonant is given in **Table 1**.

Identifying Target and Conflicting Frequency Ranges

For the identification of the target and conflicting frequency ranges, consonant confusion matrices were first collected and each of the 14 consonant syllables in the matrices was diotically presented in quiet to both ears *via* circumaural headphones (Sennheiser HDA-200, OLd Lyme, CT, United States). All stimuli were presented at the most comfortable level (MCL, range: 50–70 dB SPL). To determine the listener's MCLs, each subject was asked to rate the loudness of each of the 14 unprocessed consonants in quiet according to the Cox loudness rating scale (Cox et al., 1997). The MCL was the mean of dB SPLs, which were rated "comfortable." The confusion matrices were collected under both HPF and LPF. Each subject responded by pressing 1 of the 14 response buttons labeled in a consonant- /a/ context on a computer screen. To acquire reliable confusion matrices for each filtering scheme, each consonant was presented five times. For each presentation, the same consonant was presented sequentially, and each of these five presentations was only considered correct if the consonant presented was selected three times in a row. The order of

TABLE 1 | The lower and upper ranges of the target and conflicting frequency and time.

Consonants (duration in ms)	Onset of vowel (ms)	Target frequency (kHz)		Conflicting frequency (kHz)		Target time (ms)	
		Lower ranges	Upper ranges	Lower ranges	Upper ranges	Lower ranges	Upper ranges
/pa/ (240)	59	0.3–0.8	7.1–7.4	1.1–1.7	2–2.2	8–18	40–60
		SD+: 4, SD−: 1		SD+: 2, SD−: 4		SD+: 3, SD−: 2	
/ba/ (331)	32	0.3–1.1	4–4.5	0.6–1.2	1.8–2.5	8–15	25–30
		SD+: 4, SD−: 2		SD+: 4, SD−: 2		SD+: 3, SD−: 3	
/ta/ (338)	96	3–3.7	5–7.4	1.3–1.7	2.2–2.8	26–50	42–70
		SD+: 2, SD−: 3		SD+: 3, SD−: 3		SD+: 2, SD−: 2	
/da/ (240)	43	3–4.1	6–7.8	1.1–1.7	2.3–2.8	15–24	25–31
		SD+: 3, SD−: 3		SD+: 2, SD−: 3		SD+: 2, SD−: 3	
/ka/ (447)	100	0.9–1.4	2–2.5	5–5.7	7.2–7.8	50–65	70–80
		SD+: 4, SD−: 4		SD+: 3, SD−: 4		SD+: 3, SD−: 4	
/ga/ (348)	52	1–1.5	1.6–2	3.4–4	4.8–5.2	21–31	45–52
		SD+: 1, SD−: 2		SD+: 5, SD−: 2		SD+: 5, SD−: 2	
/ma/ (350)	112	0.3–0.6	0.8–1.3	1.2–1.6	2–2.5	38–51	65–85
		SD+: 3, SD−: 5		SD+: 3, SD−: 1		SD+: 1, SD−: 1	
/na/ (400)	107	1–1.8	1.8–2.4	0.4–0.6	0.8–1.3	22–30	71–80
		SD+: 3, SD−: 2		SD+: 3, SD−: 4		SD+: 2, SD−: 3	
/fa/ (548)	180	0.4–0.7	2–2.7	3–3.6	7.4–7.8	45–65	82–90
		SD+: 4, SD−: 3		SD+: 2, SD−: 4		SD+: 3, SD−: 3	
/va/ (349)	88	0.4–0.7	1.2–1.7	1.2–1.7	4.2–4.8	25–49	50–70
		SD+: 2, SD−: 4		SD+: 3, SD−: 3		SD+: 2, SD−: 3	
/sa/ (501)	202	3.9–5.8	7.4–7.8	3–3.6	4.5–5.5	70–100	148–175
		SD+: 4, SD−: 2		SD+: 2, SD−: 3		SD+: 2, SD−: 2	
/za/ (501)	197	3.1–4.6	7.2–7.8	3–3.8	4.2–5.4	40–55	75–95
		SD+: 2, SD−: 3		SD+: 2, SD−: 4		SD+: 2, SD−: 2	
/ʃa/ (549)	238	1.9–2.5	2.9–4.2	3.9–4.5	7.2–7.8	80–110	180–225
		SD+: 3, SD−: 3		SD+: 4, SD−: 3		SD+: 3 SD−: 2	
/ʒa/ (550)	260	1.4–2.6	2.9–3.9	5–5.6	7–7.8	30–45	65–80
		SD+: 1, SD−: 3		SD+: 2, SD−: 3		SD+: 2, SD−: 3	

The number of subjects with one SD above and below the mean of the target and conflicting ranges is indicated by SD+ and SD−, respectively. The duration of each consonant and an onset of the vowel /a/ from the beginning of each consonant are also given. The target time ranges indicate the consonant temporal duration from an onset of the vowel /a/.

which the consonants were presented was randomized over the five presentations.

A reason to measure confusion matrices with both LPF and HPF is that the low and high frequencies may affect consonant perception in different ways (Miller and Nicely, 1955). For HPF, 7 kHz was used as an initial cutoff frequency to exclude mid-to-high spectral information (i.e., 3–7 kHz) known to be useful for some fricative consonant perception (Jongman et al., 2000). For LPF, 0.1 kHz was used as an initial cutoff frequency to include minimal spectral information for consonant perception (Dubno and Levitt, 1981). The step size for the cutoff frequency of both filtering schemes was 0.1 kHz. When a participant's response was incorrect, the cutoff of the HPF was decreased by 0.1 kHz (i.e., from 7 to 6.9 kHz), and the cutoff of the LPF was increased by 0.1 kHz (i.e., from 0.1 to 0.2 kHz). The target frequency range was defined as the frequency range, in kHz, that provided the highest performance and was decreased 40% from that peak performance from both LPF and HPF conditions. For example, the recognition of /ʃa/ reaches 60% correct when the LPF cutoff is 2.1 kHz and reaches a maximum of 100% correct when the cutoff is moved from 2.1 to 2.2 kHz. So, the lower edge of the target frequency would be 2.1 kHz. When the HPF cutoff is 3.5 kHz, the recognition of /ʃa/ reaches a score of 60% correct and a maximum of 100% correct when the cutoff is moved from 3.5 to 3.4 kHz. So, the upper edge of the target frequency would be 3.5 kHz. Therefore, the final target frequency range would be 2.1–3.5 kHz (see the middle panel of **Figure 1**). About 20, 40, 60, and 80% below the peak performance were tested with five NH listeners. Decreasing 40% from the peak performance provided the highest benefit when the frequency range was boosted by 6 dB. **Figure 1** displays the spectrograms of the unprocessed (left panel), intensified target frequency and time ranges by a 6-dB gain (middle panel), and the combined intensified target ranges with the removal of conflicting ranges (right panel) for /ʃa/. Squares indicate the target frequency and time ranges. Dotted vertical lines indicate an onset of the vowel /a/.

To identify the conflicting frequency ranges, confusing consonants that were consistently perceived two times above

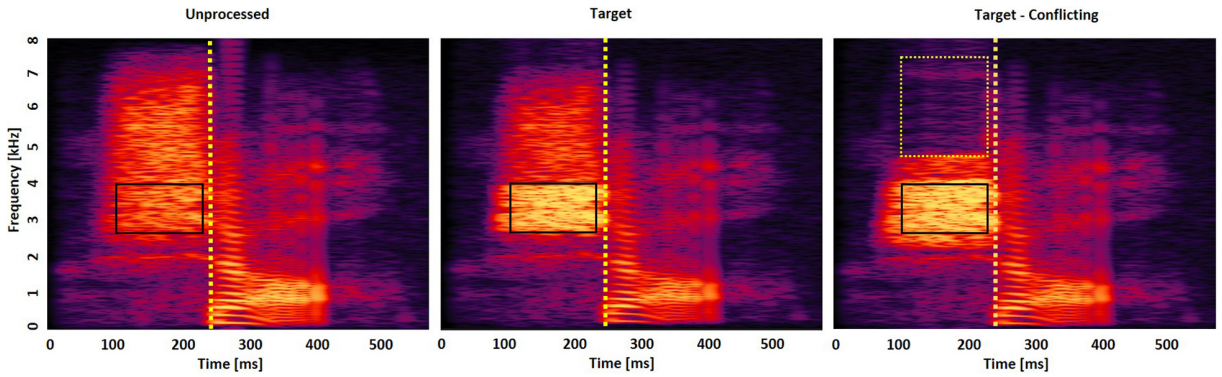


FIGURE 1 | Spectrogram of /ʃa/ consonant under the three signal processing conditions: unprocessed **(left panel)**, intensified target frequency and time ranges by + 6 dB gain (middle panel), and the combined intensified target ranges and removed conflicting ranges **(right panel)**. Solid squares indicate the target frequency and time ranges, whereas the dotted square indicates the conflicting frequency and time ranges. Dotted vertical lines indicate an onset of the vowel /a/.

the chance level performance [i.e., 2*(1/14)*100 = 14.2% points] across the cutoff frequencies were determined. This criterion was intentionally selected to choose the top two or three major confused consonants. Then, for each confused consonant, the conflicting frequency range was defined as the frequency range in kHz that created the highest scores of the confused consonant (i.e., confusion or error, against the consonant presented) and 20% less than the peak error from both LPF and HPF conditions. For example, with the consonant /ʃa/ presented, the recognition of the confused consonant /sa/ reaches 24% correct when the LPF cutoff is 4.2 kHz and a maximum of 30% correct when the cutoff is moved from 4.2 to 4.3 kHz (i.e., 24% correct is 20% below the peak 30% error). Therefore, the lower edge of the conflicting frequency would be 4.2 kHz. When the HPF cutoff is 7.6 kHz, the recognition of the confused consonant /sa/ reaches a score of 24% correct and a maximum of 30% correct when the cutoff is moved from 7.6 to 7.5 kHz. So, the upper edge of the target frequency would be 7.6 kHz. Thus, the final conflicting frequency range would be 4.2–7.5 kHz for the recognition of the consonant /ʃa/ (see the right panel of **Figure 1**). A current pilot study with five NH listeners showed that decreasing 20% from the peak error provided the highest benefit. When this study tried with a decrease of 40% or more from the peak error, more consonant confusions emerged for most of the 14 consonants when these frequency ranges were removed. No trial-by-trial feedback was provided during the test. The complete test protocol, including breaks, took approximately 2.5 h per listener.

Identifying Target and Conflicting Time Ranges

We collected consonant confusion matrices by presenting each consonant diotically in quiet. The presentation of each consonant occurred five times to acquire reliable confusion matrices for each filtering scheme. The same consonant was presented sequentially and was considered correct only if selected three times in a row. The order of consonants presented was randomized over the five presentations. The initial duration of each consonant was 3% of the total duration from the onset (i.e., the remaining 97% of the consonant was truncated out) so that minimal consonant information was presented. The duration of the consonant was increased by 1 ms when a participant's response was incorrect. The target time range was defined as a consonant segment in milliseconds from an onset of the vowel /a/ that provided the highest performance and was decreased 40% from that peak performance. A pilot study conducted by the author with five NH listeners showed that a 40% decrease from the peak provided the highest benefit when that range was boosted by 6 dB. The target time ranges were defined from the onsets of the vowel /a/ instead of defining them from the onsets of the consonants because the onsets of the vowel can be measured more reliably and precisely. An onset of the vowel /a/ was defined perceptually using a gating technique ('t Hart and Cohen, 1964). In this technique, short segments of speech were gated out from the consonant paired with the vowel /a/ and were presented in isolation. When the gate was slowly shifted in the direction of the vowel, there was a point in which one could begin to perceive an onset of the vowel. In this study, the detection of the vowel onset was verified by five adult NH listeners, and the vowel onset was accepted if all the five listeners agreed. If participants did not reach an agreement, the gating procedure were repeated with a 0.5-ms step on the segments that the five listeners disagreed upon. In **Figure 1**, the target time range for /ʃa/ is 89–193 ms, indicating a temporal duration of /ʃ/ from the onset (238 ms) of the vowel. The conflicting time ranges were not separately identified, but instead the conflicting time range was defined as the coincided target time range. A theoretical reason for the overlapped target and conflicting time ranges is that the conflicting frequency ranges would be the most detrimental factor to affect the target frequency ranges if they temporally coincide. No trial-by-trial feedback was provided during the test. The complete test protocol, including breaks, took approximately 1.5 h per listener.

Results

The identification of the target and conflicting frequency and time ranges was administered for each listener, but the mean data was presented in this study. In **Figure 2**, the mean target (filled circles) and conflicting (×symbols) frequency ranges for each

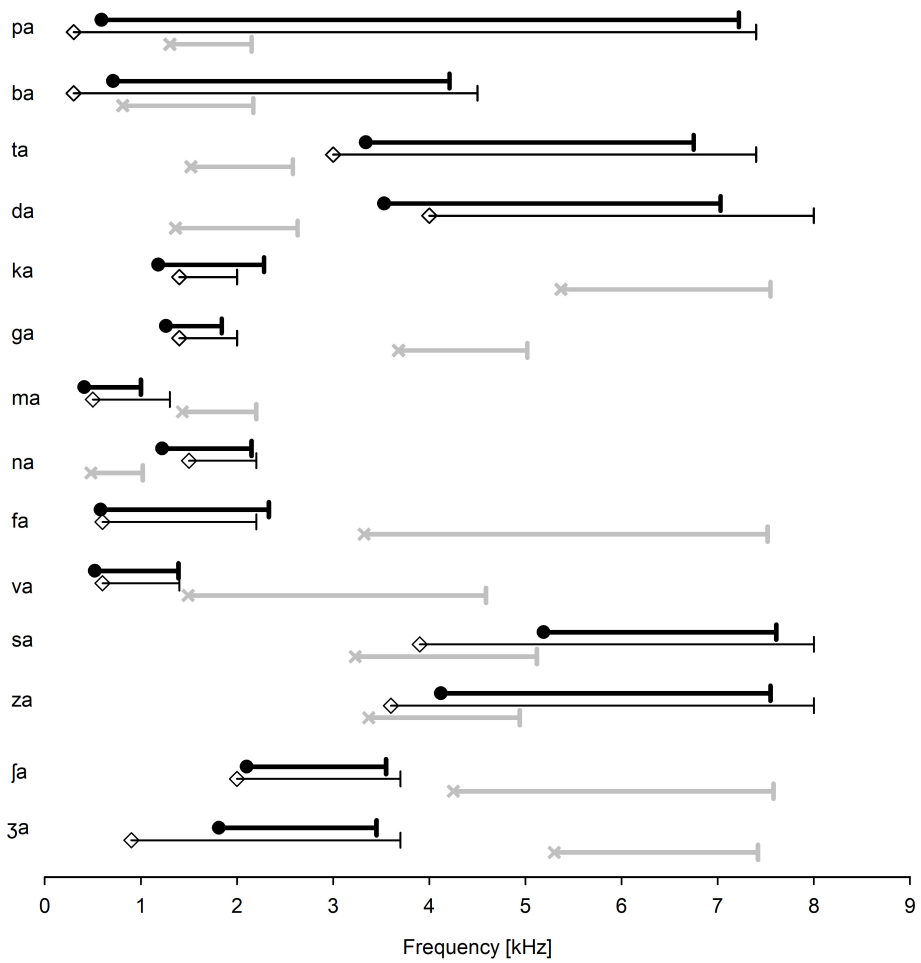


FIGURE 2 | The mean target (filled circles) and conflicting (×symbols) frequency ranges for each consonant. Three consonants (/pa, ba, za/) have a partial or complete overlap between the target and conflicting frequency ranges. For comparison, the target frequency ranges, as reported by Li et al. (2010, 2012), are also presented (open diamond symbols).

consonant were presented, along with the mean target frequency ranges (open diamond symbols) reported by Li et al. (2010, 2012). For the perception of the two bilabial stops (/pa/ and /ba/), wide ranges of spectral information were required. The target and conflicting frequency ranges fully overlapped. For the perception of the two alveolar stops (/ta/ and /da/), higher spectral ranges were required. The target frequency ranges were separated from the conflicting frequency ranges, but the conflicting frequency ranges were within similar ranges to each other. The perception of the two velar stops (/ka/ and /ga/) was dominated by mid spectral ranges. The conflicting frequency ranges were high and separated from the target frequency ranges. The perception of /ma/ and /na/ consonants required low and mid spectral ranges, respectively. The target and conflicting frequency ranges for both nasals did not overlap. The two labiodental fricatives (/fa/ and /va/) had similar lower edges of the target frequency ranges but different higher edges. Both had wide ranges of conflicting frequencies. The two alveolar fricatives (/sa/ and /za/) were characterized by high frication energy and were separated from all other consonants with high target frequency ranges above 4 kHz. Their conflicting frequency ranges had spectral energy at higher frequencies. The target and conflicting frequency ranges for /za/ were partially overlapped. For the two palatal fricatives (/ʃa/ and /ʒa/), the target frequency ranges resided at medium frication frequency energy. They were separated from all other consonants but were closer to each other, between which landed between 2 and 3 kHz. Their conflicting frequency ranges resided to be higher than 4 kHz. It should be noted that the three consonants /pa, ba, za/ had a full or partial overlap between the target and conflicting frequency ranges.

Figure 3 shows the mean target time ranges (filled circles), along with the mean target time ranges (open diamond symbols), reported by Li et al. (2010, 2012). Again, the conflicting time ranges were not separately identified. Instead, the target time ranges were used as the conflicting time ranges because the conflicting frequency ranges would be the most detrimental factor to affect the target frequency ranges if they temporally coincided. Here, the target time range was defined as the temporal

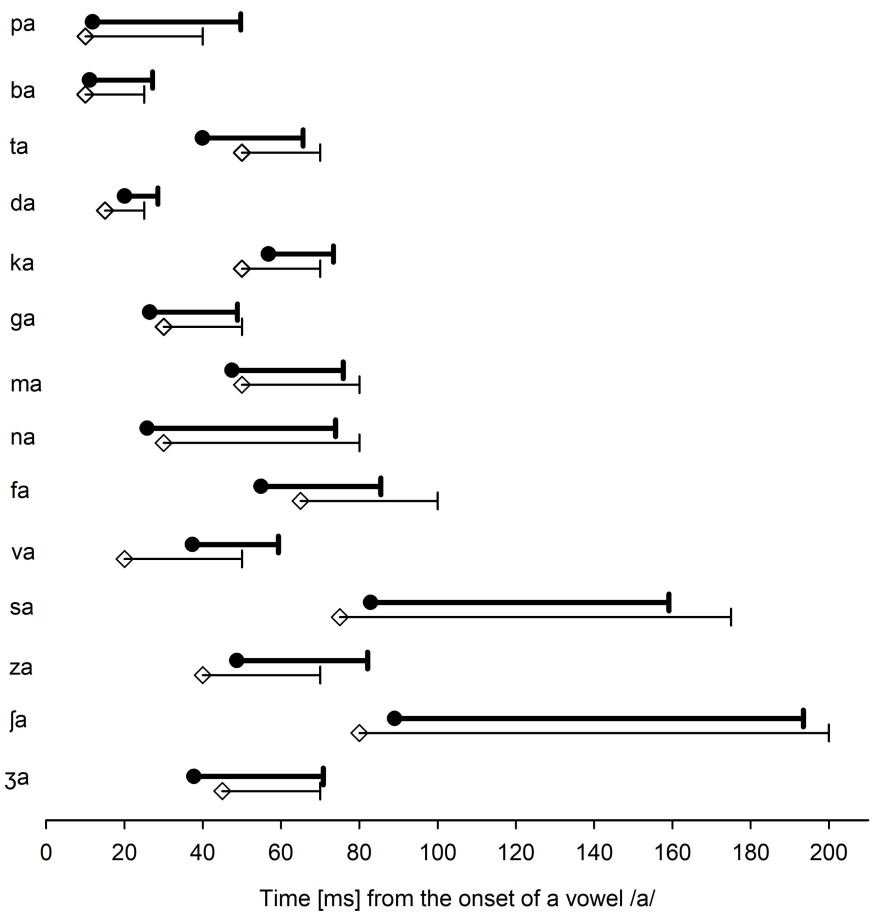


FIGURE 3 | The mean target time ranges (filled circles) for each consonant. Note that the target time ranges indicate a temporal segment from an onset of the vowel /a/. For comparison, the target time ranges, as reported by Li et al. (2010, 2012), are also presented (open diamond symbols).

duration of a consonant from the time point at the beginning of the vowel onset /a/. The target time ranges of the voiced and unvoiced stops do not overlap considerably, except for the bilabial /pa/ and /ba/ consonants. Target time ranges for voiced stops were shorter than the target ranges for unvoiced stops, except for velar consonants. The perception of /ma/ required a shorter target time (28.4 ms) than /na/ (48.2 ms), although the target time ranges partially overlapped between the nasals. Consonants /sa/ and /ʃa/ had the longest target time ranges (76.2 and 104.5 ms, respectively), which were well separated from all other consonants. Target time ranges for unvoiced fricatives (70.4 ms) were longer than those for voiced fricatives (29.5 ms).

Table 1 presents more detailed information, such as the lower and upper frequency as well as time ranges across subjects per each consonant. **Table 1** also shows the number of subjects with one SD above and below the mean of the target and conflicting frequency ranges, indicated by SD+ and SD−, respectively. The number of subjects who had SD+ and SD− ranged from 3 to 8, depending on the consonant.

To compare the target frequency ranges obtained in this study with Li et al.'s data (open diamond symbols in **Figure 2**), a two-tailed independent *t*-test was conducted. The analysis showed no significant difference in the lower target frequency ranges, $t(24) = 2.06$, $p = 0.78$ and the upper target frequency ranges, $t(24) = 2.06$, $p = 0.82$. A two-tailed independent *t*-test also showed that the target time ranges were not statistically different from Li et al.'s (2010,2012) data (2010, 2012) for the lower, $t(24) = 2.06$, $p = 0.87$, or the upper ranges, $t(24) = 2.06$, $p = 0.95$. It should, however, be noted that many subjects who had SD + and SD− (see **Table 1**) indicated a large variability in the target frequency and time ranges.

Discussion

Our results showed that the target and conflicting ranges are highly consonant-specific (**Figures 2**, **3**). The mean data obtained in this study were comparable with Li et al.'s (2010,2012) data (2010, 2012); however, this study data showed that the target and conflicting ranges were highly listener specific (**Table 1**). Stop consonants are mainly characterized by a short-duration burst from the onset (e.g., 10–20 ms), by their center frequency (wide bank, high, and medium), and their formant (particularly F2) transitions (Delattre et al., 1955; Stevens and Blumstein, 1978; Blumstein and Stevens, 1979, 1980). In this study, The target frequency ranges for stop consonants are consistent with these

typical acoustic cues. Bilabial stops had wideband clicks: 0.6–7.2 kHz for /pa/ and 0.7–4 kHz for /ba/. Alveolar stops had a high burst frequency: 3.3–6.8 kHz for /ta/ and 3.5–7 kHz for /da/. Medium burst spectral bands 1.2–2.3 and 1.3–1.8 kHz were target frequency ranges for velar stops /ka/ and /ga/, respectively. The two nasals are known to share a common feature of nasal murmur at low frequency but differed from each other in their F2 onset below 2.4 kHz (Kurowski and Blumstein, 1984; Repp, 1986; Ohde, 1994; Ohde and Ochs, 1996; Ohde et al., 2006). In this study, the results are consistent with this view: the perception of /ma/ and /na/ was dominated by low and mid spectral ranges 0.4–1.0 and 1.2–2.2 kHz, respectively.

Fricative consonants are considered a major source of perceptual error in consonant recognition in noise (Miller and Nicely, 1955; Hughes and Halle, 1956). Common characteristics of fricatives include spectral distribution of the frication noise (Hughes and Halle, 1956; Jongman et al., 2000), formant transition (Soli, 1981), overall amplitude (Behrens and Blumstein, 1988), and long duration (Baum and Blumstein, 1987; Jongman, 1989). It is known that labiodental fricatives (/f/ and /v/) show relatively flat spectra below 10 kHz with no dominating spectral peaks (McGowan and Nittrouer, 1988; Nittrouer, 2002). This may explain that non-sibilant fricatives /fa, va, θa, ðа/ are involved in more than half of the confusions at 12-dB SNR in white noise (Phatak et al., 2008). In this study, the target frequency ranges have low frication frequency energy (0.6–2.3 kHz for /fa/ and 0.5–1.4 kHz for /va/) and are within similar ranges to those reported by Li et al. (2012).

Unlike the labiodental fricatives, the sibilant alveolar consonants /sa/ and /za/ and palatal consonants /ʃa/ and /ʒa/ are seldom confused with any other consonants at 12-dB SNR (Phatak et al., 2008). Spectral cues are well defined for both alveolar and palatal consonants. In this study, the alveolar consonants were characterized by high frication energy, 5.2–7.7 kHz for /sa/ and 4.1–7.6 for /za/. These ranges are comparable with known frequency cue ranges: 4–7.5 kHz (Li and Allen, 2011), above 4 kHz (Hughes and Halle, 1956), 3.5–5 kHz (Behrens and Blumstein, 1988), and 6–8 kHz (Jongman et al., 2000). Compared to alveolar consonants, the perception of palatal fricatives /ʃa/ and /ʒa/ is known to require a lower spectral peak around 2–3.5 kHz (Miller and Nicely, 1955) and 2–4 kHz (Hughes and Halle, 1956; Behrens and Blumstein, 1988). In this study, the target frequency ranges resided at a medium frication frequency energy: 2.1–3.6 kHz for /ʃa/ and 1.8–3.5 kHz for /ʒa/, all of which are comparable to the known frequency ranges.

Regarding the target time ranges, the mean duration of the target time ranges for unvoiced stops (26.7 ms) was longer than the mean duration for voiced stops (15.7 ms), which is comparable with the mean durations reported by Li et al. (2010). The mean durations of the target time ranges for /ma/ and /na/ were 28.4 and 48.2 ms, respectively. Li et al. (2010) also reported similar durations. The mean duration of the target time ranges for unvoiced fricatives (70.4 ms) was longer than that for voiced fricatives (29.5 ms), which is also comparable with the mean duration reported by Li et al. (2010). However, the target time ranges either barely or did not overlap between the voiced and unvoiced fricatives (**Figure 2**). Although the mean duration of unvoiced fricatives is generally longer than that of the voiced fricatives, the distribution of the two overlaps considerably (Baum and Blumstein, 1987). Labiodental fricatives (/fa/ and /va/) had a shorter duration of the frication noise region compared to sibilants (/s, ʃ, z, ʒ/), indicating that the duration of the frication noise is a primary parameter to distinguish sibilant consonants from non-sibilant consonants. This finding is consistent with the findings reported by Miller and Nicely (1955). Furui conducted a time truncation experiment with Japanese consonant–vowel syllables in NH listeners (1986). In the study, consonant recognition was measured as a function of truncation position, relative to the critical point at which the syllable identification scores exceeded 80% correct. They found that consonant recognition scores rapidly decreased from 90% correct to 30% correct when the truncation point passed through the critical point 20 ms away from the critical point. This study's data, along with Furui's data, suggest that a short interval including the maximum transition position, which can be related to the perceptual critical points and bears sufficient perceptual information for syllable identification.

There are a few limitations in Experiment 1. As common speech acoustics of consonants (duration, VOT, and CVR) are also highly talker dependent, it is expected that the target and conflicting frequency ranges are highly talker dependent (Mullennix et al., 1989; Allen et al., 2003; Magnuson and Nusbaum, 2007). The target and conflicting ranges also vary depending on the preceding and following vowels (Harris, 1958; Heinz and Stevens, 1961; Blumstein and Stevens, 1980; Jongman et al., 2000; Stilp, 2020). In addition, different noise levels may have significant effects on the target and conflicting ranges (Li et al., 2010). It is understood that a more realistic identification condition for the target and conflicting ranges should include these factors; this allows the results to be more generalized. In this study, an initial intention for this experiment was to focus on the use of our identification scheme in quiet with a single talker to determine the feasibility of the identification approach studied in this paper With the proof of feasibility, an identification scheme can then be developed to take in vowels and words produced by multiple talkers and as a function of SNR as well. Larger target and conflicting data sets that have been generated and will be generated could be used to achieve the long-term goal of this study: developing algorithms for artificial intelligence-powered signal processors.

EXPERIMENT 2: MEASURE CONSONANT ENHANCEMENT

Subjects and Stimuli

The same subjects who participated in Experiment 1 participated in this experiment as well. The 14 consonants used in Experiment 1 were used in Experiment 2, however, each consonant was processed by the AI-Gram: with target frequency and time ranges enhanced by + 6 dB gain (i.e., target) and with both the target ranges enhanced and conflicting ranges removed (i.e., target-conflicting). The processed consonants were accepted as stimuli if they were perceived by five lab members at a 99% correct

level in quiet. For the three consonants (/pa, ba, za/) with overlapping target and conflicting frequency ranges (**Figure 2**), the target frequency ranges were intensified while the overlapped conflicting frequency ranges were not removed. The selection of a 6-dB gain was based on the pilot data with five NH listeners, suggesting that gains lower than 6 dB provided little or no consonant enhancement and that gains greater than 9 dB generated sound distortion.

Procedure

Subjects were seated in a single-walled sound-treated booth (Industrial Acoustics Company, North Aurora, IL, United States). Before formal testing, a 30-min familiarization session (15-min each for the target and the target-conflicting) was binaurally provided. Consonant recognition was binaurally measured in noise at −30, −20, and −10 dB SNR (speech-weighed noise) under the three signal processing conditions: the unprocessed, target, and target-conflicting. The choice of these SNRs was based on a previous study with NH listeners (Yoon et al., 2019) and was used both to validate the benefits of the AI-Gram processing in noise and to avoid a ceiling effect. Speech-shaped noise was used because the information identifying individual phonemes occurs over a very short time frame, and it was reasoned that fluctuations presented in maskers might lead to undue variability in performance. This noise masker was combined with the unprocessed and AI-Gram processed consonants to generate the designated SNRs. The sum of the speech signal and masking noise was filtered with a band-pass filter of 100–8,500 Hz before presentation. This bandwidth included the target and conflicting frequency ranges for all consonant syllables. The overall presentation level of the band-pass filtered output (i.e., speech plus noise) was scaled to the subject's MCLs (range: 50–70 dB SPL), assessed in Experiment 1. The masker commenced 500 ms before an onset of the target speech and continued for 500 ms after the target offset with cosine onset and offset ramps of 100 ms applied to the mixture. The combined speech and noise signal was diotically delivered *via* an audiometer (GSI AudioStar Pro, Eden Prairie, MN, United States) to Sennheiser HDA-200 circumaural headphones. Each consonant syllable was randomly presented ten times at each SNR. No trial-by-trial feedback was provided during the test. The complete test protocol (3 signal processing conditions × 3 SNRs × 14 consonants × 10 repetitions of each consonant), including breaks, took approximately 4 h per listener.

Results

Based on the mean (**Figure 4**) and individual consonant (**Figure 5**) analyses, there are two major findings: (1) consonant recognition improved the most with the target condition but deteriorated with the target-conflicting condition compared to the scores with the unprocessed condition and (2) the perception of seven consonants (/pa, da, ga, ma, fa, va, ʒa/) was significantly affected by signal processing, but the perception of the remaining seven consonants (/ba, ta, ka, na, sa, ʃa, za/) was not. The analyses presented in the following sections were performed with raw percent correct scores because the significance levels of the statistical analyses with transformed data sets (i.e., arcsin, log, or square root) for main effects remained unchanged, compared to the raw percent correct performances.

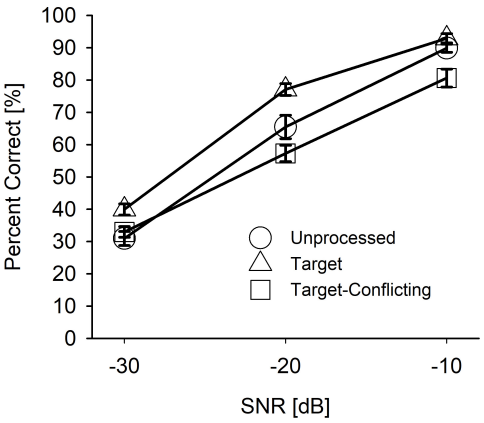


FIGURE 4 | Mean percent correct scores with SEs for each signal processing condition as a function of signal-to-noise ratio (SNR).

Mean Performance Data Analyses

Figure 4 depicts the mean percent correct with SEs for each signal processing condition as a function of SNR. Two-way repeated measure ANOVA showed a significant main effect of the signal processing, $F(2,36) = 20.68$, $p < 0.001$ and of the SNR, $F(5,36) = 498.53$, $p < 0.001$. Significant interactions between the signal processing and SNR were also observed, $F(4,36) = 9.23$, $p < 0.001$. All pairwise multiple comparisons with Bonferroni correction showed that comparisons between any two signal processing conditions were significant ($p < 0.05$), except for three comparisons: unprocessed vs. target ($p = 0.96$) at −10 dB SNR, unprocessed vs. target-conflicting ($p = 0.66$) at −30 dB SNR, and target vs. target-conflicting ($p = 0.17$) at −30 dB SNR. Multiple comparisons also showed a significant difference between any two SNRs ($p < 0.001$).

Individual Consonant by Consonant Analyses

To determine whether the effect of the signal processing on consonant recognition is consonant dependent, recognition scores were plotted with SEs per consonant (**Figure 5**). **Table 2** presents the results of the two-way repeated measures ANOVA. The results showed that the main effects of the signal processing were highly consonant-specific. Perception of seven consonants (/pa, da, ga, ma, fa, va, ʒa/) was significantly affected by the signal processing, as indicated by bold *p*-values, while the perception of the other seven consonants (/ba, ta, ka, na, sa, ʃa, za/) was not significantly affected by the signal processing. The perception of all 14 consonants was also significantly affected by SNRs. A significant interaction was observed for seven consonants (/pa, da, ga, ma, na, fa, va/).

The pairwise multiple comparisons analysis with Bonferroni correction were administered for the seven consonants (/pa, da, ga, ma, fa, va, ʒa/) which were observed for a significant main effect of the signal processing. Only pairs with significant differences were presented with a significance level in **Table 3**.

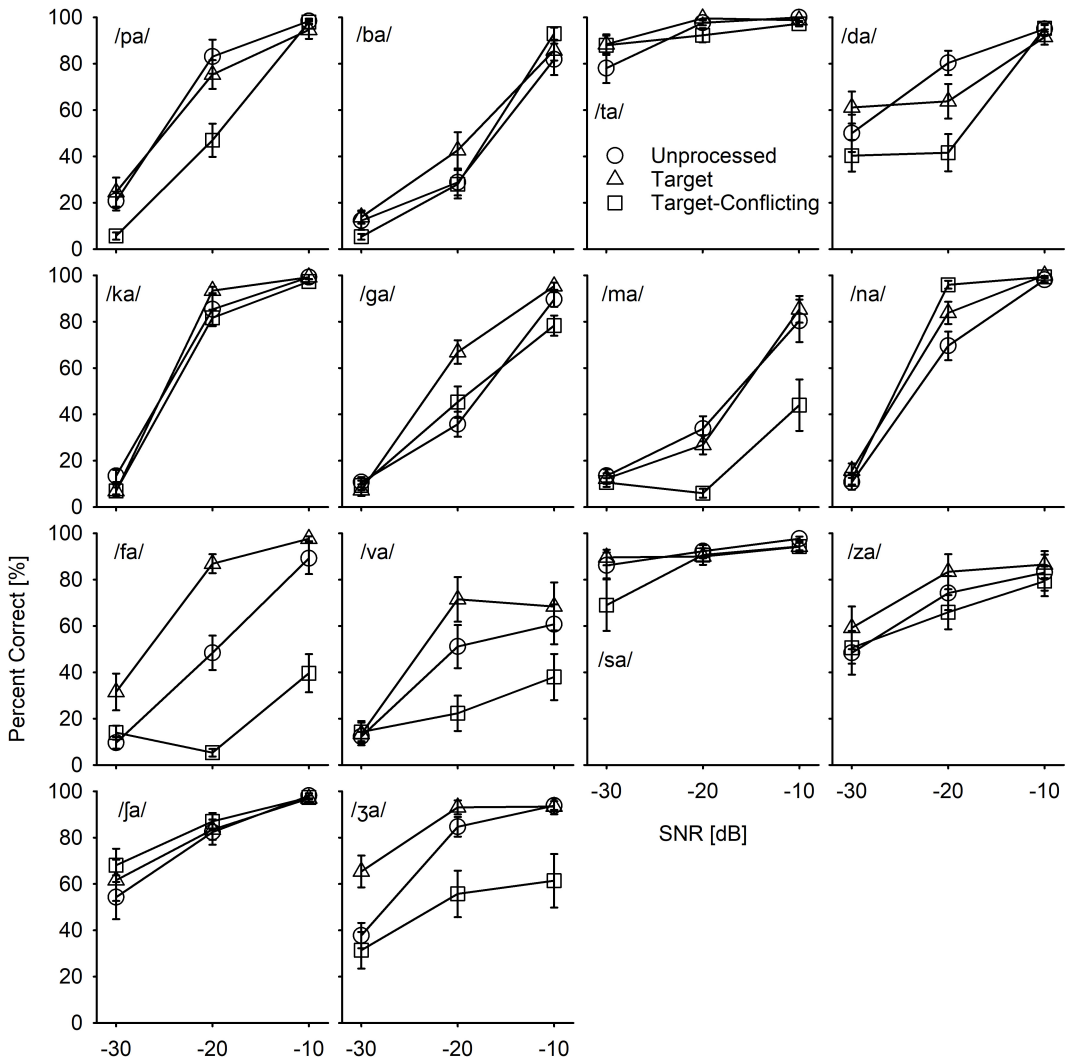


FIGURE 5 | Mean consonant recognition scores for individual consonant as a function of SNR and of the signal processing.

The results showed significant differences between the target and the target-conflicting conditions for all seven consonants over SNRs except for four consonants at a few specific SNRs between the unprocessed and the target condition. Significant differences between unprocessed and the target-conflicting conditions were observed for five consonants at a specific SNR.

Confusions Pattern Analyses

One of the goals for this current study was to define the nature of consonant enhancement or loss evoked by the AI-Gram processing on the target and conflicting ranges. In this section, key details were provided on how consonant recognition was affected by different signal processing conditions using confusion matrix analyses. Based on the statistical analyses on individual consonants (**Figure 5**), confusion patterns were presented for two exemplary consonants from the seven consonants (/pa, da, ga, ma, fa, va, ʒa/) with a significant signal processing effect and the other seven consonants (/ba, ta, ka, na, sa, za, ʃa/) with a non-significant signal processing effect. The confusion patterns for other consonants are available in the **Supplementary Material**. For these figures of confusion pattern analyses, each signal processing condition is given as a title above each panel. The consonant presented is given at the top-left corner in the left panel, and all dependent values were plotted on a logarithmic scale for a better visualization of the confused consonants as a function of SNR. The percent scores for the consonant presented are denoted as a thick curve, whereas the percent scores for the confused consonants (competitors) are indicated as a thin curve with labels. Only the top three competitors are shown to avoid the congested figures.

Figures 6 and **7** are examples of significant effects of the AI-Gram signal processing on consonant recognition. **Figure 6** shows the confusion patterns when /fa/ was presented. Percent scores of /fa/ were higher with the target condition but lower with the target-conflicting condition compared to scores with the unprocessed condition. With the unprocessed condition (left

TABLE 2 | Results of two-way ANOVA for each consonant with two factors: signal processing and signal-to-noise ratio (SNR).

	Signal processing	SNR	Interactions
/pa/	$F(2,36) = 11.5$, $p <$ **0.001**	$F(2,36) = 114.1$, **$p < 0.001$**	$F(4,36) = 4.8$, **$p = 0.003$**
/ba/	$F(2,36) = 1.1$, $p < 0.35$	$F(2,36) = 192.0$, **$p < 0.001$**	$F(4,36) = 2.2$, $p = 0.09$
/ta/	$F(2,36) = 0.4$, $p = 0.80$	$F(2,36) = 18.5$, **$p < 0.001$**	$F(4,36) = 1.3$, $p = 0.30$
/da/	$F(2,36) = 4.1$, **$p < 0.04$**	$F(2,36) = 73.1$, **$p < 0.001$**	$F(4,36) = 6.9$, **$p < 0.001$**
/ka/	$F(2,36) = 1.3$, $p < 0.31$	$F(2,36) = 562.8$, **$p < 0.001$**	$F(4,36) = 1.9$, $p = 0.14$
/ga/	$F(2,36) = 11.4$, **$p < 0.001$**	$F(2,36) = 203.3$, **$p < 0.001$**	$F(4,36) = 8.6$, **$p < 0.001$**
/ma/	$F(2,36) = 19.8$, **$p < 0.001$**	$F(2,36) = 213.1$, **$p < 0.001$**	$F(4,36) = 9.4$, **$p < 0.001$**
/na/	$F(2,36) = 3.7$, $p = 0.06$	$F(2,36) = 402.9$, **$p < 0.001$**	$F(4,36) = 4.5$, **$p = 0.005$**
/fa/	$F(2,36) = 30.0$, **$p < 0.001$**	$F(2,36) = 100.0$, **$p < 0.001$**	$F(4,36) = 16.6$, **$p < 0.001$**
/va/	$F(2,36) = 5.0$, **$p = 0.02$**	$F(2,36) = 29.1$, **$p < 0.001$**	$F(4,36) = 7.3$, **$p < 0.001$**
/sa/	$F(2,36) = 0.9$, $p = 0.39$	$F(2,36) = 9.1$, **$p = 0.002$**	$F(4,36) = 1.7$, $p = 0.17$
/za/	$F(2,36) = 1.1$, $p = 0.36$	$F(2,36) = 15.3$, **$p < 0.001$**	$F(4,36) = 0.5$, $p = 0.80$
/ʃa/	$F(2,36) = 0.4$, $p = 0.67$	$F(2,36) = 18.8$, **$p < 0.001$**	$F(4,36) = 0.7$, $p = 0.61$
/ʒa/	$F(2,36) = 8.5$, **$p = 0.003$**	$F(2,36) = 49.0$, **$p < 0.001$**	$F(4,36) = 2.0$, $p = 0.11$

A main effect of the two factors and their interactions were given for each consonant. Significant main effect was indicated by bold p-values.

panel), the combination of three competitors /ba, ma, va/ led to more than 40% errors at −30 dB SNR and continued to compete with more than 30% errors at −20 dB SNR. With the target condition (middle panel), the percent scores of /fa/ improved at the lower two SNRs compared to the unprocessed because the confusion from /ma/ and /va/ was reduced. With the target-conflicting condition (right panel), the performance of /fa/ fell to below 20% correct at the two lower SNRs because of increasing confusions with two nasals. Based on these patterns, recognition enhancement with the target condition was due to reduced confusions from /ma/, whereas recognition deterioration with the target-conflicting condition was due to increased confusion from /ma/. **Figure 7** shows the confusion patterns when /ma/ was presented. Percent scores between the unprocessed and the target condition were not statistically different, but percent scores with the target-conflicting condition were significantly lower from those with either the unprocessed or the target condition (see **Table 3**). Specifically, for both the unprocessed and the target condition, the three consonants /na, va, fa/ competed with /ma/ across SNRs. For the target-conflicting condition, the confusions from /fa/ and /va/ increased at −20 dB SNR and new confusions emerged from /ba/ at −10 dB SNR, resulting in a large decrease in /ma/ perception. Based on these results, the target condition did not help to improve the perception of /ma/. The target-conflicting condition added more confusions rather than enhancement.

TABLE 3 | Pairwise multiple comparisons between the signal processing conditions for the seven consonants, which were observed for a significant main effect of the signal processing.

Consonants	At SNR	Unprocessed vs Target	Unprocessed vs Target-conflicting	Target vs Target-conflicting
/pa/	−30 dB			**
	−20 dB		***	***
/da/	−30 dB			*
	−20 dB	***		*
/ga/	−20 dB	***		***
	−10 dB		**	**
/ma/	−20 dB		***	**
	−10 dB		***	***
/fa/	−20 dB	***		***
	−10 dB		***	***
/va/	−20 dB	*		***
	−10 dB			*
/ʒa/	−30 dB			*
	−20 dB		*	**
	−10 dB		**	**

****$p < 0.001$, **$p < 0.01$, and *$p < 0.05$. Empty cells indicate "not significant."*

The results of confusion analyses for the other five consonants (/pa, da, ga, va, ʒa/) are provided in the **Supplementary Material**. Like the confusion patterns for two consonants presented above, major competitors were similar across the signal processing condition, but confusion patterns were somewhat dependent on the signal processing. The confusions were reduced or resolved with the target condition, contributing to consonant enhancement. In contrast, the confusions were increased with the target-conflicting condition, leading to poorer consonant recognition compared to the unprocessed condition.

Figures 8, **9** are examples of non-significant effects of the AI-Gram signal processing on consonant recognition. The two consonants /na/ and /ʃa/ were intentionally chosen because their percent scores were the highest with the target-conflicting condition even though these differences were not statistically significant compared to the scores with the other two signal processing conditions (**Table 3**). When /na/ was presented, one major and consistent competitor /ma/ emerged across the signal processing condition and SNR (**Figure 8**). With the target-conflicting condition, the confusions were resolved, leading to a rise in the identification of /na/ to near-perfect scores at the higher two SNRs. When /ʃa/ was presented, /ʒa/ competed across the signal processing and SNR (**Figure 9**). Compared to the two

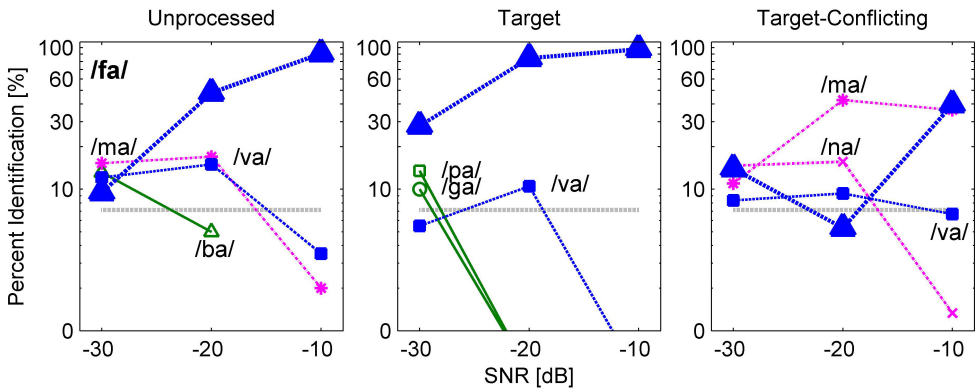


FIGURE 6 | Confusion patterns when /fa/ was presented. Each signal processing condition is given as a title above each panel. The presented consonant is given in the left panel, and all dependent values are plotted on a logarithmic scale as a function of SNR. The percent correct scores for the presented consonant are denoted as a thick curve, whereas the percent scores for each non-presented consonant or competitors are indicated as a thin curve with labels. Only up to top three competitors are shown for a better visualization of the confusion patterns.

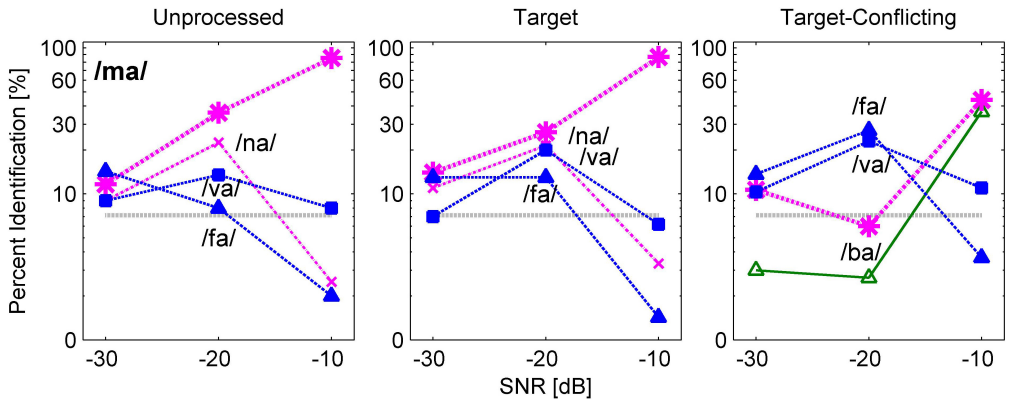


FIGURE 7 | Confusion patterns when /ma/ was presented. A description for the figure is the same as presented in **Figure 6**.

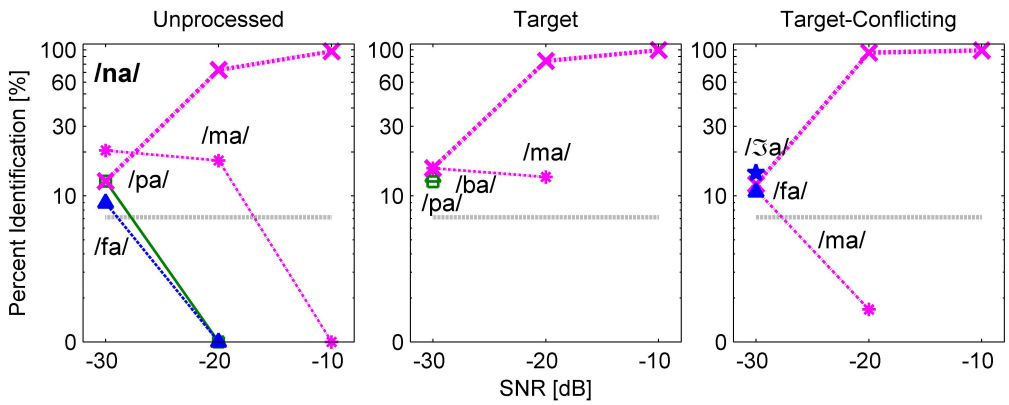


FIGURE 8 | Confusion patterns when /na/ was presented. A description for the figure is the same as presented in **Figure 6**.

other conditions, the confusions were reduced with the target-conflicting condition, resulting in an improved perception at the lower two SNRs. Based on these results, the target-conflicting condition contributed to the enhancement of recognition for at least these two consonants even though their contribution was not statistically significant.

As presented in the **Supplementary Material**, the major competitors for each of the other five consonants (/ba, ta, ka, sa,

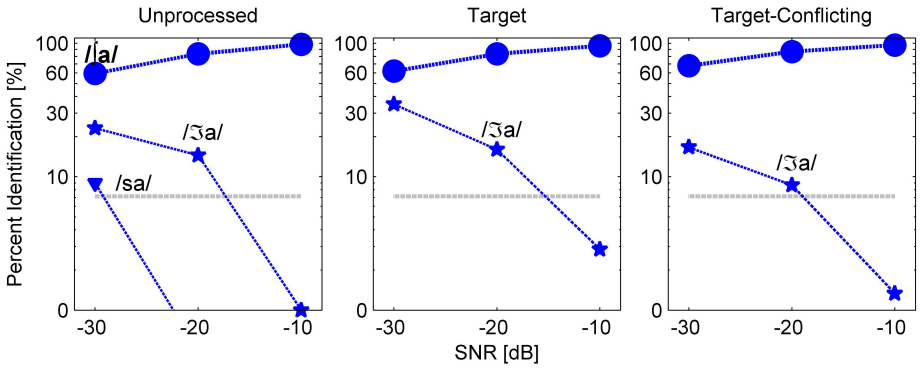


FIGURE 9 | Confusion patterns when /ʃa/ was presented. A description for the figure is the same as presented in **Figure 6**.

za/) were similar across the signal processing conditions, and the confusion patterns were different. For example, the perception of the three consonants (/ba, sa, za/) improved with the target condition but decreased with the target-conflicting condition even though these changes in percent scores were not statistically significant. However, percent scores for /ta/ and /ka/ were similar across the signal processing condition and SNR.

Acoustic Features Analyses

To determine which acoustic features (voicing, manner, and place) of consonants contributed to either consonant enhancement or loss, the percent information transmitted using information theory equations was computed (Wang and Bilger, 1973). Firstly, consonant syllables were categorized in terms of voicing, manner, and place features and then the percent correct for each of the three feature-based group consonants was computed. To obtain the percent information transmitted, the percent correct was divided by the total number of consonant syllables presented and then multiplied by 100. The results of these computations with SEs are shown in **Figure 10**. For voicing (left panel), two-way repeated measures ANOVA showed a significant main effect of the signal processing, $F(2, 36) = 21.7$, $p < 0.001$ and of SNR, $F(2, 36) = 198.7$, $p < 0.001$. The interaction effect was also significant, $F(4, 36) = 12.3$, $p < 0.001$. Mean manner information transmitted (middle panel) was significantly different over the signal processing, $F(2, 36) = 30.2$, $p < 0.001$ and SNR, $F(2, 36) = 437.9$, $p < 0.001$. Interaction was also significant, $F(4, 36) = 9.3$, $p < 0.001$. Mean place information transmitted (right panel) was significantly different over the signal processing, $F(2, 36) = 12.3$, $p < 0.001$, and SNR, $F(2, 36) = 609.8$, $p < 0.001$. Interaction was significant as well, $F(4, 36) = 3.6$, $p = 0.02$.

The results of pairwise multiple comparison analysis with the Bonferroni correction for each feature are given in **Table 4** for the signal processing factor at each SNR. The analyses showed two main findings. The target condition had helped to transmit significantly improved information of manner and place at −20 and −10 dB SNR, compared to the two other conditions. In contrast, the target-conflicting condition significantly damaged the transmission of voicing and manner information at −20 and −10 dB SNR, compared to the other two conditions. For the SNR factor, a significant difference was observed for voicing between any two SNRs ($p < 0.001$) except for −30 vs. −20 dB SNR for the target-conflicting condition ($p = 0.20$). A significant difference was also observed for both manner and place cues between any two SNRs ($p < 0.001$). Based on these results, consonant perception was enhanced under the target condition through the improved transmission of manner and place cues while consonant perception was declined under the target-conflicting condition through the decreased transmission of voicing and manner cues.

Discussion

Our mean percent correct data (**Figure 4**) reveals that consonant recognition improved the most with the target condition but deteriorated with the target-conflicting condition, compared to the scores with the unprocessed condition. Individual consonant analysis (**Figure 5**) showed that the perception of seven consonants (/pa, da, ga, ma, fa, va, ʒa/) was significantly affected by the signal processing, but the perception of the remaining seven consonants (/ba, ta, ka, na, sa, ʃa, za/) was not. Confusion analyses (**Figures 6–9**) showed similar competitors across the signal processing for each of all 14 consonants, but confusion patterns varied. Overall, the target condition had helped to reduce the confusions, resulting in improved consonant recognition, whereas the target-conflicting condition increased confusions, resulting in poorer consonant recognition compared to the unprocessed condition. Feature analyses (**Figure 10**) showed that consonant enhancement with the target condition was primarily attributed to better transmission of manner and place information. Consonant deterioration with the target-conflicting condition was primarily due to a poorer transmission of voicing and manner information. These results suggest that intensifying the target ranges is an effective way to improve consonant recognition in noise while removing the conflicting ranges negatively impacts consonant recognition.

Effects of the Signal Processing on Mean Consonant Recognition

Mean consonant enhancement was 9% points at −30 dB SNR, 11.6% points at −20 dB SNR, and 3% points at −10 dB SNR with the target condition, compared to the scores with the

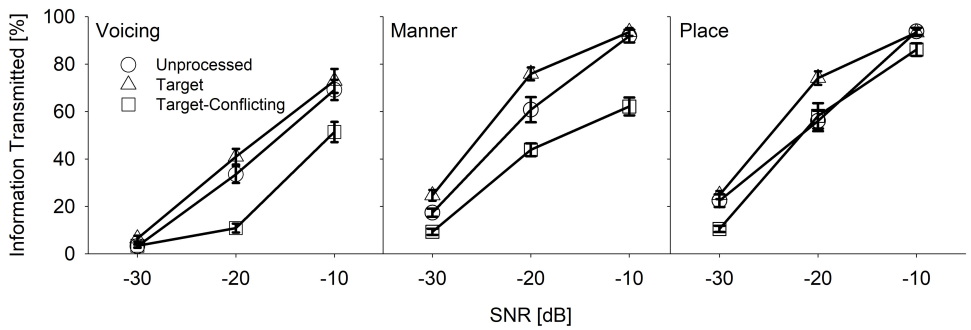


FIGURE 10 | Mean percent information transmitted for voicing, manner, and place as a function of SNR and of the signal processing.

unprocessed condition (**Figure 4**). Under the target-conflicting condition, consonant recognition deteriorated, with 8.2% points at −20 dB SNR and 9.2% points at −10 dB SNR, compared to the scores with the unprocessed condition. These findings are different from the results reported in Kapoor and Allen's (2012) study, which assessed the effects of the target condition on four stop consonants (/ta, da, ka, ga/) with 21 NH listeners over −12, −6, 0, +6, +12 dB SNR (speech-weighted noise). They reported a range of 10–70% points benefit of the target condition at SNR ≤ 0 dB, but approximately less than 7% points at SNR ≥ + 6 dB. Direct comparisons between the two studies should not be made even though the same AI-Gram processing was employed for consonant recognition. Potential reasons for these discrepancies between the two studies are discussed in a separate subsection below.

Effect of the Signal Processing on Individual Consonant Recognition

One of the major findings was that the perceptions of 12 out of 14 consonants (except for /na/ and /ʃa/) worsened under the target-conflicting condition (**Figure 5**). This finding is different from the result of Li and Allen's (2011) study (2011), which measured consonant recognition with two stops /ga/ and /ka/ at −9 and −3 dB SNR (speech-weighted noise) with three NH listeners. They tested the four signal processing conditions: unprocessed, conflicting alone (i.e., complete removal of the conflicting ranges), and two target-conflicting (i.e., complete removal of the conflicting ranges and 6 or 12 dB gain on the target ranges together). They observed the two main trends. The first trend was that the effect of the conflicting condition alone was very different between /ga/ and /ka/. Compared to the scores with the unprocessed condition, the perception of /ga/ improved by 15.9 and 14% points at −9 and −3 dB SNR, respectively, while perception for /ka/ deteriorated by 19.2 and 42.8% points. A second trend was that the target-conflicting condition had helped to improve the perception of both /ga/ and /ka/ with a range of 3.6–36.9% points over SNR compared to the scores with the unprocessed condition. As Li and Allen (2011) did not measure the effect of the target condition alone, it is unclear whether the benefit of the target-conflicting condition stemmed from the target condition alone or not. The effect of the target condition was measured, and consonant enhancement for the seven consonants was demonstrated. When the combined effect of the target and the conflicting ranges was measured, consonant perception for the 12 consonants declined (some of them were significant while others were not). The data suggest that the conflicting condition alone negatively impacted consonant recognition.

There are three possible explanations for the different outcomes between the current study and the Li and Allen's (2011) study. Firstly, the use of different SNRs may be a factor. For the current study, consonant confusions were measured at −30, −20, and −10 dB SNR (speech-weighed noise). Li and Allen collected consonant confusions at −9 and −3 dB SNR (speech-weighed noise). Secondly, the current study measured consonant confusions with 14 alternative-forced choices, whereas Li and Allen measured confusions with 6 alternative-forced choices. Finally, the target and the conflicting ranges were used, which were identified from each subject, whereas Li and Allen used averaged target and conflicting ranges. One or any combination of these factors could contribute to the differences in percent scores and perceptual confusions.

TABLE 4 | Pairwise multiple comparisons between the signal processing conditions for voicing, manner, and place.

Feature	At SNR	Unprocessed vs Target	Unprocessed vs Target-conflicting	Target vs Target-conflicting
Voicing	−20 dB		***	***
	−10 dB		***	***
Manner	−30 dB			*
	−20 dB	***	**	***
	−10 dB		***	***
Place	−30 dB			**
	−20 dB	***		***

$^{***}p < 0.001$, $^{**}p < 0.01$, and $^{*}p < 0.05$.

Acoustic Features

In terms of acoustic features, consonant enhancement with the target condition was primarily due to improved transmission of manner and place information. Consonant deterioration with the target-conflicting condition was primarily due to a

poorer transmission of voicing and manner information. To show the overall effect of the AI-Gram signal processing, the transmitted percent information were presented, which was averaged across SNRs. The percent information with the target condition was better transmitted with 4.9, 8.1, and 6.7% for voicing, manner, and place, respectively, compared to the unprocessed condition. These differences were not statistically significant at most of the SNRs (see **Table 4**). This data are comparable with the Yoon et al. (2019) study, which measured the information transmitted for the same three features with the unprocessed and the target conditions. They observed the benefit of the target condition with 3%, 10%, and 6.5% for voicing, manner, and place compared to the unprocessed condition. The percent information transmitted was also averaged over the SNR. The current study also showed that the target condition had helped to transmit more voicing (18.3%), manner (26.3%), and place (12.5%) compared to the target-conflicting condition. In contrast, the target-conflicting condition damaged the transmission of voicing (13.4%), manner (18.2%), and place (5.8%), compared to the unprocessed condition. These differences were statistically significant at most of the SNRs (see **Table 4**). No comparable data exist; direct comparisons cannot be made.

One clear trend is that the transmission of manner information was mostly enhanced with the target condition but most damaged with the target-conflicting condition. For consonants, the size of the constriction provides a clue to the manner. The first formant (F1) is known to be most affected by the size of the vocal tract constriction (Stevens and Klatt, 1974; Soli, 1981). The results of this study that F1 was enhanced with the target condition but worsened with the target-conflicting condition. It is known that the second formant frequency (F2) and F2 transition are important acoustic cues to the place of articulation for consonant recognition (Kurowski and Blumstein, 1984). A reason for the improved transmission of place information with the target condition may be due to enhanced F2 and F2 transition.

Limitations

One specific concern is the possibility that the optimal effect may occur with the target-conflicting condition if different levels of attenuations are applied to the conflicting ranges, as opposed to complete removal. This possibility is supported by the finding that consonant recognition (12 out of 14 consonants) worsened with the target-conflicting condition. In addition, in a pilot study with individuals with hearing aids or cochlear implants, maximum consonant enhancement was observed when the conflicting ranges were attenuated by −6 dB. It had been also found that attenuations between −6 and −20 dB did not make any significant difference in consonant enhancement. Our data and the pilot data with device users suggest that intensified target and attenuated conflicting ranges may facilitate a greater consonant enhancement.

Future Plan and Clinical Application

Our long-term plan is to develop an individually customized fitting scheme using an artificial intelligence-powered algorithm for bimodal users, who wear a hearing aid in one ear and a cochlear implant in the opposite ear. The current audiogram-based bimodal fitting provides highly mixed outcomes (Dorman and Gifford, 2010; Gifford et al., 2017). Some bimodal users experience interference (Mussoi and Bentler, 2017; Goupell et al., 2018). A primary reason for the inconsistent fitting outcome is the lack of knowledge regarding the exact cues driving a bimodal benefit. This limitation seriously prohibits the development of an efficient frequency fitting scheme. The results obtained from the current study serve as control data for future bimodal studies with a plan to determine the frequency and time ranges responsible for bimodal benefit and interference in consonant recognition on an individual, subject-by-subject basis. Data sets from both studies will be used to train a neural network-based deep machine learning algorithm, which can cope up with the complexity of data that will be generated. One of the key components for the development of artificial intelligence-based algorithms is the availability and volume of high-quality data inputs. High-performing deep machine learning algorithms require high-quality data to learn which data set variables are most important for maximizing algorithm accuracy and minimizing errors (Vaerenberg et al., 2011; Wang, 2017; Wathour et al., 2020). Training the deep machine learning algorithm will be effective as the target and conflicting ranges are the individualized "right" answer for each consonant to the algorithm. This testing protocol is also possibly applied to vowel confusions to see a more holistic approach to this novel target and conflicting range-based fitting procedure for bimodal hearing in the future.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Texas Tech University Health Sciences Center. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

Y-SY conceived and designed the study, conducted the experiments, analyzed the data, and wrote the draft of the manuscript.

FUNDING

This work was supported by the American Hearing Research Foundation. The funder was not involved in

the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

ACKNOWLEDGMENTS

The author thank participants for their patience and continuous support. The author also wants to thank BaileyAnn Toliver, Christine Park, Sydney Dukes, and Trisha Karnik for their editorial help.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2021.733100/full#supplementary-material

REFERENCES

Allen, J. B. (2005). Consonant recognition and the articulation index. *J. Acoust. Soc. Am.* 117(4 Pt 1), 2212–2223. doi: 10.1121/1.1856231

Allen, J. S., Miller, J. L., and DeSteno, D. (2003). Individual talker differences in voice-onset-time. *J. Acoust. Soc. Am.* 113, 544–552. doi: 10.1121/1.1528172

Baum, S. R., and Blumstein, S. E. (1987). Preliminary observations on the use of duration as a cue to syllable-initial fricative consonant voicing in English. *J. Acoust. Soc. Am.* 82, 1073–1077. doi: 10.1121/1.395382

Behrens, S., and Blumstein, S. E. (1988). On the role of the amplitude of the fricative noise in the perception of place of articulation in voiceless fricative consonants. *J. Acoust. Soc. Am.* 84, 861–867. doi: 10.1121/1.396655

Blumstein, S. E., and Stevens, K. N. (1979). Acoustic invariance in speech production: evidence from measurements of the spectral characteristics of stop consonants. *J. Acoust. Soc. Am.* 66, 1001–1017. doi: 10.1121/1.383319

Blumstein, S. E., and Stevens, K. N. (1980). Perceptual invariance and onset spectra for stop consonants in different vowel environments. *J. Acoust. Soc. Am.* 67, 648–662. doi: 10.1121/1.383890

Blumstein, S. E., Stevens, K. N., and Nigro, G. N. (1977). Property detectors for bursts and transitions in speech perception. *J. Acoust. Soc. Am.* 61, 1301–1313. doi: 10.1121/1.381433

Chen, F. (1980). *Acoustic Characteristics of Clear and Conversational Speech at Segmental Level*. Cambridge, MA: Massachusetts Institute of Technology.

Cox, R. M., Alexander, G. C., Taylor, I. M., and Gray, G. A. (1997). The contour test of loudness perception. *Ear Hear.* 18, 388–400. doi: 10.1097/00003446-199710000-00004

Delattre, P. C., Liberman, A. M., and Cooper, F. S. (1955). Acoustic loci and transitional cues for consonants. *J. Acoust. Soc.* 27, 769–773.

Dorman, M. F., and Gifford, R. H. (2010). Combining acoustic and electric stimulation in the service of speech recognition. *Int. J. Audiol.* 49, 912–919. doi: 10.3109/14992027.2010.509113

Dubno, J. R., and Levitt, H. (1981). Predicting consonant confusions from acoustic analysis. *J. Acoust. Soc. Am.* 69, 249–261. doi: 10.1121/1.385345

Ferguson, S. H. (2004). Talker differences in clear and conversational speech: vowel intelligibility for normal-hearing listeners. *J. Acoust. Soc. Am.* 116(4 Pt 1), 2365–2373. doi: 10.1121/1.1788730

Ferguson, S. H., and Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *J. Acoust. Soc. Am.* 112, 259–271. doi: 10.1121/1.1482078

Ferguson, S. H., and Morgan, S. D. (2018). Talker differences in clear and conversational speech: perceived sentence clarity for young adults with normal hearing and older adults with hearing loss. *J. Speech Lang. Hear Res.* 61, 159–173. doi: 10.1044/2017_jslhr-h-17-0082

French, N., and Steinberg, J. (1947). Factors governing the intelligibility of speech sounds. *J. Acoust. Soc. Am.* 19, 90–119.

Gifford, R. H., Davis, T. J., Sunderhaus, L. W., Menapace, C., Buck, B., Crosson, J., et al. (2017). Combined electric and acoustic stimulation with hearing preservation: effect of cochlear implant low-frequency cutoff on speech understanding and perceived listening difficulty. *Ear Hear.* 38, 539–553. doi: 10.1097/aud.0000000000000418

Goupell, M. J., Stakhovskaya, O. A., and Bernstein, J. G. W. (2018). Contralateral interference caused by binaurally presented competing speech in adult bilateral cochlear-implant users. *Ear Hear.* 39, 110–123. doi: 10.1097/aud.0000000000000470

Harris, K. S. (1958). Cues for the discrimination of American English fricatives in spoken syllables. *Lang. Speech* 1, 1–7. doi: 10.1177/002383095800100101

Hayden, R. E. (1950). The relative frequency of phonemes in general-American English. *Word* 6, 217–223.

Heinz, J. M., and Stevens, K. N. (1961). On the properties of voicelss fricative consonants. *J. Acoust. Soc. Am.* 33, 589–596. doi: 10.1121/1.1908734

Helfer, K. S. (1997). Auditory and auditory-visual perception of clear and conversational speech. *J. Speech Lang. Hear Res.* 40, 432–443. doi: 10.1044/jslhr.4002.432

Hughes, G. W., and Halle, M. (1956). Spectral properties of fricative consonants. *J. Acoust. Soc. Am.* 28, 303–310. doi: 10.1121/1.1908271

Jongman, A. (1989). Duration of frication noise required for identification of English fricatives. *J. Acoust. Soc. Am.* 85, 1718–1725. doi: 10.1121/1.397961

Jongman, A., Wayland, R., and Wong, S. (2000). Acoustic characteristics of English fricatives. *J. Acoust. Soc. Am.* 108(3 Pt 1), 1252–1263. doi: 10.1121/1.1288413

Kapoor, A., and Allen, J. B. (2012). Perceptual effects of plosive feature modification. *J. Acoust. Soc. Am.* 131, 478–491. doi: 10.1121/1.3665991

Kurowski, K., and Blumstein, S. E. (1984). Perceptual integration of the murmur and formant transitions for place of articulation in nasal consonants. *J. Acoust. Soc. Am.* 76, 383–390. doi: 10.1121/1.391139

Li, F., and Allen, J. B. (2011). Manipulation of consonants in natural speech. *IEEE Trans. Audio Speech a Lang. Process.* 19:9.

Li, F., Menon, A., and Allen, J. B. (2010). A psychoacoustic method to find the perceptual cues of stop consonants in natural speech. *J. Acoust. Soc. Am.* 127, 2599–2610. doi: 10.1121/1.3295689

Li, F., Trevino, A., Menon, A., and Allen, J. B. (2012). A psychoacoustic method for studying the necessary and sufficient perceptual cues of American English fricative consonants in noise. *J. Acoust. Soc. Am.* 132, 2663–2675. doi: 10.1121/1.4747008

Magnuson, J. S., and Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *J. Exp. Psychol. Hum. Percept. Perform.* 33, 391–409. doi: 10.1037/0096-1523.33.2.391

McGowan, R. S., and Nittrouer, S. (1988). Differences in fricative production between children and adults: evidence from an acoustic analysis of /sh/ and /s. *J. Acoust. Soc. Am.* 83, 229–236. doi: 10.1121/1.396425

Miller, G. A., and Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *J. Acoust. Soc. Am.* 27, 338–352.

Mullennix, J. W., Pisoni, D. B., and Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *J. Acoust. Soc. Am.* 85, 365–378. doi: 10.1121/1.397688

Mussoi, B. S., and Bentler, R. A. (2017). Binaural interference and the effects of age and hearing loss. *J. Am. Acad. Audiol.* 28, 5–13. doi: 10.3766/jaaa.15011

Nittrouer, S. (2002). Learning to perceive speech: how fricative perception changes, and how it stays the same. *J. Acoust. Soc. Am.* 112, 711–719. doi: 10.1121/1.1496082

Ohde, R. N. (1994). The development of the perception of cues to the [m]-[n] distinction in CV syllables. *J. Acoust. Soc. Am.* 96(2 Pt 1), 675–686. doi: 10.1121/1.411326

Ohde, R. N., Haley, K. L., and Barnes, C. W. (2006). Perception of the [m]-[n] distinction in consonant-vowel (CV) and vowel-consonant (VC) syllables produced by child and adult talkers. *J. Acoust. Soc. Am.* 119, 1697–1711. doi: 10.1121/1.2140830

Ohde, R. N., and Ochs, M. T. (1996). The effect of segment duration on the perceptual integration of nasals for adult and child speech. *J. Acoust. Soc. Am.* 100(4 Pt 1), 2486–2499. doi: 10.1121/1.417357

Phatak, S. A., and Allen, J. B. (2007). Consonant and vowel confusions in speech-weighted noise. *J. Acoust. Soc. Am.* 121, 2312–2326. doi: 10.1121/1.2642397

Phatak, S. A., Lovitt, A., and Allen, J. B. (2008). Consonant confusions in white noise. *J. Acoust. Soc. Am.* 124, 1220–1233. doi: 10.1121/1.2913251

Picheny, M. A., Durlach, N. I., and Braida, L. D. (1986). Speaking clearly for the hard of hearing. II: acoustic characteristics of clear and conversational speech. *J. Speech Hear. Res.* 29, 434–446. doi: 10.1044/jshr.2904.434

Repp, B. H. (1986). Perception of the [m]-[n] distinction in CV syllables. *J. Acoust. Soc. Am.* 79, 1987–1999. doi: 10.1121/1.393207

Revoile, S., Pickett, J. M., Holden, L. D., and Talkin, D. (1982). Acoustic cues to final stop voicing for impaired-and normal hearing listeners. *J. Acoust. Soc. Am.* 72, 1145–1154. doi: 10.1121/1.388324

Soli, S. (1981). Second formants in fricatives: acoustic consequences of fricative-vowel coarticulation. *J. Acoust. Soc. Am.* 70, 976–984.

Stevens, K. N., and Blumstein, S. E. (1978). Invariant cues for place of articulation in stop consonants. *J. Acoust. Soc. Am.* 64, 1358–1368. doi: 10.1121/1.382102

Stevens, K. N., and Klatt, D. H. (1974). Role of formant transitions in the voiced-voiceless distinction for stops. *J. Acoust. Soc. Am.* 55, 653–659. doi: 10.1121/1.1914578

Stilp, C. (2020). Acoustic context effects in speech perception. *Wiley Interdiscip. Rev. Cogn. Sci.* 11:e1517. doi: 10.1002/wcs.1517

't Hart, J., and Cohen, A. (1964). Gating techniques as an aid in speech analysis. *Lang. Speech* 7, 22–39.

Vaerenberg, B., Govaerts, P. J., de Ceulaer, G., Daemers, K., and Schauwers, K. (2011). Experiences of the use of FOX, an intelligent agent, for programming cochlear implant sound processors in new users. *Int. J. Audiol.* 50, 50–58. doi: 10.3109/14992027.2010.531294

Wang, D. (2017). Deep learning reinvents the hearing aid: finally, wearers of hearing aids can pick out a voice in a crowded room. *IEEE Spectr.* 54, 32–37. doi: 10.1109/mspec.2017.7864754

Wang, M. D., and Bilger, R. C. (1973). Consonant confusions in noise: a study of perceptual features. *J. Acoust. Soc. Am.* 54, 1248–1266. doi: 10.1121/1.1914417

Wathour, J., Govaerts, P. J., and Deggouj, N. (2020). From manual to artificial intelligence fitting: two cochlear implant case studies. *Cochlear Implants Int.* 21, 299–305. doi: 10.1080/14670100.2019.1667574

Yoon, Y. S., Riley, B., Patel, H., Frost, A., Fillmore, P., Gifford, R., et al. (2019). Enhancement of consonant recognition in bimodal and normal hearing listeners. *Ann. Otol. Rhinol. Laryngol.* 128, 139s–145s. doi: 10.1177/0003489419832625

Conflict of Interest: The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Neuroscience

SYSTEMATIC REVIEW
published: 20 December 2021
doi: 10.3389/fnins.2021.778197

Detecting Noise-Induced Cochlear Synaptopathy by Auditory Brainstem Response in Tinnitus Patients With Normal Hearing Thresholds: A Meta-Analysis

Feifan Chen[1], *Fei Zhao*[1,2], *Nadeem Mahafza*[1] *and Wei Lu*[3]*

[1] Centre for Speech and Language Therapy and Hearing Science, Cardiff School of Sport and Health Sciences, Cardiff Metropolitan University, Cardiff, United Kingdom, [2] Department of Hearing and Speech Science, Guangzhou Xinhua College, Guangzhou, China, [3] Department of Otolaryngology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China

Noise-induced cochlear synaptopathy (CS) is defined as a permanent loss of synapses in the auditory nerve pathway following noise exposure. Several studies using auditory brainstem response (ABR) have indicated the presence of CS and increased central gain in tinnitus patients with normal hearing thresholds (TNHT), but the results were inconsistent. This meta-analysis aimed to review the evidence of CS and its pathological changes in the central auditory system in TNHT. Published studies using ABR to study TNHT were reviewed. PubMed, EMBASE, and Scopus databases were selected to search for relevant literature. Studies (489) were retrieved, and 11 were included for meta-analysis. The results supported significantly reduced wave I amplitude in TNHT, whereas the alternations in wave V amplitude were inconsistent among the studies. Consistently increased V/I ratio indicated noise-induced central gain enhancement. The results indicated the evidence of noise-induced cochlear synaptopathy in tinnitus patients with normal hearing. However, inconsistent changes in wave V amplitude may be explained by that the failure of central gain that triggers the pathological neural changes in the central auditory system and/or that increased central gain may be necessary to generate tinnitus but not to maintain tinnitus.

Keywords: tinnitus, cochlear synaptopathy, hidden hearing loss, central gain, auditory brainstem response, meta-analysis

OPEN ACCESS

Edited by:
William Sedley, Newcastle University, United Kingdom

Reviewed by:
Paul Van De Heyning, University of Antwerp, Belgium
Roland Schaette, University College London, United Kingdom

***Correspondence:**
Wei Lu
luwei611@zzu.edu.cn

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

Received: *16 September 2021*
Accepted: *15 November 2021*
Published: *20 December 2021*

Citation:
Chen F, Zhao F, Mahafza N and Lu W (2021) Detecting Noise-Induced Cochlear Synaptopathy by Auditory Brainstem Response in Tinnitus Patients With Normal Hearing Thresholds: A Meta-Analysis. Front. Neurosci. 15:778197. doi: 10.3389/fnins.2021.778197

INTRODUCTION

Tinnitus is defined as a phantom sound without any corresponding external acoustic stimulus (Langguth et al., 2013). Long-term noise exposure, either occupational or recreational, is identified as the most common cause of tinnitus (Axelsson and Prasher, 2000). Tinnitus is often reported by patients with elevated hearing thresholds and, consequently, hyperactivity along the peripheral, and central auditory pathways after cochlear damage has been proposed as the primary cause (Jastreboff, 1990; Rauschecker et al., 2010; Roberts et al., 2010). It seems contradictory, therefore, that some 8–27.5% of tinnitus patients show a relatively normal performance in pure-tone audiometry (Sanchez et al., 2005; Zhao et al., 2010; Sheldrake et al., 2015). This suggests that normal hearing audiometry does not necessarily indicate normal cochlear function.

Recent research proposed that tinnitus with normal hearing thresholds may be explained by noise-induced cochlear synaptopathy (CS), which is a loss of synapses between the inner hair cells (IHCs) and auditory nerve (AN) fibers after excessive noise exposure (Kujawa and Liberman, 2009). Two types of AN fibers exist: high spontaneous discharge rate (SR) of fibers have low response thresholds, whereas low-SR fibers may be only activated by high threshold stimulation (Liberman, 1978). Notably, low-SR fibers are more vulnerable to noise exposure, by which the synaptic ribbons of IHCs could immediately and permanently be damaged [Furman et al., 2013; for review see Hickox et al. (2017)]. Since auditory brainstem response (ABR) wave I amplitudes represent the neural synchronization strength from spiral ganglion neurons (SGNs) into the auditory nerves, reduced suprathreshold wave I amplitudes could serve as a good proxy of loss or degeneration of low-SR fibers (Melcher and Kiang, 1996). Since most high-SR fibers remain intact, for hearing function in quiet, the threshold level still performs normal, which may explain normal audiograms in animal tinnitus models (Hickox et al., 2017).

Noise-induced tinnitus with normal hearing may result from increased central gain modulated by the homeostatic plasticity (Schaette and McAlpine, 2011). Homeostatic plasticity allows neurons to adjust their activity level within a dynamic range to respond to changes in synaptic inputs (Turrigiano, 1999). Thus, reduced activity of AN fibers may trigger increased central gain that enhances excitatory inputs and decreases the inhibitory inputs of downstream neurons to the ascending auditory pathway (Schaette and Kempter, 2006). This could be demonstrated by increased (or normal) ABR wave V amplitudes, which originate from the inferior colliculus (IC) at the level of brainstem (Melcher and Kiang, 1996). In addition, the ratio of wave V to wave I specifically indicates the magnitude of hyperactivity from the cochlear neurons (CN) to the IC (Schaette and McAlpine, 2011). Evidence obtained from animal studies showed increased spontaneous activity of the dorsal cochlear neurons (DCN; Middleton et al., 2011; Wu et al., 2016) and IC (Longenecker and Galazyuk, 2011) several days after noise exposure in different animal models with the sign of tinnitus, which supported the hypothesis of increased central gain to generate tinnitus. In tinnitus patients with normal hearing thresholds (TNHT), increased V/I ratios further implied the effect of central gain to generate tinnitus after deafferentation of AN fibers (Schaette and McAlpine, 2011; Gu et al., 2012; Nemati et al., 2014; Song et al., 2018; Valderrama et al., 2018).

Although the reduction of wave I amplitude in TNHT was detected by several studies (Schaette and McAlpine, 2011; Gu et al., 2012), some studies showed neither evidence of CS nor increased central gain (Guest et al., 2017; Shim et al., 2017), suggesting that ABR may not be sensitive to detect CS and/or central gain. Indeed, ABR could be influenced by some factors such as age (Grose et al., 2019), sex (McFadden and Champlin, 2000), hearing status of higher frequencies (Verhulst et al., 2016), or performance of distortion product otoacoustic emissions (DPOAE) (Bramhall et al., 2018).

On the other hand, it is also possible that low-SR fiber loss is not sufficient to generate tinnitus. Notably, tinnitus did not occur in animals where there was lower IHC ribbon loss (low-SR fiber) even with similarly reduced wave I amplitudes (Rüttiger et al., 2013). Knipper et al. (2013) hypothesized that tinnitus may result from a failure of induced central gain in the central auditory system. They suggested that it is the severe loss of high-SR fibers rather than low-SR fibers that mainly contributes to the generation of tinnitus (Knipper et al., 2020). High-SR fibers contribute to maintaining the auditory inhibitory network at central level (Singer et al., 2014). If a critical loss of high-SR fibers occurs, hyperactivity in the central auditory system may be explained by the reversal to excitation rather than disinhibition (Knipper et al., 2020). In this case, tinnitus is the result of increased neural noise rather than increased central gain, which is hypothesized to account for hyperacusis instead (Zeng, 2013). Recent studies using animal models have demonstrated a critical loss of IHC ribbons (to high- and low-SR fibers) related to reduced wave V amplitudes in animals with tinnitus-related behavior (Rüttiger et al., 2013; Singer et al., 2013). However, there is little consistent evidence in human study, despite one with tinnitus participants who had mild hearing loss (Hofmeier et al., 2018).

Although Milloy et al. (2017) reviewed the ABR findings on tinnitus patients with and without hearing loss, small numbers of studies for TNHT ($n = 5$) and missing ABR wave V data made it difficult to evidence CS and increased central gain in TNHT. There has been an increase in papers investigating CS by ABR in tinnitus patients with normal hearing since 2017. Thus, it is useful to reanalyze changes in ABR waveforms combined with new published papers. The primary aim of this study is to present a meta-analysis of ABR wave I and V amplitude to review the evidence of noise-induced CS and its possible effect on the central auditory system in tinnitus patients with normal hearing thresholds. This meta-analysis may also bring insights to two hypotheses of noise-induced tinnitus and corresponding neural effects in the central auditory system.

MATERIALS AND METHODS

Search Strategy

Searches were conducted on September 5, 2021. PubMed, EMBASE, and SCOPUS databases were selected to search for relevant literature. Search terms were designed to identify all relevant papers: (tinnitus[Title]) AND (ABR*[Title/Abstract] OR auditory brainstem response*[Title/Abstract] OR brainstem response*[Title/Abstract] OR brainstem potential*[Title/Abstract] OR electrophysiology*[Title/Abstract]). The review followed the structure recommended by PRISMA to improve the quality and reporting of meta-analyses (Moher et al., 2009).

Study Selection

Two authors (FC and NM) independently screened the title and abstract of identified papers. Since the research of noise-induced CS began to be well-conducted after the study by Kujawa and Liberman (2009), journal articles published after 2009 were added as an inclusion criterion. Only clinical human studies that utilized ABR as one of the main measurements were

TABLE 1 | Inclusion and exclusion criteria for searching.

	Detailed items
Inclusion criteria	Participants: Chronic tinnitus with normal hearing thresholds. Publication type: Peer-reviewed journals; published after 2009; in English. Outcome measure: Measured ABR wave I and V amplitudes, wave V/I and/or I/V ratio.
Exclusion criteria	Participants: pulsatile tinnitus; history of ear surgery, severe brain injury, tumors or ototoxic drug use; psychological disorders. Study design: animal studies, case reports/series, reviews, meta-analyses, conference articles, editorials. Study objective: studies investigating genetics, histology or treatment outcomes.

ABR, auditory brainstem response.

included. Literature reviews, case reports/series, meta-analyses, and animal studies were excluded. Participants (at least one group) in the included studies were required to be adults with normal audiometry and without a history of ear surgery, severe brain injury, tumors or ototoxic drug use. **Table 1** summarizes the key components of the inclusion and exclusion criteria. Since the criteria of normal hearing varied across studies, we did not define the normal hearing audiometry but listed the criteria used by the included studies (**Supplementary Table 1**).

Data Extraction

The two authors extracted the information and data independently. The second author (FZ) was involved when a discrepancy occurred. General characteristics of the studies were collated and listed in **Supplementary Table 1**, including participant characteristics (e.g., sample size, sex, and age), tinnitus characteristics (definition, pitch, and loudness matching), noise exposure history, hearing thresholds, and ABR results. ABR methodologies are summarized in **Table 2** including the device model, transducer model, polarity of stimulus, type of stimulus, duration, sound level, stimulated rate, repetition, and filters.

Quality Assessment

A critical appraisal was conducted to determine the methodological quality of the included studies using the Newcastle–Ottawa Scale (NOS) (Stang, 2010). The NOS uses a star system to assess the quality of cohort studies based on three dimensions: selection, comparability, and outcomes. The results of individual studies were classified from one to nine stars with one star representing the highest quality for each item, except for a comparability item that could be awarded two stars. A final score of 0–3 was indicated as high risk of bias, 4–6 as medium risk of bias, and 7–9 as low risk of bias.

Data Synthesis

Meta-analysis was conducted in Review Manager (RevMan, Version 5.4), The Cochrane Collaboration, 2020. The primary assessment investigated the amplitudes of ABR waves I and V and V/I ratio in normal hearing participants with or without tinnitus. Hedges's g was used, and a standardized mean difference (SMD) was calculated for the effect size with a 95% confidence interval (CI). A random effect model was chosen, and weighting of individual studies was calculated by combining the impact of the quality assessment and sample size. Similar to Cohen's *d*, 0.2, 0.5, and 0.8 of *Z* represents small, medium, and large effects, respectively. I^2 is used to estimate the heterogeneity of individual studies contributing to the pooled estimate. The degree of heterogeneity was set at low (<40%), medium (40–60%), and high (>60%).

Sensitivity analysis was performed by excluding each study, in turn, to determine the influence of each individual study on overall estimates. Subgroup analysis was performed to determine the source of any heterogeneity. Various characteristics of participants were extracted as moderators, such as age, sex, and sound level of stimuli.

RESULTS

A total of 489 publications were retrieved by the search terms. After removing duplicates, the title and abstract of 256 papers were screened. The full text of 50 papers were assessed for eligibility. Reasons for excluding papers are shown in **Figure 1**. It should be noted that although Hofmeier et al. (2018) recruited some participants with mild hearing loss (≤40-dB HL at a single frequency), the mean hearing thresholds in both groups from 0.125 to 8 kHz were within 20-dB HL. Finally, 12 studies were included for quality assessment and 11 for meta-analysis (**Figure 1**). Missing data or raw data were requested, and eight studies elicited a response.

Demographic Characteristics of Included Studies

Supplementary Table 1 summarizes the demographic characteristics in the included studies. The sample size ranged from 33 to 128 (median: 51). The mean age of the tinnitus and control groups were 35.91 and 33.09, respectively, but the range varied widely. Gender distribution in nine studies were equal or nearly balanced. One study included only female participants (Schaette and McAlpine, 2011) and another only males (Gu et al., 2012). Most tinnitus participants in Bramhall et al. (2018) were males (male vs. female: 13 vs. 2), which could be explained by the military experience in the high-noise exposure group.

Noise exposure history was evaluated in six studies, whereas two studies excluded participants with a noise-related history. However, it should be noted that no structured interview or questionnaire for lifetime noise exposure was presented, which might create risk of bias (Nemati et al., 2014; Konadath and Manjula, 2016). The recruitment of noise-exposed participants in another study relied mainly on self-reporting (Gilles et al., 2016), which could also be biased. In contrast, three studies measured occupational and leisure noise exposure using an interview (Guest et al., 2017) or questionnaire (Bramhall et al., 2018; Valderrama et al., 2018). Notably, the structural interview

TABLE 2 | ABR methodology of the included studies.

References	Device	Transducer	Stimulus type	Polarity	Duration	Sound level	Stimulated rate	Repetition	Filters
Schaette and McAlpine (2011)	Medelec Synergy T-EP system	Telephonics TDH 49 headphones	clicks	N/A	50 μs	90, 100 dB SPL	11/s	90 dB: ≥8,000 100 dB: ≥6,000	100–1500
Gu et al. (2012)	Tucker-Davis Medusa	Sennheiser, HDA-200 headphones	clicks	Condensation	100 μs	30,50,70,80 dB nHL	11/s	30 dB: 15,840 50, 70, 80 dB: 7,920	5–5,000
Nemati et al. (2014)	ICS CHARTR	earphones	clicks	Alternating		90 dB SPL	11/s	2,000	N/A
Gilles et al. (2016)	Bio-Logic Auditory Evoked Potentials	N/A	clicks	Alternating	100 μs	80 dB nHL+ 55 dB nHL masking	31/s	2,000	100–3,000
Konadath and Manjula (2016)	Biologic Navigator Pro	N/A	clicks	Rarefaction	100 μs	70 dB nHL	11.1/s	1,500	30–3,000
Guest et al. (2017)	BioSemi ActiveTwo	EARtone 3A insert earphones	clicks	N/A	N/A	102 dB peSPL	14.1/s	7,040	30–1,500
Shim et al. (2017)	Navigator Pro	ER-3A insert earphones	clicks	N/A	N/A	90 dB nHL+30 dB nHL masking	13.3/s	1,500	100–3,000
Bramhall et al. (2018)	Intelligent Hearing Systems SmartEP	N/A	4 kHz tone burst	Alternating	2 ms	80, 90, 100, 110 dB peSPL	11.1/s	80, 90, 100 dB: 2,048 110 dB: 1,024	10–1,500
Hofmeier et al. (2018)	GSI Audera	Telephonics TDH 39p headphones	clicks	N/A	100 μs	25–75 dB SPL in 10 dB steps	11.1/s	2,000	150–3,000
Song et al. (2018)	Navigator Pro	N/A	clicks	N/A	N/A	90 dB	N/A	N/A	N/A
Valderrama et al. (2018)	SmartEP with Continuous Acquisition Module	ER-3A insert earphones	clicks	Rarefaction	113 μs	108.5 dB peSPL	39.1/s	12,500	200–2,000
Joo et al. (2020)	Navigator pro	N/A	clicks	N/A	N/A	90 dB	N/A	N/A	N/A

N/A, not applicable.

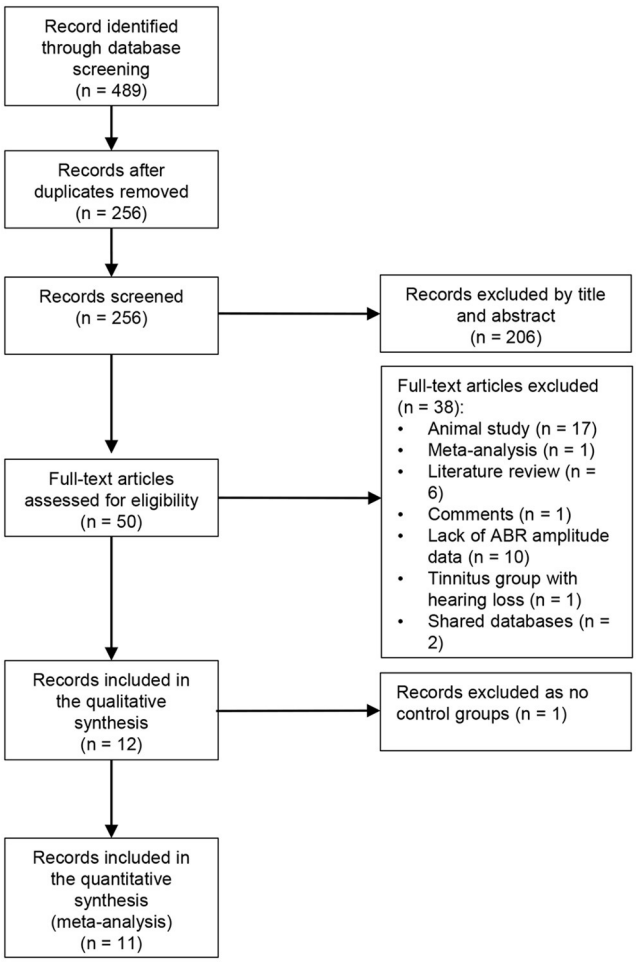


FIGURE 1 | Flow diagram of study selection (following PRISMA).

by Guest et al. (2017) applied a different strategy to estimate noise exposure dose (>80 dBA). The sound level of individual events and activities were calculated with the sum providing total lifetime noise exposure [for details, see Guest et al. (2018)].

The inclusion criteria for tinnitus participants varied between studies. Six studies used the duration of tinnitus. The type of tinnitus was defined or collected in seven studies, and no participants with pulsatile tinnitus were recruited except the study by Valderrama et al. (2018). Since pulsatile tinnitus is usually triggered by the alteration in blood flow and different from noise-induced tinnitus (Hofmann et al., 2013), it may create a risk of bias. Four studies measured the psychoacoustic characteristics of the tinnitus, including localization ($n = 3$), pitch ($n = 3$), loudness ($n = 4$), minimum masking level, and residual inhibition ($n = 1$). The functional or emotional impact of tinnitus was evaluated by different questionnaires in six studies.

Gilles et al. (2016) suggested that recreational noise exposure was the most possible cause of tinnitus, though this was based on the self-report. Although tinnitus is observed frequently in people exposed to high-level noise, the pathophysiology of noise-induced tinnitus could be different from other types that are generated by different risk factors.

Hearing thresholds were measured in all studies. Notably, in the study of Hofmeier et al. (2018), some participants had no more than 40-dB HL in each frequency, which could make it hard to exclude the confounding effect of mild outer hair cell (OHC) loss from CS to ABR waveforms. By contrast, Gu et al. (2012) did not define the normal hearing threshold range of participants or report the average hearing thresholds in both groups. Hearing status at extended high frequencies, from 9 to 16 kHz, was evaluated in five studies, with three studies reporting no significant difference from the control group (Schaette and McAlpine, 2011; Gilles et al., 2016; Guest et al., 2017). Valderrama et al. (2018) defined no more than 40-dB HL from 8 to 12.5 kHz. In OAE results, only five studies measured DPOAE or TEOAE (Nemati et al., 2014; Gilles et al., 2016; Bramhall et al., 2018; Song et al., 2018; Valderrama et al., 2018). Three used OAE results as one of the inclusion criteria, with signal-to-noise ratio >3 or 6 dB from 1 to 4 kHz used most frequently. Lack of OHC status from OAE results may make it more difficult to interpret the reasons for changes in ABR amplitudes.

Parameters Used for Auditory Brainstem Response Measurements

The details of ABR methodology used in each study are presented in **Table 2**. Eleven studies applied a click stimulus, whereas Bramhall et al. (2018) used a 4-kHz tone burst as the stimulus. Six studies reported the polarity of stimulus. Three applied condensation (Gu et al., 2012) or rarefaction (Konadath and Manjula, 2016; Valderrama et al., 2018), whereas three used alternating polarity (Nemati et al., 2014; Gilles et al., 2016; Bramhall et al., 2018). Notably, it is recommended that rarefaction should be used rather than condensation to produce enhanced amplitudes of ABR waveforms. Alternating polarity should be avoided to minimize artifacts (Hall, 2006). The polarity of the 4-kHz tone burst stimulus was reported as rarefaction by Bramhall et al. (2018), which is the same as recommended (Hall, 2006).

Durations of clicks or tone burst were reported in seven studies. Four presented clicks of 100 μs, while Schaette and McAlpine (2011) used 50 μs, and Valderrama et al. (2018) set 113 μs. Bramhall et al. (2018) used a tone burst of 2 ms. Although most studies presented the stimuli at high intensity, the sound levels of the stimuli varied between studies with dB normal hearing level (nHL), dB sound pressure level (SPL), or dB peak-equivalent SPL (peSPL) being used (**Table 2**). In order to compare the ABR results, dB SPL and dB peSPL were converted into dB nHL using the formula 0-dB nHL = 36.4 peak SPL = 29.9 peSPL (in a condition of 100 μs and 20/s; Hall, 2006), as shown in **Figure 2**. Although the relationship between stimulus level and mean wave I and V amplitude were inconsistent, V/I ratios showed a negative tendency with increase in stimulus level. It should also be noted that even the ABR results at a single stimulus level varied widely across the studies, questioning the sensitivity of wave I amplitude to detect CS.

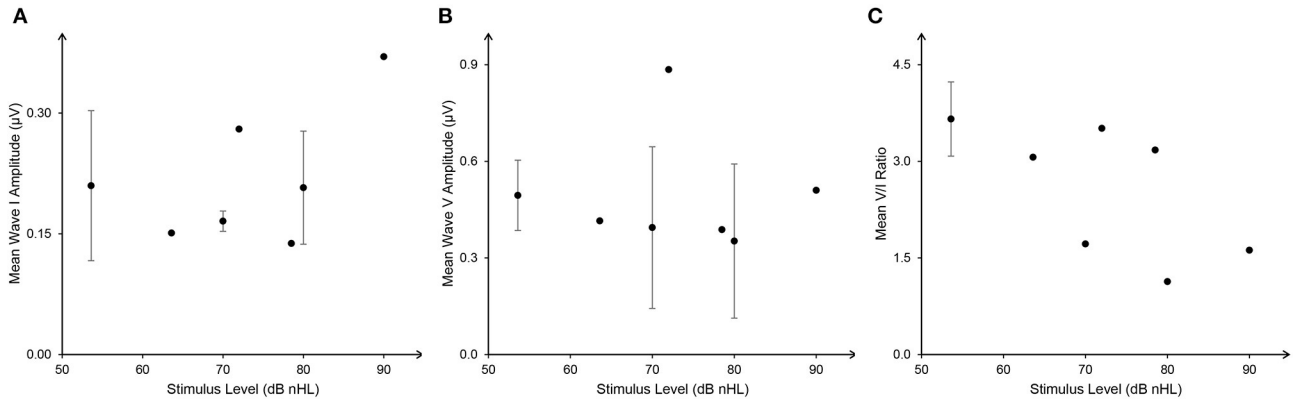


FIGURE 2 | The weighted relationship between stimulus level [dB normal hearing level (nHL)] and auditory brainstem response (ABR) wave I **(A)**, V amplitude **(B)**, and V/I ratio **(C)** (mean ± SD). ABR data with dB sound pressure level (SPL) and dB peak-equivalent sound pressure level (peSPL) were converted into dB nHL with the formula: 0-dB nHL = 36.4 peak SPL = 29.9 peSLP (in a condition of 100 μs and 20/s).

In addition, 10 studies reported stimulus rate. Eight studies used 11/s or similar. The clicks rates of the other studies were over 30/s (Gilles et al., 2016; Valderrama et al., 2018). ABR amplitude may decrease if the presentation rate increases over 31.1/s (Hall, 2006). Although 21/s was recommended by Bramhall et al. (2019) for investigating cochlear synaptopathy in a human ABR study, limited evidence from the included studies supports that recommendation.

Six studies applied between 1,500 and 2,000 sweeps, which according to Hall (2006) is sufficient to produce a confident SNR ratio for identifying wave V latency and amplitude. Thus, it seems effective to use 1,000 or 2,000 sweeps of clicks at high-level intensity (>90-dB SPL; Bramhall et al., 2019). In addition, filters were used in eight studies, with low- and high-pass filters included (**Table 2**). A low-pass filter is especially recommended to exclude the potential effect of OHC loss at higher frequencies (Bramhall et al., 2019). High-pass filters were highly variable, from 5 to 200 Hz, between studies. It should be noted that high pass over 100 Hz should be avoided (Hall, 2006). However, Gu et al. (2012) used a much lower-frequency high-pass filter (i.e., 5 Hz). The influence of using a lower-frequency high-pass filter may result in an increased amplitude of ABR waveforms, which could be contaminated by artifacts (Hall, 2006).

Quality Assessment

The results of the quality assessment of the included studies are shown in **Table 3**. Two studies had high risk of bias (Konadath and Manjula, 2016; Joo et al., 2020), while six studies had medium risk of bias, and four studies had low risk of bias. Specifically, all but two studies provided details of tinnitus participant recruitment and defined inclusion and/or exclusion criteria. The participants reported in the included studies were typical tinnitus cases, though the tinnitus characteristics varied from individual studies. Six studies, however, did not recruit controls from the community, and the definition of controls was not given in several studies. As for comparability, based on the aim of the current study, the first and second impact factors were noise exposure history and OAE status. Three studies had two stars and another three received one star. In addition, noise exposure history of both the tinnitus and control groups were measured by three studies (**Supplementary Table 1**). Furthermore, although several studies reported missing data, the reason for those in the study of Valderrama et al. (2018) included technical problems, which did not meet the standard.

Meta-Analysis

Data from the 11 included studies were extracted for meta-analysis of the ABR wave I amplitude. This included 313 tinnitus ears and 595 control ears (**Figure 3**). There was a significant difference in wave I amplitude between the tinnitus participants and controls (SMD = −0.45, 95% CI: −0.74, −0.15, $p < 0.001$), with lowered wave I amplitudes in the tinnitus participants. Although total SMD reduced to −0.25 (95% CI: −0.45, −0.06) after excluding two substudies with a large effect size [70 and 80 dB of Gu et al. (2012)], the significant difference of wave I amplitude between the two groups remained ($p < 0.05$). The result showed a large heterogeneity across the studies ($Chi^2 = 59.02$, $p < 0.001$, $I^2 = 73\%$), and 33% heterogeneity still remained when the study by Gu et al. (2012) was removed. Several reasons might account for the reduction, such as potential loss of OHCs at the higher frequencies or the combined effect of age-related CS because of a much higher mean age (42 ± 6) of the participants (**Supplementary Table 1**). Condensation polarity may decrease ABR amplitude when compared with rarefaction polarity by producing an outward direction of basilar membrane movement that is opposite to that when afferent auditory nerves activate (Hall, 2006). It should be noted that Gu et al. (2012) used a much wider bandpass filter from 5 to 5,000 Hz. As a result, they showed a larger amplitude of wave I. In addition, the mean amplitude of controls at 80-dB nHL in the study of Gu et al. (2012) was much higher than two studies that used similar stimulus levels (Gilles et al., 2016; Konadath and Manjula, 2016) (**Figure 3**).

TABLE 3 | Quality assessment of the included studies by Newcastle-Ottawa-Scale (NOS) questionnaire.

References	Case definition	Representativeness	Selection of controls	Definition of controls	Comparability	Ascertainment of exposure	Same ascertainment for controls	Non-response rate	Total stars
Schaette and McAlpine (2011)	⁂	⁂	⁂		No noise, no OAE			⁂	4
Gu et al. (2012)	⁂	⁂	⁂		No noise, no OAE			⁂	4
Nemati et al. (2014)	⁂	⁂		⁂	⁂ No noise, OAE			⁂	5
Gilles et al. (2016)	⁂	⁂	⁂	⁂	⁂⁂ Noise, OAE			⁂	7
Konadath and Manjula (2016)		⁂	⁂		No noise, no OAE			⁂	3
Guest et al. (2017)	⁂	⁂	⁂	⁂	⁂ OAE	⁂	⁂	⁂	8
Shim et al. (2017)	⁂	⁂		⁂	No noise, no OAE			⁂	4
Bramhall et al. (2018)	⁂	⁂		⁂	⁂⁂	⁂	⁂	⁂	8
Hofmeier et al. (2018)	⁂	⁂		⁂	No noise, no OAE			⁂	4
Song et al. (2018)	⁂	⁂			⁂ OAE			⁂	4
Valderrama et al. (2018)	⁂	⁂	⁂		⁂⁂	⁂	⁂		7
Joo et al. (2020)		⁂						⁂	2

The study can be awarded a maximum of two stars for comparability and one star for the other items. Total stars of individual studies range from 0 to 9.

Meta-analysis for ABR wave V amplitude included 325 tinnitus ears and 598 control ears from 11 studies (**Figure 4**). The results showed no significant difference in wave V amplitude between tinnitus participants and controls (SMD = 0.09, 95% CI: −0.30, 0.48, $p = 0.65$) and large heterogeneity ($Chi^2 = 106.33$, $p < 0.001$, $I^2 = 85\%$). Heterogeneity decreased to 43% after removing the study of Gu et al. (2012). This result could be explained by the different measurement methods used in the study by Gu et al. (2012), i.e., wave V amplitude from prestimulus baseline to peak in comparison with the measures from peak to the following trough used in other studies. It is noteworthy that alteration in wave V amplitudes may not be consistent with the central gain hypothesis, given that eight studies showed reduced wave V amplitudes or a tendency for the reduction in tinnitus participants (Knipper et al., 2020). Data in nine studies (tinnitus ears 271 and control ears 484) showed significantly increased V/I amplitude ratios in tinnitus participants (SMD = 0.23, 95% CI: 0.06, 0.39, $p < 0.05$), which is consistent with the central gain hypothesis. The results indicated no heterogeneity across the included studies ($Chi^2 = 12.46$, $p = 0.49$, $I^2 = 0\%$) (**Figure 5**).

Subgroup analysis was conducted to investigate the source of heterogeneity in wave I amplitudes. Sex, age, noise exposure history, and polarity were examined. Although there was no significant effect identified, sex may have had an influence (**Table 4**) as the overall reduction in the female tinnitus subgroup was larger than in the males (SMD = −0.53, 95% CI: −0.87, −0.19, $p < 0.05$) with no heterogeneity in the subgroups and no significant difference between the two subgroups ($Chi^2 = 2.26$, $p = 0.13$).

There was no significant difference between the young and older age groups ($Chi^2 = 0.35$, $p = 0.55$). However, wave I amplitudes in the older subgroup were much lower than the controls and with little heterogeneity ($I^2 = 0\%$), which supports the hypothesis of age-related synaptic loss in AN fibers (Sergeyenko et al., 2013). There was no significant correlation between noise exposure history and wave I amplitude ($Chi^2 = 0.43$, $p = 0.81$). While the results of the two subgroups (investigated and no history) were no different, this could be attributed to the limited sample size, and it would be unwise to neglect the effect of measurement of LNE on wave I amplitude reduction.

Notably, three studies with a younger subgroup were included in the investigated subgroup (Gilles et al., 2016; Guest et al., 2017; Bramhall et al., 2018), which may be able to explain the relatively large heterogeneity ($I^2 = 72\%$) within that subgroup (**Supplementary Table 1**). Polarity had no significant effect on wave I amplitude between the three subgroups ($p = 0.86$).

Subgroup analysis of wave V amplitudes produced no significant differences. This may indirectly suggest that two hypotheses for noise-induced tinnitus coexist (**Table 5**) and that there is a possibly combined, rather than contradictory, role of two types of CS in generating noise-induce tinnitus. Interestingly, there were also consistently increased V/I ratios in the reduced wave V amplitude

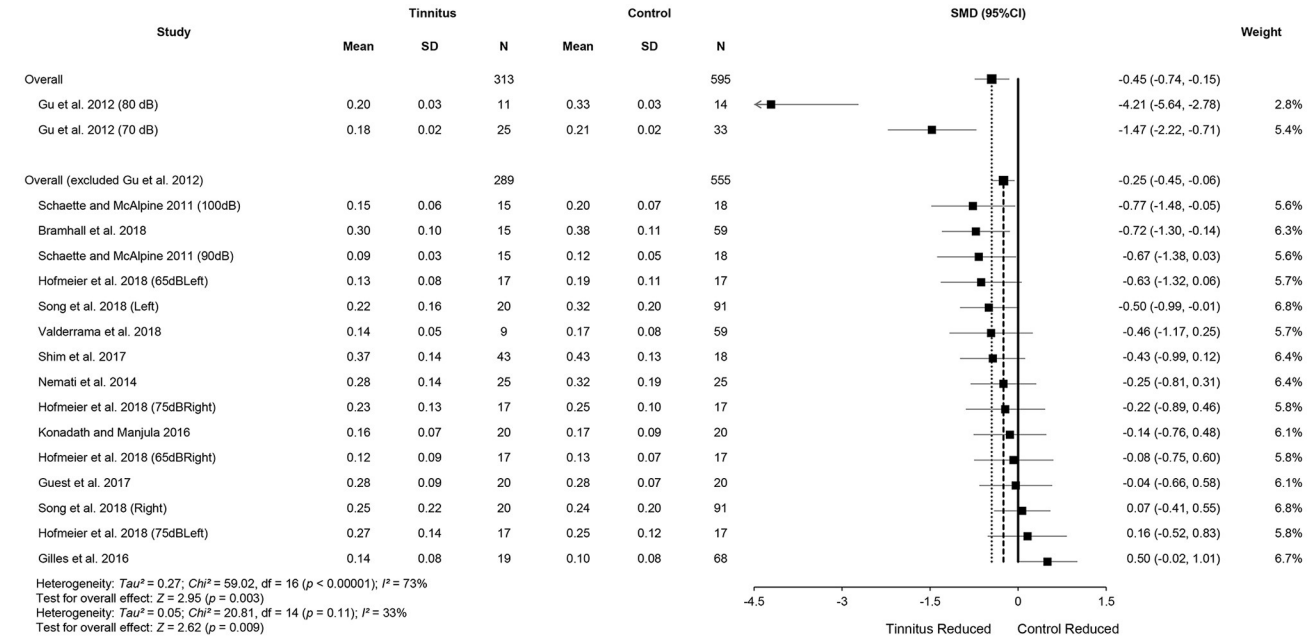


FIGURE 3 | Forest plot of 11 studies for the difference of wave I amplitudes (95% CI) between tinnitus and control participants.

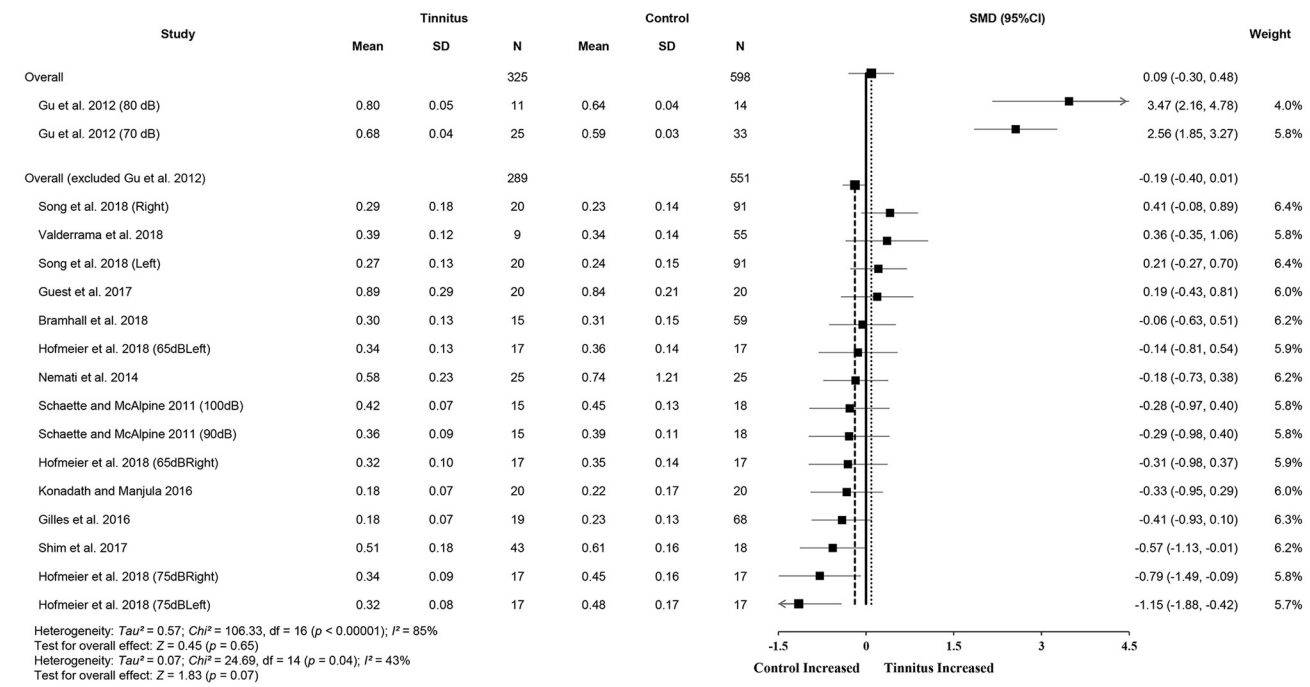


FIGURE 4 | Forest plot of 11 studies for the difference of wave V amplitudes (95% CI) between tinnitus and control participants.

subgroup (**Figure 6**), either indicating that the distinct regions contribute to increased central gain, or there is enhanced evoked activity that could be associated with hyperacusis. However, more evidence is needed to verify this speculation.

DISCUSSION

This review investigated whether the ABR changes in tinnitus patients with normal hearing are consistent across studies. The results show significantly reduced wave I amplitudes with low

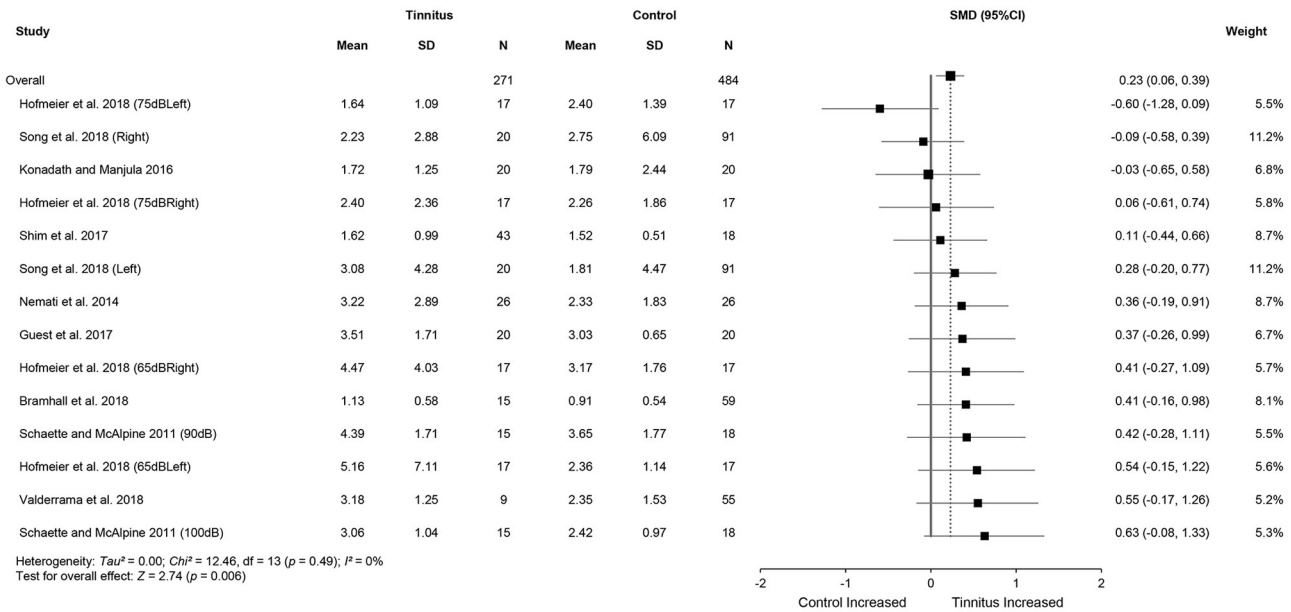


FIGURE 5 | Forest plot of nine studies for the difference of wave V/I amplitude ratios (95% CI) between tinnitus and control participants.

heterogeneity and increased wave V/I amplitude ratios. The changes in wave V amplitudes were inconsistent. No interaction was identified by subgroup analysis to explain the heterogeneity shown in reduced wave I amplitudes.

Reduced Wave I Amplitude: Loss of Low-Spontaneous Discharge Rate Fibers or High-Spontaneous Discharge Rate Fibers

Synaptic ribbon, a presynaptic structure at active zones of the IHC synapse, tethers a large number of vesicles that enable a sustained high rate of transmission to AN fibers (Glowatzki and Fuchs, 2002). Transmitter release at these synapses enables precise temporal and intensity information to be passed to the auditory neurons for the accurate coding of time and intensity (Goutman and Glowatzki, 2007). The IHCs generate action potentials and transmit them to the central auditory system *via* AN fibers (Robles and Ruggero, 2001). Acoustic trauma has been linked to the loss of synapses between the IHC and the SGN in the terminals of type I AN fibers (Kujawa and Liberman, 2009; Rüttiger et al., 2013; Singer et al., 2013) and the disorganization of synaptic vesicles in the IHC cells (Bullen et al., 2019).

Notably, the roles of two types of AN fibers differ in auditory perception and processing. High-SR fibers determine the threshold of the auditory neural response at characteristic frequencies and are important for temporal resolution (Bourien et al., 2014). In contrast, low-SR fibers are important for hearing in noise during which high-SR fibers have become saturated (Costalupes et al., 1984; Costalupes, 1985). Moreover, high-SR fibers transmit envelop cues to the cochlear nucleus [for review, see Bharadwaj et al. (2014)], whereas low-SR fibers are involved in the coding of temporal fine structure and the temporal envelope, which account for speech intelligibility in noise (Lorenzi and Moore, 2007) or speech-on-speech masking release (Christiansen et al., 2013).

The reason why low-SR fibers are more vulnerable to acoustic damage may be because of fewer mitochondria present (Knipper et al., 2015). Deafferentation of low-SR fibers parallels the lower numbers of ribbons in the IHCs but with larger size, which in turn

TABLE 4 | Subgroup analysis of the relationship (95% CI) between wave I amplitude and sex, age, noise exposure history and polarity.

Subgroup	Sample size (*n*)		SMD (95%CI)	*p*-value
	Tinnitus	Control		
Overall (sex only)	183	223	−0.35 (−0.58, −0.13)	
Sex				0.13
Male	91	112	−0.19 (−0.48, 0.11)	
Female	92	111	−0.53 (−0.87, −0.19)	
Overall (other subgroups)	289	555	−0.25 (−0.45, −0.06)	
Age				0.55
<30	54	147	−0.08 (−0.79, 0.64)	
>30	235	408	−0.30 (−0.48, −0.13)	
Noise exposure history				0.81
No history	45	45	−0.2 (−0.61, 0.22)	
Investigated	63	206	−0.16 (−0.73, 0.4)	
Not investigate	181	304	−0.4 (−0.7, −0.11)	
Polarity				0.93
Alternating	59	152	−0.15 (−0.85, 0.56)	
Rarefaction	29	79	−0.28 (−0.74, 0.19)	
Not report	201	324	−0.29 (−0.49, −0.09)	

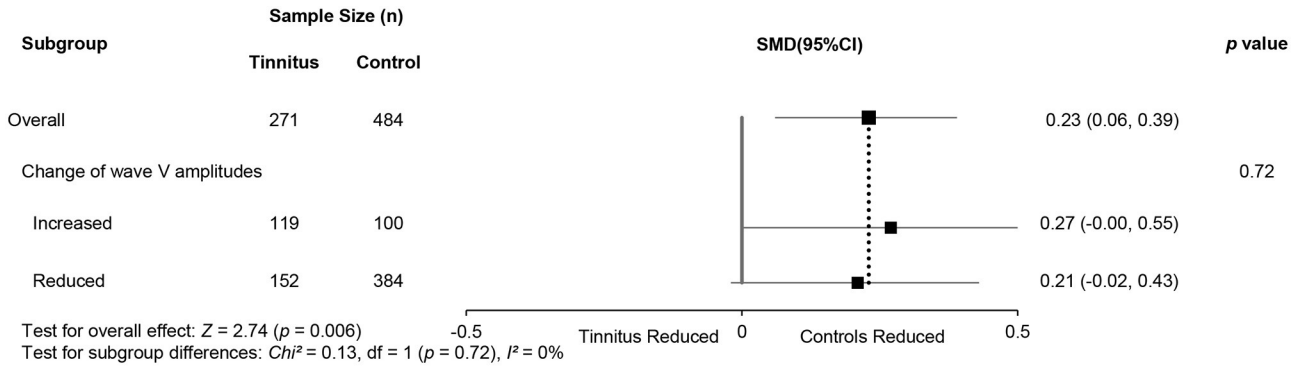


FIGURE 6 | Subgroup analysis of the relationship (95% CI) between wave V/I ratios and changes in wave V amplitudes.

TABLE 5 | Subgroup analysis of the relationship (95% CI) between wave V amplitude and sex, age, noise exposure history, and polarity.

Subgroup	Sample size (*n*)		SMD (95%CI)	*p*-value
	Tinnitus	Control		
Overall (sex only)	183	219	−0.27 (−0.54, −0.00)	
Sex				0.69
Male	91	112	−0.21 (−0.62, 0.19)	
Female	92	107	−0.33 (−0.70, 0.05)	
Overall (other subgroups)	289	551	−0.19 (−0.40, 0.01)	
Age				0.67
<30	54	147	−0.13 (−0.47, 0.22)	
>30	235	404	−0.22 (−0.48, 0.03)	
Noise exposure history				0.55
No history	45	45	−0.24 (−0.66, 0.17)	
Investigated	63	202	−0.03 (−0.37, 0.30)	
Not investigate	181	304	−0.28 (−0.61, 0.04)	
Polarity				0.82
Alternating	59	152	−0.23 (−0.54, 0.08)	
Rarefaction	29	75	−0.01 (−0.68, 0.67)	
Not report	201	324	−0.23 (−0.53, 0.07)	

are correlated with the neural degeneration of SGNs (Kujawa and Liberman, 2009; Lin et al., 2011). Consequently, several animal studies suggested that loss of synapses in low-SR fibers would result in reduced spontaneous firing rates with no elevation of hearing thresholds [Lopez-Poveda and Barrios, 2013; for review, see Aedo and Aguilar (2020)]. Apart from ABR, the presence of CS in TNHT has been indicated by other measurements, such as damaged speech perception in noise (SpiN; Gilles et al., 2016) and increased SP/AP ratio in electrocochleography (EcochG; Kara et al., 2020), both of which have been linked with higher risk of cochlear synaptopathy in subjects (Liberman et al., 2016).

However, some animal studies have proposed that severe damage to high-SR, rather than low-SR fibers, induces noise-induced tinnitus (Rüttiger et al., 2013; Singer et al., 2013). The reduction of wave I amplitude was detected after noise exposure, but tinnitus-related behaviors linked with over 80% ribbon loss in high-frequency cochlear turns (Rüttiger et al., 2013). When there is extensive loss of ribbons in the synapses to high-SR fibers, response reliability of AN fibers degenerates along with reduced spontaneous and sound-evoked activity, reflected by prolonged latency and reduced amplitudes of wave I (Buran et al., 2010). Evidence showed that behavioral thresholds still were normal in mice, even though there was 95% loss of both low and high-SR AN fibers (Chambers et al., 2016). However, as higher degrees of ribbon loss cannot maintain the precision of the spike response of the AN fiber, hearing thresholds in the extended high frequencies also elevated without damage to the OHCs (Buran et al., 2010). This was consistent with the results in TNHT after noise overexposure (Sulaiman et al., 2014; Kumar and Deepashree, 2016; You et al., 2020). To support that, envelope following responses (EFR) in quiet were measured in the tinnitus group with normal audiometry, suggesting the damage to both high-SR and low-SR fibers, though the etiology of tinnitus in this study remained unclear (Paul et al., 2017).

Regardless of the types of AN fiber loss, so far, the presence of CS in tinnitus still remains inconsistent among studies from different measurements [for review see Bramhall et al. (2019)]. However, a possible role of synaptopathy in the generation of tinnitus in humans cannot be excluded. Noise exposure showed a close interplay with aging to CS in animals (Fernandez et al., 2015; Möhrle et al., 2016), which is accordant with the higher risk of tinnitus in older people (Al-Swiahb and Park, 2016). Noise-related effect could exaggerate the age-related synaptopathy, though such an effect was expected to be "acute" (Fernandez et al., 2015). In addition, the age-related neural degeneration between SGN and IHC was observed in temporal bones in humans (Wu et al., 2019). Therefore, younger individuals are likely more resistant to noise trauma than older participants, and as a result, the presence of CS in TNHT may be combined results of noise exposure and aging.

Changed Wave V Amplitude: Central Gain or Failure of Central Gain

The current review found large heterogeneity in wave V amplitudes with increased or decreased amplitudes in tinnitus

patients. Enhanced or normal wave V amplitude with decreased wave I amplitude is consistent with the consequences of central gain following deafferentation of low-SR fibers (Schaette and McAlpine, 2011). Yet decreased amplitudes supported the "failure of central gain to restore amplitudes completely," which is hypothesized resulting from the severe loss of high-SR fibers (Knipper et al., 2020).

Several animal studies have suggested that loss of low-SR fibers reduced spontaneous and sound-evoked activities with a concomitant reduction of excitatory drive to the ascending auditory pathway (Puel et al., 1998; Wang and Green, 2011). However, homeostatic adaptation plays an important role in maintaining central gain by increasing excitation and decreasing inhibition to restore the reduced neural inputs, which could be reflected in normal or increased wave V amplitudes (Schaette and Kempter, 2006). Central gain is believed to originate between the CN and IC, where there are both excitatory and inhibitory interneurons that can interact to maintain homeostasis (Schaette and McAlpine, 2011; Auerbach et al., 2014; Sedley, 2019). Spontaneous neural activity in the central auditory system was enhanced after excessive noise exposure within a few hours or weeks and could persist in the absence of inputs (for review see Eggermont, 2017), including the DCN (Kaltenbach and Afman, 2000; Dehmel et al., 2012; Koehler and Shore, 2013), IC (Bauer et al., 2008; Mulders et al., 2010; Longenecker and Galazyuk, 2011; Manzoor et al., 2012), and the auditory cortex (AC; Sun et al., 2008; Basura et al., 2015; Eggermont, 2015; Vanneste and De Ridder, 2016). Consequently, tinnitus might be generated as a result of hyperactivity. In other words, tinnitus could be a side effect of homeostatic adaptation causing increasing spontaneous activity in the central auditory system [Schaette and Kempter, 2006, 2009; for review see Noreña (2011)].

However, the magnitude of central gain may depend on the degree of cochlear damage after noise exposure. While moderate damage produces central enhancement to compensate for reduced neural activity, severe damage to IHCs may fail to increase the spontaneous activity in the central auditory system (Schaette and Kempter, 2006). This is supported by reduced central ABR wave V amplitudes in both animal (Rüttiger et al., 2013; Singer et al., 2013; Möhrle et al., 2019) and human studies (Hofmeier et al., 2018). Loss of high-SR fibers could trigger the impairment of an inhibitory network. The development and maintenance of the fast-spiking parvalbumin-positive (PV^+) inhibitory interneurons to cortical pyramidal neurons depend on the development of high-SR fibers after hearing onset (Chumak et al., 2016). When high-SR fibers are severely damaged, the inhibitory network could be reversed into hyperexcitation rather than disinhibition, resulting in the increased spontaneous activity at the central auditory system [for review see Knipper et al. (2020)]. In addition, the PV+ interneurons actively participate in bottom–up feedforward and top–down feedback inhibition to improve sound resolution through frequency-dependent contrast amplification (Knipper et al., 2020), which is consistent with the frequency-related characteristics of residual inhibition in numerous tinnitus patients (Roberts et al., 2008). Different findings from tinnitus research may support this hypothesis. A few complaints of tinnitus from people with congenital hearing loss may point out the importance of mature high-SR fibers for the pathophysiology of tinnitus (Eggermont and Kral, 2016; Lee et al., 2017). Another evidence is that diminished high-SR fibers limited the ability to properly attenuate irrelevant stimuli over relevant information, which were reported by some tinnitus patients (Delano et al., 2007; Wittekindt et al., 2014).

As for increased wave V amplitudes in some studies, Knipper et al. (2020) regarded them as the confounding effect of hyperacusis, considering that it is also believed to result from enhanced central gain. There is a high prevalence but lack of measurement in tinnitus patients (Schecklmann et al., 2014). One neuroimaging study identified increased activation in the IC and medial geniculate body (MGB) in patients with hyperacusis but cortical activation in tinnitus and hyperacusis (Gu et al., 2010). This parallels the hypothesis of Zeng (2013) that tinnitus is the result of increased loudness by enhanced central neural noise, but hyperacusis is the result of steeper loudness growth by central gain enhancement. Möhrle et al. (2019) instead proposed that central compensation may reflect a healthy status in homeostatic adaptation processes and their ability to stabilize the discharge rate of the central auditory system, which also contradicts the hypothesis of Schaette and McAlpine (2011).

Different results of central gain in the included studies may also indicate the contribution of central gain in distinct auditory structures either from cortical or subcortical levels, though the characteristics of tinnitus may cover the hyperacusis and render it undetectable. Notably, since only a few included studies measured hyperacusis in the tinnitus group, consistently increased V/I ratios may indicate the contribution of hyperactivity triggered by either or both types of deafferentation to the generation of tinnitus or hyperacusis. In particular, central enhancement for low-level stimuli, which is mainly due to high-SR fiber loss, may provide the neural basis of tinnitus, whereas neural gain for high-level sound triggered by low-SR fiber loss may account for the generation of hyperacusis (Salvi et al., 2016). On the other hand, inconsistent wave V amplitudes may result from different subtypes of tinnitus. Although loss of high-SR fibers could explain relevant hearing impairment and prolonged latency of wave V in tinnitus patients (Hofmeier et al., 2018), the function of OHCs was not assessed. Altered ABR wave V are potentially accounted for by OHC loss around the tinnitus frequency. According to the model proposed by Schaette and Kempter (2012), the cochlear damage that reduces sound-evoked input to the central auditory system could trigger homeostatic plasticity and cause tinnitus as a side effect.

A Possible Explanation of Tinnitus With Normal Hearing Thresholds

Although numerous animal and human studies investigated how cochlear synaptopathy may generate tinnitus, a few studies considered seemingly contradictory findings of synaptopathic effect and the change in central gain in the time course of tinnitus. Animal results showed that noise exposure can result in a wide range of ribbon loss in the IHCs (30–80%), which could be from sole low-SR fiber loss to combined fiber loss (Kujawa and Liberman, 2009; Lin et al., 2011; Furman et al., 2013; Rüttiger

et al., 2013; Singer et al., 2013). Wave I amplitude could be intact after mere low-SR fiber loss (Bourien et al., 2014) and start to decrease in any point of the following damage range (Knipper et al., 2015). A previous study found that monkeys might have a much smaller degree of synaptopathy than mice when their hearing thresholds are normal, suggesting that primates could be more vulnerable to the damage to hair cells (Valero et al., 2017). On the other hand, inconsistent changes of wave V amplitude (**Figure 4**) imply that the neural effect of central gain at lower level could disappear in long-term tinnitus (all included studies recruited chronic tinnitus). Thus, it is possible that increased central gain by low-SR fiber loss may provide the necessary neural basis for tinnitus with normal hearing, including the increased spontaneous and synchronized activity at the central auditory system. When the deafferentation aggravates (up to high-SR fibers), such loss reverses the inhibitory circuit into excitatory status, modulates top–down regulation, and eventually makes tinnitus persistent.

Such hypothesis could be partially supported by the findings that alternations of noise-induced central gain may vary at different levels and during different times [for review, see Auerbach et al. (2014)]. Amplitude from the IC reduced 1-h post-noise exposure and gradually increased to normal after a week, while the evoked responses from the AC amplified immediately after noise and showed similar amplitudes as previously 1 week later (Syka et al., 1994). However, this theory cannot be used directly into the tinnitus model and need to be further verified, considering that the source of noise exposure is much more complicated in the environment.

Implications for Future Study

The present systematic review with meta-analysis attempts to gather all relevant studies and, thus, clarified whether and how the cochlear synaptopathy is involved in the mechanism of tinnitus in patients with normal hearing. The results derived from meta-analysis implied that two types of AN fiber loss may characterize different stages in the tinnitus generation and progress. It may lead to a new direction for future studies in noise-induced tinnitus.

Since cochlear synaptopathy is not able to be directly measured in the human study, it is very important to unify the definition of normal hearing thresholds in the future study, at least adopting a common one. The WHO updates the definition of normal hearing audiogram to that less than 20 dB in average of thresholds at 500, 1,000, 2,000, and 4,000 Hz (World Health Organization, 2021). Notably, since the hearing damage at high frequencies is very frequent after noise exposure, including the thresholds at 3,000, 6,000, and 8,000 Hz is highly recommended (Bramhall et al., 2019). Extended high frequencies from 12 to 16 kHz should be assessed as well. In addition, the characteristics of hyperacusis should be measured to avoid the confounding effect, since it is also believed to be triggered by increased central gain (Auerbach et al., 2014).

The latency of ABR wave I and wave V is recommended to better reveal the degree of degeneration in the AN fibers. It may also help to distinguish the high-SR fiber loss from low-SR fiber loss, considering the prolonged wave I latency when high-SR fibers were damaged (Buran et al., 2010). A future study is encouraged to recruit acute noise-induced tinnitus patients and compare with chronic tinnitus patients to find whether the synaptopathic effect on the central auditory system is different in the time course of tinnitus. A longitudinal study is highly recommended to track the possible role of CS as well as central gain in the maintenance of tinnitus. Moreover, future studies that include ABR and EEG or fMRI are expected to better investigate any cortical changes caused by CS in either the auditory or non-auditory regions.

CONCLUSION

This review highlighted a significant reduction in wave I amplitude in tinnitus patients with normal hearing thresholds. Two possible hypotheses were discussed: increased central gain triggered by low-SR fiber loss or failure of central gain caused by high-SR fiber loss. However, neither of them could solely explain the inconsistency of wave V amplitude change. Consistently increased V/I ratio may indicate the contribution to central gain in different regions, which plays an important role in the generation of tinnitus and/or hyperacusis. Further study is recommended to investigate the subcortical and cortical changes along the auditory pathway caused by noise-induced cochlear synaptopathy, which helps to reveal the roles of two mechanisms in the generation and maintenance of tinnitus.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

FC and FZ designed the study. FC and NM performed the literature search and data extraction. FC wrote the paper. FZ and WL provided critical revision of the manuscript. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

The authors thank Hannah Guest, Joaquin Valderrama, Marlies Knipper, Naomi Bramhall, Rasool Panahi, Roland Schaette, and Sreeraj Konadath for providing raw ABR data and permission to use data for meta-analysis. The authors also thank Christopher Wigham for proofreading.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2021.778197/full#supplementary-material

REFERENCES

Aedo, C., and Aguilar, E. (2020). Cochlear synaptopathy: new findings in animal and human research. *Rev. Neurosci.* 31, 605–615. doi: 10.1515/revneuro-2020-0002

Al-Swiahb, J., and Park, S. N. (2016). Characterization of tinnitus in different age groups: a retrospective review. *Noise Health* 18, 214–219. doi: 10.4103/1463-1741.189240

Auerbach, B. D., Rodrigues, P. V., and Salvi, R. J. (2014). Central gain control in tinnitus and hyperacusis. *Front. Neurol.* 5:206. doi: 10.3389/fneur.2014.00206

Axelsson, A., and Prasher, D. (2000). Tinnitus induced by occupational and leisure noise. *Noise Health* 2, 47–54. Available online at: https://www.noiseandhealth.org/text.asp?2000/2/8/47/31752

Basura, G. J., Koehler, S. D., and Shore, S. E. (2015). Bimodal stimulus timing-dependent plasticity in primary auditory cortex is altered after noise exposure with and without tinnitus. *J. Neurophysiol.* 114, 3064–3075. doi: 10.1152/jn.00319.2015

Bauer, C. A., Turner, J. G., Caspary, D. M., Myers, K. S., and Brozoski, T. J. (2008). Tinnitus and inferior colliculus activity in chinchillas related to three distinct patterns of cochlear trauma. *J. Neurosci. Res.* 86, 2564–2578. doi: 10.1002/jnr.21699

Bharadwaj, H. M., Verhulst, S., Shaheen, L., Liberman, M. C., and Shinn-Cunningham, B. G. (2014). Cochlear neuropathy and the coding of supra-threshold sound. *Front. Syst. Neurosci.* 8, 26–26. doi: 10.3389/fnsys.2014.00026

Bourien, J., Tang, Y., Batrel, C., Huet, A., Lenoir, M., Ladrech, S., et al. (2014). Contribution of auditory nerve fibers to compound action potential of the auditory nerve. *J. Neurophysiol.* 112, 1025–1039. doi: 10.1152/jn.00738.2013

Bramhall, N., Beach, E. F., Epp, B., Le Prell, C. G., Lopez-Poveda, E. A., Plack, C. J., et al. (2019). The search for noise-induced cochlear synaptopathy in humans: mission impossible? *Hearing Res.* 377:16. doi: 10.1016/j.heares.2019.02.016

Bramhall, N. F., Konrad-Martin, D., and McMillan, G. P. (2018). Tinnitus and auditory perception after a history of noise exposure: relationship to auditory brainstem response measures. *Ear Hearing* 39:881. doi: 10.1097/AUD.0000000000000544

Bullen, A., Anderson, L., Bakay, W., and Forge, A. (2019). Localized disorganization of the cochlear inner hair cell synaptic region after noise exposure. *Biol. Open* 8:38547. doi: 10.1242/bio.038547

Buran, B. N., Strenzke, N., Neef, A., Gundelfinger, E. D., Moser, T., and Liberman, M. C. (2010). Onset coding is degraded in auditory nerve fibers from mutant mice lacking synaptic ribbons. *J. Neurosci.* 30, 7587–7597. doi: 10.1523/JNEUROSCI.0389-10.2010

Chambers, A. R., Resnik, J., Yuan, Y., Whitton, J. P., Edge, A. S., Liberman, M. C., et al. (2016). Central gain restores auditory processing following near-complete cochlear denervation. *Neuron* 89, 867–879. doi: 10.1016/j.neuron.2015.12.041

Christiansen, C., MacDonald, E. N., and Dau, T. (2013). Contribution of envelope periodicity to release from speech-on-speech masking. *J. Acoust. Soc. Am.* 134, 2197–2204. doi: 10.1121/1.4816409

Chumak, T., Rüttiger, L., Lee, S. C., Campanelli, D., Zuccotti, A., Singer, W., et al. (2016). BDNF in lower brain parts modifies auditory fiber activity to gain fidelity but increases the risk for generation of central noise after injury. *Mol. Neurobiol.* 53, 5607–5627. doi: 10.1007/s12035-015-9474-x

Costalupes, J. A. (1985). Representation of tones in noise in the responses of auditory nerve fibers in cats. I. Comparison with detection thresholds. *J. Neurosci.* 5, 3261–3269. doi: 10.1523/JNEUROSCI.05-12-03261.1985

Costalupes, J. A., Young, E. D., and Gibson, D. J. (1984). Effects of continuous noise backgrounds on rate response of auditory nerve fibers in cat. *J. Neurophysiol.* 51, 1326–1344. doi: 10.1152/jn.1984.51.6.1326

Dehmel, S., Pradhan, S., Koehler, S., Bledsoe, S., and 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, 1660–1671. doi: 10.1523/JNEUROSCI.4608-11.2012

Delano, P. H., Elgueda, D., Hamame, C. M., and Robles, L. (2007). Selective attention to visual stimuli reduces cochlear sensitivity in chinchillas. *J. Neurosci.* 27, 4146–4153. doi: 10.1523/JNEUROSCI.3702-06.2007

Eggermont, J. J. (2015). The auditory cortex and tinnitus – a review of animal and human studies. *Eur. J. Neurosci.* 41, 665–676. doi: 10.1111/ejn.12759

Eggermont, J. J. (2017). Effects of long-term non-traumatic noise exposure on the adult central auditory system. Hearing problems without hearing loss. *Hear. Res.* 352, 12–22. doi: 10.1016/j.heares.2016.10.015

Eggermont, J. J., and Kral, A. (2016). Somatic memory and gain increase as preconditions for tinnitus: insights from congenital deafness. *Hear. Res.* 333, 37–48. doi: 10.1016/j.heares.2015.12.018

Fernandez, K. A., Jeffers, P. W. C., Lall, K., Liberman, M. C., and Kujawa, S. G. (2015). Aging after noise exposure: acceleration of cochlear synaptopathy in "recovered" ears. *J. Neurosci.* 35, 7509–7520. doi: 10.1523/JNEUROSCI.5138-14.2015

Furman, A. C., Kujawa, S. G., and Liberman, M. C. (2013). Noise-induced cochlear neuropathy is selective for fibers with low spontaneous rates. *J. Neurophysiol.* 110, 577–586. doi: 10.1152/jn.00164.2013

Gilles, A., Schlee, W., Rabau, S., Wouters, K., Fransen, E., and Van de Heyning, P. (2016). Decreased speech-in-noise understanding in young adults with tinnitus. *Front. Neurosci.* 10:288. doi: 10.3389/fnins.2016.00288

Glowatzki, E., and Fuchs, P. A. (2002). Transmitter release at the hair cell ribbon synapse. *Nat. Neurosci.* 5, 147–154. doi: 10.1038/nn796

Goutman, J. D., and Glowatzki, E. (2007). Time course and calcium dependence of transmitter release at a single ribbon synapse. *Proc. Natl. Acad. Sci. U. S. A.* 104, 16341–16346. doi: 10.1073/pnas.0705756104

Grose, J. H., Buss, E., and Elmore, H. (2019). Age-related changes in the auditory brainstem response and suprathreshold processing of temporal and spectral modulation. *Trends Hear.* 23:2331216519839615. doi: 10.1177/2331216519839615

Gu, J. W., Halpin, C. F., Nam, E. C., Levine, R. A., and Melcher, J. R. (2010). Tinnitus, diminished sound-level tolerance, and elevated auditory activity in humans with clinically normal hearing sensitivity. *J. Neurophysiol.* 104, 3361–3370. doi: 10.1152/jn.00226.2010

Gu, J. W., Herrmann, B. S., Levine, R. A., and Melcher, J. R. (2012). Brainstem auditory evoked potentials suggest a role for the ventral cochlear nucleus in tinnitus. *J. Assoc. Res. Otolaryngol.* 13, 819–833. doi: 10.1007/s10162-012-0344-1

Guest, H., Dewey, R. S., Plack, C. J., Couth, S., Prendergast, G., Bakay, W., et al. (2018). The Noise Exposure Structured Interview (NESI): an instrument for the comprehensive estimation of lifetime noise exposure. *Trends Hear.* 22:2331216518803213. doi: 10.1177/2331216518803213

Guest, H., Munro, K. J., Prendergast, G., Howe, S., and Plack, C. J. (2017). Tinnitus with a normal audiogram: relation to noise exposure but no evidence for cochlear synaptopathy. *Hear. Res.* 344, 265–274. doi: 10.1016/j.heares.2016.12.002

Hall, J. W. (2006). New handbook for auditory evoked responses. *Alcohol. Clin. Exp. Res.* 22, 868–875.

Hickox, A. E., Larsen, E., Heinz, M. G., Shinobu, L., and Whitton, J. P. (2017). Translational issues in cochlear synaptopathy. *Hear. Res.* 349, 164–171. doi: 10.1016/j.heares.2016.12.010

Hofmann, E., Behr, R., Neumann-Haefelin, T., and Schwager, K. (2013). Pulsatile tinnitus: imaging and differential diagnosis. *Dtsch. Arztebl. Int.* 110, 451–458. doi: 10.3238/arztebl.2013.0451

Hofmeier, B., Wolpert, S., Aldamer, E. S., Walter, M., Thiericke, J., Braun, C., et al. (2018). Reduced sound-evoked and resting-state BOLD fMRI connectivity in tinnitus. *NeuroImage Clin.* 20, 637–649. doi: 10.1016/j.nicl.2018.08.029

Jastreboff, P. J. (1990). Phantom auditory perception (tinnitus): mechanisms of generation and perception. *Neurosci. Res.* 8, 221–254. doi: 10.1016/0168-0102(90)90031-9

Joo, J. W., Jeong, Y. J., Han, M. S., Chang, Y. S., Rah, Y. C., and Choi, J. (2020). Analysis of auditory brainstem response change, according to tinnitus duration, in patients with tinnitus with normal hearing. *J. Int. Adv. Otol.* 16, 190–196. doi: 10.5152/iao.2020.7951

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. *Hear. Res.* 140, 165–172. doi: 10.1016/S0378-5955(99)00197-5

Kara, E., Aydin, K., Akbulut, A. A., Karakol, S. N., Durmaz, S., Yener, H. M., et al. (2020). Assessment of hidden hearing loss in normal hearing individuals with and without tinnitus. *J. Int. Adv. Otol.* 16:87. doi: 10.5152/iao.2020.7062

Knipper, M., Panford-Walsh, R., Singer, W., Rüttiger, L., and Zimmermann, U. (2015). Specific synaptopathies diversify brain responses and hearing

disorders: you lose the gain from early life. *Cell Tissue Res.* 361, 77–93. doi: 10.1007/s00441-015-2168-x

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. *Progr. Neurobiol.* 111, 17–33. doi: 10.1016/j.pneurobio.2013.08.002

Knipper, M., van Dijk, P., Schulze, H., Mazurek, B., Krauss, P., Scheper, V., et al. (2020). The neural bases of tinnitus: lessons from deafness and cochlear implants. *J. Neurosci.* 40, 7190–7202. doi: 10.1523/JNEUROSCI.1314-19.2020

Koehler, S. D., and Shore, S. E. (2013). Stimulus timing-dependent plasticity in dorsal cochlear nucleus is altered in tinnitus. *J. Neurosci.* 33, 19647–19656. doi: 10.1523/JNEUROSCI.2788-13.2013

Konadath, S., and Manjula, P. (2016). Auditory brainstem response and late latency response in individuals with tinnitus having normal hearing. *Intractable Rare Dis. Res.* 5, 262–268. doi: 10.5582/irdr.2016.01053

Kujawa, S. G., and Liberman, M. C. (2009). Adding insult to injury: cochlear nerve degeneration after "temporary" noise-induced hearing loss. *J. Neurosci.* 29, 14077–14085. doi: 10.1523/JNEUROSCI.2845-09.2009

Kumar, U. A., and Deepashree, S. R. (2016). Personal music systems and hearing. *J. Laryngol. Otol.* 130, 717–729. doi: 10.1017/S0022215116001031

Langguth, B., Kreuzer, P. M., Kleinjung, T., and De Ridder, D. (2013). Tinnitus: causes and clinical management. *Lancet Neurol.* 12, 920–930. doi: 10.1016/S1474-4422(13)70160-1

Lee, S. Y., Nam, D. W., Koo, J. W., De Ridder, D., Vanneste, S., and Song, J. J. (2017). No auditory experience, no tinnitus: lessons from subjects with congenital- and acquired single-sided deafness. *Hear. Res.* 354, 9–15. doi: 10.1016/j.heares.2017.08.002

Liberman, M. C. (1978). Auditory-nerve response from cats raised in a low-noise chamber. *J. Acoust. Soc. Am.* 63, 442–455. doi: 10.1121/1.381736

Liberman, M. C., Epstein, M. J., Cleveland, S. S., Wang, H., and Maison, S. F. (2016). Toward a differential diagnosis of hidden hearing loss in humans. *PLoS ONE* 11:e0162726. doi: 10.1371/journal.pone.0162726

Lin, H. W., Furman, A. C., Kujawa, S. G., and Liberman, M. C. (2011). Primary neural degeneration in the Guinea pig cochlea after reversible noise-induced threshold shift. *J. Assoc. Res. Otolaryngol.* 12, 605–616. doi: 10.1007/s10162-011-0277-0

Longenecker, R. J., and Galazyuk, A. V. (2011). Development of tinnitus in CBA/CaJ mice following sound exposure. *J. Assoc. Res. Otolaryngol.* 12, 647–658. doi: 10.1007/s10162-011-0276-1

Lopez-Poveda, E. A., and Barrios, P. (2013). Perception of stochastically undersampled sound waveforms: a model of auditory deafferentation. *Front. Neurosci.* 7:124. doi: 10.3389/fnins.2013.00124

Lorenzi, C., and Moore, B. C. J. (2007). "Role of temporal envelope and fine structure cues in speech perception: A review," in *Proceedings of the International Symposium on Auditory and Audiological Research*, 1, 263–272. Available online at: https://proceedings.isaar.eu/index.php/isaarproc/article/view/2007-25

Manzoor, N. F., Licari, F. G., Klapchar, M., Elkin, R. L., Gao, Y., Chen, G., et al. (2012). Noise-induced hyperactivity in the inferior colliculus: its relationship with hyperactivity in the dorsal cochlear nucleus. *J. Neurophysiol.* 108, 976–988. doi: 10.1152/jn.00833.2011

McFadden, D., and Champlin, C. A. (2000). Comparison of auditory evoked potentials in heterosexual, homosexual, and bisexual males and females. *J. Assoc. Res. Otolaryngol.* 1, 89–99. doi: 10.1007/s101620010008

Melcher, J. R., and Kiang, N. Y. (1996). Generators of the brainstem auditory evoked potential in cat. III: identified cell populations. *Hear. Res.* 93, 52–71. doi: 10.1016/0378-5955(95)00200-6

Middleton, J. W., Kiritani, T., Pedersen, C., Turner, J. G., Shepherd, G. M., and Tzounopoulos, T. (2011). Mice with behavioral evidence of tinnitus exhibit dorsal cochlear nucleus hyperactivity because of decreased GABAergic inhibition. *Proc. Natl. Acad. Sci. U. S. A.* 108, 7601–7606. doi: 10.1073/pnas.1100223108

Milloy, V., Fournier, P., Benoit, D., Noreña, A., and Koravand, A. (2017). Auditory brainstem responses in tinnitus: a review of who, how, and what? *Front. Aging Neurosci.* 9:237. doi: 10.3389/fnagi.2017.00237

Moher, D., Liberati, A., Tetzlaff, J., and Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann. Internal Med.* 151, 264–269. doi: 10.7326/0003-4819-151-4-200908180-00135

Möhrle, D., Hofmeier, B., Amend, M., Wolpert, S., Ni, K., Bing, D., et al. (2019). Enhanced central neural gain compensates acoustic trauma-induced cochlear impairment, but unlikely correlates with tinnitus and hyperacusis. *Neuroscience* 407, 146–169. doi: 10.1016/j.neuroscience.2018.12.038

Möhrle, D., Ni, K., Varakina, K., Bing, D., Lee, S. C., Zimmermann, U., et al. (2016). Loss of auditory sensitivity from inner hair cell synaptopathy can be centrally compensated in the young but not old brain. *Neurobiol. Aging* 44, 173–184. doi: 10.1016/j.neurobiolaging.2016.05.001

Mulders, W. H., Seluakumaran, K., and Robertson, D. (2010). Efferent pathways modulate hyperactivity in inferior colliculus. *J. Neurosci.* 30, 9578–9587. doi: 10.1523/JNEUROSCI.2289-10.2010

Nemati, S., Faghih Habibi, A., Panahi, R., and Pastadast, M. (2014). Cochlear and brainstem audiologic findings in normal hearing tinnitus subjects in comparison with non-tinnitus control group. *Acta Med. Iran* 52, 822–826.

Noreña, A. J. (2011). An integrative model of tinnitus based on a central gain controlling neural sensitivity. *Neurosci. Biobehav. Rev.* 35, 1089–1109. doi: 10.1016/j.neubiorev.2010.11.003

Paul, B. T., Bruce, I. C., and 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. doi: 10.1016/j.heares.2016.11.010

Puel, J. L., Ruel, J., Gervais d'Aldin, C., and Pujol, R. (1998). Excitotoxicity and repair of cochlear synapses after noise-trauma induced hearing loss. *Neuroreport* 9, 2109–2114. doi: 10.1097/00001756-199806220-00037

Rauschecker, J. P., Leaver, A. M., and Mühlau, M. (2010). Tuning out the noise: limbic-auditory interactions in tinnitus. *Neuron* 66, 819–826. doi: 10.1016/j.neuron.2010.04.032

Roberts, L. E., Eggermont, J. J., Caspary, D. M., Shore, S. E., Melcher, J. R., and Kaltenbach, J. A. (2010). Ringing ears: the neuroscience of tinnitus. *J. Neurosci.* 30, 14972–14979. doi: 10.1523/JNEUROSCI.4028-10.2010

Roberts, L. E., Moffat, G., Baumann, M., Ward, L. M., and Bosnyak, D. J. (2008). Residual inhibition functions overlap tinnitus spectra and the region of auditory threshold shift. *J. Assoc. Res. Otolaryngol.* 9, 417–435. doi: 10.1007/s10162-008-0136-9

Robles, L., and Ruggero, M. A. (2001). Mechanics of the mammalian cochlea. *Physiol. Rev.* 81, 1305–1352. doi: 10.1152/physrev.2001.81.3.1305

Rüttiger, L., Singer, W., Panford-Walsh, R., Matsumoto, M., Lee, S. C., Zuccotti, A., et al. (2013). The reduced cochlear output and the failure to adapt the central auditory response causes tinnitus in noise exposed rats. *PLoS ONE* 8:e57247. doi: 10.1371/journal.pone.0057247

Salvi, R., Sun, W., Ding, D., Chen, G. D., Lobarinas, E., Wang, J., et al. (2016). Inner hair cell loss disrupts hearing and cochlear function leading to sensory deprivation and enhanced central auditory gain. *Front. Neurosci.* 10:621. doi: 10.3389/fnins.2016.00621

Sanchez, T. G., Medeiros, I. R. T. d., Levy, C. P. D., Ramalho, J. d. R. O., and Bento, R. F. (2005). Tinnitus in normally hearing patients: clinical aspects and repercussions. *Brazil. J. Otorhinolaryngol.* 71, 427–431. doi: 10.1016/S1808-8694(15)31194-0

Schaette, R., and Kempter, R. (2006). Development of tinnitus-related neuronal hyperactivity through homeostatic plasticity after hearing loss: a computational model. *Eur. J. Neurosci.* 23, 3124–3138. doi: 10.1111/j.1460-9568.2006.04774.x

Schaette, R., and Kempter, R. (2009). Predicting tinnitus pitch from patients' audiograms with a computational model for the development of neuronal hyperactivity. *J. Neurophysiol.* 101, 3042–3052. doi: 10.1152/jn.91256.2008

Schaette, R., and Kempter, R. (2012). Computational models of neurophysiological correlates of tinnitus. *Front. Syst. Neurosci.* 6:34. doi: 10.3389/fnsys.2012.00034

Schaette, R., and McAlpine, D. (2011). Tinnitus with a normal audiogram: physiological evidence for hidden hearing loss and computational model. *J. Neurosci.* 31, 13452–13457. doi: 10.1523/JNEUROSCI.2156-11.2011

Schecklmann, M., Landgrebe, M., and Langguth, B. (2014). Phenotypic characteristics of hyperacusis in tinnitus. *PLoS ONE* 9:e86944. doi: 10.1371/journal.pone.0086944

Sedley, W. (2019). Tinnitus: does gain explain? *Neuroscience* 407, 213–228. doi: 10.1016/j.neuroscience.2019.01.027

Sergeyenko, Y., Lall, K., Liberman, M. C., and Kujawa, S. G. (2013). Age-related cochlear synaptopathy: an early-onset contributor to auditory functional decline. *J. Neurosci.* 33, 13686–13694. doi: 10.1523/JNEUROSCI.1783-13.2013

Sheldrake, J., Diehl, P. U., and Schaette, R. (2015). Audiometric characteristics of hyperacusis patients. *Front. Neurol.* 6:105. doi: 10.3389/fneur.2015.00105

Shim, H. J., An, Y.-H., Kim, D. H., Yoon, J. E., and Yoon, J. H. (2017). Comparisons of auditory brainstem response and sound level tolerance in tinnitus ears and non-tinnitus ears in unilateral tinnitus patients with normal audiograms. *PLoS ONE* 12:e0189157. doi: 10.1371/journal.pone.0189157

Singer, W., Panford-Walsh, R., and Knipper, M. (2014). The function of BDNF in the adult auditory system. *Neuropharmacology* 76, 719–728. doi: 10.1016/j.neuropharm.2013.05.008

Singer, W., Zuccotti, A., Jaumann, M., Lee, S. C., Panford-Walsh, R., Xiong, H., et al. (2013). Noise-induced inner hair cell ribbon loss disturbs central arc mobilization: a novel molecular paradigm for understanding tinnitus. *Mol. Neurobiol.* 47, 261–279. doi: 10.1007/s12035-012-8372-8

Song, K., Shin, S. A., Chang, D. S., and Lee, H. Y. (2018). Audiometric profiles in patients with normal hearing and bilateral or unilateral tinnitus. *Otol. Neurotol.* 39, e416–e421. doi: 10.1097/MAO.0000000000001849

Stang, A. (2010). Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. *Eur. J. Epidemiol.* 25, 603–605. doi: 10.1007/s10654-010-9491-z

Sulaiman, A. H., Husain, R., and Seluakumaran, K. (2014). Evaluation of early hearing damage in personal listening device users using extended high-frequency audiometry and otoacoustic emissions. *Eur. Arch. Otorhinolaryngol.* 271, 1463–1470. doi: 10.1007/s00405-013-2612-z

Sun, W., Zhang, L., Lu, J., Yang, G., Laundrie, E., and Salvi, R. (2008). Noise exposure-induced enhancement of auditory cortex response and changes in gene expression. *Neuroscience* 156, 374–380. doi: 10.1016/j.neuroscience.2008.07.040

Syka, J., Rybalko, N., and Popelár, J. (1994). Enhancement of the auditory cortex evoked responses in awake guinea pigs after noise exposure. *Hear. Res.* 78, 158–168. doi: 10.1016/0378-5955(94)90021-3

Turrigiano, G. G. (1999). Homeostatic plasticity in neuronal networks: the more things change, the more they stay the same. *Trends Neurosci.* 22, 221–227. doi: 10.1016/S0166-2236(98)01341-1

Valderrama, J. T., Beach, E. F., Yeend, I., Sharma, M., Van Dun, B., and Dillon, H. (2018). Effects of lifetime noise exposure on the middle-age human auditory brainstem response, tinnitus and speech-in-noise intelligibility. *Hear. Res.* 365, 36–48. doi: 10.1016/j.heares.2018.06.003

Valero, M. D., Burton, J. A., Hauser, S. N., Hackett, T. A., Ramachandran, R., and Liberman, M. C. (2017). Noise-induced cochlear synaptopathy in rhesus monkeys (*Macaca mulatta*). *Hear. Res.* 353, 213–223. doi: 10.1016/j.heares.2017.07.003

Vanneste, S., and De Ridder, D. (2016). Deafferentation-based pathophysiological differences in phantom sound: tinnitus with and without hearing loss. *NeuroImage* 129, 80–94. doi: 10.1016/j.neuroimage.2015.12.002

Verhulst, S., Jagadeesh, A., Mauermann, M., and Ernst, F. (2016). Individual differences in auditory brainstem response wave characteristics: relations to different aspects of peripheral hearing loss. *Trends Hear.* 20:2331216516672186. doi: 10.1177/2331216516672186

Wang, Q., and Green, S. H. (2011). Functional role of neurotrophin-3 in synapse regeneration by spiral ganglion neurons on inner hair cells after excitotoxic trauma *in vitro*. *J. Neurosci.* 31, 7938–7949. doi: 10.1523/JNEUROSCI.1434-10.2011

Wittekindt, A., Kaiser, J., and Abel, C. (2014). Attentional modulation of the inner ear: a combined otoacoustic emission and EEG study. *J. Neurosci.* 34, 9995–10002. doi: 10.1523/JNEUROSCI.4861-13.2014

World Health Organization. (2021). *World Report on Hearing*.

Wu, C., Martel, D. T., and Shore, S. E. (2016). Increased synchrony and bursting of dorsal cochlear nucleus fusiform cells correlate with tinnitus. *J. Neurosci.* 36, 2068–2073. doi: 10.1523/JNEUROSCI.3960-15.2016

Wu, P. Z., Liberman, L. D., Bennett, K., de Gruttola, V., O'Malley, J. T., and Liberman, M. C. (2019). Primary neural degeneration in the human cochlea: evidence for hidden hearing loss in the aging ear. *Neuroscience* 407, 8–20. doi: 10.1016/j.neuroscience.2018.07.053

You, S., Kong, T. H., and Han, W. (2020). The effects of short-term and long-term hearing changes on music exposure: a systematic review and meta-analysis. *Int. J. Environ. Res. Public Health* 17:62091. doi: 10.3390/ijerph17062091

Zeng, F. G. (2013). An active loudness model suggesting tinnitus as increased central noise and hyperacusis as increased nonlinear gain. *Hear. Res.* 295, 172–179. doi: 10.1016/j.heares.2012.05.009

Zhao, F., Manchaiah, V. K. C., French, D., and Price, S. M. (2010). Music exposure and hearing disorders: an overview. *Int. J. Audiol.* 49, 54–64. doi: 10.3109/14992020903202520

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 04 January 2022
doi: 10.3389/fnins.2021.784721

Objective Recognition of Tinnitus Location Using Electroencephalography Connectivity Features

Zhaobo Li[1], Xinzui Wang[1,2], Weidong Shen[3], Shiming Yang[3], David Y. Zhao[4], Jimin Hu[5], Dawei Wang[5], Juan Liu[1], Haibing Xin[1], Yalun Zhang[1], Pengfei Li[1], Bing Zhang[1], Houyong Cai[1], Yueqing Liang[1] and Xihua Li[1]*

[1] Jihua Laboratory, Foshan, China, [2] Suzhou Institute of Biomedical Engineering and Technology, Chinese Academy of Sciences, Suzhou, China, [3] Department of Otolaryngology, Head and Neck Surgery, Chinese PLA General Hospital, Institute of Otolaryngology, Beijing, China, [4] BetterLifeMedical, Suzhou, China, [5] Jiangsu Testing and Inspection Institute for Medical Devices, Nanjing, China

OPEN ACCESS

Edited by:
Patrick Krauss,
University of Erlangen Nuremberg, Germany

Reviewed by:
Stefan Schoisswohl,
University Medical Center Regensburg, Germany
Konstantin Tziridis,
University Hospital Erlangen, Germany

***Correspondence:**
Xinzui Wang
wangxz@jihualab.com

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

Received: 28 September 2021
Accepted: 16 November 2021
Published: 04 January 2022

Citation:
Li Z, Wang X, Shen W, Yang S, Zhao DY, Hu J, Wang D, Liu J, Xin H, Zhang Y, Li P, Zhang B, Cai H, Liang Y and Li X (2022) Objective Recognition of Tinnitus Location Using Electroencephalography Connectivity Features. Front. Neurosci. 15:784721. doi: 10.3389/fnins.2021.784721

Purpose: Tinnitus is a common but obscure auditory disease to be studied. This study will determine whether the connectivity features in electroencephalography (EEG) signals can be used as the biomarkers for an efficient and fast diagnosis method for chronic tinnitus.

Methods: In this study, the resting-state EEG signals of tinnitus patients with different tinnitus locations were recorded. Four connectivity features [including the Phase-locking value (PLV), Phase lag index (PLI), Pearson correlation coefficient (PCC), and Transfer entropy (TE)] and two time-frequency domain features in the EEG signals were extracted, and four machine learning algorithms, included two support vector machine models (SVM), a multi-layer perception network (MLP) and a convolutional neural network (CNN), were used based on the selected features to classify different possible tinnitus sources.

Results: Classification accuracy was highest when the SVM algorithm or the MLP algorithm was applied to the PCC feature sets, achieving final average classification accuracies of 99.42 or 99.1%, respectively. And based on the PLV feature, the classification result was also particularly good. And MLP ran the fastest, with an average computing time of only 4.2 s, which was more suitable than other methods when a real-time diagnosis was required.

Conclusion: Connectivity features of the resting-state EEG signals could characterize the differentiation of tinnitus location. The connectivity features (PCC and PLV) were more suitable as the biomarkers for the objective diagnosing of tinnitus. And the results were helpful for clinicians in the initial diagnosis of tinnitus.

Keywords: tinnitus location, objective recognition, resting-state EEG, connectivity features, deep learning algorithms

INTRODUCTION

Tinnitus refers to patient's perception of sound in the ear without any external sound or electrical stimulation. Clinically, tinnitus is divided into subjective and objective tinnitus, and most patients report subjective tinnitus (Smit et al., 2015). There are 5–15% of people in the world who have experienced tinnitus, among them, 1–3% of tinnitus patients' everyday life is seriously affected and need medical treatment (Tunkel et al., 2014; Gallus et al., 2015). Because tinnitus is a subjective perception for patients, its clinical detection and treatment are significant challenges (Pan et al., 2009). A multidisciplinary European guideline for tinnitus points out uniformity in the assessment and treatment of adult patients with subjective tinnitus (Cima et al., 2019). However, for the initial diagnosis of tinnitus, an efficient and objective method for recognizing and diagnosing tinnitus is still urgently needed.

Tinnitus is a health condition, not a disease. It is a symptom of pathological neurological activity that manifests as an unwanted auditory hallucination (Henry et al., 2020), and the complex pathophysiologic mechanism of tinnitus is still poorly understood.

On a sensory level, tinnitus is assumed to be caused by cochlear damage, but many tinnitus patients have normal audiograms, that is, there are no direct signs of cochlear damage (Schaette and McAlpine, 2011; Eggermont and Roberts, 2012). And the severance of the auditory nerve does not eliminate the sensation of tinnitus, and tinnitus is also not associated with the excessive activity of the auditory nerve (Müller et al., 2003). In rat experiments, the degree of reorganization of the auditory cortex is related to the intensity of tinnitus (Engineer et al., 2011). Tinnitus may originate in the dorsal cochlear nucleus (DCN) (Krauss et al., 2016), which is the auditory brainstem. The pathological changes in the activity of spontaneous neurons in the auditory brainstem can drive the reorganization of the auditory cortex (Rauschecker et al., 2010). A survey of rock musicians who felt transient tinnitus after practicing in a loud band (∼120 dB SPL for ∼2 h), showed that the gamma frequency power of the right temporal cortex was significantly increased during the appearance of transient tinnitus (Schlee et al., 2011), and importantly, which was not correlated with the degree of hearing loss (Ohlemiller et al., 2011).

On a macroscopic level, tinnitus was related to abnormally oscillating brain activity patterns (Schoisswohl et al., 2021), and neuroimaging studies also showed that there was a hyperactive auditory cortex during tinnitus (Mirz et al., 1999; Cunnane, 2019). Weisz et al. (2005) found that there was a significant enhancement of delta (1–4 Hz) and gamma (40–90 Hz) frequency power and a concomitant reduction of alpha (8–12 Hz) activity in tinnitus patients, and these changes mainly occurred in the temporal area. This statement was very consistent with the thalamocortical dysrhythmia model (TCD) (Llinás et al., 2005; Vanneste et al., 2018). A decrease in alpha power was associated with an increase in cortical excitability (Sauseng et al., 2005; Klimesch et al., 2007; Romei et al., 2008), the alpha desynchronization observed in chronic tinnitus reflected the release of inhibition, and thus facilitated the synchronization of neuronal activity. The synchronization by loss of inhibition model (SLIM) also supported this view, which explained the enhanced synchronization of auditory activity by a reduction of cortical inhibition (Weisz et al., 2007). However, the changes in brain activity that accompany tinnitus were not limited to the auditory cortices, a global brain model of tinnitus pointed out the phase coupling between the anterior cingulate cortex (ACC) and the right frontal lobe and the phase coupling between the ACC and the right parietal lobe were related to tinnitus. At the same time, in participants with a shorter tinnitus history, the gamma network was mainly distributed in the left temporal cortex, while in participants with a longer tinnitus duration, the gamma network was more widely distributed throughout the cortex (Schlee et al., 2011). This also reminded us of the potential relationship between the global brain function network connection and tinnitus.

Schlee et al. (2008) showed the existence of a tinnitus network by using a magnetoencephalogram (MEG) and found a wide range of abnormal functional connections in the brain. When conducting brain network anatomy research, we should focus on the spatial information interaction between brain regions and nodes (Zhang X. et al., 2021). Different brain regions did not work independently of each other, instead, they were connected in various long-distance networks (Bressler and Menon, 2010). One study on tinnitus patients found aberrant functional connectivity within the default mode network (DMN) (Cai et al., 2019). Chronic tinnitus patients had abnormal functional connectivity networks originating from ACC to other brain regions that were associated with specific tinnitus characteristics (Chen et al., 2018). Therefore, the brain functional network information may be a potential biomarker for tinnitus. Though those existing studies have evidenced the abnormal functional network in tinnitus patients, whether functional connection information can be used as the biomarker for tinnitus is less probed in previously reported studies.

In past studies, the time-domain features (rhythm signals) and frequency-domain features (power spectral density, PSD) of EEG signals were usually used as biomarkers to distinguish tinnitus and non-tinnitus, but these features were not obvious, and the accuracy of the identification was as high as 87% (Wang et al., 2017; Vanneste et al., 2018). We hope that functional connectivity features can better reflect the pathological features of tinnitus, and it will greatly improve the accuracy of the identification of tinnitus by using functional connectivity features as the biomarker.

At the same time, among studies that reported the localization of tinnitus-related signals, in about half of tinnitus patients, the percept occurred in the middle of the head or both ears. For others, tinnitus was perceived to be mainly on the left side, and some patients even feel that their tinnitus came from outside of the head (Baguley et al., 2013). The tinnitus might be heard as unilateral or bilateral (Meikle and Taylor-Walsh, 1984), some studies pointed out unilateral tinnitus was more commonly associated with underlying disease processes than bilateral tinnitus (Chari and Limb, 2018). Therefore, determining the location of tinnitus was one of the most important steps for the treatment of tinnitus in the cortex (Lee et al., 2019), because this information determined the type or order of treatments that the patient was given (Herraiz, 2008), and which might

also enable us to probe the neurophysiology of tinnitus in ways heretofore not considered (Searchfield et al., 2015). If we can identify tinnitus symptoms and determine the location of tinnitus by using functional connectivity features, it will be of great help for clinicians.

Our research aim is to find the optimal solution for the objective recognition of possible tinnitus sources. As far as we know, there is no reference for the selection of functional connectivity features in the field of tinnitus recognition. We considered three functional connectivity features that were widely used in the field of emotion recognition (Lee and Hsieh, 2014; Gao et al., 2020; Moon et al., 2020), including the PCC, PLV, and TE. At the same time, the PLI was also included by us as a connected feature. PLV and PLI can measure the phase synchronization between time series nodes, and PCC and TE can measure the correlation between time series nodes, both of which can represent the connection status of network functions. At the same time, we also selected the mean and standard deviation of the rhythm of the six bands of EEG, such as δ delta (1–3.5 Hz), θ theta (4–7.5 Hz), α alpha (8–12 Hz), β beta (12.5–30 Hz), low γ gamma (30.5–48 Hz) and high γ(52–90 Hz), as well as the PSD, as the comparison features. We selected four classification algorithms, included two SVM, a MLP, and a CNN, to combine with the extracted features.

MATERIALS AND METHODS

Participants

Participants with chronic tinnitus were recruited from the Otolaryngology-Head and Neck Surgery Department of Chinese PLA General Hospital and other hospitals from October 2018 to July 2019. Participants without tinnitus and hearing loss were recruited from the general population.

All participants in the tinnitus group had signs of definitive tinnitus for at least 3 months, and they also had normal hearing on the conventional hearing tests. Normal hearing was defined as follows: (1) A pure tone audiometric (PTA) threshold of 25 dB hearing level (HL) or better for all octave frequencies from 250 to 8,000 Hz; (2) transient evoked otoacoustic emission (OAE) with a signal-to-noise ratio (SNR) $>$ 5 dB and a distortion product OAE with an SNR $>$ 3 dB on OAE tests; (3) a waves I-III inter-peak latency $<$ 2.4 ms and a wave V latency $<$ 6.2 ms on 90 dB nHL click-evoked auditory brainstem response (ABR) tests; and (4) a normal tympanic membrane on otoscopy (Ahn et al., 2017). It had been proposed that the people with tinnitus who showed normal audiograms could have hidden hearing loss defined as damage to the auditory nerve fibers (Gollnast et al., 2017; Tziridis et al., 2021). Despite these possibilities, all patients with tinnitus in this study showed normal latencies in waves I-III of the ABRs and normal OAEs, which usually indicated the integrity of the peripheral auditory nerves (Møller et al., 1981; Møller and Jannetta, 1982) and normal function of the cochlear hair cells (Kemp, 1978; Mills and Rubel, 1994). To reduce the possibility of including patients with hidden hearing loss, and to ensure the cognitive abilities between patients and healthy participants were comparable, we included subjects who (1) were between 20 and 50 years old; (2) had no current or previous history of vertigo, Meniere's disease, noise exposure, hyperacusis, or psychiatric problems; (3) had no history of head trauma or central nervous system disease and no anxiety or depression; (4) were not exposed to ototoxic drugs; and (5) had no complex or poorly defined tinnitus. Additionally, all patients completed a tinnitus questionnaire, which included a visual analog scale (VAS) and the Tinnitus Handicap Inventory (THI) questionnaire.

Forty-two participants with valid EEG data were selected for the study, which included 10 healthy participants, 12 bilateral tinnitus patients, 8 right-sided tinnitus patients, and 12 left-sided tinnitus patients (**Table 1**). The difference between the age and sex or duration of all participants was analyzed using the between-group $t\sim$test, and $p > 0.05$. There was not a significant difference between these groups in terms of age and sex or duration. All participants were informed about the background and purpose of this study, and they all gave written informed consent. This study complied with the ethical principles of the Declaration of Helsinki.

Electroencephalography Recordings

The 64-channel Neuroscan device recorded the EEG signals. The electrode position was set according to the international 10–20 electrode distribution. The sampling rate was 1,000 Hz, and the impedances were kept below 10 k**Ω**. During the EEG recording, the participants were asked to stay still, awake, and close their eyes in a sound and electrically shielded room. Each participant's resting-state EEG signals were recorded for an estimated 10 min. According to reports, the global connectional properties in the brain stabilize with acquisition times as little as 5 min (van Dijk et al., 2010). And to ensure the availability of data, after acquiring the EEG data, the subject was asked whether he\she was sleepy during the resting state.

Data Preprocessing

We used MATLAB-EEGLAB v14.1.2 toolbox for preprocessing. First, we performed interpolation processing on the insufficient lead data and deleted abnormally fluctuating time-period signals. Then, EEG signals were filtered by 50 Hz power frequency and bandpass-filtered to 0.5–90 Hz. Some research methods directly performed 0.5–48 hz band-pass filtering and ignored the beneficial components in the high-frequency signal, but the high-frequency γ rhythm (52–90 hz) was also related to tinnitus (Pierzycki et al., 2016). Next, baseline correction and REST re-reference of EEG signals (Yao, 2001) were performed. Last, the interference signals in EEG signals include ocular electricity, electrocardiogram (ECG), and other artifacts were filtered out by independent component analysis (ICA). The preprocessed signals were segmented at an interval of 10 s to expand the sample capacity.

Thalamocortical dysrhythmia (TCD) was a model proposed to explain divergent neurological disorders (Vanneste et al., 2018), which utilized the resting-state EEG signals of the left and right auditory cortex (AUD) areas to perform oscillation analysis of

TABLE 1 | Overview of the patients with tinnitus.

Number	Sex	Age (year)	Localization	Duration (患病时间)	Tinnitus frequency (kHz)
1	Female	21	Both ears	2 y	4
2	Male	25	Both ears	3 y	0.5
3	Male	23	Both ears	3 y	8
4	Male	43	Both ears	1 y 6 m	4
5	Male	50	Both ears	6 y	4
6	Male	40	Both ears	8 y	4
7	Male	32	Both ears	1 y	4
8	Male	29	Both ears	0 y 3 m	4
9	Female	25	Both ears	0 y 6 m	0.5
10	Female	31	Both ears	5 y	4
11	Female	36	Both ears	5 y	4
12	Female	48	Both ears	7 y	4
13	Female	37	Right ear	4 y	0.125
14	Female	44	Right ear	0 y 4 m	8
15	Female	50	Right ear	1 y	3
16	Male	28	Right ear	0 y 3 m	4
17	Female	45	Right ear	0 y 3 m	8
18	Male	42	Right ear	1 y	0.5
19	Male	30	Right ear	3 y	4
20	Female	44	Right ear	1 y	8
21	Male	31	Left ear	1 y	0.5
22	Female	31	Left ear	0 y 6 m	0.125
23	Female	45	Left ear	0 y 6 m	0.125
24	Male	37	Left ear	0 y 3 m	0.125
25	Male	42	Left ear	0 y 6 m	0.2
26	Male	25	Left ear	0 y 3 m	6
27	Female	38	Left ear	0 y 3 m	0.15
28	Female	25	Left ear	0 y 4 m	1
29	Female	46	Left ear	1 y	2.175
30	Female	45	Left ear	5 y	6
31	Male	48	Left ear	1 y	4
32	Female	28	Left ear	0 y 5 m	4

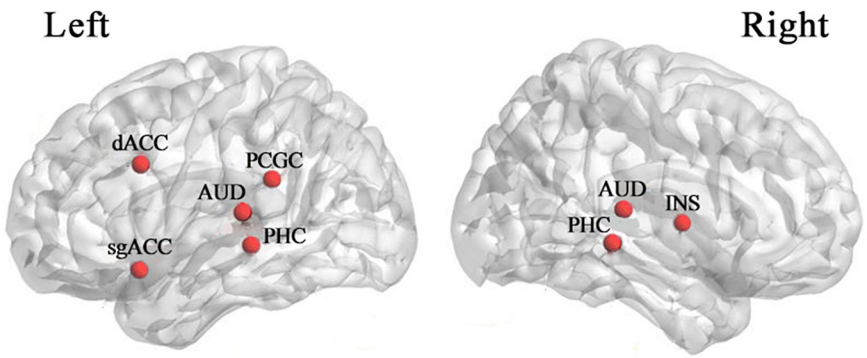


FIGURE 1 | Cortical areas are associated with tinnitus (Vanneste et al., 2018). dACC dorsal anterior cingulate cortex, sgACC subgenual anterior cingulate cortex, PCGC posterior cingulate cortex, AUD auditory cortex, PHC parahippocampus, INS insula.

tinnitus disorders (**Figure 1**), including the parahippocampus (PHC), the dorsal anterior cingulate cortex (dACC), the subgenual anterior cingulate cortex (sgACC), the posterior cingulate cortex (PCGC) and the right insula (rINS). Imaging studies in patients with tinnitus had shown functional and structural abnormalities distributed in the auditory areas of the brain, including the cingulate cortex, PHC, and INS, and there was an increase in the activation of ACC and INS

(Elgoyhen et al., 2015). Therefore, this study also selected the channel data of the auditory cortex on the left and right sides of the brain to be included in the next calculation and analysis, as shown in **Figure 2**, the channels including FT7, FC5, C5, T7, TP7, CP5, FC6, FT8, C6, T8, CP6, TP8. The efficiency of data processing can be greatly improved by reducing the calculation of channel data, and this targeted reduction will not make us lose important reference information. In the future, we may apply this research method to clinical diagnosis. As we all know, the 12-channel EEG acquisition experiment is easier than the 64-channel acquisition experiment, and the simplified method can also improve the efficiency of clinical diagnosis.

Connectivity Measures

Four different connectivity measures were used in this study: PLV, PLI, PCC, and TE, which denoted the connection characteristics between different EEG channel signals. This study also considered the time-frequency domain features, including the rhythm and PSD.

Phase Locking Value

The PLV (Lachaux et al., 1999) denotes the phase synchronization between two signals, which is calculated as the average of the absolute phase difference. The value range of PLV is from 0 to 1, indicating perfectly independent or perfectly synchronization

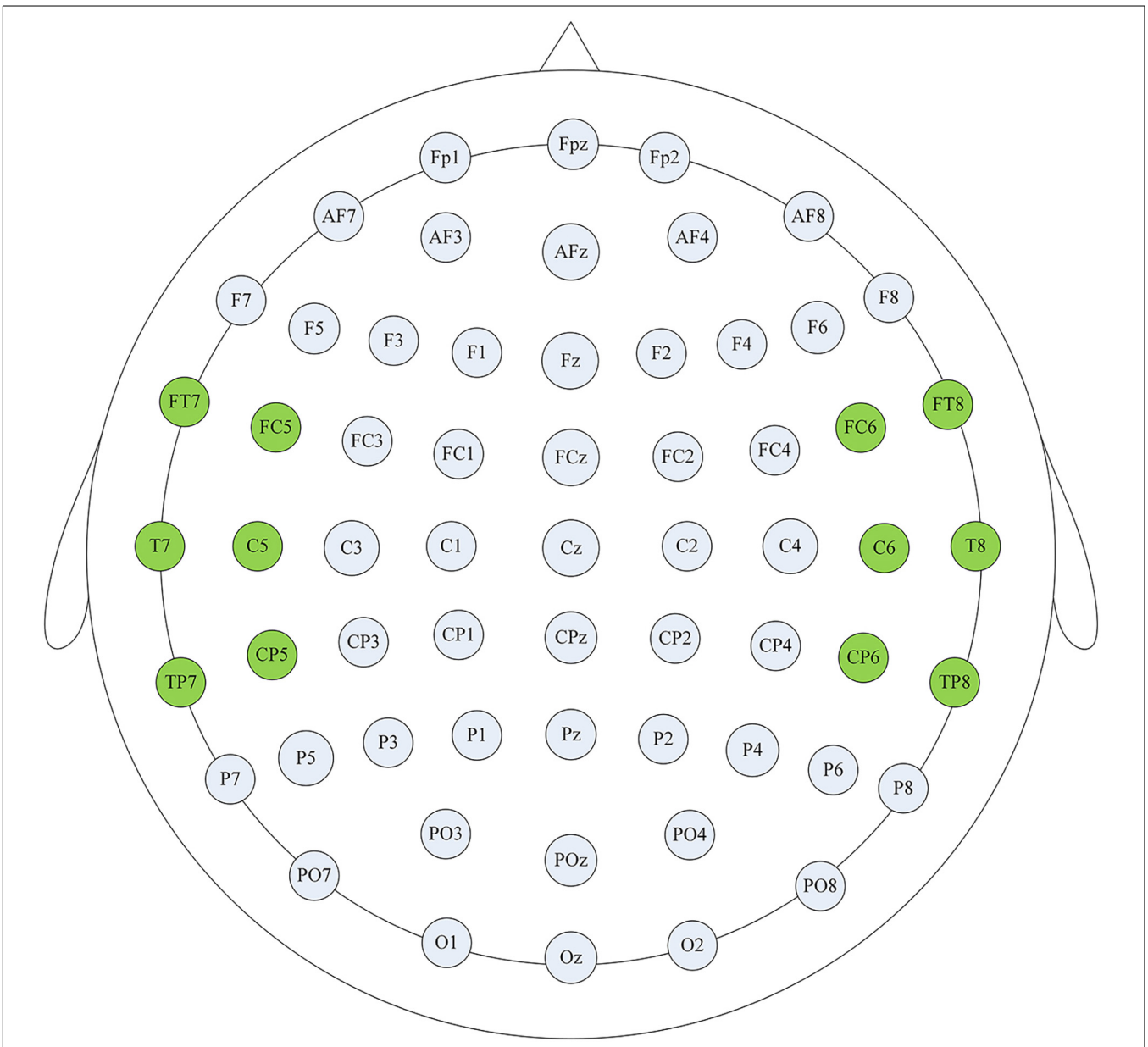


FIGURE 2 | Channel selection display. The highlighted channels cover the auditory cortex on the left and right sides of the brain, and these channel data were used for subsequent analysis to improve the efficiency of clinical diagnosis.

TABLE 2 | CNN-LSTM model architecture.

Layers	Types	Activation function	Output shape	Size of kernels	Filters/units	Stride	Rate of dropout
0	Input	–	1 × 66	–	–	–	–
1	1DCNN	Relu	1 × 16	3 × 1	16	1	–
2	1DCNN	Relu	1 × 32	3 × 1	32	1	–
3	LSTM	Tanh	1 × 32	–	32	–	–
4	LSTM	Tanh	32	–	32	–	–
5	Batch normalization	–	32	–	–	–	–
6	Flatten	–	32	–	–	–	–
7	Dense	Relu	16	–	16	–	–
8	Dropout	–	16	–	–	–	0.2
9	Dense	Relu	8	–	8	–	–
10	Dropout	–	8	–	–	–	0.2

of the two signals, respectively. PLV is also an undirected connectivity feature.

$$PLV\left(i,k\right)=\frac{1}{T}\left|\sum_{t=1}^{T}\left(\varphi_i^t-\varphi_k^t\right)\right|$$

Where φ^t is the phase of the signal at time t.

Phase Lag Index

PLI (Stam et al., 2007) is also a measure of phase synchronization. PLI between the signals is defined as follows:

$$PLI\left(i,k\right)=\frac{1}{T}\sum_{t=1}^{T}sgn\left(Im(Z_i(t)Z_k(t)*)\right)$$

where $Z_i(t)$ and $Z_k(t)$ for t = 1, …, T is the time-frequency representations of i and k signal in a given frequency band, note that Z_i and Z_k are complex-valued, *Sgn* denotes the sign function and the superscript, * denotes complex conjugation.

Pearson Correlation Coefficient

The PCC measures the linear relationship between two signals as a continuous number in [−1,1]. PCC values of −1 and 1 correspond to perfectly negative and positive linear relationships, respectively, and a PCC value of zero indicates that the two signals are not correlated. Let $x_i \quad \left\{x_i^1, x_i^2, \ldots\ldots, x_i^T\right\}$ indicate an EEG signal of the ith electrode, where T is the time length of the signal. The PCC of two signals $x_i$and $x_k$is calculated as:

$$PCC\left(i,k\right)=\frac{\frac{1}{T}\sum_{t=1}^{T}\left[\left(x_i^t-\mu_i\right)*\left(x_k^t-\mu_k\right)\right]}{\sigma_i*\sigma_k}$$

where μ and σ are the mean and standard deviation of the signal, respectively.

Transfer Entropy

The TE (Schreiber, 2000) describes the directed flow of information from a signal x_i to another signal x_k:

$$TE\left(i\rightarrow k\right)$$
$$=\frac{1}{T-1}\sum_{t=1}^{T-1}\left[p\left(x_i^t,x_k^t,x_k^{t+1}\right)*\log_2\left(\frac{p\left(x_k^{t+1}\mid x_i^t,x_k^t\right)}{p\left(x_k^{t+1}\mid x_k^t\right)}\right)\right]$$

A TE value of zero indicates that there is no causal relationship between the two signals, and TE belongs to the directed connectivity feature.

Classification Methods

We selected four different machine learning algorithms to avoid the classification differences caused by the algorithm itself. The first two were the support vector machines (SVM) with a linear kernel function and two different cross-validation methods to improve the classification accuracy of the model, one was leave-one-out cross-validation (Loo-CV), and the other was 10-fold cross-validation (10-CV); The third was a multi-layer perception network (MLP), with parameter settings: the optimizer was "adam", the alpha was 10^{-7}, the hidden layer nodes were 200 for each of the two layers, the maximum number of iterations was 380, the initial learning rate was 0.001, and other parameters were the default parameter settings; The fourth was convolutional neural network (CNN) with long short-term memory model (CNN-LSTM), the layer details and parameters used for the CNN-LSTM model were shown in **Table 2**. All classification algorithms developed in the open-source Python libraries Scikit-Learn and TensorFlow.

In this study, a total of 2,312 valid segment data were calculated (including data from the healthy control group and three types of tinnitus patients), and the category calibration was completed for each valid segment, 1,622 groups were randomly selected as the training set, and 690 groups as the validation set. Each group of classification tests was repeated 10 times, and the results were the average of the classification accuracy and model computing time, respectively. The computer configuration: CPU Core I5 8th generation processor, GPU AMD R620, the highest frequency 1.8 GHz, and the running memory 16 GB. It is necessary to explain that the difference in calculator configuration will bring about differences in computing power and model computing time.

Statistical Analysis

The one-factorial ANOVAs and the Kruskal-Wallis ANOVAs (non-parametric) methods were used in the statistical analysis in this article. We used the one-factorial ANOVAs to analyze the recognition accuracy of the same feature combined with different

machine learning algorithms (**Figure 3**) which could test whether there were significant differences between different classification methods. At the same time, the Kruskal-Wallis ANOVAs (non-parametric) were utilized to evaluate the calculation time for different approaches.

It is known from the literature that low-frequency tinnitus (LFT) is perceived and processed differently from high-frequency tinnitus (HFT) (Zhang et al., 2020). However, in the tinnitus participants recruited in this study, we did not deliberately distinguish between patients with LFT or HFT. Among the tinnitus participants recruited in this study, there were two LFT patients and ten HFT patients in the bilateral tinnitus group; seven LFT patients and five HFT patients in the left-sided tinnitus group; and two LFT patients and five HFT patients in the right-sided tinnitus group. To explore whether the EEG data of patients with LFT or HFT has an impact on the diagnostic methods of this study, we selected the left-sided tinnitus group with an almost balanced LFT and HFT for testing. We combined the LFT and HFT in the left-sided tinnitus group with the data of the same healthy control group into a new data set and then used the method in this study to classify tinnitus disorders. Next, we used the one-factorial ANOVAs method to evaluate the difference between the LFT and HFT during using different combinations of methods (**Figure 4**).

We also did a statistical analysis of the two connectivity features PCC and PLV with excellent recognition accuracy, and the results are shown in **Figure 5**. Both the green block and the blue block indicated that there were significant differences in the feature data between the healthy control group and the tinnitus group in this area, and the green block was the difference area that both existed in the three tinnitus groups.

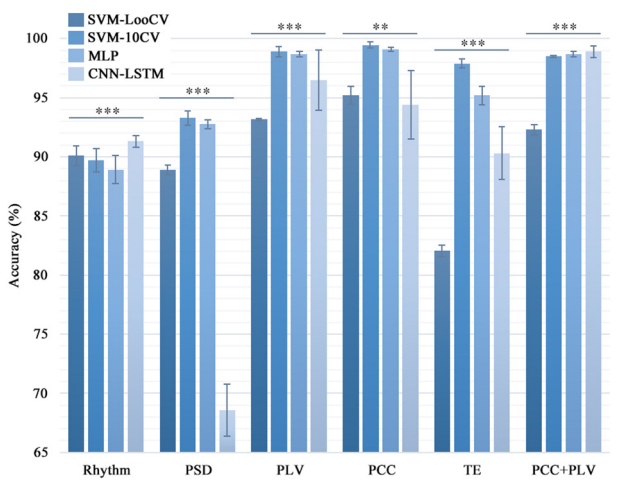


FIGURE 3 | Comparison of the classification accuracy in different combinations of classifiers and features. Four machine learning algorithms were used to calculate the six feature data sets one by one to distinguish the healthy group and the three types of tinnitus patients. For the same feature set, the classification accuracy of the four models were significantly different (*$p < 0.05$, **$p < 0.01$, ***$p < 0.005$). And the standard deviation showed that the SVM and MLP models were more stable than the CNN model.

RESULTS

Results for Classification

Table 3 shows the results of the SVM-LooCV, SVM-10CV, MLP, CNN-LSTM algorithms when classifying data based on Rhythm, PSD, PLI, PLV, PCC, TE features alone or the combination of PCC and PLV features. **Figure 3** shows classification accuracy using four methods with different features (PLI is not mentioned here because its identification result is lower than other features). As can be seen from **Table 3**, classification accuracy is better when using connectivity features, especially PLV, PCC, and TE. The SVM-10CV classifier based on the PCC feature shows the highest classification accuracy (99.42%), and included the PLV feature achieves an accuracy of 98.9%, at the same time, the MLP classifier based on the PCC feature achieves an accuracy of 99.1% and included the PLV feature achieves an accuracy of 98.7%. These four combinations of classifier and alone feature are the best result so far.

At the same time, the statistical analysis of the calculation results shows that there are significant differences in the recognition accuracy of different classification models ($p < 0.01$). The standard deviation can show that the SVM and MLP models are more stable than the CNN model. The possible reason is that CNN needs a larger sample size to make the experimental results more accurate and stable.

Comparing the classification results from **Table 3** and **Figure 3**, it is shown that the use of the PLV or PCC features results in about 6.5% higher classification accuracy than the PSD feature, regardless of the classification algorithm, and the difference is even greater than Rhythm feature. Thus, PLV and PCC show to be more suitable than other features when used to classify tinnitus.

In addition to the classification accuracy, there is another important factor for judging the quality of a classifier, which is computing time. As can be seen from **Table 3**, there are significant differences in the average calculation time of the four model algorithms ($p < 0.001$). The computing time of SVM-LooCV and CNN-LSTM is similar, and they are the longest in the four classifiers. Compared with them, the computing efficiency of SVM-10CV is improved by 2.4 times. And MLP runs the fastest, with an average computing time of only 4.2 s, which is 52 times that of CNN-LSTM and 120 times that of SVM. Thus, MLP is more suitable than other methods when a real-time classification is required.

There was a difference in the recognition accuracy between LFT patients and HFT patients, as shown in **Figure 4**. Among the groups that showed significant differences ($p < 0.05$), only when the SVM-10CV combined with PLI, the recognition accuracy of HFT patients was lower than LFT patients. And the other groups were all HFT patients with better recognition results, which were 1.92% higher on average. The MLP classifier based on the PLV or PCC feature had the best comprehensive performance in this article. The MLP classifier combined with PLV features had very significant differences in the classification results of LFT and HFT, and the classification accuracy of HFT was 0.54% higher than that of LFT. Although the average classification accuracy

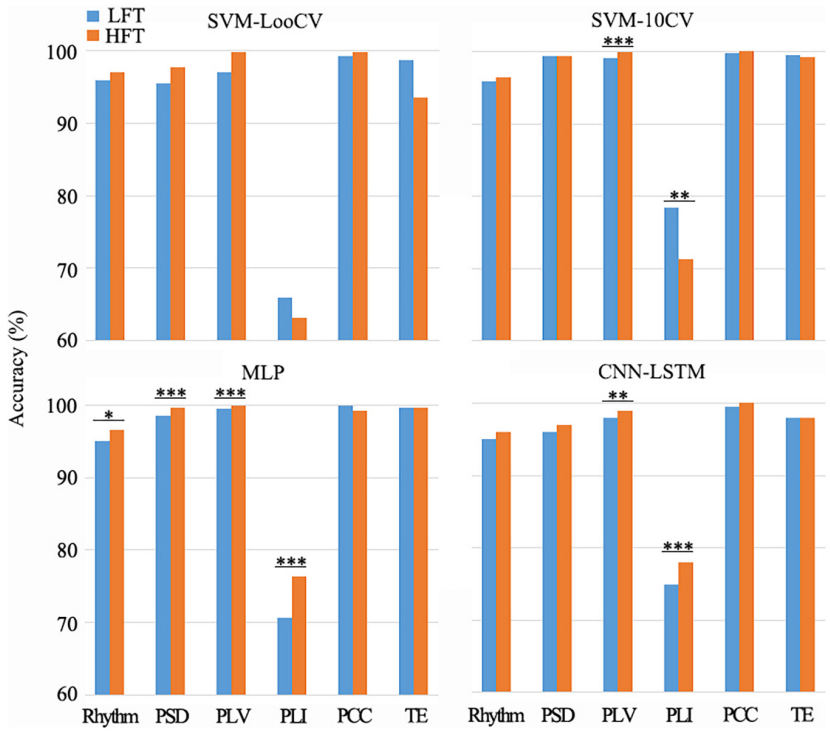


FIGURE 4 | Differences between LFT and HFT in left-sided tinnitus patients in the combination of different methods. There were some combinations that showed a significant difference between LFT and HFT (*$p < 0.05$, **$p < 0.01$, ***$p < 0.005$). Among the groups that showed significant differences ($p < 0.05$), only when the SVM-10CV combined with PLI, the recognition accuracy of HFT was lower than LFT. And the other groups were all HFT with better recognition results.

of LFT patients was higher than that of HFT patients, there was no significant difference between them during the MLP classifier based on the PCC feature. In summary, the impact of LFT and HFT in this method was very small, and would not affect the identification of the location of tinnitus with low-frequency and high-frequency. However, this did not mean that we denied the difference between LFT and HFT, and it was also worthy of attention.

The calculation idea of SVM is to find the most suitable "linear segmenter," also called hyperplane, for the input data set. This process generally needs to project the data set into a high-dimensional space, and the kernel function keeps working to find a hyperplane that can converge. It is more suitable for two classification problems, but for multi-class classification problems, it takes a long time. The calculation idea of the MLP network is to map a set of input vectors to a set of output vectors. It is a process of continuous simplification through the calculation of non-linear activation functions, so the calculation speed of the MLP is the fastest. In addition, it also has a back-propagation mechanism, which can continuously optimize the weight coefficients, and ultimately make the prediction error smaller and smaller. CNN is more commonly used in the classification of two-dimensional data, such as image data. However, the EEG data this time is one-dimensional data, which does not fully utilize the operating capabilities of the model. This may also be the reason for the poor performance of the model. During the analysis of the connectivity feature matrix, we assume that if the two-dimensional data of the matrix is used as the feature input in the future, it is likely to improve the classification accuracy and calculation efficiency of tinnitus again.

Relationship of Connectivity Features and Tinnitus

In this paper, the classification accuracy of features PLV and PCC are excellent and stable. Therefore the values of PLV and PCC between each of the 12 selected channels are used to construct functional connection matrices for subsequent analysis. Since PLV and PCC are both undirected connection characteristics, there is only one value between the two channels, so in **Figure 6**, the upper left area represents the PLV matrices, and the lower right area represents the PCC matrices. Visual inspection shows clear differences in the PLV matrices and PCC matrices.

In **Figure 5**, the significant differences in most regions show the effectiveness of the two connectivity features (PLV and PCC) as biomarkers for tinnitus. In addition to areas with significant differences in common, there are many areas with significant differences in a single disease. Which also improves the classification accuracy of different localizations of tinnitus.

We use the method of multidimensional cluster statistics (MCS) to analyze the connectivity feature value from another

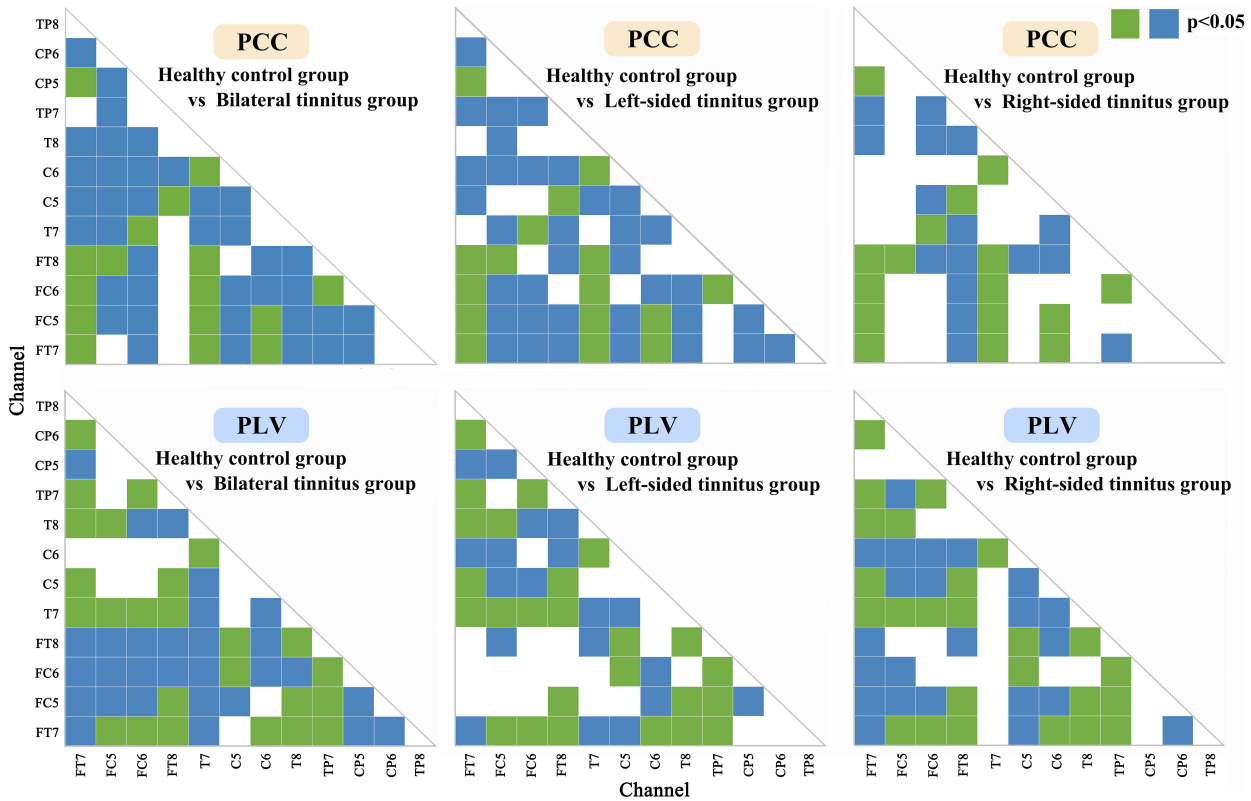


FIGURE 5 | Statistical analysis of Phase lock value and Pearson correlation coefficient matrix. Both the green block and the blue block indicated that there were significant differences ($p < 0.05$) in the feature data between the healthy control group and the tinnitus group in this area, and the green block was the significant difference area that both existed in the three tinnitus groups.

TABLE 3 | Classification accuracy and computing time using four methods with different features.

Features	SVM-LooCV		SVM-10CV		MLP		CNN-LSTM	
	Accuracy	Times	Accuracy	Times	Accuracy	Times	Accuracy	Times
Rhythm	90.09	480	89.7	420	88.9	4.6	91.3	502
PSD	88.88	540	93.27	150	92.76	3.5	68.5	533
PLV	93.16	366	98.9	142	98.7	3	96.5	474
PLI	45.65	600	54.7	220	52.9	6.8	43	570
PCC	95.24	300	99.42	104	99.1	3	94.4	578
TE	82.09	600	97.9	390	95.2	5.2	90.3	461
PLV+PCC	92.3	800	98.5	120	98.7	3.3	98.9	470

perspective, the MCS allows for a comparison of clusters of data points in multidimensional space, and which can quantify the similarities and dissimilarities of cortical activation patterns across recording channels (Krauss et al., 2018). First, multidimensional scaling (MDS) is used to reduce the dimension of the connection matrix (Burt and Torgerson, 1959; Kruskal, 1964a,b), then we visualize the data, and we calculate the Euclidean distance between points (channels) and perform clusters division finally (**Figure 7**). First, we look at the cluster analysis of the PLV. There are three similar connectivity activations (SCA) between the channels on both sides of the brain in the healthy control group. Compared with the healthy control group, the SCA between the channels on both sides of the brain is reduced to one in the bilateral tinnitus group, and there is one SCA on the right side of the brain at the same time. In the left-sided tinnitus group, there are two SCA on the right side of the brain, and there is also one SCA on the left side of the brain. In the right-sided tinnitus group, there are two SCA on the left side of the brain. Although it contains one channel on the right side of the brain, most of the similar channels are located on the left side of the brain. Then, we look at the cluster analysis of the PCC. There are two SCA on the left side of the brain and one SCA between both sides of the brain in the healthy control group;

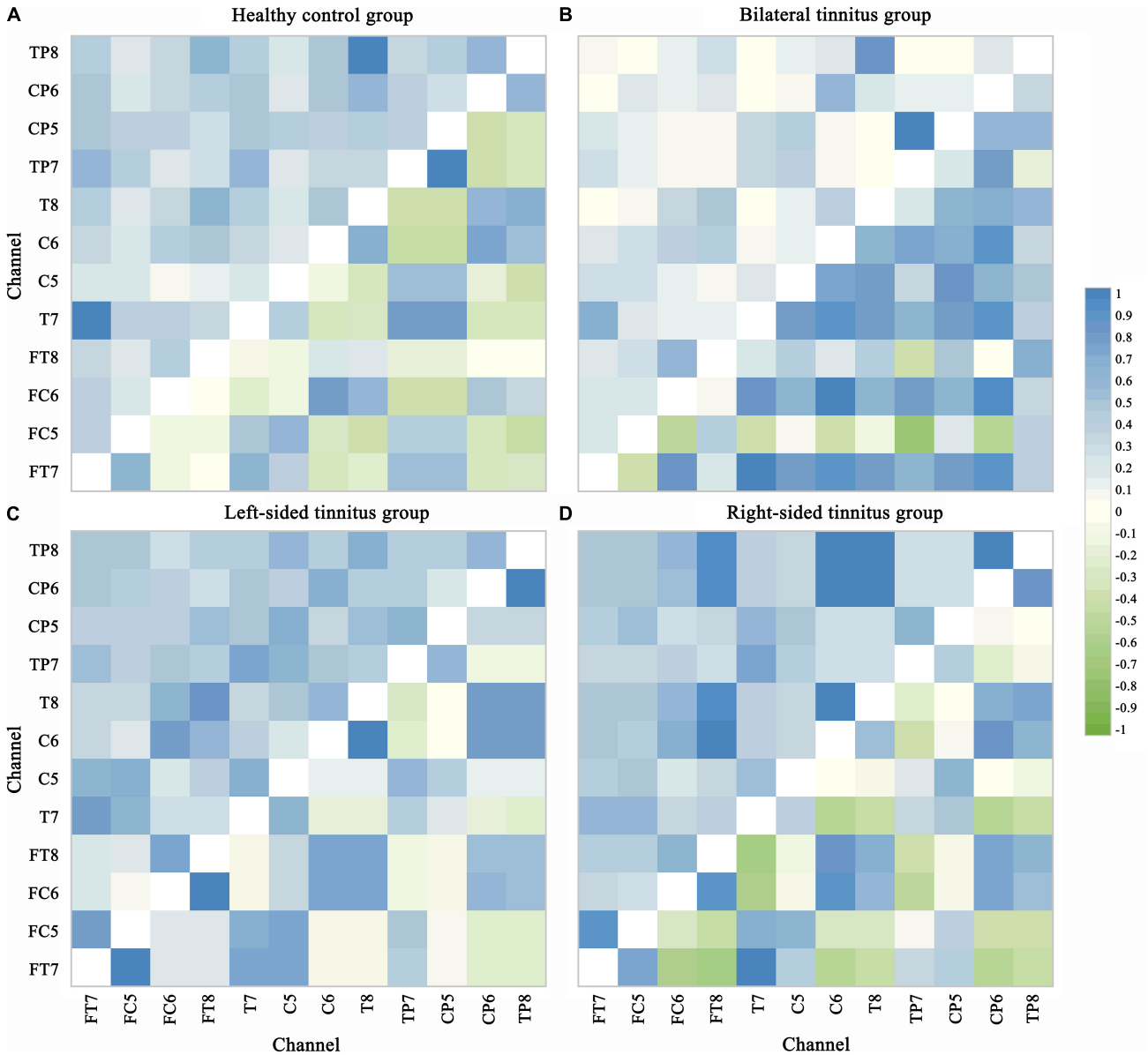


FIGURE 6 | Phase lock value and Pearson correlation coefficient matrix. PLV and PCC are both undirected connection characteristics, there is only one value between the two channels. The upper left area represents the PLV matrices and the lower right area represents the PCC matrices in each picture. Visual inspection shows clear differences in the PLV matrices and PCC matrices. The value range of PLV is from 0 to 1, indicating perfectly independent or perfectly synchronization of the two signals, respectively. PCC values of –1 and 1 correspond to perfectly negative and positive linear relationships, respectively, and a PCC value of 0 indicates that the two signals are not correlated. **(A)** Healthy control group. **(B)** Bilateral tinnitus group. **(C)** Left-sided tinnitus group. **(D)** Right-sided tinnitus group.

The binaural tinnitus group also has two SCA between both sides of the brain, and the similarity is stronger, compared to the healthy control group, the SCA on the left side of the brain disappears. Compared with the healthy control group, the left-sided tinnitus group has two new SCA on the right side of the brain, but the SCA between both sides of the brain disappears. Compared with the left-sided tinnitus group, the right-sided tinnitus group has stronger SCA on the right side of the brain, and the SCA between both sides of the brain also appears.

DISCUSSION

Two cross-validation methods were selected to match the SVM algorithm model which greatly improve the classification accuracy of the SVM classifier, but they also increased the calculation time, especially the Loo-CV. The SVM model with the Loo-CV always used almost 100% data in each training. It was very effective for small sample classification problems, but when the sample size became larger (for example, this experiment had 2,300 data samples), the calculation time of the model

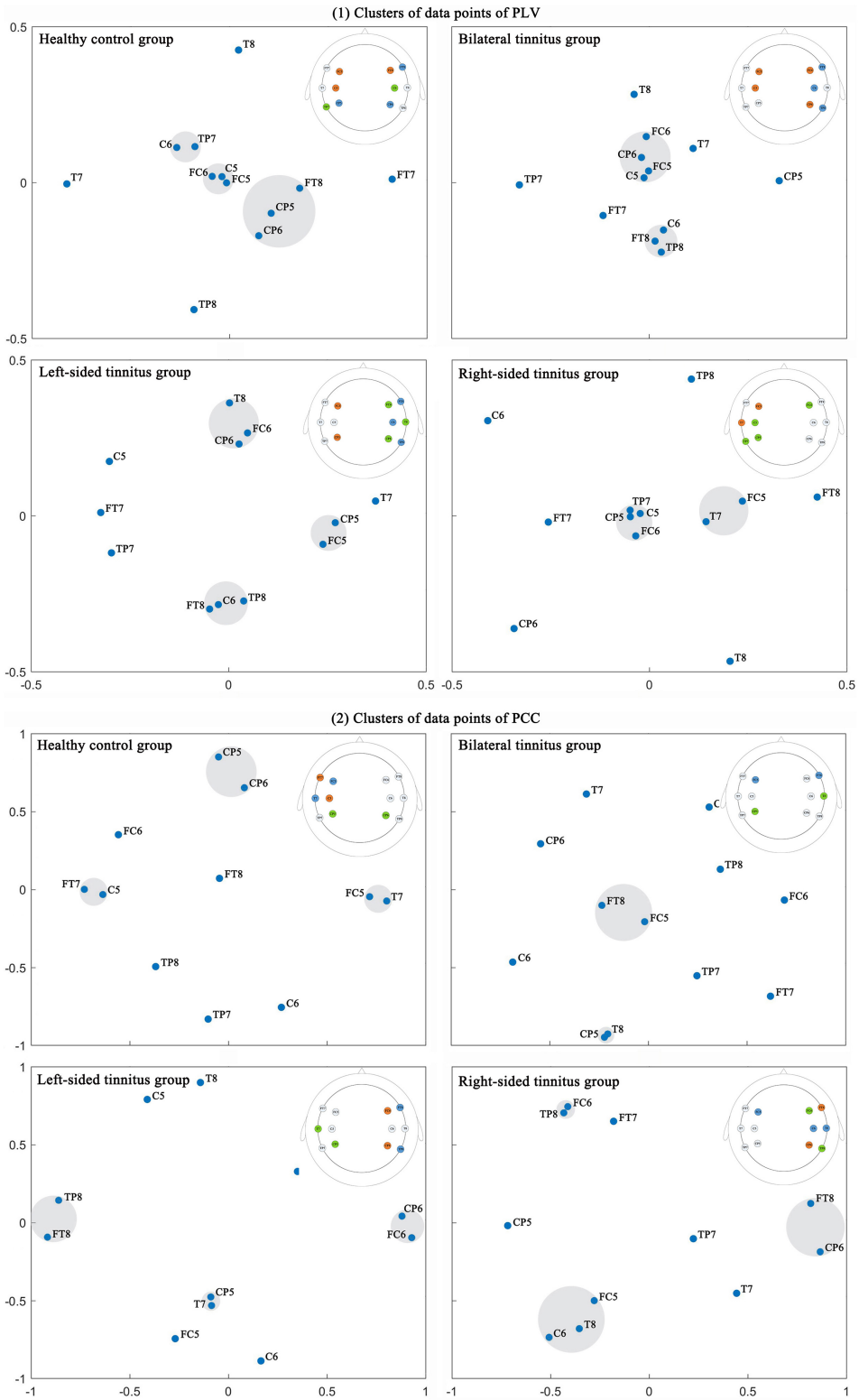


FIGURE 7 | The data point clusters of multidimensional scaling of connectivity feature value were obtained by multidimensional cluster statistics. The highlighted area was a cluster of data points. The channels in the same cluster were highlighted with the same color in the brain topographic map, and which is in the upper right area of each picture. (1) Clusters of data points of PLV; (2) Clusters of data points of PCC.

would be greatly increased, which also meant that the model was not suitable for real-time diagnostic evaluation. Compared with SVM, the classification efficiency of MLP was higher, mainly because the core solution of the classification model was different. The former was to find the global optimal solution, while the latter was the local optimal solution, so the calculation speed of the MLP was faster. At the same time, the classification accuracy of MLP was also close to SVM, which was more suitable for real-time diagnostic evaluation. Similarly, the CNN model usually performed multi-layer pooling of data, the computational efficiency was not too high, and the running time was close to the worst SVM-LooCV. The statistical results also showed that the classification accuracy obtained by different methods had significant differences ($p < 0.01$). In other words, the four machine learning models had significant differences. And MLP model performed excellently in the four model algorithms in terms of classification accuracy and computing time.

The classification accuracy of PCC or PLV features in the four classification models was better than other features, which verified that the functional connectivity feature, especially PCC or PLV, was more suitable as a biomarker of tinnitus. However, the result of the PCC+PLV was not better than the accuracy of a single feature, and there was a reduction. In the future, we should adopt a more effective combination mechanism to improve the classification accuracy, for example, extracting the more relevant parts of the single feature for combination. In the left-sided tinnitus group, there is a significant difference in the classification accuracy of LFT patients and HFT patients, but the difference in the recognition accuracy of the two is very small. And it will not affect the diagnosis of tinnitus location. In our research on the identification of tinnitus location, the difference between LFT and HFT patients can be ignored.

A recent study on the neural connection of tinnitus recruited a group of special tinnitus patients who can actively turn on the state of tinnitus through physical stimulation or sound stimulation (Zhang J. et al., 2021). The study noted that the occurrence of unilateral tinnitus would have an effect on the EEG signals in the contralateral brain area by locating the EEG signals source, and the symptoms of bilateral tinnitus mainly affected the EEG signals in the left side of the brain. Still, this conclusion might have individual differences because the data was only from one participant. Some research supported that the changes in the EEG signals in the left auditory area were related to the formation of tinnitus and did not distinguish the effects of unilateral and bilateral tinnitus (Ashton et al., 2007; Schecklmann et al., 2013).

In this study, the cluster analysis of the PLV and PCC showed that bilateral tinnitus mainly affected the connectivity of the auditory cortex on both sides of the brain and the connectivity of the right or left auditory cortex. The left-sided tinnitus mainly affected the connectivity of the right auditory cortex. And the right-sided tinnitus affected the connectivity of the left auditory cortex, and it also affected the connectivity of the right auditory cortex. An interesting finding was that tinnitus patients always had increased causal connectivity on the right side of the brain (Cai et al., 2020).

As far as we know, there is no previous study on the objective classification of tinnitus location through neural network algorithms, and this study also confirms that functional connectivity features, especially PCC or PLV, are more efficient in identifying the location of tinnitus. This study provides evidence for the effectiveness of the functional connectivity features of resting-state EEG as biomarkers for tinnitus.

In addition, an epidemiologic perspective on tinnitus pointed out that the prevalence of tinnitus among military veterans was relatively high. In a tinnitus screening survey for veterans, the prevalence rate reached 63%, and the participants were young veterans with an average age of 34.8, who had been out of the army for less than 3 years. A more noteworthy problem was that only approximately 20% of them would seek clinical intervention, and most of them respected tinnitus as being "not a problem" or "a small problem" (Henry et al., 2020). If we realize the rapid detection and objective diagnosis of tinnitus, it will greatly improve the life happiness index of military veterans, and this is especially true for the general population.

This study also has certain limitations. Considering that we want to carry out the objective diagnosis of tinnitus more conveniently in the future, we utilize 12-channel EEG on the auditory cortex on both sides of the brain for analytical research, because the 12-channel EEG acquisition experiment is simpler and faster than 64-channel. Although the 12-channel EEG signal can show obvious differences, the connectivity of the auditory regions on both sides of the brain cannot fully represent the functional network connectivity of the whole brain. In addition, the audiogram measured in this study is 8 khz. Although extending to higher frequencies above 8 khz may improve our ability to detect invisible hearing loss, the clinical significance of the high-frequency hearing test is still unclear (Balatsouras et al., 2005; Schmuziger et al., 2007). Finally, our work only researches tinnitus patients without hearing loss, but it is necessary to conduct further experiments on tinnitus patients with hearing loss to confirm the superiority of connectivity features in distinguishing tinnitus patients from healthy people.

CONCLUSION

This study showed that connectivity features, especially PLV and PCC, could be biomarkers of tinnitus location in the resting-state EEG signals. Classification accuracy was highest when the SVM-10CV algorithm or the MLP algorithm was applied to the PCC feature sets, achieving final average classification accuracies of 99.42 and 99.1%, respectively. And based on the PLV feature, the classification result was also particularly good. Together, these results confirmed the feasibility of this method and the method could also meet the needs of objective diagnosis of tinnitus location.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and

accession number(s) can be found: https://pan.baidu.com/s/1MOw__KoFKZQWpRVzA7_6WQ, password:jhl6.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Ethics Committee of Chinese PLA General Hospital. The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

ZL: study design, data analysis, result interpretation, manuscript drafting, and revision. XW: study design, result interpretation, manuscript drafting, and revision. WS and SY: conceptualization and methodology. DZ: funding acquisition. JH and DW: supervision. JL and HX: data curation. YZ: visualization and investigation. PL: software and validation. BZ and HC: resources. YL and XL: experiment management. All authors had contributions on preparing this manuscript.

FUNDING

This study was supported by the Jihua Laboratory of China (Grant Nos. X201221XD200 and X190341TD190) and the National Key Research and Development Program of China (Grant Nos. 2019YFC0121300 and 2019YFC0121303).

ACKNOWLEDGMENTS

We are thankful to all participants for their time and patience, and we thank all authors for their valuable support in data acquisition and study management.

REFERENCES

Ahn, M. H., Hong, S. K., and Min, B. K. (2017). The absence of resting-state high-gamma cross-frequency coupling in patients with tinnitus. *Hear. Res.* 356, 63–73. doi: 10.1016/j.heares.2017.10.008

Ashton, H., Reid, K., Marsh, R., Johnson, I., Alter, K., and Griffiths, T. (2007). High frequency localised "hot spots" in temporal lobes of patients with intractable tinnitus: a quantitative electroencephalographic (QEEG) study. *Neurosci. Lett.* 426, 23–28. doi: 10.1016/j.neulet.2007.08.034

Baguley, D., McFerran, D., and Hall, D. (2013). Tinnitus. *Lancet* 382, 1600–1607. doi: 10.1016/S0140-6736(13)60142-7

Balatsouras, D. G., Homsioglou, E., and Danielidis, V. (2005). Extended high-frequency audiometry in patients with acoustic trauma. *Clin. Otolaryngol.* 30, 249–254. doi: 10.1111/j.1365-2273.2005.00984.x

Bressler, S. L., and Menon, V. (2010). Large-scale brain networks in cognition: emerging methods and principles. *Trends Cogn. Sci.* 14, 277–290. doi: 10.1016/j.tics.2010.04.004

Burt, C., and Torgerson, W. S. (1959). Theory and methods of scaling. *Biometrika* 46, 493. doi: 10.2307/2333553

Cai, Y., Li, J., Chen, Y., Chen, W., Dang, C., Zhao, F., et al. (2019). Inhibition of brain area and functional connectivity in idiopathic sudden sensorineural hearing loss with tinnitus, based on resting-state EEG. *Front. Neurosci.* 13:851. doi: 10.3389/fnins.2019.00851

Cai, Y., Xie, M., Su, Y., Tong, Z., Wu, X., Xu, W., et al. (2020). Aberrant functional and causal connectivity in acute tinnitus with sensorineural hearing loss. *Front. Neurosci.* 14:592. doi: 10.3389/fnins.2020.00592

Chari, D. A., and Limb, C. J. (2018). Tinnitus. *Med. Clin. N. Am.* 102, 1081–1093. doi: 10.1016/j.mcna.2018.06.014

Chen, Y. C., Liu, S., Lv, H., Bo, F., Feng, Y., Chen, H., et al. (2018). Abnormal resting-state functional connectivity of the anterior cingulate cortex in unilateral chronic tinnitus patients. *Front. Neurosci.* 12:9. doi: 10.3389/fnins.2018.00009

Cima, R. F. F., Mazurek, B., Haider, H., Kikidis, D., Lapira, A., Noreña, A., et al. (2019). A multidisciplinary European guideline for tinnitus: diagnostics, assessment, and treatment. *HNO* 67, 10–42. doi: 10.1007/s00106-019-0633-7

Cunnane, M. B. (2019). Imaging of tinnitus. *Neuroimaging Clin. N. Am.* 29, 49–56. doi: 10.1016/j.nic.2018.09.006

Eggermont, J. J., and Roberts, L. E. (2012). The neuroscience of tinnitus: understanding abnormal and normal auditory perception. *Front. Syst. Neurosci.* 6:53. doi: 10.3389/fnsys.2012.00053

Elgoyhen, A. B., Langguth, B., de Ridder, D., and Vanneste, S. (2015). Tinnitus: perspectives from human neuroimaging. *Nat. Rev. Neurosci.* 16, 632–642. doi: 10.1038/nrn4003

Engineer, N. D., Riley, J. R., Seale, J. D., Vrana, W. A., Shetake, J. A., Sudanagunta, S. P., et al. (2011). Reversing pathological neural activity using targeted plasticity. *Nature* 470, 101–104. doi: 10.1038/nature09656

Gallus, S., Lugo, A., Garavello, W., Bosetti, C., Santoro, E., Colombo, P., et al. (2015). Prevalence and determinants of tinnitus in the Italian adult population. *Neuroepidemiology* 45, 12–19. doi: 10.1159/000431376

Gao, Y., Wang, X., Potter, T., Zhang, J., and Zhang, Y. (2020). Single-trial EEG emotion recognition using Granger Causality/Transfer Entropy analysis. *J. Neurosci. Methods* 346:108904. doi: 10.1016/j.jneumeth.2020.108904

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. *Front. Neurol.* 8:31. doi: 10.3389/fneur.2017.00031

Henry, J. A., Reavis, K. M., Griest, S. E., Thielman, E. J., Theodoroff, S. M., Grush, L. D., et al. (2020). Tinnitus: an epidemiologic perspective. *Otolaryngol. Clin. N. Am.* 53, 481–499. doi: 10.1016/j.otc.2020.03.002

Herraiz, C. (2008). Assessing the cause of tinnitus for therapeutic options. *Expert Opin. Med. Diagn.* 2, 1183–1196. doi: 10.1517/17530059.2.10.1183

Kemp, D. T. (1978). Stimulated acoustic emissions from within the human auditory system. *J. Acoust. Soc. Am.* 64, 1386–1391. doi: 10.1121/1.382104

Klimesch, W., Sauseng, P., and Hanslmayr, S. (2007). EEG alpha oscillations: the inhibition–timing hypothesis. *Brain Res. Rev.* 53, 63–88. doi: 10.1016/j.brainresrev.2006.06.003

Krauss, P., Metzner, C., Schilling, A., Tziridis, K., Traxdorf, M., Wollbrink, A., et al. (2018). A statistical method for analyzing and comparing spatiotemporal cortical activation patterns. *Sci. Rep.* 8, 5433. doi: 10.1038/s41598-018-23765-w

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. *Front. Neurosci.* 10:597. doi: 10.3389/fnins.2016.00597

Kruskal, J. B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29, 1–27. doi: 10.1007/BF02289565

Kruskal, J. B. (1964b). Nonmetric multidimensional scaling: a numerical method. *Psychometrika* 29, 115–129. doi: 10.1007/BF02289694

Lachaux, J.-P., Rodriguez, E., Martinerie, J., and Varela, F. J. (1999). Measuring phase synchrony in brain signals. *Hum. Brain Mapp.* 8, 194–208. doi: 10.1002/(sici)1097-0193(1999)8:4<194::aid-hbm4>3.0.co;2-c

Lee, H. Y., Kim, S. J., Chang, D. S., and Shin, S. A. (2019). Tinnitus in the side with better hearing. *Am. J. Otolaryngol. Head Neck Med. Surg.* 40, 400–403. doi: 10.1016/j.amjoto.2019.02.009

Lee, Y. Y., and Hsieh, S. (2014). Classifying different emotional states by means of eegbased functional connectivity patterns. *PLoS One* 9:e95415. doi: 10.1371/journal.pone.0095415

Llinás, R., Urbano, F. J., Leznik, E., Ramírez, R. R., and van Marle, H. J. F. (2005). Rhythmic and dysrhythmic thalamocortical dynamics: GABA systems and the edge effect. *Trends Neurosci.* 28, 325–333. doi: 10.1016/j.tins.2005.04.006

Meikle, M., and Taylor-Walsh, E. (1984). Characteristics of tinnitus and related observations in over 1800 tinnitus clinic patients. *J. Laryngol. Otol.* 98, 17–21. doi: 10.1017/S1755146300090053

Mills, D. M., and Rubel, E. W. (1994). Variation of distortion product otoacoustic emissions with furosemide injection. *Hear. Res.* 77, 183–199. doi: 10.1016/0378-5955(94)90266-6

Mirz, F., Brahe Pedersen, C., Ishizu, K., Johannsen, P., Ovesen, T., Stødkilde-Jørgensen, H., et al. (1999). Positron emission tomography of cortical centers of tinnitus. *Hear. Res.* 134, 133–144. doi: 10.1016/S0378-5955(99)00075-1

Møller, A. R., and Jannetta, P. J. (1982). Evoked potentials from the inferior colliculus in man. *Electroencephalogr. Clin. Neurophysiol.* 53, 612–620. doi: 10.1016/0013-4694(82)90137-7

Møller, A. R., Jannetta, P., Bennett, M., and Møller, M. B. (1981). Intracranially recorded responses from the human auditory nerve: new insights into the origin of brain stem evoked potentials (BSEPs). *Electroencephalogr. Clin. Neurophysiol.* 52, 18–27. doi: 10.1016/0013-4694(81)90184-X

Moon, S. E., Chen, C. J., Hsieh, C. J., Wang, J. L., and Lee, J. S. (2020). Emotional EEG classification using connectivity features and convolutional neural networks. *Neural Netw.* 132, 96–107. doi: 10.1016/j.neunet.2020.08.009

Müller, M., Klinke, R., Arnold, W., and Oestreicher, E. (2003). Auditory nerve fibre responses to salicylate revisited. *Hear. Res.* 183, 37–43. doi: 10.1016/S0378-5955(03)00217-X

Ohlemiller, K. K., Rybak Rice, M. E., Rellinger, E. A., and Ortmann, A. J. (2011). Divergence of noise vulnerability in cochleae of young CBA/J and CBA/CaJ mice. *Hear. Res.* 272, 13–20. doi: 10.1016/j.heares.2010.11.006

Pan, T., Tyler, R. S., Ji, H., Coelho, C., Gehringer, A. K., and Gogel, S. A. (2009). Changes in the tinnitus handicap questionnaire after cochlear implantation. *Am. J. Audiol.* 18, 144–151. doi: 10.1044/1059-0889(2009/07-0042)

Pierzycki, R. H., McNamara, A. J., Hoare, D. J., and Hall, D. A. (2016). Whole scalp resting state EEG of oscillatory brain activity shows no parametric relationship with psychoacoustic and psychosocial assessment of tinnitus: a repeated measures study. *Hear. Res.* 331, 101–108. doi: 10.1016/j.heares.2015.11.003

Rauschecker, J. P., Leaver, A. M., and Mühlau, M. (2010). Tuning out the noise: limbic-auditory interactions in tinnitus. *Neuron* 66, 819–826. doi: 10.1016/j.neuron.2010.04.032

Romei, V., Rihs, T., Brodbeck, V., and Thut, G. (2008). Resting electroencephalogram alpha-power over posterior sites indexes baseline visual cortex excitability. *NeuroReport* 19, 203–208. doi: 10.1097/WNR.0b013e3282f454c4

Sauseng, P., Klimesch, W., Stadler, W., Schabus, M., Doppelmayr, M., Hanslmayr, S., et al. (2005). A shift of visual spatial attention is selectively associated with human EEG alpha activity. *Eur. J. Neurosci.* 22, 2917–2926. doi: 10.1111/j.1460-9568.2005.04482.x

Schaette, R., and McAlpine, D. (2011). Tinnitus with a normal audiogram: physiological evidence for hidden hearing loss and computational model. *J. Neurosci.* 31, 13452–13457. doi: 10.1523/JNEUROSCI.2156-11.2011

Schecklmann, M., Landgrebe, M., Poeppl, T. B., Kreuzer, P., Männer, P., Marienhagen, J., et al. (2013). Neural correlates of tinnitus duration and Distress: a positron emission tomography study. *Hum. Brain Mapp.* 34, 233–240. doi: 10.1002/hbm.21426

Schlee, W., Lorenz, I., Hartmann, T., Müller, N., Schulz, H., and Weisz, N. (2011). "A global brain model of tinnitus," in *Textbook of Tinnitus*, eds A. R. Møller, B. Langguth, D. DeRidder, and T. Kleinjung (New York, NY: Springer), 161–169. doi: 10.1007/978-1-60761-145-5_20

Schlee, W., Weisz, N., Bertrand, O., Hartmann, T., and Elbert, T. (2008). Using auditory steady state responses to outline the functional connectivity in the tinnitus brain. *PLoS One* 3:e3720. doi: 10.1371/journal.pone.0003720

Schmuziger, N., Patscheke, J., and Probst, R. (2007). An assessment of threshold shifts in nonprofessional pop/rock musicians using conventional and extended high-frequency audiometry. *Ear Hear.* 28, 643–648. doi: 10.1097/AUD.0b013e31812f7144

Schoisswohl, S., Schecklmann, M., Langguth, B., Schlee, W., and Neff, P. (2021). Neurophysiological correlates of residual inhibition in tinnitus: hints for trait-like EEG power spectra. *Clin. Neurophysiol.* 132, 1694–1707. doi: 10.1016/j.clinph.2021.03.038

Schreiber, T. (2000). Measuring information transfer. *Phys. Rev. Lett.* 85, 461–464. doi: 10.1103/PhysRevLett.85.461

Searchfield, G. D., Kobayashi, K., Proudfoot, K., Tevoitdale, H., and Irving, S. (2015). The development and test-retest reliability of a method for matching perceived location of tinnitus. *J. Neurosci. Methods* 256, 1–8. doi: 10.1016/j.jneumeth.2015.07.027

Smit, J. v, Janssen, M. L. F., Schulze, H., Jahanshahi, A., van Overbeeke, J. J., Temel, Y., et al. (2015). Deep brain stimulation in tinnitus: current and future perspectives. *Brain Res.* 1608, 51–65. doi: 10.1016/j.brainres.2015.02.050

Stam, C. J., Nolte, G., and Daffertshofer, A. (2007). Phase lag index: assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources. *Hum. Brain Mapp.* 28, 1178–1193. doi: 10.1002/hbm.20346

Tunkel, D. E., Bauer, C. A., Sun, G. H., Rosenfeld, R. M., Chandrasekhar, S. S., Cunningham, E. R., et al. (2014). Clinical practice guideline: tinnitus. *Otolaryngol. Head Neck Surg. (U. S.)* 151, S1–S40. doi: 10.1177/0194599814545325

Tziridis, K., Forster, J., Buchheidt-Dörfler, I., Krauss, P., Schilling, A., Wendler, O., et al. (2021). Tinnitus development is associated with synaptopathy of inner hair cells in Mongolian gerbils. *Eur. J. Neurosci.* 54, 4768–4780. doi: 10.1111/ejn.15334

van Dijk, K. R. A., Hedden, T., Venkataraman, A., Evans, K. C., Lazar, S. W., and Buckner, R. L. (2010). Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. *J. Neurophysiol.* 103, 297–321. doi: 10.1152/jn.00783.2009

Vanneste, S., Song, J. J., and de Ridder, D. (2018). Thalamocortical dysrhythmia detected by machine learning. *Nat. Commun.* 9, 1103. doi: 10.1038/s41467-018-02820-0

Wang, S. J., Cai, Y. X., Sun, Z. R., Wang, C. D., and Zheng, Y. Q. (2017). "Tinnitus EEG classification based on multi-frequency bands," in *Proceedings of the International Conference on Neural Information Processing*, (Berlin: Springer), 788–797.

Weisz, N., Dohrmann, K., and Elbert, T. (2007). The relevance of spontaneous activity for the coding of the tinnitus sensation. *Prog. Brain Res.* 166, 61–70. doi: 10.1016/S0079-6123(07)66006-3

Weisz, N., Moratti, S., Meinzer, M., Dohrmann, K., and Elbert, T. (2005). Tinnitus perception and distress is related to abnormal spontaneous brain activity as measured by magnetoencephalography. *PLoS Med.* 2:e153. doi: 10.1371/journal.pmed.0020153

Yao, D. (2001). A method to standardize a reference of scalp EEG recordings to a point at infinity. *Physiol. Meas.* 22, 693–711. doi: 10.1088/0967-3334/22/4/305

Zhang, J., Huang, S., Nan, W., Zhou, H., Wang, J., Wang, H., et al. (2021). Switching tinnitus-on: maps and source localization of spontaneous EEG. *Clin. Neurophysiol.* 132, 345–357. doi: 10.1016/j.clinph.2020.10.023

Zhang, J., Zhang, Z., Huang, S., Zhou, H., Feng, Y., Shi, H., et al. (2020). Differences in clinical characteristics and brain activity between patients with low- and high-frequency tinnitus. *Neural Plast.* 2020:5285362. doi: 10.1155/2020/5285362

Zhang, X., Jiang, Y., Zhang, S., Li, F., Pei, C., He, G., et al. (2021). Correlation analysis of EEG brain network with modulated acoustic stimulation for chronic tinnitus patients. *IEEE Trans. Neural Syst. Rehabil. Eng.* 29, 156–162. doi: 10.1109/TNSRE.2020.3039555

Conflict of Interest: DZ was employed by the company BetterLifeMedical.

The remaining 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 18 March 2022
doi: 10.3389/fnins.2022.816735

Dynamic Transitions Between Brain States Predict Auditory Attentional Fluctuations

Hirohito M. Kondo[1], Hiroki Terashima[2], Takahiro Ezaki[3], Takanori Kochiyama[4], Ken Kihara[5] and Jun I. Kawahara[6]*

[1] School of Psychology, Chukyo University, Nagoya, Japan, [2] NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation, Atsugi, Japan, [3] Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan, [4] Brain Activity Imaging Center, ATR-Promotions, Seika, Japan, [5] Department of Information Technology and Human Factors, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan, [6] Department of Psychology, Hokkaido University, Sapporo, Japan

Achievement of task performance is required to maintain a constant level of attention. Attentional level fluctuates over the course of daily activities. However, brain dynamics leading to attentional fluctuation are still unknown. We investigated the underlying mechanisms of sustained attention using functional magnetic resonance imaging (fMRI). Participants were scanned with fMRI while performing an auditory, gradual-onset, continuous performance task (gradCPT). In this task, narrations gradually changed from one to the next. Participants pressed a button for frequent Go trials (i.e., male voices) as quickly as possible and withheld responses to infrequent No-go trials (i.e., female voices). Event-related analysis revealed that frontal and temporal areas, including the auditory cortex, were activated during successful and unsuccessful inhibition of predominant responses. Reaction-time (RT) variability throughout the auditory gradCPT was positively correlated with signal changes in regions of the dorsal attention network: superior frontal gyrus and superior parietal lobule. Energy landscape analysis showed that task-related activations could be clustered into different attractors: regions of the dorsal attention network and default mode network. The number of alternations between RT-stable and erratic periods increased with an increase in transitions between attractors in the brain. Therefore, we conclude that dynamic transitions between brain states are closely linked to auditory attentional fluctuations.

Keywords: brain dynamics, hearing, sustained attention, attentional fluctuation, gradual-onset continuous performance task, reaction time (RT), energy landscape analysis, functional magnetic resonance imaging (fMRI)

OPEN ACCESS

Edited by:
Andreas K. Maier, University of Erlangen Nuremberg, Germany

Reviewed by:
Han Lv, Capital Medical University, China
Chang-Hao Kao, Yale University, United States

****Correspondence:***
Hirohito M. Kondo
kondo@lets.chukyo-u.ac.jp

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

Received: *17 November 2021*
Accepted: *23 February 2022*
Published: *18 March 2022*

Citation:
Kondo HM, Terashima H, Ezaki T, Kochiyama T, Kihara K and Kawahara JI (2022) Dynamic Transitions Between Brain States Predict Auditory Attentional Fluctuations. Front. Neurosci. 16:816735. doi: 10.3389/fnins.2022.816735

INTRODUCTION

Sustained attention is essential to adaptive behaviors such as driving or listening to a lecture. Focused states of attention are often disturbed by external events and internal factors. Reduction of one's attention level leads to "mind wandering," cognitive errors, and even serious accidents (Edkins and Pollock, 1997; Cheyne et al., 2006; Smallwood and Schooler, 2015). From the perspective of vigilance decrements, human error research has investigated mistakes that occur only rarely over

long periods of time (Mackworth, 1948; Esterman and Rothlein, 2019). However, this classical method is not sensitive to momentary changes in attentional states. One sophisticated study used a gradual-onset continuous performance task (gradCPT) to demonstrate moment-to-moment fluctuations of sustained attention (Esterman et al., 2013).

The gradCPT is an experimental paradigm that was designed for assessing temporal dynamics of sustained attention. Stimuli, such as visual images, in this task are presented continuously, overlapping one another. Participants quickly respond to frequent Go trials and withhold responses to infrequent No-go trials. Importantly, the false alarm (FA) rate of responses is closely linked to variability of reaction times (RTs) (Esterman et al., 2013). Previous studies have also found neural correlates of vigilant and sustained attention. The core network related to vigilant attention includes the frontal, parietal, and subcortical areas (Langner and Eickhoff, 2013). For the gradCPT, the dorsal attention network (DAN) (Corbetta and Shulman, 2002, 2011), including the frontal eye field and intraparietal sulcus, was activated during erratic responses, whereas the default mode network (DMN) (Fox et al., 2005), including the medial prefrontal cortex (mPFC) and precuneus (PCu), was activated during stable responses (Esterman et al., 2013; Fortenbaugh et al., 2018). In contrast, another study using multi-voxel pattern analysis revealed that activations in the DAN and DMN did not distinguish attentional periods (Rosenberg et al., 2015). Thus, findings of previous studies using visual gradCPTs have been mixed, and different approaches are needed to gain new insights into the properties of sustained attention.

This functional magnetic resonance imaging (fMRI) study focused on temporal dynamics of brain activity, particularly in an auditory domain. We used an auditory gradCPT, the performance of which was correlated with visual gradCPT performance (Terashima et al., 2021). Participants indicated using a button press whether a voice was male or female for each trial, while listening to a sequence of narrations. For each participant, we obtained time-series data of RT variability from the auditory gradCPT. Intra-individual variability reflects the efficiency of attentional resources assigned to cognitive demands (Stuss et al., 2003; Klein et al., 2006). The fluctuations of RTs occurred between a stable period and an erratic period, which are termed "in the zone" and "out of the zone" periods (Esterman et al., 2013). We computed the variance time course (VTC) of RTs and specified VTC-related brain regions that are relevant to maintenance of attentional levels.

We further checked whether dynamic features of brain activations explained auditory gradCPT performance: FA rate, correct rejection (CR) rate, sensitivity (d'), and the number of alternations between in- and out-of-the-zone periods. To probe an interaction between the DAN and DMN, the following brain areas were chosen as regions of interest (ROIs): the superior frontal gyrus (SFG) including the frontal eye field, superior parietal lobule (SPL) including the intraparietal sulcus, mPFC, PCu, and superior temporal gyrus (STG). A previous study using energy landscape analysis showed that dynamic transitions of frontal-area and visual-area states could predict individual differences in bistable visual perception (Watanabe et al., 2014). This indicates that perceptual organization is related to underlying brain dynamics. Thus, it is reasonable to assume such a link between task performance and brain dynamics. Brain dynamics are displayed as a series of stays and transitions between different attractors in the energy landscape (Watanabe et al., 2014; Watanabe and Rees, 2017; Ezaki et al., 2018). We examined how dynamic transitions between task-related activities contribute to auditory gradCPT performance.

MATERIALS AND METHODS

Participants

Twenty-nine participants (15 males and 14 females; mean ± SD age = 25.5 ± 4.4 years, range 20–35 years) were recruited for the present study. They were right-handed, healthy Japanese people with normal hearing. According to *a priori* power analyses with a power of 0.8 (α-level = 0.05), we required at least 29 participants to detect significant correlations (r = 0.5; bivariate normal model). The present study was approved by the Research Ethics and Safety Committees of Chukyo University and ATR-Promotions (approval nos. RS20-017 and AN21-056). Experimental procedures were implemented in accordance with Ethical Guidelines for Medical and Biological Research Involving Human Subjects. All participants gave written informed consent after the procedures were fully explained to them.

Stimuli and Procedures

Stimuli of the auditory gradCPT consisted of sequential narrations by a male (90%) and female (10%) that gradually changed from one to the next (Terashima et al., 2021). Narrations of ten males and ten females chosen from a language database, excluding Japanese narrations, were randomly presented through a run (**Figure 1**). Thus, using phonetic features of stimuli, for each trial, participants judged whether a voice was male or female. The sound pressure level of all narrations was adjusted to a comfortable listening level. Stimuli were delivered through plastic tubes and headphones (Hitachi Advanced Systems, Yokohama, Japan). The same stimuli were not repeated in succession.

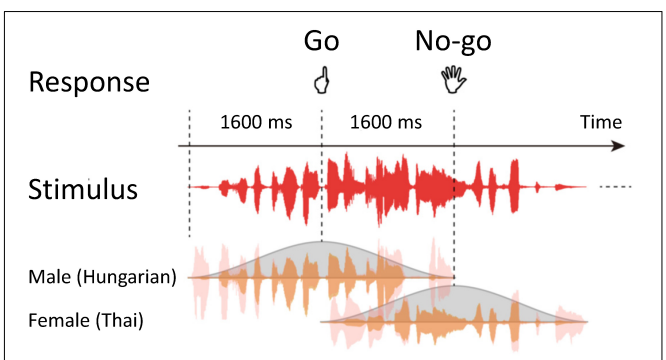


FIGURE 1 | Schematic representation of the auditory gradual-onset, continuous performance task (gradCPT). Stimuli temporally overlapped. Participants judged the genders of voices for each trial.

Participants were instructed to press a button as quickly and accurately as possible if they heard male narrations (Go trials) and to refrain from responding for female narrations (No-go trials). They responded with their left index fingers to avoid affecting the activity of language areas in the left hemisphere. A response deadline was implicit in the task because the current stimulus was replaced by the next stimulus within 1.6 s of stimulus onset asynchrony.

We obtained blood-oxygen-level-dependent (BOLD) signals during auditory gradCPT. An fMRI session comprised two experimental 410-s runs, including 250 trials for each. The duration of a single stimulus was 3.2 s. The order of the stimuli was randomized across runs. Practice trials preceded the experimental runs to familiarize participants with the task. We managed stimulus presentation and response collection using Presentation software (Neurobehavioral Systems, Berkeley, CA, United States).

Imaging Data Acquisition

We scanned participants during the task using a MAGNETOM Prisma (Siemens, Munich, Germany), a 3T MRI scanner with a body coil as a transmitter and a 20-channel head coil as a receiver. We placed small, comfortable, elastic pads on both sides of participants' heads to minimize head motion. For an assessment of cortical thickness and volume, we acquired three-dimensional anatomical images of the whole brain with a T1-weighted MPRAGE sequence: repetition time (TR) = 2,250 ms; echo time (TE) = 3.06 ms; inversion time = 900 ms; flip angle = 9 deg; 208 sagittal slices; matrix size = 256 mm × 256 mm; isotropic voxel size = 1 mm^3.

For each run, we acquired 205 volumes using the multi-band, echo-planar imaging (EPI) sequence. Functional images sensitive to the BOLD response covered the whole brain: 72 consecutive slices parallel to the plane of the anterior-posterior commissure. A T2*-weighted EPI sequence was used with the following parameters: TR/TE = 2000/30 ms; flip angle = 80 deg; multiband acceleration factor = 3; partial Fourier = 6/8; matrix size = 100 × 100; voxel size = 2 mm × 2 mm × 2 mm. At the beginning of an fMRI session, we acquired a B0 field map to correct for geometric distortions: TR/TE1/TE2 = 750/5.17/7.63 ms; flip angle = 50 deg; matrix size = 100 × 100; 72 slices in the same orientation and geometry as the EPI sequence.

Behavioral Data Analysis

We defined an RT as the relative time from stimulus onset to button press (Esterman et al., 2013; Terashima et al., 2021). The time window to assign the response to each stimulus was set from 70% of the appearance phase to 40% of the disappearing phase. That is, we assigned all button presses in the time window to the corresponding trial. We used the shortest RT for a subsequent analysis when finding multiple button presses in a single time window. On the basis of the response assignment, we classified all trials into hit, miss, FA, or CR trials. For each participant, we calculated d' and the median of RTs. We focused on trial-by-trial RT variability because we were also interested in the fluctuation of sustained attention within participants. We computed z-scored RTs in each run, transformed them into absolute values, and called the time-series data the VTC. Using a median split, we divided performance into low- and high-variability epochs, which were termed in- and out-of-the-zone periods, respectively. RTs that were too quick or too slow could be considered a signature of attentional fluctuation. We smoothed the VTCs with a Gaussian kernel at full width at half maximum of 7 s. Statistical analyses were carried out using IBM SPSS Statistics (ver. 25).

Imaging Data Analysis

We analyzed fMRI data using SPM12[1] and in-house codes implemented in MATLAB R2019a (MathWorks, Natick, MA, United States). For each run, we discarded the five initial images to achieve steady-state equilibrium between radio-frequency pulsing and relaxation. For preprocessing, we performed slice-timing correction, realigned all functional images to correct for head motion, and unwarped them to remove dynamic EPI distortions (movement-by-susceptibility interactions). Head-motion correction included 24 realignment parameters, white matter, and cerebrospinal fluid as nuisance covariates. For three participants, maximum values of head movements were greater than either 1.5-mm transformations or 1.5-deg rotations within each run. We excluded these data from subsequent analyses, leaving 26 participants. A value (mean ± SD) of framewise displacement was 0.090 ± 0.039. We used a B0 field map processed using the FieldMap toolbox of SPM12 (Andersson et al., 2001; Hutton et al., 2002). All functional images were normalized to Montreal Neurological Institute (MNI) space, resampled to a voxel size of 2 mm × 2 mm × 2 mm, and smoothed with an isotopic Gaussian kernel of 6 mm full width at half maximum.

Using a general linear model, we entered the four types of trials into a design matrix (Worsley and Friston, 1995). Each trial type was embedded as a stick function. Trial-related regressors were convolved with a canonical hemodynamic response function (HRF). A high-pass filter with a cut-off period of 128 s was used to remove an artifact of the low-frequency trend. We calculated serial autocorrelation from pooled active voxels with a maximum likelihood procedure. Autocorrelation was applied to whiten the data of the design matrix (Friston et al., 2002). We obtained FA- and CR-related statistical maps at the first level and performed random-effects analyses to identify brain activations at the second level.

We constructed another design matrix to examine changes in BOLD signals corresponding to the VTC. We estimated the amplitude-modulated, non-smoothed VTC that was convolved with a canonical HRF. The time-delayed VTC was downsampled to 0.5 Hz (i.e., TR = 2 s) and was used as the regressor of the design matrix. We conducted random-effects analyses using VTC-related statistical maps.

Energy Landscape Analysis

On the basis of automated anatomical labeling (Tzourio-Mazoyer et al., 2002), we defined five ROIs as follows: SFG, SPL,

[1] http://www.fil.ion.ucl.ac.uk/spm

mPFC, PCu, and STG. To reduce the number of signals for a robust energy landscape analysis, we averaged BOLD signals of ROIs in the right and left hemispheres and fed the signals into the energy landscape analysis. Following procedures used in previous studies (Ezaki et al., 2017, 2018), we first binarized each of the five signals and concatenated the data of all participants. Second, appearance probabilities of brain activity patterns were fitted using the pairwise maximum entropy model (Boltzmann distribution; see section "Pairwise Maximum Entropy Model"). Third, using "energy" values defined with the fitting function, we constructed an energy landscape representation of activity patterns (see section "Construction of Energy Landscape"). Fourth, on the basis of the energy landscape, we divided activity patterns into discrete states, each of which corresponds to a basin of an energy local minimum. Finally, using the list of activity patterns in each discrete state, we obtained a coarse-grained representation of the original time series.

Pairwise Maximum Entropy Model

For each session, participant, and ROI, we computed the average value of the BOLD signal, which was then used as a threshold to binarize the signal into -1 (inactive) or $+1$ (active). For each volume, the brain state of the five ROIs was represented by an activity pattern $\sigma = (\sigma_1, \sigma_2, ..., \sigma_5)$, where σ_i $(i = 1, ..., 5)$ denotes the activity of ith ROI (inactive: $\sigma_i = -1$, active: $\sigma_i = 1$). We computed the empirical appearance probability of each of the $2^5 (= 32)$ activity patterns by counting the number of appearances. This probability distribution was fitted with the following pairwise maximum entropy model:

$$P(\sigma) = \frac{\exp(-E(\boldsymbol{\sigma}))}{\sum_{\boldsymbol{\sigma}} \exp(-E(\boldsymbol{\sigma}))},$$

where $E(\boldsymbol{\sigma}) = -\sum_{i=1}^{5} h_i \sigma_i - \sum_{i=1}^{5} \sum_{\substack{j=1 \\ j \neq i}}^{5} J_{ij} \sigma_i \sigma_j$ is a function to compute the energy value defined for each activity pattern. We tuned the parameters of the model (h_i and J_{ij}) using the gradient ascent algorithm that maximizes the likelihood function (Ezaki et al., 2017). The accuracy of fitting (r_D) was sufficiently high (0.997). We used these energy values in the following analyses.

Construction of Energy Landscape

Here, we show that two activity patterns $(\boldsymbol{\sigma} =)\boldsymbol{\alpha}$ and $(\boldsymbol{\sigma} =)\boldsymbol{\beta}$ are neighbors if and only if they differ only at a single ROI, i.e., if the Hamming distance between these activity patterns is equal to 1. Thus, each of the 32 activity patterns had five neighboring activity patterns. The algorithm to compute the energy landscape was as follows. (i) We selected an activity pattern from a list of 32 possible activity patterns. (ii) We checked the energy value of the selected activity pattern and its neighboring activity patterns and moved to one of these patterns that had the minimum energy value. (iii) We repeated (ii) until the current activity pattern was selected as the next move, which meant that that pattern was a local minimum, having a smaller energy value than that of its neighbors. (iv) We recorded this sequence of moves. (v) These procedures from (i) to (iv) were performed for all initial activity patterns. The resultant paths defined sets of activity patterns that belong to a basin of energy local minimum activity patterns. In this fashion, a single energy local minimum was associated with each activity pattern. A set of activity patterns that were associated with an energy local minimum is termed the basin of the local minimum. Using this labeling, we mapped BOLD signals to the time series of the label of basins. It should be noted that transition rates of the brain state between basins are a common measure used in the literature (Watanabe et al., 2014; Watanabe and Rees, 2017; Ezaki et al., 2018).

RESULTS

Behavioral Performance

Behavioral performance (mean ± SD) for the auditory gradCPT generally reached a satisfactory level: 74.1 ± 14.6% for hit rate, 23.8 ± 11.7% for FA rate, and 1.56 ± 0.65 for d'. Due to scanner noise, d' observed in the present study was worse than that (2.85 ± 0.82) obtained from an experiment outside the scanner: $t = 6.42$, $p < 0.001$, Cohen's $d = 1.77$. The number of alternations between in- and out-of-the-zone periods was 32.2 ± 3.4 throughout 800-s sessions. Correlations between the task performance and alternation numbers did not reach statistical significance: $|r| < 0.36$, $p > 0.07$. The pattern of the results is consistent with previous findings obtained in a laboratory environment (Terashima et al., 2021).

Imaging Results

We first specified trial-based brain activations. For both FA and CR trials, we found activations of the auditory cortex (Brodmann area: BA 42/41), STG (BA 22), right dorsolateral PFC (BA 46), supplementary motor area (BA 6), and premotor area (BA 6) (**Table 1**). Activated areas largely overlapped between FA and CR trials. Specifically, the activity of the auditory cortex was extended from the primary auditory area to the secondary auditory area, including the lateral area and posterior area (**Figure 2**). Local maxima of auditory-related activations were positioned at the lateral area, regardless of trial type, although additional activations were found in the anterior area of the auditory cortex during FA trials. Cognitive subtraction analyses revealed that significant activations did not survive, in contrast to FA vs. CR trials. Thus, for auditory gradCPT, unsuccessful inhibition of prepotent responses shares similar mental processes with successful inhibition of prepotent responses.

We computed the VTC for each run and produced regressors corresponding to the VTC convolved with an HRF (**Figures 3A,B**). The pattern of VTC-related activations differed from that of trial-based activations: the frontal eye field (BAs 8 and 6) and intraparietal sulcus (BA7) (**Figure 3C**). Local maxima were positioned in the coordinates (−26, 6, 56; $T = 4.22$) and (28, 6, 54; $T = 5.06$) for the SFG; (−12, −72, 58; $T = 4.36$) and (18, −66, 56; $T = 5.02$) for the intraparietal sulcus. There was no region negatively correlated with VTCs. Taken collectively,

TABLE 1 | Brain regions activated during false alarm and correct rejection trials.

Brain region	BA	*X*	*Y*	*Z*	*T*-value
False alarm trials					
Prefrontal cortex	R46	38	50	24	4.65
Premotor area	R6	54	10	44	5.38
Supplementary motor area	R6	2	2	62	5.42
Superior temporal gyrus	L22	−56	−28	10	6.63
	R22	64	−24	6	5.58
Auditory cortex	L42	−54	−26	14	5.94
	R42	56	−36	10	7.27
Correct rejection trials					
Prefrontal cortex	R46	42	50	20	5.40
Premotor area	R6	52	8	44	5.69
Supplementary motor area	R6	2	6	60	6.64
Superior temporal gyrus	L22	−54	−20	4	9.78
	R22	54	−26	8	9.56
Auditory cortex	L42	−60	−34	12	5.98
	R42	64	−26	14	8.03
Supramarginal gyrus	R40	52	−38	50	7.78

Activations of local maxima are significant at $p < 0.001$ (uncorrected; $T > 3.47$). Coordinates (x, y, z) are in MNI stereotaxic space. BA, Brodmann area; L, left; R, right.

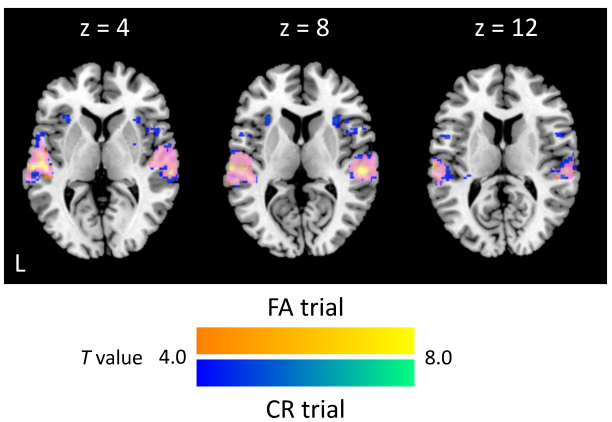


FIGURE 2 | Extent and intensity of auditory-related activations. Activated areas during false alarm (FA) and correct rejection (CR) trials are colored yellow and blue, respectively, whereas overlapped activations are colored magenta ($p < 0.05$, corrected at the cluster level; $k = 279$ voxels).

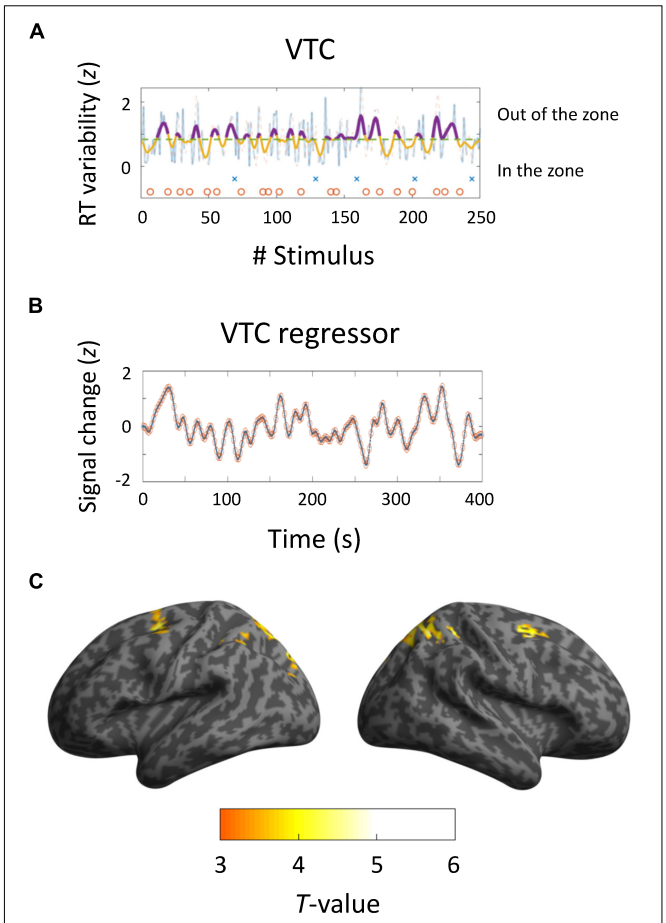


FIGURE 3 | Brain activations derived from variance time courses (VTCs). **(A,B)** Behavioral and neuroimaging results of a representative participant. VTCs were computed using RTs for 250 trials. Dashed lines show median VTCs. The area below the line represents in-the-zone period, whereas the area above the line indicates out-of-the-zone period. Crosses and circles indicate false alarm and correct rejection trials, respectively. VTC regressors were convolved with a canonical hemodynamic response function and downsampled to 0.5 Hz. **(C)** Activated areas related to VTCs ($p < 0.05$, corrected at the cluster level; $k = 71$).

these results indicate that DAN regions are closely linked with modulation of sustained auditory attention.

Brain Dynamics

We performed energy landscape analysis to characterize multivariate dynamics of the five ROIs. BOLD signals were binarized into −1 or 1, by which an activity pattern of these ROIs was represented by one of the possible 32 states. Based on binarized data, these activity patterns were clustered into "basins" of frequently visited states (local minima) (**Figures 4A,B**). We found four local minima. Specifically, two local minima (1 and 4) showed synchronized activity of ROIs (all active and inactive, respectively). Energy landscape analysis revealed two types of activity patterns in ROIs, whereas a general pattern of positive correlations between ROIs was found in terms of functional connectivity analysis (**Figure 4C**). These two activity patterns were classified into DAN and DMN regions. The activity pattern of the STG was associated with those of the DAN regions. Thus, we focused on dynamics between the two brain states (local minima 1 and 4).

We investigated whether auditory gradCPT performance was explained by the transition rate between the two brain states. We computed the transition rate by counting the number of direct transitions from basins 1 to 4 and from 4 to 1. The transition rate was positively correlated with the number of alternations between in- and out-of-the-zone periods: $r = 0.49$, $p = 0.011$ (**Figure 5**). When we calculated the transition rate between all basins, the correlation result was blurred: $r = 0.18$, $p = 0.38$. There was no

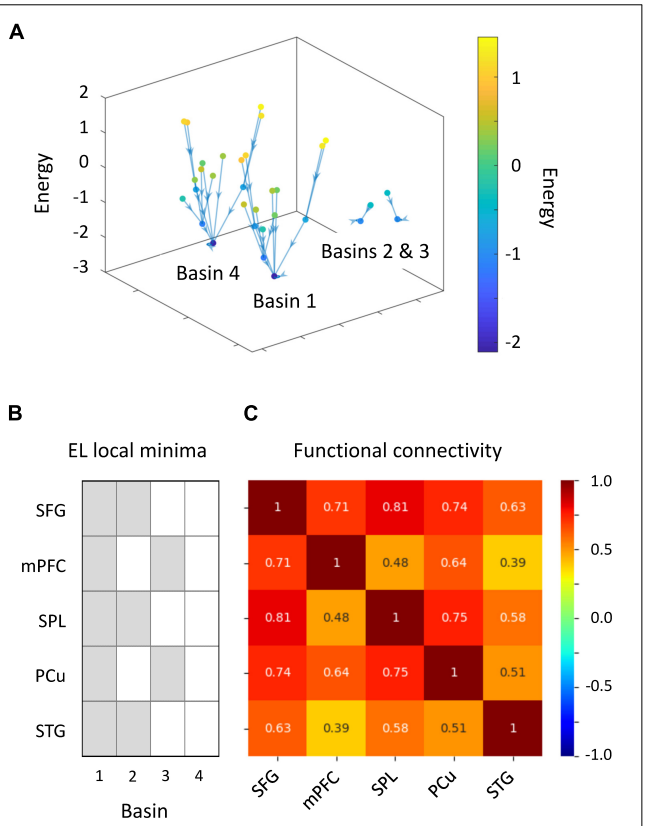


FIGURE 4 | Results of energy landscape (EL) analysis. **(A)** Constructed EL. BOLD signals obtained from five regions of interests (ROIs) were binarized. Activity patterns were computed using the appearance probability of binarized states and categorized into four basins in EL. **(B)** ROI classification based on the disconnectivity graph. The EL transition rate was derived from activity patterns of active and inactive ROIs. **(C)** Comparison of EL local minima and functional connectivity.

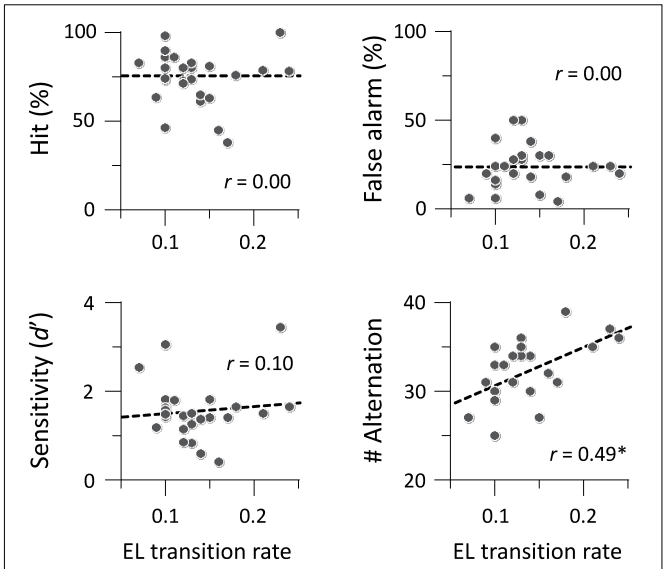


FIGURE 5 | Brain-behavior relationship between transition rate and task performance ($N = 26$). The number of alternations between in- and out-of-the-zone periods was correlated with the transition rate between local minima 1 and 4 in energy landscape (EL) analysis. $^*p < 0.05$ (Bonferroni correction for multiple comparisons).

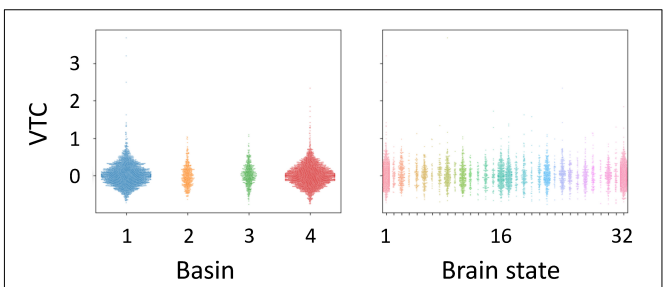


FIGURE 6 | Variance time course (VTC) values identified for each basin and each brain state in the energy landscape. VTC values shown here were subtracted by the median VTC value of each run. Positive values correspond to out-of-the-zone periods.

correlation between the transition rate and other performance measures: hit rate, FA rate, and d' ($r < 0.10$, $p > 0.61$). We did not find any systematic change in VTC values between basins and between brain states (**Figure 6**). In other words, there was no specific activity pattern that could be characterized by in- and out-of-the-zone periods. The results suggest that activations of the DAN and DMN regions, in addition to the STG, are involved in fluctuations in sustained auditory attention, although DAN and DMN activations are functionally differentiated.

DISCUSSION

We provide insights into dynamic brain mechanisms underlying sustained auditory attention. Event-related fMRI analyses demonstrated that frontal and temporal activations overlapped between successful and unsuccessful inhibition of predominant responses. RT variability of the auditory gradCPT was correlated with signal changes of DAN regions, including the SFG and SPL. The activation pattern of response inhibition differed from that of RT variability. Functional connectivity analysis confirmed that task-related activations were generally correlated. The energy landscape analysis produced major and minor brain states to further characterize the dynamics of these regions. Dynamics of DAN regions were synchronized with those of DMN regions in terms of major brain states, whereas dynamics of the DAN and DMN regions were functionally separable in terms of minor brain states. The number of alternations between in- and out-of-the-zone periods increased with an increase in the transition rate between the two major brain states. Taking VTC-related activity into account, we can imagine that neural mechanisms of sustained attention are probably dissociated from those of task performance itself. Therefore, we conclude that dynamic transitions between brain states are closely linked with auditory attentional fluctuations at the behavioral level.

Consistent with previous findings (Terashima et al., 2021), our results showed that alternations between successful and erratic

periods ranged from 25 to 50 s. This suggests that attentional fluctuation is less susceptible to experimental environments and sensory disturbance. Intriguingly, lapses of attention during a Go/No-go task occur every 15–40 s (Vaurio et al., 2009), whereas temporal dynamics of perceptual switching have a shorter timescale, from several to 10 s (Pressnitzer and Hupé, 2006; Kondo et al., 2012). Spontaneous switching in multistable perception is based on formation and selection of perceptual objects, such as auditory streams (Gutschalk et al., 2005; Kondo and Kashino, 2009) and verbal forms (Sato et al., 2004; Kondo and Kashino, 2007). Thus, different biological rhythms are probably involved in perceptual organization and sustained attention, even though mental processes of administered tasks share the same auditory modality.

Auditory functions are supported by hierarchical processing in the brain. From a histological view, the auditory cortex comprises six subregions (Rivier and Clarke, 1997). We found that lateral and posterior areas of the auditory cortex were activated during changes in phonetic categories, that is, different genders of voices. Auditory-related activations extended to the lateral and posterior STG. Previous studies have demonstrated that the lateral area of the auditory cortex is associated with perceptual formation of speech envelope-modulated noises (Mummery et al., 1999) and triplet-tone sequences (Gutschalk et al., 2005; Kondo and Kashino, 2009). In addition, speech-related stimuli with phonetic cues, regardless of intelligibility, induce STG activity (Scott et al., 2000). The STG is also activated by phonetic reorganization of a stimulus word (Kondo and Kashino, 2007). Thus, the secondary auditory area is critical in change detection for both speech and non-speech stimuli. However, VTC-related activity was found in the SFG and SPL, but not in the auditory cortex and STG. It has been argued that selective attention to auditory perceptual organization is implemented in the intraparietal sulcus (Cusack, 2005). Volitional control to multistable perception is associated with the balance between neural inhibition and excitation in the intraparietal cortex (Kondo et al., 2018). These findings indicate that one of the DAN regions contributes to controlled attention in auditory domain. Thus, it is possible that neural correlates of auditory perceptual organization differ from those of auditory selective attention.

In general, achievement of task performance is required to maintain a constant level of attention. A classical study proposed that the attention system has three major functions: (a) orienting to events, (b) detecting signals for conscious processing, and (c) maintaining an alert state (Posner and Petersen, 1990). Several researchers have indicated that the ability to remain alert over time (sustained attention) is not necessarily the same as the ability to quickly change to an alert state (selective attention) (Petersen and Posner, 2012; Tang et al., 2015; Fortenbaugh et al., 2017). A resource-control model has been proposed to account for the nature of sustained attention (Thomson et al., 2015). This model postulates that a gradual reduction in task performance reflects a bias such that attentional resources are assigned to a default state. For the auditory gradCPT, it should be noted that RTs do not decrease monotonically, but vary with time. Trial-by-trial variability of task performance is frequently observed in humans even when a task is constant (Gilden, 2001). This may be the reason why VTC-related regions include DAN regions, rather than DMN regions.

Dynamic states of different attractors, including DAN, DMN, and STG regions, were linked to attentional fluctuation in the auditory gradCPT. Specifically, the transition rate between brain states was positively correlated with the number of alternations between in- and out-of-the-zone periods. This indicates that dynamics of brain states, rather than strength of brain activity, are important to predict moment-to-moment attentional levels. However, it is difficult to interpret what mental processes are directly reflected in such a brain state. We would like to emphasize that coarse-grained brain dynamics are important to explain fluctuations of sustained auditory attention.

Activity patterns of DAN and DMN regions were generally synchronized during the auditory gradCPT, but functionally separable. Dynamic states of brain regions did not differ between in- and out-of-the-zone periods. In contrast, it is well known that intrinsic BOLD signals of DAN regions are anti-correlated with those of DMN regions during rest (Fox et al., 2005). Previous studies using energy landscape analysis have revealed that dynamic transitions of brain states can predict individual differences in visual perception and personality traits. Transitions to frontal-area or visual-area-dominant states are associated with perceptual switching of a bistable structure-from-motion stimulus (Watanabe et al., 2014). Individuals with autism spectrum disorder have infrequent transitions between unstable intermediate states of brain networks (Watanabe and Rees, 2017). Thus, future studies should investigate how brain states change between task-positive and task-negative paradigms.

Brain activations time-locked to FA and CR trials were found in the right dorsolateral PFC, motor-related areas, and auditory-related areas. Meta-analysis studies have argued that the right inferior frontal cortex (Aron et al., 2004) and supplementary motor area (Simmonds et al., 2008) are responsible for response inhibition in Go/No-go tasks. From a methodological perspective, procedures of standard CPTs are similar to those of Go/No-go tasks that assess the ability to suppress unwanted actions or predominant responses (Ogg et al., 2008; Koizumi et al., 2018). In the dual-network model for top-down control (Dosenbach et al., 2007), the fronto-parietal network initiates and adjusts adaptive control on a trial-by-trial basis, whereas the cingulo-opercular network maintains a mental set throughout an entire task epoch. In addition, most studies on vigilance have investigated a linear decrease in brain activity (Olsen et al., 2013). It should be noted that the present study focused on moment-to-moment attentional fluctuations. Energy landscape analysis showed that temporal dynamics of DAN regions are synchronized with those of DMN regions, but that they are functionally separable from each other.

Various researchers have devoted their efforts to investigating transient attention, such as visual search (Treisman and Gelade, 1980), attentive object tracking (Pylyshyn and Storm, 1988), and attentional blink (Raymond et al., 1992). We focused on brain dynamics of sustained attention during auditory gradCPT. The frontal and temporal areas were responsible for intermittent response inhibition, whereas the DAN, DMN, and STG regions were involved in time-series RT variability.

Energy landscape analysis revealed that dynamic transitions between brain states were closely linked to attentional fluctuation. Our findings can yield new insights into various types of research not only on perception and attention, but also on vigilance and mind wandering.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Research Ethics and Safety Committees of Chukyo University and ATR-Promotions. The participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

HMK, HT, KK, and JIK: conceptualization. HMK, HT, TE, and TK: data curation, formal analysis, methodology, resources, and writing—original draft. HMK, TE, KK, and JIK: funding acquisition. HMK and TK: investigation. HMK: project administration and supervision. HMK and TE: writing—review and editing. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by JSPS KAKENHI grants (nos. 17K04494 and 20H01789).

ACKNOWLEDGMENTS

We thank Yasuhiro Shimada for his help with data collection.

REFERENCES

Andersson, J. L., Hutton, C., Ashburner, J., Turner, R., and Friston, K. (2001). Modeling geometric deformations in EPI time series. *Neuroimage* 13, 903–919. doi: 10.1006/nimg.2001.0746

Aron, A. R., Robbins, T. W., and Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends Cogn. Sci.* 8, 170–177. doi: 10.1016/j.tics.2004.02.010

Cheyne, J. A., Carriere, J. S., and Smilek, D. (2006). Absent-mindedness: lapses of conscious awareness and everyday cognitive failures. *Conscious Cogn.* 15, 578–592. doi: 10.1016/j.concog.2005.11.009

Corbetta, M., and Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* 3, 201–215. doi: 10.1038/nrn755

Corbetta, M., and Shulman, G. L. (2011). Spatial neglect and attention networks. *Annu. Rev. Neurosci.* 34, 569–599. doi: 10.1146/annurev-neuro-061010-113731

Cusack, R. (2005). The intraparietal sulcus and perceptual organization. *J. Cogn. Neurosci.* 17, 641–651. doi: 10.1162/0898929053467541

Dosenbach, N. U., Fair, D. A., Miezin, F. M., Cohen, A. L., Wenger, K. K., Dosenbach, R. A., et al. (2007). Distinct brain networks for adaptive and stable task control in humans. *Proc. Natl. Acad. Sci. U. S. A.* 104, 11073–11078. doi: 10.1073/pnas.0704320104

Edkins, G. D., and Pollock, C. M. (1997). The influence of sustained attention on railway accidents. *Accid. Anal. Prev.* 29, 533–539. doi: 10.1016/s0001-4575(97)00033-x

Esterman, M., and Rothlein, D. (2019). Models of sustained attention. *Curr. Opin. Psychol.* 29, 174–180. doi: 10.1016/j.copsyc.2019.03.005

Esterman, M., Noonan, S. K., Rosenberg, M., and Degutis, J. (2013). In the zone or zoning out? Tracking behavioral and neural fluctuations during sustained attention. *Cereb. Cortex* 23, 2712–2723. doi: 10.1093/cercor/bhs261

Ezaki, T., Sakaki, M., Watanabe, T., and Masuda, N. (2018). Age-related changes in the ease of dynamical transitions in human brain activity. *Hum. Brain Mapp.* 39, 2673–2688. doi: 10.1002/hbm.24033

Ezaki, T., Watanabe, T., Ohzeki, M., and Masuda, N. (2017). Energy landscape analysis of neuroimaging data. *Philos. Trans. A Math. Phys. Eng. Sci.* 375:20160287. doi: 10.1098/rsta.2016.0287

Fortenbaugh, F. C., DeGutis, J., and Esterman, M. (2017). Recent theoretical, neural, and clinical advances in sustained attention research. *Ann. N. Y. Acad. Sci.* 1396, 70–91. doi: 10.1111/nyas.13318

Fortenbaugh, F. C., Rothlein, D., McGlinchey, R., DeGutis, J., and Esterman, M. (2018). Tracking behavioral and neural fluctuations during sustained attention: a robust replication and extension. *Neuroimage* 171, 148–164. doi: 10.1016/j.neuroimage.2018.01.002

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., and Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl. Acad. Sci. U. S. A.* 102, 9673–9678. doi: 10.1073/pnas.0504136102

Friston, K. J., Glaser, D. E., Henson, R. N., Kiebel, S., Phillips, C., and Ashburner, J. (2002). Classical and Bayesian inference in neuroimaging: applications. *Neuroimage* 16, 484–512. doi: 10.1006/nimg.2002.1091

Gilden, D. L. (2001). Cognitive emissions of 1/f noise. *Psychol. Rev.* 108, 33–56. doi: 10.1037/0033-295x.108.1.33

Gutschalk, A., Micheyl, C., Melcher, J. R., Rupp, A., Scherg, M., and Oxenham, A. J. (2005). Neuromagnetic correlates of streaming in human auditory cortex. *J. Neurosci.* 25, 5382–5388. doi: 10.1523/JNEUROSCI.0347-05.2005

Hutton, C., Bork, A., Josephs, O., Deichmann, R., Ashburner, J., and Turner, R. (2002). Image distortion correction in fMRI: a quantitative evaluation. *Neuroimage* 16, 217–240. doi: 10.1006/nimg.2001.1054

Klein, C., Wendling, K., Huettner, P., Ruder, H., and Peper, M. (2006). Intra-subject variability in attention-deficit hyperactivity disorder. *Biol. Psychiatry* 60, 1088–1097. doi: 10.1016/j.biopsych.2006.04.003

Koizumi, A., Lau, H., Shimada, Y., and Kondo, H. M. (2018). The effects of neurochemical balance in the anterior cingulate cortex and dorsolateral prefrontal cortex on volitional control under irrelevant distraction. *Conscious Cogn.* 59, 104–111. doi: 10.1016/j.concog.2018.01.001

Kondo, H. M., and Kashino, M. (2007). Neural mechanisms of auditory awareness underlying verbal transformations. *Neuroimage* 36, 123–130. doi: 10.1016/j.neuroimage.2007.02.024

Kondo, H. M., and Kashino, M. (2009). Involvement of the thalamocortical loop in the spontaneous switching of percepts in auditory streaming. *J. Neurosci.* 29, 12695–12701. doi: 10.1523/JNEUROSCI.1549-09.2009

Kondo, H. M., Kitagawa, N., Kitamura, M. S., Koizumi, A., Nomura, M., and Kashino, M. (2012). Separability and commonality of auditory and visual bistable perception. *Cereb. Cortex* 22, 1915–1922. doi: 10.1093/cercor/bhr266

Kondo, H. M., Pressnitzer, D., Shimada, Y., Kochiyama, T., and Kashino, M. (2018). Inhibition-excitation balance in the parietal cortex modulates volitional control for auditory and visual multistability. *Sci. Rep.* 8:14548. doi: 10.1038/s41598-018-32892-3

Langner, R., and Eickhoff, S. B. (2013). Sustaining attention to simple tasks: a meta-analytic review of the neural mechanisms of vigilant attention. *Psychol. Bull.* 139, 870–900. doi: 10.1037/a0030694

Mackworth, N. H. (1948). The breakdown of vigilance during prolonged visual search. *Q. J. Exp. Psychol.* 1, 6–21. doi: 10.1080/17470214808416738

Mummery, C. J., Ashburner, J., Scott, S. K., and Wise, R. J. (1999). Functional neuroimaging of speech perception in six normal and two aphasic subjects. *J. Acoust. Soc. Am.* 106, 449–457. doi: 10.1121/1.427068

Ogg, R. J., Zou, P., Allen, D. N., Hutchins, S. B., Dutkiewicz, R. M., and Mulhern, R. K. (2008). Neural correlates of a clinical continuous performance test. *Magn. Reson. Imaging* 26, 504–512. doi: 10.1016/j.mri.2007.09.004

Olsen, A., Ferenc Brunner, J., Evensen, K. A., Garzon, B., Landrø, N. I., and Håberg, A. K. (2013). The functional topography and temporal dynamics of overlapping and distinct brain activations for adaptive task control and stable task-set maintenance during performance of an fMRI-adapted clinical continuous performance test. *J. Cogn. Neurosci.* 25, 903–919. doi: 10.1162/jocn_a_00358

Petersen, S. E., and Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annu. Rev. Neurosci.* 35, 73–89. doi: 10.1146/annurev-neuro-062111-150525

Posner, M. I., and Petersen, S. E. (1990). The attention system of the human brain. *Annu. Rev. Neurosci.* 13, 25–42. doi: 10.1146/annurev.ne.13.030190.000325

Pressnitzer, D., and Hupé, J. M. (2006). Temporal dynamics of auditory and visual bistability reveal common principles of perceptual organization. *Curr. Biol.* 16, 1351–1357. doi: 10.1016/j.cub.2006.05.054

Pylyshyn, Z. W., and Storm, R. W. (1988). Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spat. Vis.* 3, 179–197. doi: 10.1163/156856888x00122

Raymond, J. E., Shapiro, K. L., and Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: an attentional blink? *J. Exp. Psychol. Hum. Percept. Perform.* 18, 849–860. doi: 10.1037//0096-1523.18.3.849

Rivier, F., and Clarke, S. (1997). Cytochrome oxidase, acetylcholinesterase, and NADPH-diaphorase staining in human supratemporal and insular cortex: evidence for multiple auditory areas. *Neuroimage* 6, 288–304. doi: 10.1006/nimg.1997.0304

Rosenberg, M. D., Finn, E. S., Constable, R. T., and Chun, M. M. (2015). Predicting moment-to-moment attentional state. *Neuroimage* 114, 249–256. doi: 10.1016/j.neuroimage.2015.03.032

Sato, M., Beciu, M., Lævenbruck, H., Schwartz, J. L., Cathiard, M. A., Segebarth, C., et al. (2004). Multistable representation of speech forms: a functional MRI study of verbal transformations. *Neuroimage* 23, 1143–1151. doi: 10.1016/j.neuroimage.2004.07.055

Scott, S. K., Blank, C. C., Rosen, S., and Wise, R. J. S. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain* 123, 2400–2406. doi: 10.1093/brain/123.12.2400

Simmonds, D. J., Pekar, J. J., and Mostofsky, S. H. (2008). Meta-analysis of Go/No-go tasks demonstrating that fMRI activation associated with response inhibition is task-dependent. *Neuropsychologia* 46, 224–232. doi: 10.1016/j.neuropsychologia.2007.07.015

Smallwood, J., and Schooler, J. W. (2015). The science of mind wandering: empirically navigating the stream of consciousness. *Annu. Rev. Psychol.* 66, 487–518. doi: 10.1146/annurev-psych-010814-015331

Stuss, D. T., Murphy, K. J., Binns, M. A., and Alexander, M. P. (2003). Staying on the job: the frontal lobes control individual performance variability. *Brain* 126, 2363–2380. doi: 10.1093/brain/awg237

Tang, Y. Y., Hölzel, B. K., and Posner, M. I. (2015). The neuroscience of mindfulness meditation. *Nat. Rev. Neurosci.* 16, 213–225. doi: 10.1038/nrn3916

Terashima, H., Kihara, K., Kawahara, J. I., and Kondo, H. M. (2021). Common principles underlie the fluctuation of auditory and visual sustained attention. *Q. J. Exp. Psychol.* 74, 705–715. doi: 10.1177/1747021820972255

Thomson, D. R., Besner, D., and Smilek, D. (2015). A resource-control account of sustained attention: evidence from mind-wandering and vigilance paradigms. *Perspect. Psychol. Sci.* 10, 82–96. doi: 10.1177/1745691614556681

Treisman, A. M., and Gelade, G. (1980). A feature-integration theory of attention. *Cognit. Psychol.* 12, 97–136. doi: 10.1016/0010-0285(80)90005-5

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15, 273–289. doi: 10.1006/nimg.2001.0978

Vaurio, R. G., Simmonds, D. J., and Mostofsky, S. H. (2009). Increased intra-individual reaction time variability in attention-deficit/hyperactivity disorder across response inhibition tasks with different cognitive demands. *Neuropsychologia* 47, 2389–2396. doi: 10.1016/j.neuropsychologia.2009.01.022

Watanabe, T., and Rees, G. (2017). Brain network dynamics in high-functioning individuals with autism. *Nat. Commun.* 8:16048. doi: 10.1038/ncomms16048

Watanabe, T., Masuda, N., Megumi, F., Kanai, R., and Rees, G. (2014). Energy landscape and dynamics of brain activity during human bistable perception. *Nat. Commun.* 5:4765. doi: 10.1038/ncomms5765

Worsley, K. J., and Friston, K. J. (1995). Analysis of fMRI time-series revisited–again. *Neuroimage* 2, 173–181. doi: 10.1006/nimg.1995.1023

Conflict of Interest: HT was employed by NTT Corporation. TK was employed by ATR-Promotions.

The remaining 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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers in Neuroscience

ORIGINAL RESEARCH
published: 23 March 2022
doi: 10.3389/fnins.2022.818686

Juxtaposing Medical Centers Using Different Questionnaires Through Score Predictors

Clara Puga[1], Miro Schleicher[1], Uli Niemann[1], Vishnu Unnikrishnan[1], Benjamin Boecking[2], Petra Brueggemann[2], Jorge Simoes[3], Berthold Langguth[3], Winfried Schlee[3], Birgit Mazurek[2] and Myra Spiliopoulou[1]*

[1] Knowledge Management & Discovery Lab, Otto-von-Guericke University Magdeburg, Magdeburg, Germany, [2] Tinnitus Center, Charité-Universitaetsmedizin Berlin, Corporate Member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany, [3] Department of Psychiatry and Psychotherapy, University of Regensburg, Regensburg, Germany

Background: Chronic tinnitus is a clinically multidimensional phenomenon that entails audiological, psychological and somatosensory components. Previous research has demonstrated age and female gender as potential risk factors, although studies to this regard are heterogeneous. Moreover, whilst recent research has begun to identify clinical "phenotypes," little is known about differences in patient population profiles at geographically separated and specialized treatment centers. Identifying such differences might prevent potential biases in joint randomized controlled trials (RCTs) and allow for population-specific treatment adaptations.

Method: Two German tinnitus treatment centers were compared regarding pre-treatment data distributions of their patient population bases. To identify overlapping as well as center-specific factors, juxtaposition-, similarity-, and meta-data-based methods were applied.

Results: Between centers, significant differences emerged. One center demonstrated some predictive power of the patients of the other center with regard to questionnaire score after treatment, indicating similarities in treatment response across center populations. Furthermore, adherence to the completion of the questionnaires was found to be an important factor in predicting post-treatment data.

Discussion: Differential age and gender distributions per center should be considered as regards RCT design and individualized treatment planning.

Keywords: tinnitus, networks, similarity, socio-demographics, adherence, predictive modeling

OPEN ACCESS

Edited by:
Andreas K. Maier, University of Erlangen Nuremberg, Germany

Reviewed by:
Marc Fagelson, East Tennessee State University, United States
Tijana Bojić, University of Belgrade, Serbia
Alessandra Fioretti, European Hospital, Italy

****Correspondence:***
Clara Puga
clara.puga@ovgu.de

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

Received: *19 November 2021*
Accepted: *01 February 2022*
Published: *23 March 2022*

Citation:
Puga C, Schleicher M, Niemann U, Unnikrishnan V, Boecking B, Brueggemann P, Simoes J, Langguth B, Schlee W, Mazurek B and Spiliopoulou M (2022) Juxtaposing Medical Centers Using Different Questionnaires Through Score Predictors. Front. Neurosci. 16:818686. doi: 10.3389/fnins.2022.818686

1. INTRODUCTION

Tinnitus can be described as a phantom auditory perception (Jastreboff, 1990). When investigating tinnitus prevalence, age and gender are commonly analysed variables. While age is considered to be an important factor in predicting tinnitus severity, the effect of gender does not reach consensus yet among the tinnitus research community (Biswas and Hall, 2020). There is also no conclusive evidence linking tinnitus or patient characteristics to treatment outcome (Schlee et al., 2021).

The goal of this study is to expand on previous research by investigating not only tinnitus patient characteristics within a clinical center, but

also whether patients from one center are predictive of patient post-treatment data from another center. Therefore, we investigate how transferable the knowledge gathered in one center is to the other. In summary, the following research questions are tackled:

RQ1: To what extent do age and gender distributions within clinical centers reflect the age and gender distributions of the general population?
RQ2: How are the similarities and differences in both centers with respect to age, gender, and questionnaire scores?
RQ3: To what extent are pre-treatment data of one center predictive of post-treatment data of patients from the same center and from the other center? Does gender improve post-treatment data predictions?

To investigate the interplay between the age variables from the general German population and the tinnitus patients (RQ1), we firstly use a method to juxtapose the two distributions (age in Germany and age in the center) using a kernel density estimation plot. This method helps highlighting the parts of the distributions that are very similar and those which are not. We also analyse the differences in age distributions per gender and per center. As a result, the age distribution of one center's female tinnitus patients is compared to the age distribution of the other center's female tinnitus patients. The same reasoning is used to the male tinnitus patients. Gender is also compared between centers and with the distribution in Germany.

Our approach to uncover similarities in distributions of questionnaire scores by gender and clinical center (RQ2) is network-based. We design the network with nodes as patients and edges as a distance function that measures the difference between the questionnaire scores of each pair of patients. We use the same representation as in a recent work (Puga et al., 2021). After mapping the data into a network representation, we use the netLSD distance to compute and compare the distance across networks. These distances can then be rated, with shorter distances suggesting higher similarity between patients.

Finally, we train models to predict post-treatment questionnaire data (i) within each center and then (ii) with the other center's data (RQ3). These models are computed with the complete sample and per gender. There are two main goals for such analysis: (i) to detect if a center is predictive of the other and (ii) to detect if same-gender patients from one center are predictive of the same-gender patients of the other center. The adherence behavior of the patients to filling out the questionnaires is included as different sets of features in the models. Hence, the extent to which these adherence features improve the prediction of the post-treatment score is also investigated.

The remainder of this article is organized as follows: section 2 presents the dataset properties used for the analyses, section 3 describes the methodology to find the relationship of age and gender with tinnitus prevalence organized per research question, section 4 shows the results of our methods to respond to each research question. In section 5 the results are discussed with respect to the available literature and section 6 summarizes the main findings.

TABLE 1 | Number of patients and treatment with no missing values at t_0, t_1 and t_0 and t_1.

Center	Time	No. of patients		No. of treatments
		f	m	
UHREG	t_0	260	473	19
	t_1	51	108	16
	t_0 and t_1	24	46	9
CHA	t_0	1,828	1,994	1
	t_1	916	885	1
	t_0 and t_1*	852	807	1

500 randomly selected patients were used for the analysis.

2. MATERIALS

The data used in this study are from patients with chronic tinnitus who were admitted to the University Hospital of Regensburg and to the Tinnitus Center, Charité - Universitätsmedizin Berlin. The Regensburg University Hospital collected the data between January 3, 2016 and May 28, 2020. The data from the Tinnitus Center, Charité - Universitätsmedizin Berlin were gathered between January 1, 2011 and October 15, 2015. Despite the fact that both datasets contain data from many questionnaires and socio-demographic data, some variables appear only in one dataset and vice versa. **Table 1** shows the number of patients with available data at different time points as well as the number of different treatments that were assigned. Two time points were considered: t_0 and t_1. The former denotes the time point at admission, while the latter represents the time at the final visit of the patient to the clinical center. For the sake of simplicity, University Hospital of Regensburg is denoted by UHREG and the Tinnitus Center, Charité - Universitätsmedizin Berlin is denoted by CHA.

Specific data pre-processing steps are required depending on the type of analysis performed. The three research questions require different filters. I.e., for RQ1 we consider all patients from both centers who filled out the age and gender question among all questionnaires, at admission. RQ2 and RQ3 require a match on the questionnaire data that are being compared. This phase is critical for the juxtaposition of the medical centers' questionnaires and ensuring the feasibility of the comparison.

Table 2 summarizes the available questionnaires per center. In order to compute the treatment outcome, the scores at admission and after treatment are used. As a consequence, the shared questionnaire (in this case, TQ) has to be available at both time points (t_0 and t_1) in order to learn a model that predicts the treatment outcome.

3. METHODS

3.1. Our Approach to Answer RQ1

To juxtapose the centers to each other and to the general population, we combine statistical testing and visualization tools.

According to RQ1, we compare the age distributions of our two samples s_1 and s_2 (one per center) for each gender separately and correct it for multiple testing over the 6 comparisons that are

TABLE 2 | Questionnaire categories and the available questionnaire data per center.

Category	Questionnaire	CHA	UHREG	Citation
Tinnitus distress	TQ	✓	✓	Goebel and Fichter, 2005
	TLQ	✓		Goebel and Hiller, 1998
	THI		✓	Jacobson and Newman, 1990
	TFI		✓	Meikle et al., 2012; Henry et al., 2016
	TBF12		✓	Greimel et al., 1999
	CGI		✓	Zeman et al., 2011
Physical strain	BI	✓		Brähler and Scheer, 1979
Depressivity	ADSL	✓		Hautzinger and Bailer, 1993
	BSF	✓		Hoerhold et al., 1993
	MDI		✓	Gislén et al., 2003
Stress	PSQ	✓		Levenstein et al., 1993; Fliege et al., 2005
Quality of life	SF8	✓		Beierlein et al., 2012
Coping	SWOP	✓		Scholler et al., 1999
Socio-demographics	SOZK	✓		
	[age, gender]	✓	✓	

TQ, tinnitus questionnaire; TL, tinnitus localization and quality questionnaire; THI, tinnitus handicap inventory; TFI, tinnitus functional index; TBF12, tinnitus impairment questionnaire; CGI, clinical global impression; BI, Berlin complaint inventory; ADSL, general depression scalelong form; BSF, Berlin mood questionnaire; MDI, major depression inventory; PSQ, perceived stress questionnaire; SF8, short-form 8 health survey; SWOP, self-efficacy- optimism-pessimism scale; SOZK, socio-demographics questionnaire.

performed, using Bonferroni correction. In particular, we apply the Shapiro-Wilk test (Shapiro and Wilk, 1965) to check whether the samples follow a normal distribution: if yes, student's test (Gosset, 1908) is applied on the means of the samples; if not, MannWhitney U test (Neuhäuser, 2011) is conducted instead. All tests are performed with $\alpha = 0.05$.

To visualize the distributions, we use kernel density estimation plots. In particular, we plot the two age distributions—of the German population and of the tinnitus patients from both clinical centers. For comparison purposes, we computed the percentage of people at a certain age for both data sets.

Secondly, we identified the regions of the age distributions that have similar shapes. I.e., we plot both distributions in a kernel density estimation plot and then interpret the zones where both distributions behave similarly—when there is a high ratio of German population for a certain age and a high percentage of tinnitus patients too.

This visual tool enables the identification of intervals of age in which both distributions show a high or low percentage of people. On the other hand, we may also identify age intervals in which one distribution has a high percentage of people and the other has lower values. The goal of this analysis is to identify age intervals in both populations (residents in Germany and tinnitus patients) where the percentage of people inside that interval is similar and where it differs. By identifying these differences and similarities, we explore the extent to which the two distributions agree. Concerning gender, the female and male ratio in Germany and in each clinical center are plotted using a bi-directional bar plot.

3.2. Our Approach to Answer RQ2

3.2.1. Modeling Center Similarity With Respect to Questionnaire Score and Gender

A network-based analysis is performed to compute and interpret the similarity of tinnitus patients per gender and clinical center. In this network, nodes represent patients and edges represent the distance between them. The questionnaire score used to compute the edge weights is the TQ score. As a first step we standardize the questionnaire scores, as in Equation (1).

$$score'_{p_{i,X}} = \frac{q_{p_{i,X}} - \mu_{q,X}}{\sigma_{q,X}} \tag{1}$$

where $p_{i,X}$ denotes a patient i from the clinical center X, $q_{p_{i,X}}$ is the questionnaire score of patient $p_{i,X}$, $\mu_{q,X}$ and $\sigma_{q,X}$ are the mean and standard deviation of the questionnaire scores q in clinical center X, respectively.

The higher the score difference between patients, the weaker the connection between them. To account for that, a transformation $1/x$ is applied as shown in Equation (2), representing the edge weight.

$$w_{p_i,p_j,X} = \frac{1}{|score'_{p_i,X} - score'_{p_j,X}|} \tag{2}$$

where $w_{p_i,p_j,X}$ as the edge weight between patient p_i and p_j in clinical center X.

For the graph (or network) comparison, Tsitsulin et al. (2018) states that there are three main approaches: (i) direct methods, (ii) kernel methods, and (iii) statistical representations. Their approach falls in a different category, which is based on spectral representation. Tsitsulin et al. (2018) introduce the NetLSD (network laplacian spectral descriptor) method which creates a vector for each network using the "heat equation." The difference between these vectors is computed and the final distance is the NetLSD metric. In comparison to the other approaches this method is able to meet three properties simultaneously: permutation invariance, scale-adaptivity, and size-invariance.

The permutation invariance property guarantees that the distance between two isomorphic graphs is equal to zero.

TABLE 3 | Sets of adherence features over all questionnaires, shared questionnaires and categories.

Name	Description
Adherence set 1	average of the adherence values over all questionnaires of a center, i.e. $\frac{\sum_{I \in \mathcal{A}_{center}} adh\_I}{\lvert\mathcal{A}_{center}\rvert}$
Adherence set 2	set of distinct adherence values for each of the shared questionnaires, i.e. $\{adh\_I \lvert I \in \mathcal{S}\}$
Adherence set 3	set of distinct adherence values of the questionnaires in all categories of a center, i.e. $\{adh\_I \lvert I \in \cup_{c \in \mathcal{C}_{center}} Q_c\}$, where Q_c is the set of questionnaires in category c of the center
Adherence set 4	set of distinct adherence values of the questionnaires of the shared categories, i.e. $\{adh\_I \lvert I \in \cup_{c \in \mathcal{V}} Q_c\}$, where Q_c is the set of questionnaires in the shared category c

The scale-adaptivity property is based on the representation including both local and global graph properties. Finally, the size-invariance property takes into account the magnitude of the graphs and can distinguish between graphs with comparable features but different magnitudes.

Datasets from different clinical centers vary in size and hence each graph that represents a different clinical center is of different size. The size-invariance property of NetLSD is critical in this work since it enables graph comparison between graphs of different sizes.

3.2.2. Modeling Adherence

Since the centers use partially different questionnaires, we first organize the questionnaires into topical categories that reflect the (co-)morbidity they capture, independently of the questionnaire(s) in use. Then, we model *adherence with respect to a set of items I* for a patient x, denoted as $adh\_I(x)$ as the percentage of items from I answered by patient x over the total number of items in I, i.e., $|I|$. The set of items I can be a questionnaire or a topical category that encompasses more than one questionnaire. The adherence values are computed for each patient x, but in the following, we skip (x) from the notation for simplicity.

Let $\mathcal{A}_{center}$ be the set of questionnaires in a center and $\mathcal{S} = \cap_{center}\mathcal{A}_{center}$ be the set of shared questionnaires between centers. Similarly, let $\mathcal{C}_{center}$ be the set of categories in a center, where a category encompasses more than one questionnaire. Finally, let $\mathcal{V} = \cap_{center}\mathcal{C}_{center}$ be the set of topical categories common to all centers.

The setups—hereafter named as "adherence sets"—are subsets that represent different adherence rates expressed in percent.

On this basis, we generate sets of adherence features, as depicted on **Table 3** and described hereafter. It is stressed that all sets refer to adherence at registration, i.e., at t_0, since we intend to use these features to augment the original "basic set" of features used in the post-treatment predictors, i.e., at t_1.

"Adherence set 1" is the average adherence behavior in percent over all questionnaires at t_0. I.e., for each patient, all the adherence rates for each questionnaire at t_0 are calculated as the answered questions divided by the total number of questions of the respective questionnaire. This information serves as foundation to calculate the average adherence rate by summing up the percentages and divide this sum by the number of rates. This is a rough summary of the adherence behavior of each patient and can be compared among the centers ignoring, for, e.g., different sets of questionnaires. Since the current task is a prediction task at t_1, the subset must be limited to the visit at t_0. Otherwise, the train set would contain information regarding the target variable (post-treatment data).

"Adherence set 2" includes the average adherence rates for each questionnaire calculated by the answer behavior of a patient at t_0. It should be noted, however, that in this subset only the adherence rates of the questionnaires that are common in both centers to be compared can be used. The rest are excluded from this subset. The reason is on the one hand the availability of the questionnaire and on the other hand the feasibility of the comparison for the cross-center prediction.

"Adherence set 3" assigns all the available questionnaires at a center to a category. For instance, tinnitus distress (TD) groups all tinnitus-related questionnaires. The adherence rate is calculated as the average of the questionnaire adherence rates per category. This is also exclusively for t_0. In order to enable cross-center prediction, only categories available in both centers are included.

"Adherence set 4" is the average of the categories of subset 3 after removing the categories that have no counterpart in the opposite center.

3.3. Our Approach to Answer RQ3

To understand the relationship between the pre-treatment data and the post-treatment data within centers, a regressor is trained in the data of each center with the target being the post-data treatment of interest. In our case, this refers to the TQ score at t_1.

In order to extend the comprehension of the similarity between patients across centers, we fit a regression model in one center and use it to predict the score after treatment of the patients from the other center. Aside from the previously stated adherence features, we introduce a new drill-down criterion to the experiments: gender. As a result, we utilize the models trained using patients from the CHA center to predict post-treatment data of patients from the UHREG center, per gender.

The regression models used are the following: linear regressor (LR), lasso, ridge, and support vector regressor (SVR). When developing these predictors, data are divided into a test and train set, and a 10-fold cross validation is performed inside the train set to choose the best model and its hyperparameters (with a grid search approach). For evaluation, standard measures, such as MAE (mean absolute error), MSE (mean squared error), and R^2 (explained variance) are used.

Firstly, we train regressor models to predict the TQ score after treatment (TQ_{t_1}), per center and with different features. In total, 5 experiments are performed, per center. The first and the one with the least number of features included is constituted by age, gender, and TQ_{t_0}—the basic set. Then, we add adherence features to this set. Our main goal was to understand which set of features were most

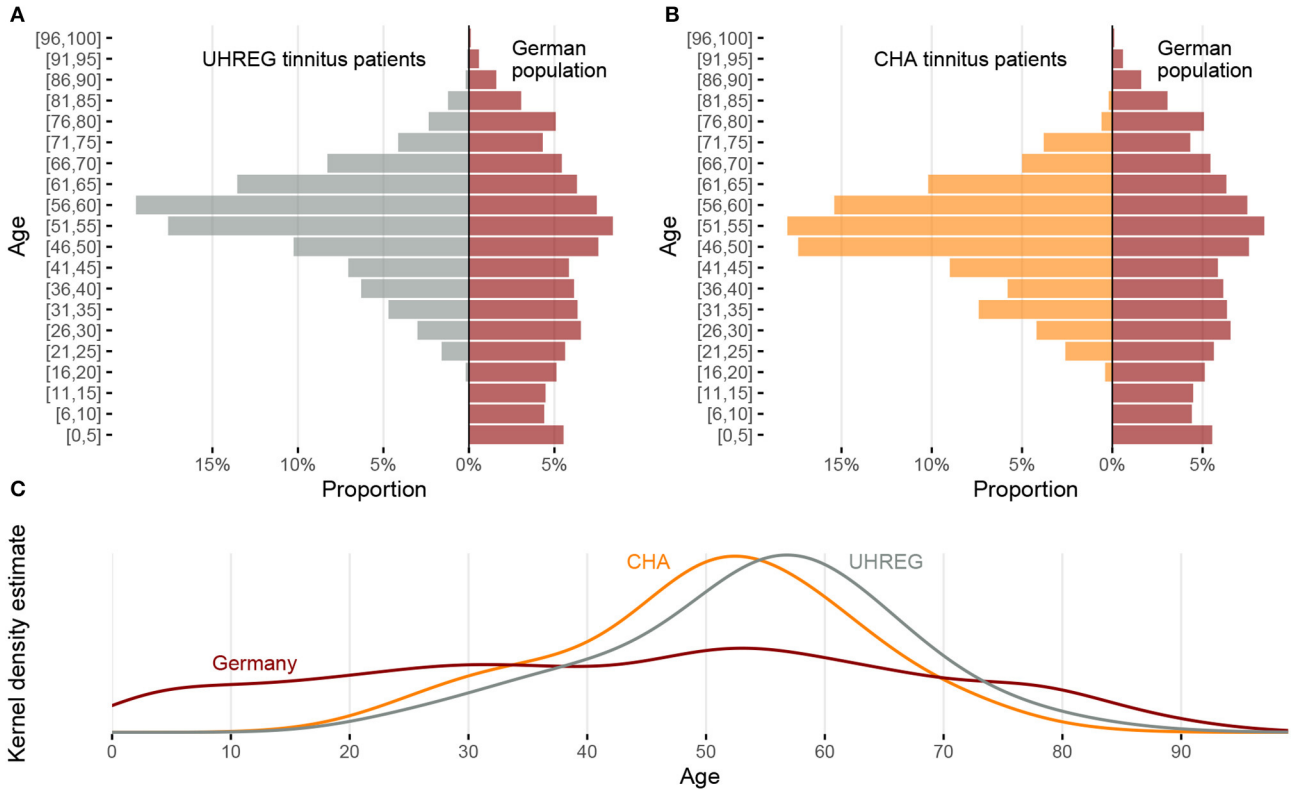


FIGURE 1 | Comparison of age distribution between the two clinical centers and the German population. **(A)** Relative frequency by age intervals for UHREG (left) and the German population (right). **(B)** Relative frequency by age intervals for CHA (left) and the German population (right). **(C)** Density distributions of age for UHREG, CHA and the German population.

predictive of post-treatment TQ score, for within-center and between-center predictions.

4. RESULTS

4.1. Comparison of Age and Gender Distributions in Germany and in Each Clinical Center

4.1.1. Age

Despite the fact that demographics variables were analysed previously in tinnitus research (Seydel et al., 2013; Niemann et al., 2020), it was not yet explored that the age distribution may be influenced by another variable—for instance, the fact that we have less people with 70+ years in the population (due to the fact that life expectancy is within this range in Germany).

Hence, we propose to analyse the age distribution of tinnitus patients from the clinical centers along with the age distribution in Germany. Statistics from the German population age were gathered from a public data source, namely the German Federal Statistical Office ("Statistisches Bundesamt")[1].

Figure 1 shows the distribution of age in Germany and in both clinical centers and the kernel density estimate of the age distributions.

[1]https://www.destatis.de, site accessed in March 2020.

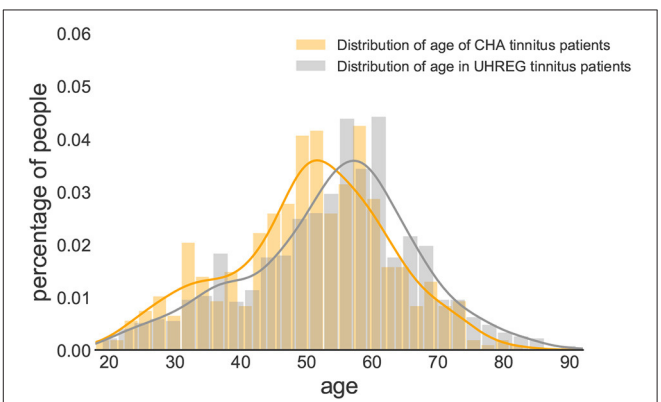


FIGURE 2 | Age distribution comparison between clinical centers.

In the histograms of **Figures 1A,B** and in the kernel density plots of **Figure 1C**, we see that the centers of all three distributions are between 50 and 60 years but the distributions are very different. The curves of the two clinics cross the curve of the German population in the interval between 30 and 40 years of age and a little earlier than 80 years of age.

Figures 2, **3** show the distribution of age per center and gender, respectively. The descriptive statistics of

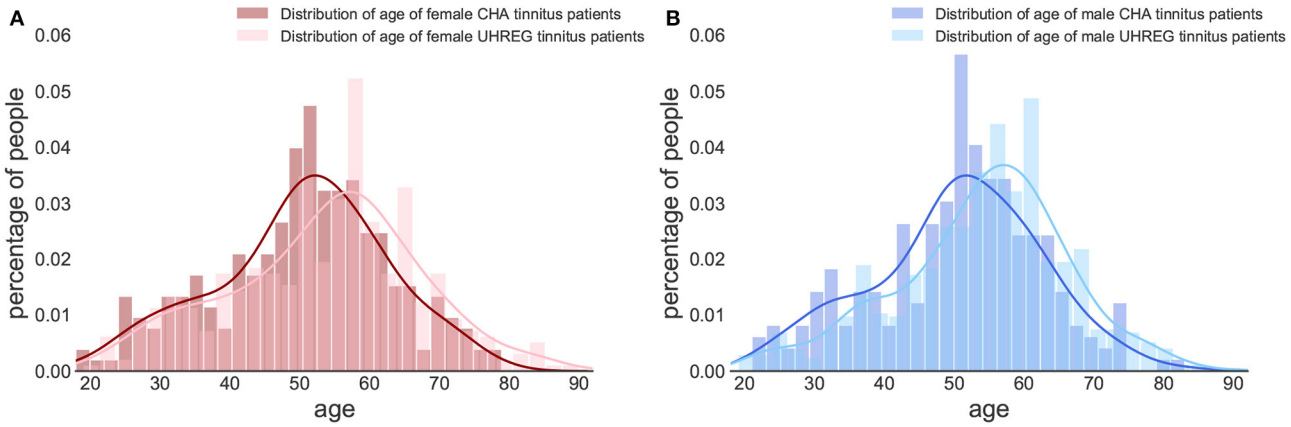


FIGURE 3 | Age distribution per gender and clinical center. **(A)** Female tinnitus patients. **(B)** Male tinnitus patients.

TABLE 4 | Descriptive statistics of age distributions.

Data	N	min	max	mean	SD (σ)	p-value (Shapiro)
age_{uhreg}	1,087	19	91	53.7	12.9	$3.796 * 10^{-8}$
age_{cha}	500	18	83	50.3	12.2	$0.001 * 10^{-1}$
$age_{uhreg,f}$	397	19	90	53.5	13.8	0.001
$age_{uhreg,m}$	690	19	91	53.9	12.5	$2.080 * 10^{-6}$
$age_{cha,f}$	260	18	79	50.3	12.4	0.006
$age_{cha,m}$	240	21	83	50.4	12.1	0.021

SD, standard deviation; f, female; m, male; N, number of data points.

these distributions are shown in **Table 4**, per center and per gender.

We perform statistical tests to compare the age distributions between centers and determine whether there is statistical evidence to confirm that they are different. Given that the hypotheses of normality (Shapiro Wilk test) are rejected for all subsets in **Table 4**, we elect the Mann-Whitney test to compare location measures between samples. In particular, we test (1) whether female (first line of **Table 5**) vs. male tinnitus patients (second line) featured similar age distributions in the two centres and (2) whether female and male patients featured similar age distributions within each centre (3rd line for UHREG, 4th line for CHA) and, finally, whether female tinnitus patients of one centre featured similar age distributions as male patients in the other center (last two lines).

As can be seen in the table, it can be stated that female and male patients within the same center follow the same age distribution (H_0 cannot be rejected). The other null hypotheses are rejected. Based on these findings and complementing it with the analysis of the histograms, we can infer that the age distribution in UHREG is consistently higher than the one in CHA, regardless of gender.

4.1.2. Gender

Regarding gender, **Figure 4** shows the distributions of gender in Germany [from the 'Statistisches Bundesamt' (see text footnote 1)] and in both clinical centers. We can see that distribution of gender in CHA is close to the one in Germany. In contrast, UHREG has a higher percentage of male individuals than the general German population.

4.2. Clinical Center Similarity With Respect to TQ Score Per Gender

A network-based analysis is carried as a method to capture the similarity between patients, with respect to the TQ score. Four networks are generated: one per gender and one per center.

The four networks that represent distinct groups of tinnitus patients are illustrated in **Figure 5**. The purple networks represent TQ score data from female patients, whereas the blue networks represent TQ score data from male patients. In this analysis, the emphasis is on the TQ scores and similarity is modeled with respect to this variable.

As previously stated, the nodes represent the patients, while the edges connecting them show the similarity between them. Denser areas in each network reflect patients who are similar to one another with respect to their TQ score, whereas darker and thicker edges indicate a stronger connection between patients and therefore a high similarity.

The NetLSD score provides a measure of the distance between networks. The lower this score, the higher similarity between the networks. Given the purpose of this paper, we focus on the difference of the distance values only: **Figure 5** illustrates

TABLE 5 | Medians, Mann-Whitney U-statistic and *p*-value of a Mann-Whitney two-sided test for comparison of two samples.

Sample 1 (s_1)	Sample 2 (s_2)	Median (s_1)	Median (s_2)	U (statistic)	p-value
$age_{uhreg,f}$	$age_{cha,f}$	55	51	59233.0	$< 0.01^*$
$age_{uhreg,m}$	$age_{cha,m}$	55	51	97328.5	$< 0.01^*$
$age_{uhreg,m}$	$age_{uhreg,f}$	55	55	138974.5	0.69
$age_{cha,m}$	$age_{cha,f}$	51	51	31163.0	0.98
$age_{uhreg,m}$	$age_{cha,f}$	55	51	105214.0	$< 0.01^*$
$age_{uhreg,f}$	$age_{cha,m}$	55	51	54711.5	$< 0.01^*$

*An asterisk * indicates statistical significance after Bonferroni correction of the critical value. Therefore, the* $p_{crit} = \alpha / n_{comparisons} = 0.05/6 \approx 0.008$ *.* $age_{uhreg,f}$ *denote the age of female tinnitus patients in UHREG,* $age_{cha,f}$ *the age of female tinnitus patients in CHA,* $age_{uhreg,m}$ *denotes the age of male tinnitus patients in UHREG and* $age_{cha,m}$ *the age of male tinnitus patients in CHA.*

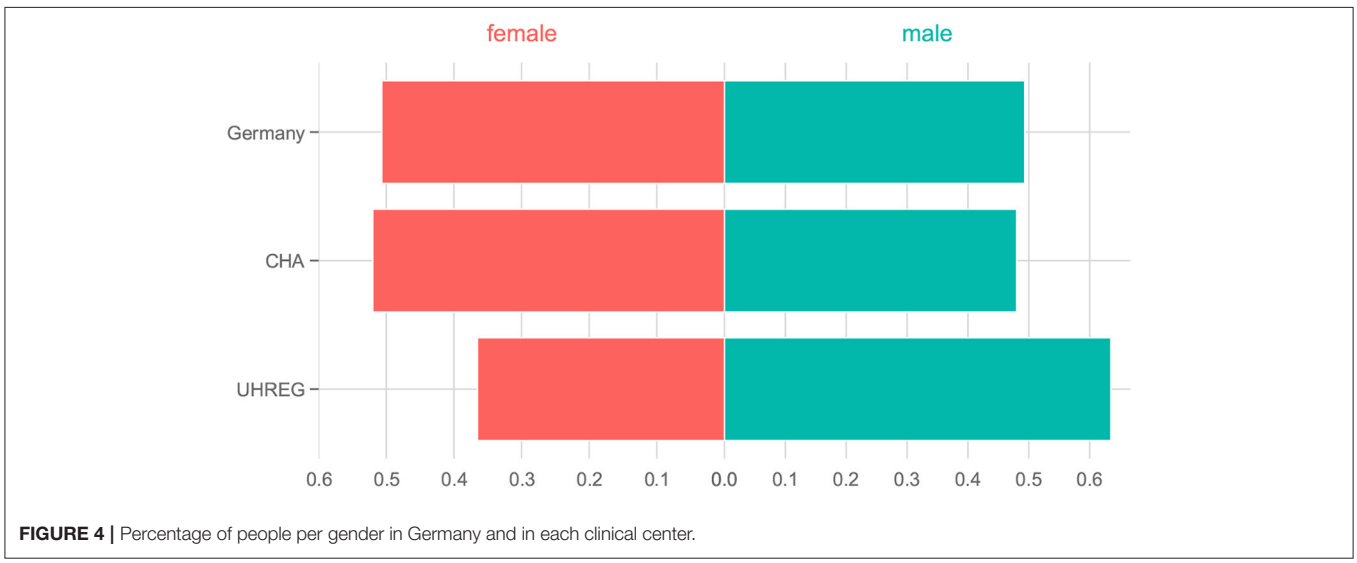


FIGURE 4 | Percentage of people per gender in Germany and in each clinical center.

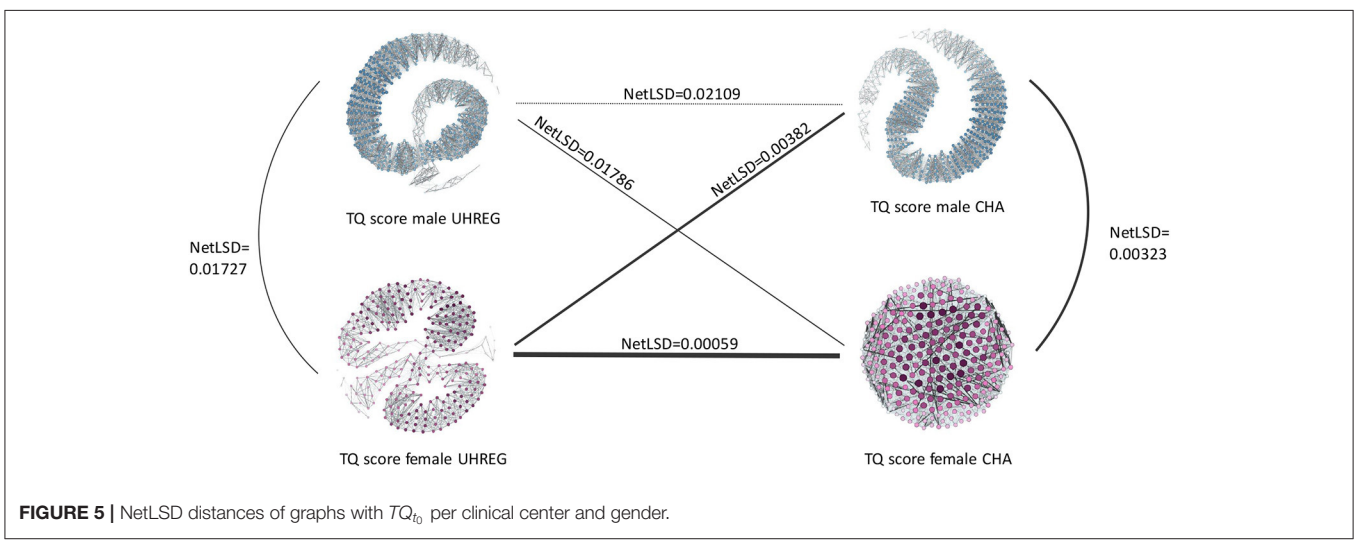


FIGURE 5 | NetLSD distances of graphs with TQ_{t_0} per clinical center and gender.

that, compared to the respective other networks, female patient networks in UHREG and CHA are the most similar, and male patient networks in UHREG and CHA the most dissimilar with respect to the TQ score.

Irrespective of the findings in **Table 5**, which demonstrated statistically significant differences in the age distributions of female patients within centres, the female patients' networks were the most similar according to the TQ score. By contrast, male

tinnitus patients, who also featured significant differences in the age distributions, differed most in terms of their TQ scores.

The blank regions of the graphs indicate that there is no edge linking nodes on opposite sides of the graph. This occurs following the graph pruning phase, during which non-statistically significant edges are deleted. These empty areas differ amongst graphs due to their diverse characteristics; some have more statistically significant edges than others and hence are more connected.

4.3. Adherence Feature Sets

In order to capture the diverse spectrum of patient behavior in each center, all available and applicable records are used to get the adherence meta-information. **Table 2** summarizes the available questionnaires per center. This is a necessary step before computing the adherence sets, since a common feature space (in this case, pre-treatment data) is required to compute some of the sets.

The basic set of features is considered to be the common pre-treatment data between centers. In the case of the current study, this corresponds to age, gender and TQ_{t_0}. For both centers the meta-information is calculated. "Adherence set 1" includes for both CHA and UHREG just a single feature with an average adherence rate. "Adherence set 2" includes the common questionnaires.

The only shared questionnaire between CHA and UHREG is the TQ. Therefore, this subset includes only one feature: the average adherence rate for TQ at t_0. "Adherence set 3" focus on the average adherence rate of common categories. There are two shared categories: tinnitus distress (TD) and depressivity (D), although both centers, except for the TQ, have different sets of questionnaires per category. "Adherence set 4" is the average of both common categories, which means in fact the average of the underlying questionnaires at t_0.

Table 6 shows the average and the standard deviation (SD) of the adherence features of each center. One important remark is that there are some patients with missing values for the gender from the clinical center CHA. This is also shown in **Table 6** and is of importance for the prediction task.

4.4. Prediction of Questionnaire Score After Treatment

Table 7 shows the predictions of the TQ_{t_1} per center, using different feature sets. These sets differ in their features. The age, gender, and TQ_{t_0} are denoted as basic set of features. Then, four adherence feature sets are added - sets 1, 2, 3 and 4. The target variable is set to be the TQ score at t_1, which we refer to as TQ_{t_1}. Patients with missing values for the used features (for example, missing values of gender in CHA as reported in **Table 6**) are excluded from this study.

For UHREG, the ridge regressor achieves the best results using the basic feature set and and the "adherence feature set 1". For CHA, the SVR (support vector regressor) regressor achieves the best results using the basic feature and "adherence set 4".

To predict UHREG patients' treatment outcome, three models are trained with three different datasets: (i) all CHA patients, (ii) female CHA patients only and (iii) male CHA patients only. These models are then tested on UHREG patient data. **Table 8** illustrates the results.

We can conclude that the female tinnitus patients from UHREG are the most predictable among all learners from CHA trained with all their patients. The basic set and the "adherence 2" set of features are the most successful in terms of R^2.

In the second part of **Table 8**, only female patients from CHA are used to train a model and this model is used to predict the female patients from UHREG. In this example, the basic feature set provides the highest R^2.

Finally, a learner is trained only on the male patients from CHA and used to predict the TQ_{t_1} on patients from UHREG. The best set of features is the basic set with the "adherence set 3" set of features. This means that these features were the ones that best predicted TQ_{t_1} on patients from UHREG and can thus be deemed the most useful for understanding the post-treatment TQ score (in comparison to the other set of features).

5. DISCUSSION

In tinnitus research, age and gender have been two variables of interest to analyse. Recently, Niemann et al. (2020) showed that women present a higher level of depression and tinnitus-related distress. Another study by Seydel et al. (2013) recognized age and gender as the most relevant factors to predict tinnitus distress. Rodrigo et al. (2021) investigated the impact of several features for the success of internet-based CBT (cognitive behavioral therapy) on tinnitus patients. In this study, age and gender were used as features but the feature that proved the highest impact on the outcome of the treatment was the education level of the patients. Their findings indicated that patients with a higher level of education were more likely to succeed after treatment.

In the present study, we found that the age distribution of the general German population is partly reflected in the age distributions of the two centers' tinnitus patients. All three distributions reflect a drop in the age window 75+. An explanation could be that elderly patients are less mobile, especially those in rural areas. Another explanation could be that the likelihood of other conditions increases with age - which might be associated with a perception of tinnitus as relatively less distressing. These are hypotheses and more investigation is needed to understand this pattern of chronic elderly tinnitus patients. Middle-aged people, on the other hand, were considerably over-represented in the tinnitus centers relative to the general German population. Therefore, the percentage of patients that seek medical care for tinnitus within middle-aged people is higher than for the other age ranges and it cannot be explained by the German population characteristics.

Within, but not between centers, same-gender patients were found to differ significantly in age. While this is true for age, the results from the netLSD distances show that female patients from different centers are, among all pairs of networks, the most similar according to

TABLE 6 | Average and SD of adherence features.

Center	Adherence set	Gender			
		f	m	NA	all
CHA	Adherence set 1	99.2% ± 5.3%	99.2% ± 5.5%	55.9% ± 24.6%	97.7% ± 10.5%
	Adherence set 2	100.0% ± 0.0%	100.0% ± 0.0%	14.3% ± 35.9%	97.0% ± 17.0%
	Adherence set 3	TD: 99.8% ± 2.9% D: 98.8% ± 7.7%	TD: 99.8% ± 2.9% D: 98.9% ± 8.2%	TD: 50.0% ± 22.4% D: 76.2% ± 25.6%	TD: 98.1% ± 10.4% D: 98.1% ± 10%
	Adherence set 4	99.3% ± 4.6%	99.4% ± 5.2%	63.1% ± 20.3%	98.1% ± 9%
UHREG	Adherence set 1	66.7% ± 21.4%	68.5% ± 19.9%		67.9% ± 20.3%
	Adherence set 2	100.0% ± 0.0%	99.9% ± 0.3%		99.9% ± 0.2%
	Adherence set 3	TD: 62.5% ± 19.8% D: 87.5% ± 33.8%	TD: 65.2% ± 17.6% D: 84.8% ± 36.3%		TD: 64.3% ± 18.3% D: 85.7% ± 35.2%
	Adherence set 4	75.0% ± 25.7%	75.0% ± 25.9%		75.0% ± 25.6%

TD, tinnitus distress; D, depressivity; f, female; m, male.

TABLE 7 | Prediction of TQ score at t_1 with and without adherence feature sets.

Center	Metric	Features	LR	LASSO	Ridge	SVR
UHREG	MAE	Basic set	8.896	13.3	9.4	13.5
	MSE	($N = 70$)	123.4	228.2	157.3	273.3
	R^2		0.674	0.585	0.392	0.481
	MAE	Basic set + adherence set 1	9.426	10.490	**9.0**	12.0
	MSE	($N = 70$)	165.148	199.0	**117.7**	227.4
	R^2		0.495	0.349	**0.714**	0.431
	MAE	Basic set + adherence set 2	12.9	8.3	11.5	10.0
	MSE	($N = 70$)	234.6	114.2	193.8	170.1
	R^2		0.443	0.646	0.391	0.545
	MAE	Basic set + adherence set 3	11.9	9.3	8.9	10.6
	MSE	($N = 70$)	233.5	115.8	144.7	199.8
	R^2		0.180	0.683	0.585	0.299
	MAE	Basic set + adherence set 4	11.2	8.5	8.6	8.6
	MSE	($N = 70$)	187.6	146.6	132.9	116.2
	R^2		0.406	0.527	0.607	0.688
CHA	MAE	Basic set	6.220	6.680	6.955	6.760
	MSE	($N = 500$)	62.388	76.331	79.239	80.148
	R^2		0.804	0.742	0.766	0.720
	MAE	Basic set + adherence set 1	6.3	7.8	6.9	7.4
	MSE	($N = 218$)	65.5	99.2	86.8	89.3
	R^2		0.829	0.745	0.771	0.780
	MAE	Basic set + adherence set 2	7.1	8.2	6.8	5.7
	MSE	($N = 218$)	85.9	104.1	77.3	54.2
	R^2		0.761	0.743	0.692	0.823
	MAE	Basic set + adherence set 3	7.3	6.8	7.3	6.3
	MSE	($N = 218$)	78.6	73.6	82.2	71.5
	R^2		0.702	0.771	0.817	0.729
	MAE	Basic set + adherence set 4	7.2	7.0	6.8	**6.8**
	MSE	($N = 218$)	90.4	78.7	70.8	**73.2**
	R^2		0.749	0.761	0.810	**0.839**

The error metrics presented are as follows: MAE (mean absolute error); MSE (mean squared error); R^2 (explained variance) and the regressors trained are as follows: LR (linear regressor), LASSO, Ridge and SVR (support vector regressor). In bold are the results of the regressor trained with the set of features that achieved the highest R^2, per center.

their TQ score. Another finding was the fact that age of female and male patients from different centers were not statistically different.

Another intriguing finding was that models trained with adherence features outperformed baseline models that did not include these features. The model trained with female CHA

TABLE 8 | Prediction of *TQ* at t_1 on UHREG tinnitus patients with model trained on CHA tinnitus patients (results from the model with the highest R^2 are shown).

Model trained on	Predicted on	Adherence	MAE	MSE	R^2
CHA all	UHREG all ($N = 70$)	Basic set	9.9	158.7	0.514
($N = 500$)	**UHREG female ($N = 24$)**		**9.7**	**154.1**	**0.533**
	UHREG male ($N = 46$)		10.0	161.0	0.500
	UHREG all ($N = 70$)	Basic set + adherence set 1	9.9	160.5	0.510
	UHREG female ($N = 24$)		**9.6**	**153.8**	**0.532**
	UHREG male ($N = 46$)		10.1	164.0	0.493
	UHREG all ($N = 70$)	Basic set + adherence set 2	10.0	160.4	0.517
	UHREG female ($N = 24$)		**9.7**	**153.5**	**0.535**
	UHREG male ($N = 46$)		10.1	164.0	0.503
	UHREG all ($N = 70$)	Basic set + adherence set 3	10.0	162.0	0.505
	UHREG female ($N = 24$)		**9.7**	**154.8**	**0.529**
	UHREG male ($N = 46$)		10.2	165.9	0.487
	UHREG all ($N = 70$)	Basic set + adherence set 4	10.1	162.9	0.505
	UHREG female ($N = 24$)		**9.8**	**155.6**	**0.530**
	UHREG male ($N = 46$)		10.2	166.7	0.487
CHA female		Basic set	**9.4**	**148.8**	**0.544**
($N = 260$)		Basic set + adherence set 1	9.5	157.0	0.519
	UHREG female ($N = 24$)	Basic set + adherence set 2	9.8	153.9	0.535
		Basic set + adherence set 3	11.0	217.2	0.420
		Basic set + adherence set 4	9.9	167.1	0.492
CHA male		Basic set	10.2	166.8	0.490
($N = 240$)		Basic set + adherence set 1	10.1	163.1	0.496
	UHREG male ($N = 46$)	Basic set + adherence set 2	10.1	162.5	0.496
		Basic set + adherence set 3	**9.8**	**161.5**	**0.498**
		Basic set + adherence set 4	10.1	164.6	0.489

In bold are the results from the results predicted in the subset and the features that achieved the highest R^2.

tinnitus patients was more predictive of female UHREG patients than the model trained with male CHA patients and tested on male UHREG patients. As a result, UHREG female patients are better predicted with CHA data than male patients. This means that treatment-improvement rates and patterns improve more similarly in females than males, when centers are compared. This result can be cross-checked with the NetLSD distance results. In this analysis, female tinnitus patients from different centers were also found to be more similar than male patients with respect to questionnaire score at t_0. Hence, we can conclude that both results from different methods agree in terms of similarity. Despite the heterogeneity of clinical centers and the fact that these are preliminary results, there may be an indication that patients across clinical centers share similar characteristics. It is worth noting that the amount of available data for analysis varies between clinical centers, which may have an impact on how representative they are of their patients. As a result, we consider the reported findings to be preliminary and the analysis should be expanded to larger datasets.

The adherence features that summarize the information the most, "adherence 1" and "adherence 4," produced the greatest improvement in the R^2 value in the gender-agnostic intra-center analysis. In the inter-center prediction using all the data from the training center, however, "adherence 2" outperforms the other adherence sets. Nonetheless, the "basic" set outperforms the other 3 adherence features for all, as well as for female and male. Thus, in the cross-center prediction scenario, information on adherence in the common questionnaire(s) seems to be a useful addition.

6. CONCLUSION

In this article, we performed various analyses to compare tinnitus patients from two different large clinical centers. These comparisons were carried mainly focusing on the questionnaire scores before and after treatment, socio-demographics and adherence to the filling of the questionnaires. For that, visualization and prediction methods were implemented along with a network-based representation of the data.

The main findings can be organized into three. The first one being that the distribution of age in Germany agrees, in some age ranges, to the ones from the tinnitus patients from both centers. This indicates good and representative reach of the specialist treatment centers in offering care for their target populations. The second, that female tinnitus patients from one center (CHA) are more predictive of the female patients of the other center (UHREG) than male patients. This result is complemented by the fact that our network-based approach to compute the similarity between networks also agreed that

female tinnitus patients were more similar across centers than male patients, with respect to treatment score at t_0. The third, that including meta-information about the adherence of the patients to the filling of the questionnaires improved the baseline predictions of post-treatment data.

The evaluation's findings could be supported further by finding datasets with more overlapping questionnaires but also closer in terms of sample size. Finding such datasets will be a future task. Incorporating information about gender and adherence could improve the prediction task. Future research should place these findings to the test by applying intra- and inter-center predictions to other centers.

DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: The datasets for this article are not publicly available because no consent of the patients to publish their data was obtained. Requests to access these datasets should be directed to BM, birgit.mazurek@charite.de for the CHA dataset and WS, winfried.schlee@gmail.com for the UHREG dataset.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of the University Medicine Charité Berlin for the CHA Dataset and the Ethics Committee of the University Regensburg for the UHREG Dataset. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

CP and MSc designed and performed the data analysis and wrote the original draft. CP designed and performed the network-based analysis. MSc modeled adherence. CP and UN designed the visualizations. UN optimized the visualizations on interpretability. VU supported on questionnaire interpretation and data preparation. BB, PB, and BM curated the dataset with patients admitted to the Tinnitus Center, Charité – Universitätsmedizin Berlin. JS, BL, and WS worked on data curation and model interpretation. BB, PB, BM, WS, and MSp supervised the data analysis and reviewed and edited the manuscript. BM and MSp led the project. All authors contributed to the article and approved the submitted version.

FUNDING

Part of this project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement number 848261.

REFERENCES

Beierlein, V., Morfeld, M., Bergelt, C., Bullinger, M., and Brähler, E. (2012). Messung der gesundheitsbezogenen lebensqualität mit dem sf-8. *Diagnostica*. 58, 3. doi: 10.1026/0012-1924/a000068

Biswas, R., and Hall, D. A. (2020). "Prevalence, incidence, and risk factors for tinnitus," in *The Behavioral Neuroscience of Tinnitus* (Cham: Springer International Publishing), 3–28.

Brähler, E., and Scheer, J. W. (1979). Scaling of psychosomatic by means of the giessen inventory (gbb)(author's transl). *Psychotherapie Medizinische Psychologie* 29, 14.

Fliege, H., Rose, M., Arck, P., Walter, O. B., Kocalevent, R.-D., Weber, C., et al. (2005). The perceived stress questionnaire (psq) reconsidered: validation and reference values from different clinical and healthy adult samples. *Psychosomatic Med.* 67, 78–88. doi: 10.1097/01.psy.0000151491.80178.78

Gislén, A., Dacke, M., Kröger, R. H., Abrahamsson, M., Nilsson, D.-E., and Warrant, E. J. (2003). Superior underwater vision in a human population of sea gypsies. *Curr. Biol.* 13, 833–836. doi: 10.1016/s0960-9822(03)00290-2

Goebel, G., and Fichter, M. (2005). "Psychiatrische komorbidität bei tinnitus," in *Tinnitus* (Heidelberg: Springer), 137–150.

Goebel, G., and Hiller, W. (1998). *Tinnitus-Fragebogen:(TF); ein Instrument zur Erfassung von Belastung und Schweregrad bei Tinnitus; Handanweisung*. Hogrefe, Verlag für Psychologie.

Gosset, W. S. (1908). The probable error of a mean. *Biometrika* 6, 1–25.

Greimel, K. V., Leibetseder, M., Unterrainer, J., and Albegger, K. (1999). Ist tinnitus messbar? methoden zur erfassung tinnitusspezifischer beeinträchtigungen und präsentation des tinnitus-beeinträchtigungs-fragebogens (tbf-12). *HNO* 47, 196–201.

Hautzinger, M., and Bailer, M. (1993). *ADS, Allgemeine Depressions Skala. Manual.* Technical Report, Beltz Test.

Henry, J. A., Griest, S., Thielman, E., McMillan, G., Kaelin, C., and Carlson, K. F. (2016). Tinnitus functional index: development, validation, outcomes research, and clinical application. *Hearing Res.* 334, 58–64. doi: 10.1016/j.heares.2015.06.004

Hoerhold, M., Klapp, B. F., and Schimmack, U. (1993). Testing the invariance and hierarchy of a multidimensional model of mood by means of repeated measurement with student and patient sample. *Z Med. Psychol.* 1, 27–35.

Jacobson, G. P., and Newman, C. W. (1990). The development of the dizziness handicap inventory. *Arch. Otolaryngol. Head Neck Surg.* 116, 424–427.

Jastreboff, P. J. (1990). Phantom auditory perception (tinnitus): mechanisms of generation and perception. *Neurosci. Res.* 8, 221–254.

Levenstein, S., Prantera, C., Varvo, V., Scribano, M. L., Berto, E., Luzi, C., et al. (1993). Development of the perceived stress questionnaire: a new tool for psychosomatic research. *J. Psychosomatic Res.* 37, 19–32.

Meikle, M. B., Henry, J. A., Griest, S. E., Stewart, B. J., Abrams, H. B., McArdle, R., et al. (2012). The tinnitus functional index: development of a new clinical measure for chronic, intrusive tinnitus. *Ear Hearing* 33, 153–176. doi: 10.1097/AUD.0b013e31822f67c0

Neuhäuser, M. (2011). *Wilcoxon–Mann–Whitney Test*. Berlin: Springer, 1656–1658.

Niemann, U., Boecking, B., Brueggemann, P., Mazurek, B., and Spiliopoulou, M. (2020). Gender-specific differences in patients with chronic tinnitus—baseline characteristics and treatment effects. *Front. Neurosci.* 14, 1–11. doi: 10.3389/fnins.2020.00487

Puga, C., Niemann, U., Unnikrishnan, V., Schleicher, M., Schlee, W., and Spiliopoulou, M. (2021). "Discovery of patient phenotypes through multi-layer network analysis on the example of tinnitus," in *2021 IEEE 8th International Conference on Data Science and Advanced Analytics (DSAA)* (Porto), 1–10.

Rodrigo, H., Beukes, E. W., Andersson, G., and Manchaiah, V. (2021). Exploratory data mining techniques (decision tree models) for examining the impact of internet-based cognitive behavioral therapy for tinnitus: Machine learning approach. *J. Med. Internet Res.* 23, e28999. doi: 10.2196/28999

Schlee, W., Langguth, B., Pryss, R., Allgaier, J., Mulansky, L., Vogel, C., et al. (2021). *Using Big Data to Develop a Clinical Decision Support System for Tinnitus Treatment*. Cham: Springer International Publishing, 175–189.

Scholler, G., Fliege, H., and Klapp, B. F. (1999). Questionnaire of self-efficacy, optimism and pessimism: reconstruction, selection of items and validation of an instrument by means of examinations of clinical samples. *Psychotherapie Psychosomatik Medizinische Psychologie* 49, 275–283.

Seydel, C., Haupt, H., Olze, H., Szczepek, A. J., and Mazurek, B. (2013). Gender and chronic tinnitus. *Ear and Hearing* 34, 661–672. doi: 10.1097/AUD.0b013e31828149f2

Shapiro, S. S., and Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika* 52, 591–611.

Tsitsulin, A., Mottin, D., Karras, P., Bronstein, A., and Müller, E. (2018). "NetLSD: hearing the shape of a graph," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (London), 2347–2356.

Zeman, F., Koller, M., Figueiredo, R., Aazevedo, A., Rates, M., Coelho, C., et al. (2011). Tinnitus handicap inventory for evaluating treatment effects: which changes are clinically relevant? *Otolaryngol. Head Neck Surg.* 145, 282–287. doi: 10.1177/0194599811403882

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.

frontiers | Frontiers in Neuroscience

ORIGINAL RESEARCH
published: 11 April 2022
doi: 10.3389/fnins.2022.836834

Predicting Ecological Momentary Assessments in an App for Tinnitus by Learning From Each User's Stream With a Contextual Multi-Armed Bandit

Saijal Shahania[1], Vishnu Unnikrishnan[1], Rüdiger Pryss[2], Robin Kraft[3], Johannes Schobel[4], Ronny Hannemann[5], Winny Schlee[6] and Myra Spiliopoulou[1]**

[1] Knowledge Management and Discovery Lab, Otto-von-Guericke University Magdeburg, Magdeburg, Germany, [2] Institute of Clinical Epidemiology and Biometry, University of Würzburg, Würzburg, Germany, [3] Institute of Databases and Information Systems, Ulm University, Ulm, Germany, [4] DigiHealth Institute, Neu-Ulm University of Applied Sciences, Neu-Ulm, Germany, [5] Sivantos GmbH - WS Audiology, Erlangen, Germany, [6] Department of Psychiatry and Psychotherapy, University of Regensburg, Regensburg, Germany

OPEN ACCESS

Edited by:
Patrick Krauss,
University of Erlangen Nuremberg, Germany

Reviewed by:
Konstantin Tziridis,
University Hospital Erlangen, Germany
Richard Gault,
Queen's University Belfast, United Kingdom

***Correspondence:**
Saijal Shahania
saijal.shahania@ovgu.de
Myra Spiliopoulou
myra@ovgu.de

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

Received: 16 December 2021
Accepted: 16 February 2022
Published: 11 April 2022

Citation:
Shahania S, Unnikrishnan V, Pryss R, Kraft R, Schobel J, Hannemann R, Schlee W and Spiliopoulou M (2022) Predicting Ecological Momentary Assessments in an App for Tinnitus by Learning From Each User's Stream With a Contextual Multi-Armed Bandit. Front. Neurosci. 16:836834. doi: 10.3389/fnins.2022.836834

Ecological Momentary Assessments (EMA) deliver insights on how patients perceive tinnitus at different times and how they are affected by it. Moving to the next level, an mHealth app can support users more directly by predicting a user's next EMA and recommending personalized services based on these predictions. In this study, we analyzed the data of 21 users who were exposed to an mHealth app with non-personalized recommendations, and we investigate ways of predicting the next vector of EMA answers. We studied the potential of entity-centric predictors that learn for each user separately and neighborhood-based predictors that learn for each user separately but take also similar users into account, and we compared them to a predictor that learns from all past EMA indiscriminately, without considering which user delivered which data, i.e., to a "global model." Since users were exposed to two versions of the non-personalized recommendations app, we employed a Contextual Multi-Armed Bandit (CMAB), which chooses the best predictor for each user at each time point, taking each user's group into account. Our analysis showed that the combination of predictors into a CMAB achieves good performance throughout, since the global model was chosen at early time points and for users with few data, while the entity-centric, i.e., user-specific, predictors were used whenever the user had delivered enough data—the CMAB chose itself when the data were "enough." This flexible setting delivered insights on how user behavior can be predicted for personalization, as well as insights on the specific mHealth data. Our main findings are that for EMA prediction the entity-centric predictors should be preferred over a user-insensitive global model and that the choice of EMA items should be further investigated because some items are answered more rarely than others. Albeit our

CMAB-based prediction workflow is robust to differences in exposition and interaction intensity, experimentators that design studies with mHealth apps should be prepared to quantify and closely monitor differences in the intensity of user-app interaction, since users with many interactions may have a disproportionate influence on global models.

Keywords: contextual multi-armed bandits, mHealth for tinnitus, EMA prediction, similarity-based prediction, prediction on sparse data, prediction on time series with gaps, prediction in mHealth data

1. INTRODUCTION

According to De Ridder et al. (2021), "The capacity to measure the incidence, prevalence, and impact will help in identification of human, financial, and educational needs required to address acute tinnitus as a symptom but chronic tinnitus as a disorder." Mobile health apps for tinnitus have the potential of assisting patients in self-assessment of their condition and of delivering insights on tinnitus heterogeneity to the medical experts, as reported by Probst et al. (2017), Cederroth et al. (2019), and Pryss et al. (2019, 2021) among others. This is particularly the case for mHealth apps that collect Ecological Momentary Assessments (EMA): several studies on mHealth tinnitus apps have demonstrated that EMA recordings deliver insights on tinnitus stages during the day and on the interplay of personal traits and severity of tinnitus symptoms (see, e.g., Probst et al., 2017; Mehdi et al., 2020; Unnikrishnan et al., 2020a; Jamaludeen et al., 2021).

Modern mHealth apps are able to deliver suggestions to the app users, exploiting knowledge about the users' prior behavior (Martínez-Pérez et al., 2014) or behavior change (Mao et al., 2020), and they contribute also to clinical decision support (Watson et al., 2019). In the context of tinnitus, TinnitusTipps[1] delivers information toward "health literacy" and suggestions that promote well-being, e.g., suggestions on physical exercising. Personalization of such suggestions implies taking account of a user's personal traits and needs, and has the potential of improving user experience and of anticipating undesirable developments in the user's condition. However, personalization demands the ability to learn from past EMA and to predict future EMA.

In contrast to passive recordings, e.g., of ambient noise, EMA recordings demand action by the app user. Some studies, as by Probst et al. (2017), concentrate on users who have many recordings. While such studies are beneficial for the acquisition of insights on how tinnitus is experienced in general, they contribute less toward personalized services, which need to learn and predict for each user, even for users who interact rarely with the app and deliver too few data. Schleicher et al. (2020) model the intensity of the interaction with the app as adherence and attempt to identify adherence patterns. Other studies, as Unnikrishnan et al. (2020a) and (Prakash et al., 2021) and the earlier version of this work (Shahania et al., 2021), investigate ways of learning from users who deliver very little data. The main objective of these studies is to provide tools that predict future EMA of a user, thus forming a basis for the prediction of undesirable events and for the design of personalized services. The main challenge of such methods is the scarcity of data for some users.

Technically speaking, a method that learns from all the data of all the users induces a "global model." It abstracts from the idiosyncracies in the EMA recordings of each user and attempts to predict the future condition of a user from all that is known on the past EMA of all users. In contrast, a method that learns from the data of a single user, builds a model peculiar to that user, capturing the user's idiosyncracies and exploiting them for future predictions. Such a method is bound to learn from as little data as are available for each user, but suffers from the obvious disadvantage that some users have too few EMA recordings for any reliable prediction of their future EMA. For such users, it seems reasonable to exploit data of other, similar users for predictive modeling. Indeed, in our earlier works (Unnikrishnan et al., 2020a, 2021), we have shown that methods which exploit similarity among users can predict future EMA recordings of users with little data. We term a model learned on the data of a user and users similar to him/her as a "local model," to highlight the fact that such a model exploits only data in the user's vicinity (in a multi-dimensional space spanned over the static data of the users and their EMA recordings). A model learned on the data of the user without his/her neighbors is a special case of local model that considers only a single user.

There is a well-known separation of methods into idiographic ones that learn for each individual separately and nomothetic ones that aim to explain the common traits of all individuals. Without entering the conceptual debate of idiographic vs nomothetic approaches (cf. the elaboration of Hermans, 1988 among many), we point out that from the technical perspective, a nomothetic approach builds a global model, an idiographic approach builds a local model for each single individual, while an approach that exploits information on a user and the users similar to her is a nomothetic approach that builds a local model by concentrating only on few users that are similar to a given user. Since our methods use the same machine learning instruments independently on whether they build global or local models, we avoid the distinction between nomothecy and idiography hereafter.

In this study, we build upon our earlier works on learning local models from little data (Unnikrishnan et al., 2020a), on comparing methods that learn from all data of all individuals to methods that learn local models (Unnikrishnan et al., 2019) and on frameworks for such comparisons (Shahania et al., 2021). We present a complete framework for studying how methods

[1]https://tinnitustipps.lenoxug.de/

designed for little data can contribute to predicting the condition of a user interacting with a mHealth for chronic tinnitus. In a nutshell, we want to predict the next vector of answers in the best possible way. We investigate the following research questions:

1. **RQ 1:** To what extent can methods that learn from little-data deliver good predictions?
   · *RQ 1a:* How to exploit little-data for predictions?
   · *RQ 1b:* How to orchestrate the invocation of methods building local models, so that they are only invoked when their prediction is good?
   · *RQ 1c:* How to measure the superiority of a little-data method over a method that learns from all the data?
2. **RQ 2:** What factors influence the superiority of local vs. global models?
   · *RQ 2a:* How do different configurations of little-data methods affect their performance?
   · *RQ 2b:* How do different behavioral phenotypes affect the performance of local vs global models?

We address these research questions on the data of a clinical study involving patients with chronic tinnitus: the participants used an mHealth app to record daily EMA, and received regular non-personalized tips on good practices for living with tinnitus.

To answer RQ 1, we consider for streams. Their purpose is to capture the different interaction behavior of the entities (here: mHealth app users), which causes the variability of the data stream per entity. Hereby, the global model exploits the data stream of all entities, whereas the local models only use data of one entity (entity-centric local model) or an entity and its nearest neighbors.

To ensure that we chose the best configuration depending on the varying data streams, we incorporate the global and local models into a contextual multiarmed bandit (CMAB). This bandit employs a strategy to select one model for each time point based on past rewards and additional meta information of the entity. For RQ 2, we derive an evaluation procedure along with the factors that are responsible for selecting each type of model for different data streams. These factors include the history length of the entities, their personal traits, their time of arrival in the system, and the temporal proximity of the predictions.

This article is organized as follows. Section 2 covers the materials used for our study. Section 3 describes the state of the art techniques present in stream learning and describes techniques for the creation of multi-armed bandits. We are using those for defining our proposed approach. This section also contains details about the need to define global and local models to tackle our problem. Sections 4.2 and 4.3 depicts the results on RQ 1 and RQ 2 followed up with the discussion of the insights, findings, insights, and future work in Section 5.

2. MATERIALS

For our work, we have considered the data from a clinical study with the mHealth app TinnitusTipps, conducted in 2018/2019. The study was approved by the Ethical Review Board of the University Clinic Regensburg. The ethical approval number is 17-544-101. All study participants provided informed consent. For more information on the inclusion/exclusion criteria of the participants, please refer to Schlee et al. (2022).

The TinnitusTipps app (see text footnote 1) was developed by computer scientists, psychologists, and Sivantos GmbH—WS Audiology (a company specialized in hearing aids). As we detail in our earlier work (Shahania et al., 2021), TinnitusTipps uses tips to promote "health literacy," and engaging features to stimulate user involvement.

The study participants were split into three groups that differed on their exposition to the functionalities of TinnitusTipps. The tips were not personalized, so they were always delivered at random, independently of the group of the user.

- *Group A:* The users of the mHealth app received the tips for the entire project duration of 4 months ($n = 11$).
- *Group B:* In the first 2 months, the users received no tips. They only got the normal "TrackYourTinnitus" function (see Kraft et al., 2020), i.e., they only answered questionnaires. In the second half of the project, in months 3 and 4, they received the tips ($n = 10$).
- *Group Y:* This group is more similar to Group A, in the sense that they received tips from the beginning ($n = 13$). The difference between groups A and Y is that the participants have been supervised by different persons who followed slightly different protocols. For this reason, we skipped group Y and concentrated on groups A and B that followed the same study protocol.

The study protocol encompasses two questionnaires, one to be filled at registration and one EMA questionnaire to be filled at least once a day. The registration questionnaire is the "Tinnitus Sample Case History Questionnaire" TSCHQ (Langguth et al., 2007), consisting of 31 items. The EMA questionnaire consists of the 8 items depicted on **Table 1**.

We removed all users of group Y and those belonging to no group, i.e., we considered groups A and B only. An extensive exploratory analysis of the data of these users appears in the following section. It includes the distribution of the values of the answers to the EMA questionnaire items for each group, a

TABLE 1 | Items S01 to S08 from the EMA questionnaire of TinnitusTipps—the range of values for the answers of the items S02-S07 is between 0 and 100 (first two columns shown also in Shahania et al., 2021).

Item	Question description	Short description
S01	Do you perceive the tinnitus right now?	–
S02	How loud is your tinnitus right now?	Tinnitus loudness
S03	How distressed are you by your tinnitus right now?	Tinnitus distress
S04	How well do you hear right now?	–
S05	How much are you limited by your hearing right now?	–
S06	How stressed do you feel right now?	stress
S07	How exhausted do you feel right now?	–
S08	Are you wearing a hearing aid right now?	–

For S01 and S08, we have binary answer (yes/no).

correlation analysis among the EMA items independently of the groups, and a correlation analysis among the EMA items for each group.

Before addressing the research questions, we performed an exploratory analysis reported hereafter. For the RQs themselves, we skipped the categorical items S01 and S08. Due to space limitations, for some results on the RQs, we report only on the items S02, S03, S07 (cf. 3rd column of the **Table 1**). All results are in the **Supplementary Material**.

3. OUR CMAB-BASED METHOD

The architectural design for this article depends on several factors, which are depicted in **Figure 1**. At each time point, we want to predict the vector of answers. We have the large architecture to do so but in the end CMAB chooses the arm that is expected to be the best prediction. The expectation of quality of the prediction is quantified as reward. This aids us in deciding when local models are preferred over global models based on the behavior of this bandit (cf. RQ 1b). Therefore, the core factors of investigation are as follows:

1. Data representation : Depending on the structure and features of the data, the bandit has to be designed accordingly. User's answers to the items constitute a multidimensional vector of features. We distinguish between representations for an ensemble of bandits, where each ensemble member has a predictor for one of the items, and a whole vector bandit, where all the items are predicted as one vector by a single bandit.
2. Arms as prediction models : The bandits choose the arm to pull depending on the expected reward, informally the quality of the prediction. We consider three arms: one arm considers only the past data of the user for which the prediction must be made (entity-centric arm), one arm additionally considers the past data of the user's k nearest neighbors (neighborhood arm), and one arm considers the past data of all users (global arm). Each arm of the bandit can be handled in the same way or rules can be made to trigger the arms differently.
3. Sampling strategy: Since an arm is chosen based on the expected rewards, past rewards of when an arm was chosen are sampled in a heuristics called sampling strategy.
4. Context: If the bandit makes use of additional meta-information regarding the current sample we call it context. In our case, the context of observation is the group, to which the entity referred by this observation belongs. We consider one *contextual bandit* which considers as context the group (A or B) to which the user belongs, i.e., considers only data from users of this group when building the predictors (cf. point 2). We also consider *simple bandit* that has no context, i.e., does not take the groups into account.

In summary, we employed 4 different configurations which are based on the variations of the context and data representations. These configurations are as follows:

1. Contextual Whole Vector Bandit: CWB
2. Simple Whole Vector Bandit: SWB
3. Contextual Ensemble Bandit: CEB
4. Simple Ensemble Bandit: SEB.

The SEB configuration means that our bandit does not use any additional information to decide which arm to use but instead completely relies on the sampling strategy being employed but for each item, there is a different bandit provided. Thus, the SEB indirectly has some contextual knowledge about the items but none about the complete sample. On the other hand, CWB configuration uses meta-information about the sample for its decision but provides only a single output for all the items and due that has no contextual knowledge about them. Furthermore, the SWB configuration is a single bandit that uses no contextual information about either the samples or the items, whereas the CEB is provided with the context of both things.

3.1. Data Representation: Entities and Their Observations

The data are comprised of two main components: the entities and their stream of observations. Hereby every entity is a user of the system, which is represented by observations in the form of 6-dimensional multivariate time series (cf. items of the EMA questionnaire in **Table 1**). The time series of the user is the sequence of calendar times when we have seen the user. t_0 is the first time and t_n is the last time when we see the user. Let u be a user, and $t_{u,0}$, $t_{u,1}$, ..., t_{u,n_u} be the time steps at which the user was observed. These are calendar time steps, i.e., they contain a date and clock as `hh:mm:ss`. An entity might interact with the app in different intervals, therefore each entity contains a different amount of observations and the time between observations might be hours to days. Therefore an observation o_i at a time steps t_i is an element of the time series for a particular entity and might contain null (NaN) values. For us a time point t_i is not a specific point of date, but instead an answer is given from any entity, i.e., we sort the observations by their date of appearance and label the sorted list as time steps. Therefore there is no missing time step. But since the time step t_i is dependent on the entity's interaction with the mHealth app, the difference between time steps t_{i-1} and t_i can range between a day to a week. Furthermore, at a particular t_i, if there is a NULL value corresponding to some questions, we replace it with the value from the prior time step t_{i-1}.

As discussed in point 1 the output of the bandit has two major options namely whole vector and ensemble bandit. The whole vector bandit means that we provide a single output, i.e., a 6-dimensional vector, for all the items of the observation. In the ensemble setting, we instead use one bandit per item that returns a scalar output. It would be possible to combine both approaches, i.e., combine certain items for one bandit but not all. We did not investigate this option for this article.

3.2. Arms as Prediction Models

As mentioned earlier, we predict the upcoming observations using a bandit. For this purpose, we have to decide the definition of the arms of the bandit which in our case are models used for prediction. To better support the different types of entities and their behavior including the number of observations per entity

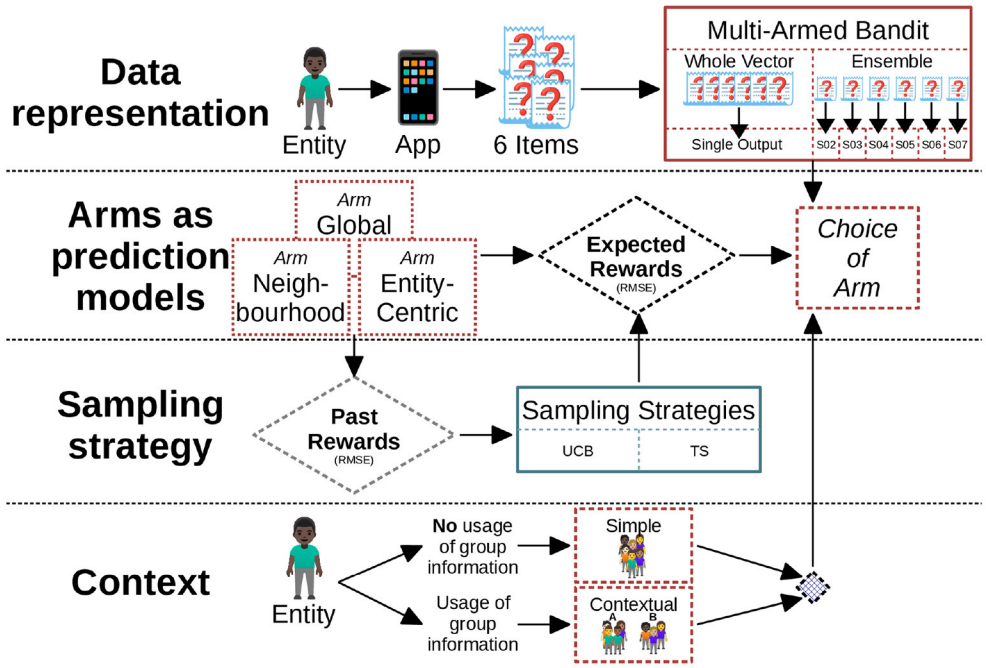


FIGURE 1 | The figure shows the outline of the system used to predict the answers given by the entities. The entire architecture is divided into 4 main parts. The first part shows the representation of the data and the output. The second part is about the prediction models as the arms of the bandit followed by the sampling strategies and the context for the bandit. We have both ensemble and whole vector bandit for each contextual and simple bandits. RMSE here refers to root mean square error.

and thus exploiting also little data for our predictions (cf. RQ 1a), several choices are made:

1. The arm setup based on the entity's past observations (history) [RQ 1a],
2. the learning algorithm for the predictor [RQ 1a],
3. the performance measure for the arm [RQ 1b],
4. and the reward function for the bandit [RQ 1c].

3.2.1. Arm Setup Based on the Entity's History (for (RQ 1a)

We design the bandit with 3 arms and hence have 3 different setups in place for prediction based on the number of entities and their respective observations. The settings are as follows:

1. *The global arm*: This is modeled based on the history of all the entities.
2. *The entity-centric arm*: This is modeled separately for each entity based on a history of only the entity itself.
3. *The neighborhood arm*: It is designed based on the history of the entity itself and its k nearest neighbors.

With the above settings, it is possible to focus on different properties of the data. On the one hand, the entity-centric arm is potentially better in exploiting the personal traits of an entity but information may not be always recent as there might be time gaps between the observations from a particular entity (cf. Section 3.1). On the other hand, the global arm can capture this temporal proximity but may over-generalize in predicting certain observations of entities. Hence, we introduce additionally the neighborhood arm which acts as a trade-off between the global and the entity-centric arms, capturing both the temporal proximity and the local behavioral patterns. Since both entity-centric and neighborhood arms capture the local behavioral patterns, we refer to them as local arms.

Based on how the arms are designed, all of them need a different number of observations before they can be used, i.e., the global arm needs only a fixed number of observations m, whereas the entity-centric arm requires n observations from the respective user. Therefore, the neighborhood arm additionally needs also observations from all its nearest neighbors, i.e., $k \cdot n$. This implies that at the beginning not all arms would be equally available for the bandit, especially the neighborhood arm. We will keep the arm setting fixed in a rule-based manner. This means that arms can only be used when enough observations have been processed for the model to start predicting. Hence, the global arm is used as a fallback option when there are not enough observations available for the local arms. For our experiments, we need to set the respective number of observations m and n. Due to the global arm needing to generalize compared to local arms m should be sufficiently larger than n. Additionally, n needs to be smaller than the minimum history length in the data which is 12 in our case for entity 25. Based on these constraints, we chose m and n as 30 and 5, respectively, from prior experimentation. Similarly, k is set

to 4 as it provided the best results for almost all the items based on preliminary analysis.

3.2.2. Choice of a Learning Algorithm for the Predictor (for RQ 1a)

As discussed in the Section 3.1, the data constitute a stream of observations, so we considered following regression algorithms for streams:

1. Hoeffding Tree (Hulten et al., 2001)
2. Hoeffding Adaptive Tree Regressor
3. Adaptive Random Forest Regressor (Gomes et al., 2017).

The only difference between the normal Hoeffding Tree and the adaptive version is that for drift detection it uses ADWIN (Bifet and Gavaldá, 2007), which is an adaptive sliding window algorithm. In the preliminary experiments all the three algorithms performed mostly at par with each other but we decided to make use of the drift detection of the adaptive version. It is assumed that this will allow us to capture the drifts in data of the users, if any, in the future. The detailed descriptions of the algorithms can be found in the **Supplementary Material**.

3.2.3. Performance Measure for the Arm (for RQ 1b)

The algorithms explored for prediction in Section 3.2.2 are regression models, therefore, the squared error is picked as the evaluation criterion. Since in other related stream classification papers (Unnikrishnan et al., 2020a) the root mean squared error (RMSE as in Equation 1) has been employed, we decided to go forward with the same.

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i}(x_i - \hat{x_i})^2} \tag{1}$$

For every target output and every time-step, RMSE value is calculated, where x_i is the actual value and $\hat{x_i}$ is the predicted value, while N is the number of observations. The RMSE value is averaged over all target outputs to understand the overall performance per time step.

3.2.4. Choice of Reward Function for the Bandit (for RQ 1c)

Most of this articles in the related work used 1 − error from the models as their reward. At each time step t_i, the bandit decides which arm to use for prediction based on past rewards, where the reward is the RMSE over all the answers to the items (cf. Equation 1), normalized on the upper bound of the error U, defined as the largest value minus the smallest value that each target can acquire. For our application scenario, $U = 100 - 0 = 100$, since all items we chose for our analysis have a value range from 0 to 100 (cf. **Table 1**).

3.3. Choice of Sampling Strategies for the Bandit

After the arms have been defined for the bandit, we need to decide the heuristic to choose an arm for the prediction of an observation. This choice is dependent on the sampling algorithms employed by the bandit. The inspiration for the sampling algorithms was taken from the related work explored. To compare various strategies Upper Confidence Bound (UCB) and Thompson Sampling (TS) were used to conduct the experiments. UCB is based on a simple principle that uncertainty regarding the arm is directly proportional to the importance of the exploration of an arm. So if the arm is very uncertain, UCB chooses this and picks the corresponding reward to this arm and makes the arm less uncertain. This goes on until the uncertainty is below some decided threshold. For our experimentation UCB1 was used which is formulated mathematically as follows:

$$Q(a) + \sqrt{\frac{2log(t)}{N_t(a)}} \tag{2}$$

where $Q(a)$ is the average reward of arm a for each round, t is the total number of rounds "played" thus far and $N_t(a)$ is the number of times arm a was selected thus far.

For each round/iteration/time-step, we play the arm that maximizes Equation (2). The first term in the Equation (2) controls exploitation, i.e., choosing the arm where the average is the largest, whereas the second term controls the exploration, hence trying to maintain the balance between both.

TS on the other hand makes use of modeling the rewards as probability distributions to sample the expected reward of an arm instead of using the average expectation like UCB does (Equation 2). Hence, the exploration is handled *via* the sampling process instead of an explicit representation in the formula. The advantage over UCB according to the survey is that already good arms are more likely to be exploited without forcing exploration based on uncertainty. Generally, this allows TS to converge faster to an optimal solution meaning the one that always chooses the model with the maximum reward. But this highly depends on the correct modeling of the reward. For our analysis, we considered normal distribution for sampling.

Additionally, to evaluate the applied strategies we have also considered the following sampling methods:

1. optimal: chooses an arm that outputs the maximum reward out of all;
2. worst: chooses an arm that outputs the minimum reward out of all;
3. random: chooses any arm randomly.

With the sampling strategies, we decided on the last building block for the bandit so that it can invoke the different arms when they are potentially most beneficial (cf. RQ 1b). This of course also is dependent on the arm itself and the reward function we have chosen for its representation.

4. RESULTS

4.1. Exploratory Data Analysis

To acquire better insights on the data distribution of the EMA of group A and group B participants, we first performed a univariate analysis on the user's answer to each EMA item and then we computed the correlation matrix for the EMA items, using heatmap as a basis for each group. Also, a comprehensive

TABLE 2 | Number of times users answer zero for S01 in relation to their total number of observations overall in their first 2 months and next 2 months of the study for group A and group B.

Group A					Group B				
	First 2 Months		Next 2 Months			First 2 Months		Next 2 Months	
User	Total Observations	Perc. of 0s	Total Observations	Perc. of 0s	User	Total Observations	Perc. of 0s	Total Observations	Perc. of 0s
17	108	69.4%	62	74.2%	20	132	0.0%	84	0.0%
18	50	22.0%	35	0.0%	24	102	0.0%	80	0.0%
19	20	20.0%	0	0.0%	25	11	72.7%	1	100%
22	157	0.0%	130	0.0%	29	169	34.9%	149	1.3%
23	129	0.8%	20	0.0%	35	190	0.0%	134	0.0%
28	150	49.3%	125	48.8%	42	58	34.5%	21	4.8%
30	101	0.0%	27	0.0%	47	165	1.2%	135	0.0%
31	104	18.3%	99	1.0%	48	173	0.6%	153	0.0%
33	39	12.8%	0	0.0%	51	259	0.4%	168	0.6%
40	24	16.7%	8	0.0%	52	45	26.7%	9	11.1%
43	50	0.0%	18	0.0%	-	-	-	-	-

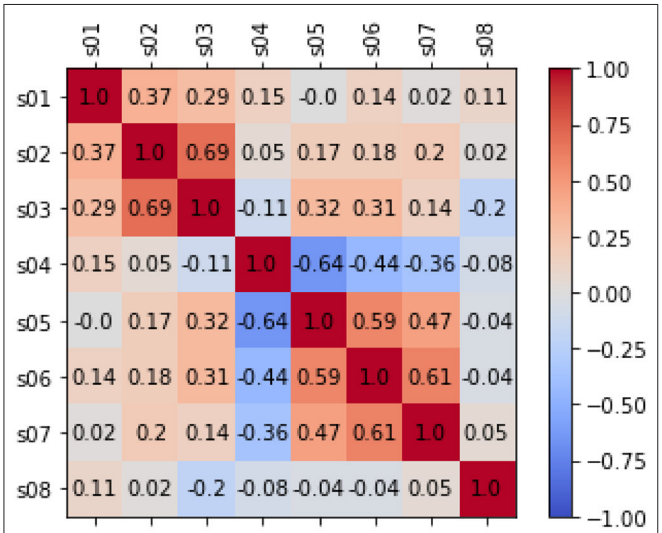


FIGURE 2 | The figure shows correlation analysis on the EMA items based on Pearson's correlation. When there is no correlation between 2 variables (when a correlation is 0 or near 0) the color is gray. The darkest red means there is a perfect positive correlation, while the darkest blue means there is a perfect negative correlation. Additionally, the numbers inside each cell represents the absolute value of the correlation.

analysis on the item S01 has been done to understand the role of perceiving tinnitus to the answers given by the user.

4.1.1. S01 Analysis

To understand the differences between the tinnitus and non-tinnitus perception times of the user, the values of item S01 were analyzed for both the groups (cf. **Table 2**). In principle, Group B is more active than Group A. Independently of the group, we have fewer observations from the users in the next 2 months of the study. Users 17, 28 from Group A and users 29, 42 from Group B are the ones who do not perceive tinnitus most of the time. Group B users mostly said so in the first 2 months while the Group A users said so in both the first and the next 2 months of their study.

4.1.2. Univariate Analysis

For the univariate analysis, the box plots and distributions for each of the items are used to see how the overall spread is for the answers between 0 and 100. We observe that for all the variables except S04 for both the groups, the likelihood of having the value 0 is high. For variable S04, the values usually are between 20 and 100. For Group A the mode of S04 is around 80 and for Group B we have a trimodal distribution with modes around 50, 80, and 100. For S02 the largest mode is around 80 for both the groups but for Group A most of the values lie around 40. For S03 we have multi-modal distribution with a mode at around 30 and 85 for Groups A and B, respectively. For the items S05, S06, and S07, there is a strong shift toward the value 0 in Group B. On the contrary for Group A, the values lie on an average between 20 and 40, but also have a mode at 0. Furthermore, we inspected the value distributions of the individual users for each item and found that each user has their own "preferred" range of values. This strengthens our expectation that local models (idiographic ones and those based on neighborhood) will be predictive for some users.

4.1.3. Bi-variate Analysis

For the bivariate analysis among the EMA items, we put the observations of both groups together and temporarily ignored that some users contributed more observations than others. The heatmap of the analysis is depicted in **Figure 2**. We found that (independently of the users), tinnitus loudness (S02), and tinnitus distress (S03) are positively correlated and so are stress (S06) and exhaustion (S07). The item S04 on hearing is particular in that higher values are better: accordingly, it

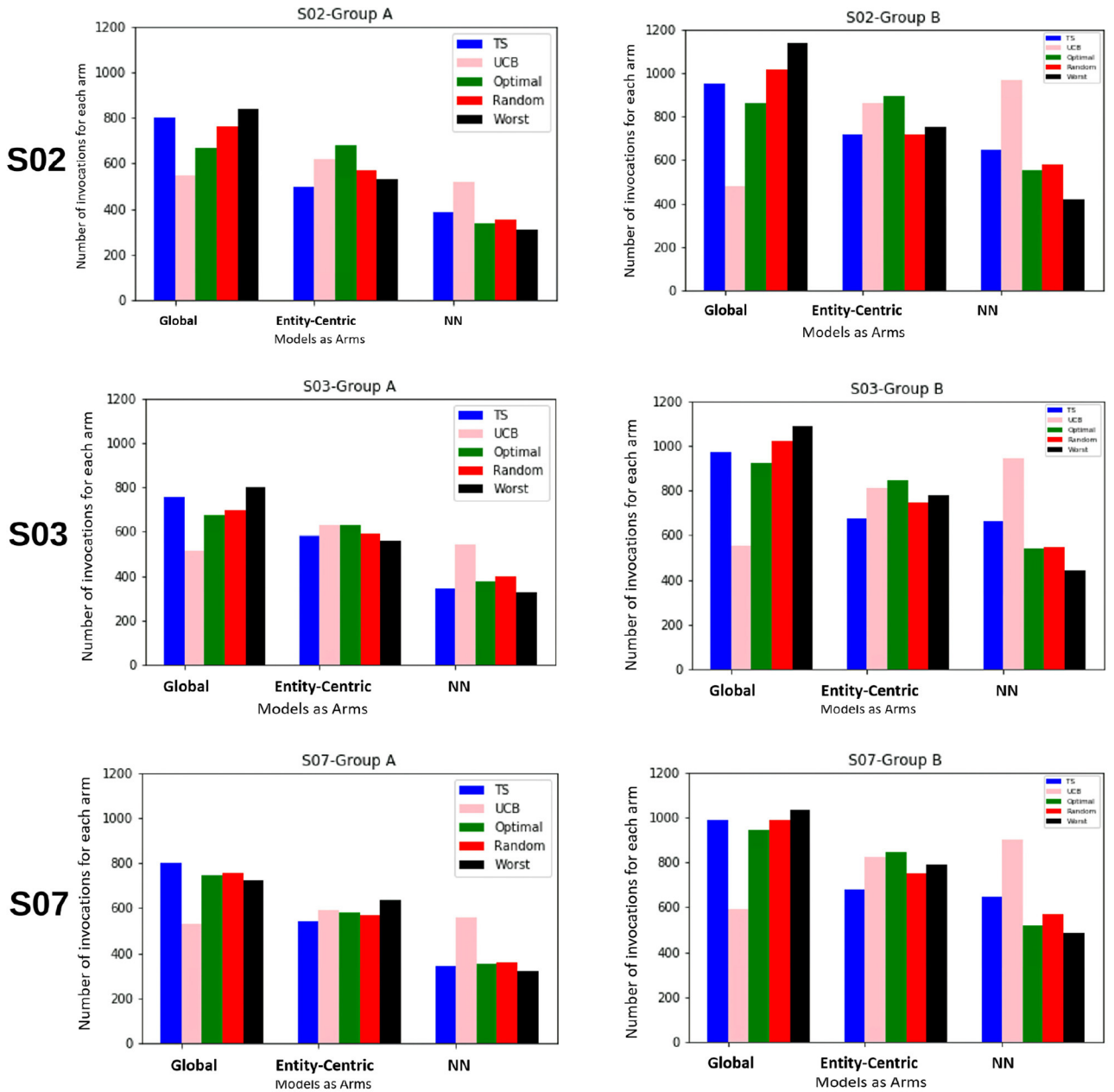


FIGURE 3 | The figure shows for the ensemble bandit, the number of times each arm is invoked for the items by different samplers for both the groups. Five samplers as discussed in Section 3.3 above have been compared here.

stands in negative correlation to stress (S06), exhaustion (S07), and limitations because of hearing (S05). There is a significant correlation between all the above correlations based on the p values before the Bonferroni correction.

4.2. Experiments for RQ 1

In the first set of experiments, we are trying to evaluate if the methods taking into account the local behavioral patterns are superior to the global method. For this purpose, we are looking at how often the global, entity-centric, and neighborhood arms (NN) are chosen, therefore, their percentage of invocations for the ensemble bandit (cf. **Figure 3**). Here, we see that for both the groups, TS follows approximately the same trend as that of the optimal sampler, whereas UCB has a very random behavior. For group B, we see that the chances of each model getting selected are approximately equally likely, whereas for group A the likelihood of neighborhood arm being chosen is comparatively low.

Also, to be able to compare the whole vector bandit (cf. **Figure 4**) to that of an ensemble bandit, we are looking at the number of invocations for each type of arm (global, entity-centric, neighborhood) for both the configurations. Since the latter version has invocations for each item, we are calculating an average number of invocations for each arm to get an overview of its behavior. For that purpose, all the invocations were summed up and divided by the total number of items, i.e., 6 (cf. **Table 1**). Additionally, we are identifying the relative number of invocations for each item using a division by the sum of all the invocations which is 1680 and 2308 for groups A and B, respectively. We then can proceed to compare UCB in **Table 3** for the ensemble bandit and the whole vector bandit. Similarly for, TS comparisons are made between the ensemble and the whole vector bandit in **Table 4**. We observe that the behavior of the bandits is very similar on average within the sampling strategy. On the other hand, when we compare UCB and TS we observe the preference of TS toward global arms whereas UCB prefers the local ones. To see which strategy is less optimal we compare the number of invocations to the optimal sampler (cf. **Table 5**) and see that TS has a behavior similar to that of the optimal one. This indicates the TS is more suitable for solving the problem since it aligns with the optimal strategy in all the configurations.

Furthermore, to better understand the convergence and therefore the learning behavior of the bandits, the average cumulative reward for each group is plotted. This provides an overview of the increase of rewards over time and therefore the increase in the quality of the models and the bandits' performance. This can be seen in **Figure 5**.

The findings that we derive from the average cumulative reward view of the **Figure 5** are as follows:

1. For the random sampling strategy rewards increases with every iteration/time-step.
2. UCB and TS have a similar trend as that of a random strategy.
3. The investigated samplers are better than the worst samplers except at the very beginning where they do not have enough data to exploit.
4. Comparing the performances for group A and group B, we see that the lines for group B are closer to each other as compared to that for A. This indicates that the predictability for group B is slightly higher than that for group A.
5. This claim is supported by the fact that for group B on average cumulative rewards are higher than that for group A, which indicates fewer errors while predicting entities in group B.

To compare the behavior of the bandit irrespective of the groups to which users belong, the same set of experiments are repeated without considering the group information. The same trends as that of the contextual multi-armed bandit are observed. **Figure 6B** shows that TS, UCB, and random samplers perform at par with each other. To see how this bandit behaves for the users in both groups A and B, the average cumulative rewards per iteration/time-step was plotted. We see in **Figure 7** bandit learns well for both group A and group B users irrespective of the difference in the number of observations in both the groups.

In summary, it can be observed that local arms are preferred in many time-steps. There is no clear pattern in the choice between global and local arms but we found differences in the preference when using different sampling strategies. Additionally, the predictability of the groups varies slightly based on average cumulative rewards. To better understand how the bandit chooses between global and local arms, we investigate potential criteria in the next subsection.

4.3. Experiments for RQ 2

To be able to understand the behavior of the bandit, i.e., choice of the arms and development of the reward, we investigate different factors. These factors include the history length of the entities, their personal traits, their time of arrival in the system, and the temporal proximity of the predictions.

Initially, we are investigating the average error of each entity per item and prediction models underlying the arms. We observe in **Table 6** that there is no clear boundary on the history length of the user where it can be clearly stated that local models are performing better on average than global models. Only entity 25 with a history length of 12 displays a clear preference to global models for all the items. Furthermore, also in **Table 6** in the last column where we have the average error of all items per entity for different prediction models, this preference of entity 25 toward the global model can be confirmed. For entity 19 with a history length of 20, the global model performs on par with the neighborhood-based model. Beyond that, for most entities, the local models seem to perform better for all items except item S07, but their difference in performance to the global models is mostly negligible.

The behavior of S07 cannot be explained by the history length as a factor. Since the global model often performs better than the local models there seems to be a property that local models miss. One of these properties is the temporal proximity of the predictions. As can be seen in **Table 1** the item S07 asks about how exhausted does the entity feel. This exhaustion might be dependent on external factors such as weather, time of day, etc., and, therefore, is more dependent upon the prior predictions rather than the entity itself.

The arrival time of the entity is interesting since the global model had more instances to predict already if the entity arrives later. Therefore, the global model might be already a better predictor for the late-arriving entities than the early birds. This cannot be proven since there seems to be no correlation between the arrival time of the entity and the preference toward the global model. This can be seen in **Figure 8** where the choice of the arms is more dependent on the history length than the arrival time of the entity.

Since the analysis is only based on the individual arms of the bandit, we additionally checked the performance of the whole vector bandit to its arms. In **Table 7**, we compare the errors of optimal sampling strategy to that of TS and see that TS performs on average worse or equal to the best performing arm. This cannot be said about the optimal

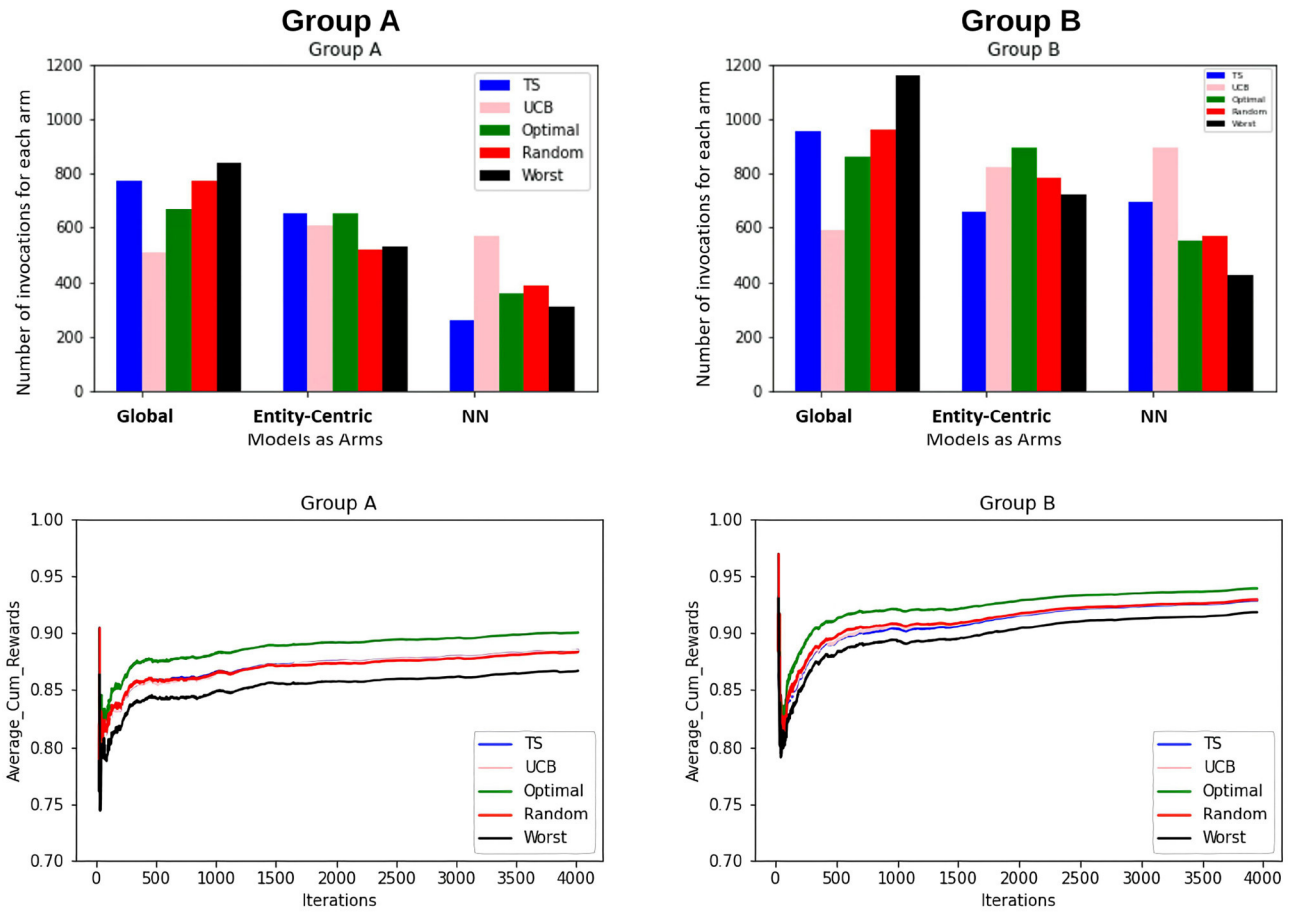


FIGURE 4 | The figure shows the combined view of both the groups for the whole vector bandit w.r.t the number of times each arm is invoked for the items by different samplers and the average cumulative reward per iteration/time-step for those items. Five samplers as discussed in Section 3.3 above have been compared here.

strategy though which specially for shorter history lengths perform better than its components. Additionally, for the longer history lengths, the error of the optimal strategy is close to the best arms. But for entity 22, which is also the first to arrive in the system, the optimal strategy is similarly better as for short history lengths. This might indicate that the initial predictions for entity 22 are better in a bandit setting but over time the performance of the individual arms averages out to be similar to that of an optimal sampler. Hence, the effect of the choice between the models is marginal at best, resulting in similar bandits and similar behavior as for a purely random model.

Additionally, we investigated if there might be a cold start problem for the bandits, i.e., initializing the bandit with 0 rewards and invocations and, hence, having no information about the expected performance of each arm. To deal with this, the optimal strategy is used to kick start each of the samplers. Meaning for the prediction of the first 70 observations, the optimal strategy is employed to initialize UCB, random, and Thompson sampler with the resulting number of invocations and sum of rewards. The observed trend is similar to that in **Figure 3**. The same experiment is repeated by taking the first 140 observations instead of 70 for initialization purposes and still the trend is the same. This experiment further supports the claim made about similar behavior of UCB, TS, and random samplers.

In conclusion, we have shown that all 4 factors have influence on the behavior of the bandit and the quality of the predictions. The biggest influence has been on the history length and the temporal proximity of the predictions which will be discussed further in Section 5.

5. DISCUSSION

5.1. Technical Perspective

As mentioned in Section 3.2, we incorporated local and global models as arms for CMAB which managed the invocations of those models to prove their effectiveness for RQ 1. A clear trend can be seen in **Tables 3**, **4**, which prefers local models two third of the time for UCB and more than half for TS, whereas the latter is close in behavior to an optimal sampling strategy. The superiority of the local models is not necessarily large though.

TABLE 3 | Number of invocations for the global and local models by ensemble predictors and whole vector predictor using UCB as a sampling strategy.

		Group A				Group B			
		Global	**Entity**	**Neighborhood**	**Total**	**Global**	**Entity**	**Neighborhood**	**Total**
S02	*Absolute*	545	617	518	1680	480	863	965	2308
	Relative	32%	37%	31%	100%	21%	37%	42%	100%
S03	*Absolute*	511	627	542	1680	554	810	944	2308
	Relative	30%	37%	32%	100%	24%	35%	41%	100%
S04	*Absolute*	485	584	611	1680	601	822	885	2308
	Relative	29%	35%	36%	100%	26%	36%	38%	100%
S05	*Absolute*	512	569	599	1680	576	833	899	2308
	Relative	30%	34%	36%	100%	25%	36%	39%	100%
S06	*Absolute*	467	642	571	1680	661	847	800	2308
	Relative	28%	38%	34%	100%	29%	37%	35%	100%
S07	*Absolute*	531	590	559	1680	588	822	898	2308
	Relative	32%	35%	33%	100%	25%	36%	39%	100%
Total	*Absolute*	3051	3629	3400	-	3460	4997	5391	-
	Average	508.5	604.8	566.7	-	576.7	832.8	898.5	-
	Relative	30%	36%	34%	-	25%	36%	39%	-
Whole Vector	*Absolute*	506	607	567	1680	589	825	894	2308
	Relative	30%	36%	34%	100%	26%	36%	39%	100%

TABLE 4 | Number of invocations for the global and local models by ensemble predictors and whole vector predictor using Thompson Sampling as a sampling strategy.

		Group A				Group B			
		Global	**Entity**	**Neighborhood**	**Total**	**Global**	**Entity**	**Neighborhood**	**Total**
S02	*Absolute*	802	494	384	1680	947	716	645	2308
	Relative	48%	29%	23%	100%	41%	31%	28%	100%
S03	*Absolute*	755	581	344	1680	972	671	665	2308
	Relative	45%	35%	20%	100%	42%	29%	29%	100%
S04	*Absolute*	756	544	380	1680	930	692	686	2308
	Relative	45%	32%	23%	100%	40%	30%	30%	100%
S05	*Absolute*	793	577	310	1680	962	706	640	2308
	Relative	47%	34%	18%	100%	42%	31%	28%	100%
S06	*Absolute*	790	517	373	1680	967	704	637	2308
	Relative	47%	31%	22%	100%	42%	31%	28%	100%
S07	*Absolute*	798	541	341	1680	987	677	644	2308
	Relative	48%	32%	20%	100%	43%	29%	28%	100%
Total	*Absolute*	4694	3254	2132	-	5765	4166	3917	-
	Average	782.3	542.3	355.3	-	960.8	694.3	652.8	-
	Relative	47%	32%	21%	-	42%	30%	28%	-
Whole Vector	*Absolute*	780	628	272	1680	970	649	689	2308
	Relative	46%	37%	16%	100%	42%	28%	30%	100%

Only minor differences could be observed in Section 4.2. The sampling strategies including worst and optimal all are relatively close together in performance, meaning that the choice of models is not that significant. This might also be due to the late activation of local models as described in Section 3.2.1, since the fallback would be always the global model. A better strategy on when to chose each arm might still prove to be beneficial.

Additionally from **Figures 3–5** we see that the context used in the CMAB does not make a big difference. Hence, the context of the items might be irrelevant to the problem or for the arms that were chosen. The sampling strategies might not be able to pick up relevant differences or models might be too close in performance. Furthermore, the predictability for group B is slightly higher than for group A for the CMAB. This might mean, that users of group B either formed a more distinct profile due to them exploring the app more, or their behavior is just simpler and therefore easier to predict. In both cases, the local arms might easier pick up the personal traits of each entity. Additionally, there is a difference in the usage of the context by the strategies. TS shows a significant difference in the choice of the neighborhood arm across groups.

TABLE 5 | Number of invocations for the global and local models by ensemble predictors and whole vector predictor using optimal sampler as a sampling strategy.

		Group A				Group B			
		Global	**Entity**	**Neighborhood**	**Total**	**Global**	**Entity**	**Neighborhood**	**Total**
S02	*Absolute*	667	678	335	1680	862	895	551	2308
	Relative	40%	40%	20%	100%	37%	39%	24%	100%
S03	*Absolute*	675	627	378	1680	924	843	541	2308
	Relative	40%	37%	23%	100%	40%	37%	23%	100%
S04	*Absolute*	686	636	358	1680	914	910	484	2308
	Relative	41%	38%	21%	100%	40%	39%	21%	100%
S05	*Absolute*	714	611	355	1680	948	839	521	2308
	Relative	43%	36%	21%	100%	41%	36%	23%	100%
S06	*Absolute*	728	612	340	1680	887	893	528	2308
	Relative	43%	36%	20%	100%	38%	39%	23%	100%
S07	*Absolute*	747	582	351	1680	945	843	520	2308
	Relative	44%	35%	21%	100%	41%	37%	23%	100%
Total	*Absolute*	4217	3746	2117	-	5480	5223	3145	-
	Average	702.8	624.3	352.8	-	913.3	870.5	524.17	-
	Relative	42%	37%	21%	-	40%	38%	23%	-
Whole Vector	*Absolute*	670	653	357	1680	861	893	554	2308
	Relative	40%	39%	21%	100%	37%	39%	24%	100%

This implies TS can make use of homogeneity of group B. These findings are not significantly visible in the optimal sampler but indications can be found. The difference between the groups is not shared by the simple bandit as seen in **Figure 7**. Therefore, we can assume that context is important overall, but the given context is not sufficient, yet.

To understand which factors influence the performance between the models (RQ 2), we looked at the four parameters in Section 4.3. In **Table 7**, we found a clear dependency on the history length of the entities. This is on one hand due to the training time of the local models but also the fact that errors average out over time. Additionally, we checked if the arrival time of an entity in the system influences the choice of arms since the global arm will have a smaller error and higher reward at some point. Such a correlation does not exists, since **Figure 8** only shows a dependency on the history length of the entity. The factors temporal proximity of the predictions and the personal traits of the entities might be in a trade-off between each other since local models might be able to better capture the personal traits but might not have recent predictions due to the behavior of the entities. In **Table 6**, most items for either local model perform better than the global one. We assume these items to be personal and therefore dependent on the traits of the entity. This seems to not be true for S07 where the global model oftentimes outperforms the local models. We explained in Section 4.3 that this might be due to external circumstances like weather and time. Due to that, context about the items might be beneficial for a CMAB.

As far as threats to validity are concerned, there is no guarantee that the multi-armed bandits converged to an optimal solution. This could be due to the lack of a fitting sampling strategy and the fact, that all models get better over time. For our experiments we assumed the data to follow the normal distribution and TS was designed accordingly. This assumption can be too stringent and hence we might just have a sub-optimal sampling strategy. Second, from a data perspective, there might be too few entities in the dataset, overall along with varying history lengths as entities with a small history might not be properly integrated by the bandit. This might have had a negative impact on the bandits and the models. This was further worsened due to the removal of entities without the group (context) information.

For future work, one main point should be therefore on the data collection process. Entities could be presented with the fact the recommendations from the mHealth app can be improved if they better work with the system, because for more active entities, their answers are easier to predict. Since the group as a context for the bandit did not provide useful insights, collecting additional meta-information about the lifestyle or similar information about the user might separate the groups inside the bandit better. Additionally, it might be helpful to get feedback from the entities on the aspect where the difference of the predictions to the actual answers are large. This could be directly incorporated into the reward function of the bandit and hence improving the sampling strategies. Given more time, the sampling strategies could have been enhanced by including information about the recent behavior of each user, because their differences might punish the rewards of yet not fully trained local models.

5.2. Mobile Health Perspective

From the perspective of optimal EMA prediction for personalized services, our findings demonstrate a clear trend: the optimal sampler of CMAB invokes the entity-centric model comparably often as the global model when it comes to predictions of the

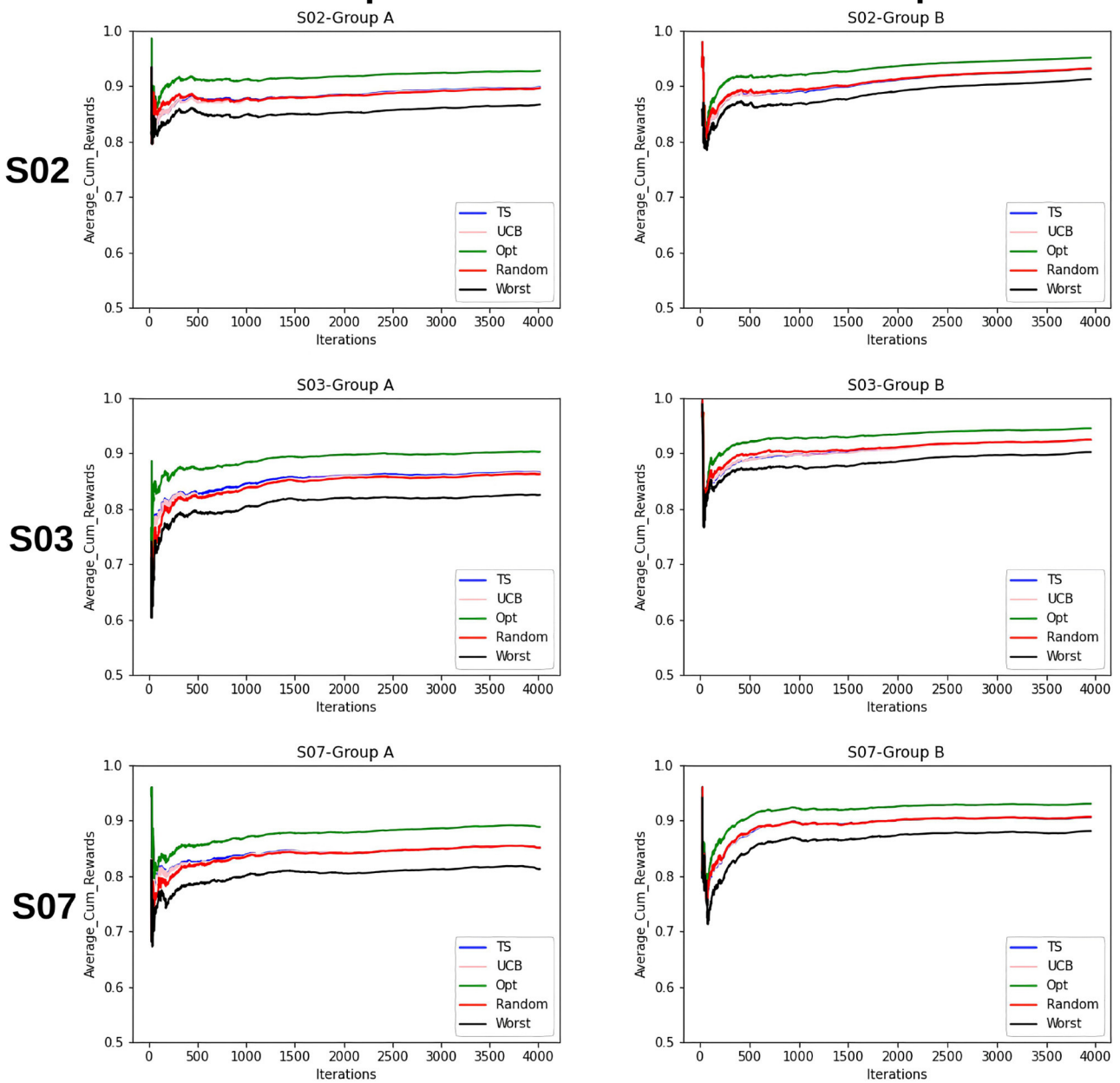


FIGURE 5 | Average cumulative reward per iteration/time-step for the items for Group A and B. Here, the first column represents selections by Group A and the second column by Group B. Five samplers as discussed above have been compared here.

individual EMA through the ensemble (cf. **Figure 3** and **Table 5**), while for the whole-vector prediction it invokes the entity-centric model equally or more often than the global one (cf. upper part of **Figure 4**). Our findings on how often the optimal sampler of the CMAB invokes each of the three models show a clear preference toward the entity-centric model. This means that predictions should be preferably based on the recordings of a user oneself, especially for users who (alike the participants of group B) deliver many recordings. Another formulation of this finding is that if an mHealth service provider has the option of invoking both a global model and an entity-centric one, then the latter should be preferred for users who interact intensively with the app—as soon as they start doing so.

One might argue that since the global model is used as often as the entity-centric model by the optimal sampler of the CMAB, then the two models are equally good. Unfortunately, this conclusion is not permitted. Rather, the global model is used so often, because many users contribute too few EMA,

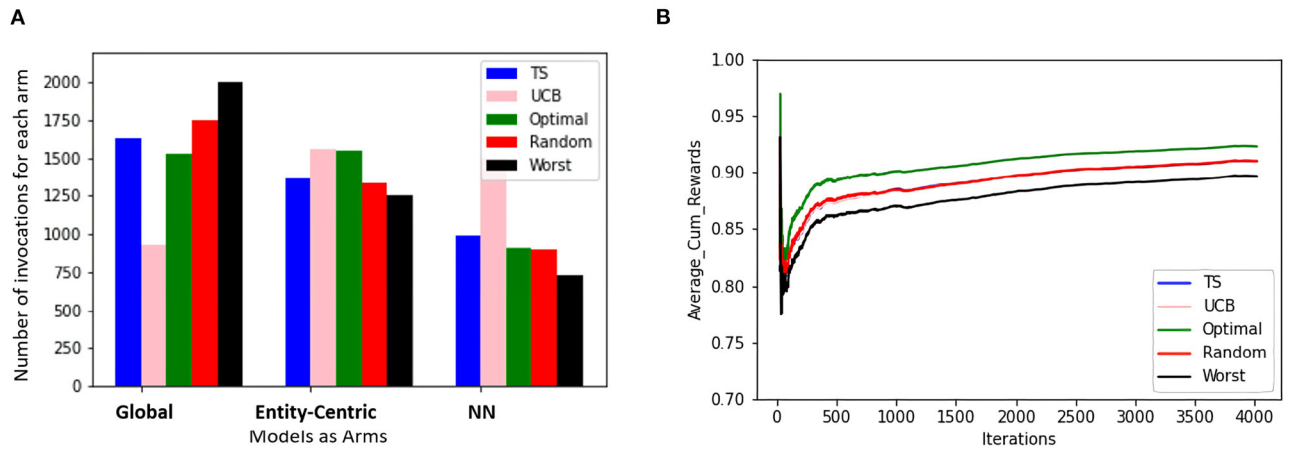


FIGURE 6 | The figure shows the view for the simple bandit wherein the subfigure **(A)** shows the number of times each arm is invoked for all the items by different samplers for every entity. The subfigure **(B)** shows the average cumulative reward per iteration/time-step for all the entities. In both cases, 5 samplers have been compared wherein different colors represent different samplers.

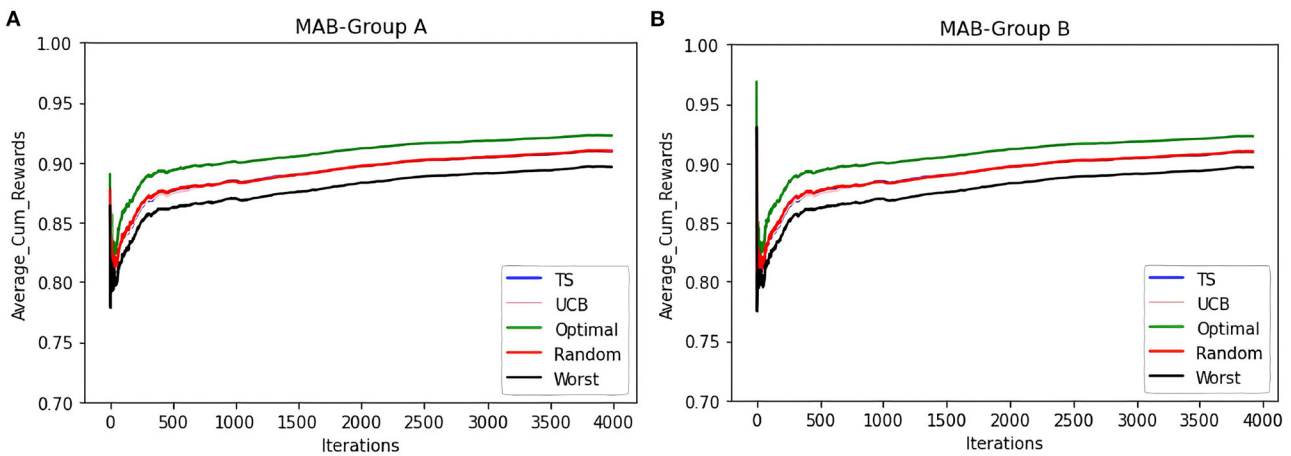


FIGURE 7 | To see how the average cumulative reward in **Figure 6B** behaves for different groups, we split the behavior for both the groups which is shown in the subfigures **(A,B)**. *Y*-axis shows the average cumulative reward and the *X*-axis shows the time-steps and different colors represent 5 samplers that are compared here.

TABLE 6 | Average error for each item across all the time steps using global, entity-centric, and neighborhood prediction models.

Entities	Hist. length	S02			S03			S07			Total		
		Global	EC	NN	Global	EC	NN	Global	EC	NN	Global	EC	NN
25	12	15.65	40.71	18.21	14.00	34.49	13.44	10.58	22.56	13.41	13.01	26.66	14.73
19	20	14.89	18.40	18.14	22.28	39.18	22.90	7.74	19.69	10.05	12.89	16.52	12.54
40	34	17.82	14.51	16.68	20.49	17.37	18.99	12.75	9.07	9.98	17.07	14.46	15.45
30	128	5.91	5.43	4.36	7.22	5.65	5.06	7.02	10.11	9.80	6.84	6.18	5.75
29	322	12.38	12.42	10.74	11.69	12.54	10.97	8.53	13.62	11.69	10.38	10.38	9.07
22	490	7.46	6.69	6.63	15.83	16.52	15.54	7.64	17.86	15.61	10.47	10.95	10.17

For demonstration purposes, we include only items S02, S03, and S07.

hence no personal, entity-centric models are available for them. As can be seen in the upper right subfigure of **Figure 4**, the optimal sampler invokes the entity-centric model is invoked more often than the global model for the users of group B, i.e., for the group the users of which contribute many EMA.

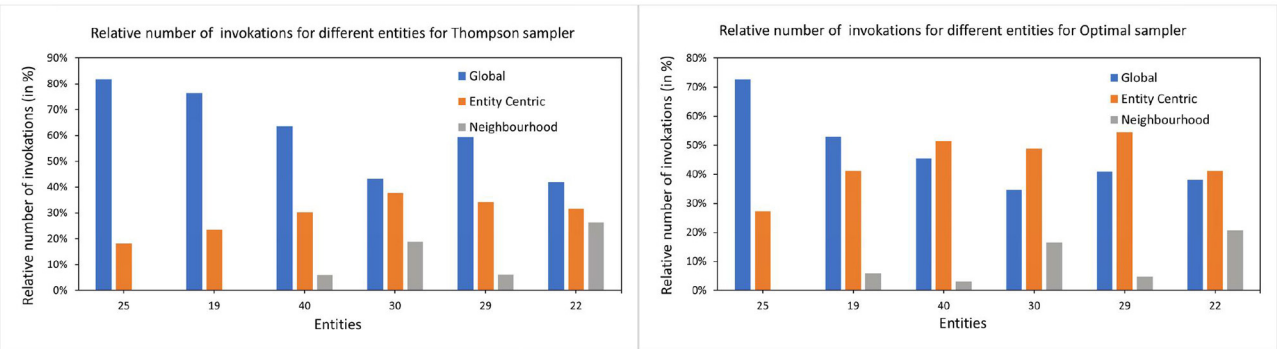


FIGURE 8 | The relative number of invocations for global and local arms. The left subfigure shows the number of times each arm is invoked for all the items by TS samplers for selected entities. *Y*-axis shows the relative number of invocations and the *X*-axis are the entities. The right subfigure shows the same for the optimal sampler. For demonstration purposes, we have chosen 6 entities with varying history lengths.

For the design of personalized services, as anticipated, e.g., in Vogel et al. (2021), this finding is rather disappointing, since it means that knowledge from other users is not very useful when it comes to prediction. This finding is also in contrast to the whole body of research on collaborative filtering for recommendation engines, where knowledge about similar users is exploited to predict a user's future preferences (Ricci et al., 2010), and to the advances on the potential of mHealth apps for decision support through prediction (Martínez-Pérez et al., 2014). However, our finding is less surprising when placed in the context of EMA: cutting-edge prediction technologies on temporal data build upon large amounts of recordings, see e.g. the sizes of the timestamped data sets used in Bellogín and Sánchez (2017); large data sets can be accumulated easily through sensor signals, as for the glycose predictor proposed by Pérez-Gandía et al. (2018), but are less easy to accumulate when users deliver EMA at their own discretion. Indeed, in our earlier work (Schleicher et al., 2020) on EMA recordings with the mHealth app TrackYourTinnitus (for short: TYT), we found that the majority of the users contributed less than three EMA in total.

Then, should we avoid learning from similar users for prediction? Since the neighborhood-based model is invoked comparatively rarely by the optimal sampler of the CMAB (cf. **Table 5**, **Figure 3**, and upper part of **Figure 4**), one is tempted to conclude that this model is inferior to the other two. However, the low number of invocations can also be explained by the limitations of the concrete study: the number of users is small in total, the period of observation is short, and the neighborhood-based model demands 5 observations per neighbor, in order to start making predictions. Hence, this model is available less often than the other two models. This means that if the population of users were larger and more EMA were available for some of them, then the aforementioned finding might be reversed. There are indeed indications in that direction: our earlier analyses on users of EMA-based mHealth apps for tinnitus (Unnikrishnan et al., 2019, 2020a, 2021) and diabetes (Unnikrishnan et al., 2020b) demonstrate that it is possible to exploit the data of users who deliver many EMA in order to do high-quality predictions for users who deliver few EMA (or are at the beginning of their interaction with the app). Nonetheless, choosing appropriate data to inform a neighborhood-based predictor is challenging (Unnikrishnan et al., 2021), not least because dependencies between past and current recordings do not generalize for the whole population of users. More research is needed to find ways of exploiting information on similar users for recommendations, when the available data are very sparse.

TABLE 7 | Combined average error for each item across all the time steps for the complete bandit using optimal and Thompson sampling strategies to that of its arms.

Entity-Id	Group	History length	Optimal	TS	Global	EC	NN
25	B	12	11.72	13.75	13.01	26.66	14.73
19	A	20	11.21	12.87	12.89	16.52	12.54
40	A	34	13.22	15.59	17.07	14.46	15.45
30	A	128	5.45	6.61	6.84	6.18	5.75
29	B	322	8.68	10.18	10.38	10.38	9.07
22	A	490	8.91	10.32	10.47	10.95	10.17

The users shown are chosen at random to have a mix of varying history lengths.

5.3. Experimentation Perspective and Insights for Tinnitus Research

Our investigation on the relative performance of global and local models is based on an experimental study. Hence, some of our findings are of relevance to the design of experiments involving mHealth app users, at least in the context of tinnitus research and of design of mHealth apps for tinnitus users.

First of all, there are differences in the predictability of different EMA items: for the prediction of EMA item S07, the optimal sampler selects the global model more often than

the entity-centric one; for item S02, the entity-centric model is equally often selected for group A and more often selected for group B. This means that the global model, which is wholly insensitive to which user has contributed which EMA answer, delivers the best prediction (and is thus selected by the optimal sampler) as often as does the entity-centric model of that user.

Further, there are EMA items that were answered more often than others. More research is needed to shed light to the (un)popularity of some EMA items and to investigate whether some items can be consistently predicted from others; in our earlier work, we provide some indication to this end, by identifying Granger causalities among EMA items answered by TrackYourTinnitus users (Jamaludeen et al., 2021).

The findings on the answers to the EMA item S01 were remarkable. While the majority of users consistently answered that they perceived tinnitus every time they were asked, some users consistently answered that they do not, and some of the latter consistently answered that they do after they entered the second phase of the experiment. Since the number of users in this study was small, we do not attempt any conclusion from this finding. However, the presence of 4 out of 21 users with so different behavior with respect to this item shows that further research is needed, and in a larger pool of users, to better understand the role of this item.

The tips considered in the mHealth app TinnitusTipps were not personalized toward users, nor aligned to a user's previous answers to specific items. Hence, the effect of the tips on the answers to the EMA items cannot be assessed. However, group A was exposed to tipps from the beginning of the study, while group B was exposed only after the first two months. More analysis is needed to understand whether this can have resulted in the observed differences between group A and group B with respect to the number of contributed EMA recordings.

The optimal sampler of our CMAB demonstrated that the choice among models is influenced by the group (A vs. B). The difference between the two groups was that group B delivered more EMA recordings from the beginning on. Since the number of recordings influences the quality of the entity-centric predictor and of the neighborhood-based predictor, and (as a matter of fact) the quality of any multi-level model, it is advisable to quantify interaction intensity (cf. Schleicher et al., 2020) and to incorporate it into the learning model.

DATA AVAILABILITY STATEMENT

The data is part of the TinnitusTipps app developed by Sivantos, data can be shared upon request to the authors. Requests should be directed to WS: winfried.schlee@gmail.com.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee from the University Regensburg: 17-544-101. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

SS designed, implemented, and evaluated the technical components under the supervision of MS building upon and extending prior results of VU and MS created a conceptual model of the whole approach. VU contributed with ideas and insights from previous studies on the same data. RP, RH, and WS designed the mHealth app. RK and JS implemented it under the guidance of RP and provided data and instructions on data usage. WS lead the design of the two-armed mHealth study that resulted in the data set used in this investigation. RH delivered comments and feedback on the purpose of the app components and RP on the architecture and functionalities. SS and MS wrote the paper together. All other authors contributed with comments and feedback.

FUNDING

This work was partially funded by the CHRODIS PLUS Joint Action, which has received funding from the European Union, in the framework of the Health Programme (2014-2020), Grant Agreement 761307 Implementing good practices for chronic diseases. This work was partially inspired by the European Union's Horizon 2020 Research and Innovation Programme, Grant Agreement 848261 Unification of treatments and Interventions for Tinnitus patients (UNITI). The development of the TinnitusTipps mHealth app was partially financed by Sivantos GmbH–WS Audiology. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article, or the decision to submit it for publication.

ACKNOWLEDGMENTS

We would like to thank all reviewers for their comments, which lead to a much broader perspective of the paper and to additional contributions. We particularly thank the reviewers for their suggestion to investigate the role of EMA item S01; this suggestion leads to a remarkable result. Although we could not draw conclusions from it, we believe that the publication itself will lead to further investigation on the issue we identified.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.836834/full#supplementary-material

REFERENCES

Bellogín, A., and Sánchez, P. (2017). "Revisiting neighbourhood-based recommenders for temporal scenarios," in *RecTemp@ RecSys* (Madrid). 40–44.

Bifet, A., and Gavaldá, R. (2007). "Learning from time-changing data with adaptive windowing," in *Proc. of the 2007 SIAM Int. Conf. on Data Mining (SDM'07)*, vol. 7 (Minneapolis, MN), 443–448.

Cederroth, C. R., Gallus, S., Hall, D. A., Kleinjung, T., Langguth, B., Maruotti, A., et al. (2019). Towards an understanding of tinnitus heterogeneity. *Front. Aging Neurosci.* 11, 53. doi: 10.3389/fnagi.2019.00053

De Ridder, D., Schlee, W., Vanneste, S., Londero, A., Weisz, N., Kleinjung, T., et al. (2021). "Tinnitus and tinnitus disorder: theoretical and operational definitions (an international multidisciplinary proposal)," in *Progress in Brain Research* (Amsterdam: Elsevier BV). 1–25

Gomes, H. M., Bifet, A., Read, J., Barddal, J. P., Enembreck, F., Pfahringer, B., et al. (2017). Adaptive random forests for evolving data stream classification. *Mach. Learn.* 106, 1469–1495. doi: 10.1007/s10994-017-5642-8

Hermans, H. J. (1988). On the integration of nomothetic and idiographic research methods in the study of personal meaning. *J. Personality* 56, 785–812. doi: 10.1111/j.1467-6494.1988.tb00477.x

Hulten, G., Spencer, L., and Domingos, P. (2001). "Mining time-changing data streams," in *Proc. of the 7th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD '01)* (New York, NY: ACM), 97–106.

Jamaludeen, N., Unnikrishnan, V., Pryss, R., Schobel, J., Schlee, W., and Spiliopoulou, M. (2021). "Circadian conditional granger causalities on ecological momentary assessment data from an mhealth app," in *2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS)* (Aveiro: IEEE), 354–359.

Kraft, R., Stach, M., Reichert, M., Schlee, W., Probst, T., Langguth, B., et al. (2020). Comprehensive insights into the trackyourtinnitus database. *Procedia Comput. Sci.* 175, 28–35. doi: 10.1016/j.procs.2020.07.008

Langguth, B., Goodey, R., Azevedo, A., Bjorne, A., Cacace, A., Crocetti, A., et al. (2007). Consensus for tinnitus patient assessment and treatment outcome measurement: tinnitus research initiative meeting, regensburg, july 2006. *Progr. Brain Res.* 166, 525–536. doi: 10.1016/S0079-6123(07)66050-6

Mao, X., Zhao, X., and Liu, Y. (2020). mhealth app recommendation based on the prediction of suitable behavior change techniques. *Decis. Support Syst.* 132, 113248. doi: 10.1016/j.dss.2020.113248

Martínez-Pérez, B., de la Torre-Díez, I., López-Coronado, M., Sainz-de Abajo, B., Robles, M., and García-Gómez, J. M. (2014). Mobile clinical decision support systems and applications: a literature and commercial review. *J. Med. Syst.* 38, 1–10. doi: 10.1007/s10916-013-0004-y

Mehdi, M., Dode, A., Pryss, R., Schlee, W., Reichert, M., and Hauck, F. J. (2020). Contemporary review of smartphone apps for tinnitus management and treatment. *Brain Sci.* 10, 867. doi: 10.3390/brainsci10110867

Pérez-Gandía, C., García-Sáez, G., Subías, D., Rodríguez-Herrero, A., Gómez, E. J., Rigla, M., et al. (2018). Decision support in diabetes care: the challenge of supporting patients in their daily living using a mobile glucose predictor. *J. Diabetes Sci. Technol.* 12, 243–250. doi: 10.1177/1932296818761457

Prakash, S., Unnikrishnan, V., Pryss, R., Kraft, R., Schobel, J., Hannemann, R., et al. (2021). Interactive system for similarity-based inspection and assessment of the well-being of mhealth users. *Entropy* 23, 1695. doi: 10.3390/e23121695

Probst, T., Pryss, R. C., Langguth, B., Rauschecker, J. P., Schobel, J., Reichert, M., et al. (2017). Does tinnitus depend on time-of-day? an ecological momentary assessment study with the TrackYourTinnitus application. *Front. Aging Neurosci.* 9, 253. doi: 10.3389/fnagi.2017.00253

Pryss, R., Langguth, B., Probst, T., Schlee, W., Spiliopoulou, M., and Reichert, M. (2021). Smart mobile data collection in the context of neuroscience. *Front. Neurosci.* 15, 618. doi: 10.3389/fnins.2021.698597

Pryss, R., Probst, T., Schlee, W., Schobel, J., Langguth, B., Neff, P., et al. (2019). Prospective crowdsensing versus retrospective ratings of tinnitus variability and tinnitus–stress associations based on the trackyourtinnitus mobile platform. *Int. J. Data Sci. Anal. (JDSA)* 8, 327–338. doi: 10.1007/s41060-018-0111-4

Ricci, F., Rokach, L., and Shapira, B. (2010). *Recommender Systems Handbook*. New York, NY: Springer. doi: 10.1007/978-0-387-85820-3_1

Schlee, W., Neff, P., Simoes, J., Langguth, B., Schoisswohl, S., Steinberger, H., et al. (2022). Smartphone-guided educational counseling and self-help for chronic tinnitus. *Preprints* (2022) 2022010469. doi: 10.20944/preprints202201.0469.v1

Schleicher, M., Unnikrishnan, V., Neff, P., Sim es, J., Probst, T., Schlee, W., et al. (2020). Understanding adherence to the recording of ecological momentary assessments in the example of tinnitus monitoring. *Sci. Rep.* 10, 1–13. doi: 10.1038/s41598-020-79527-0

Shahania, S., Unnikrishnan, V., Pryss, R., Kraft, R., Schobel, J., Hannemann, R., et al. (2021). "User-centric vs whole-stream learning for EMA prediction," in *Proceedings of the IEEE Symposium on Computer Based Medical Systems (CBMS'2021)* (Aveiro: IEEE)

Unnikrishnan, V., Beyer, C., Matuszyk, P., Niemann, U., Schlee, W., Ntoutsi, E., et al. (2019). Entity-level stream classification: exploiting entity similarity to label the future observations referring to an entity. *Int. J. Data Sci. Anal.* 9, 1–15. doi: 10.1007/s41060-019-00177-1

Unnikrishnan, V., Schleicher, M., Shah, Y., Jamaludeen, N., Schobel, J., Kraft, R., et al. (2020a). The effect of non-personalised tips on the continued use of self-monitoring mhealth applications. *Brain Sci.* 10, 924. doi: 10.3390/brainsci10120924

Unnikrishnan, V., Shah, Y., Schleicher, M., Fernández-Viadero, C., Strandzheva, M., Velikova, D., et al. (2021). "Love thy neighbours: a framework for error-driven discovery of useful neighbourhoods for one-step forecasts on ema data," in *2021 IEEE 34th International Symposium on Computer-Based Medical Systems (CBMS)* (Aveiro: IEEE), 295–300.

Unnikrishnan, V., Shah, Y., Schleicher, M., Strandzheva, M., Dimitrov, P., Velikova, D., et al. (2020b). "Predicting the health condition of mhealth app users with large differences in the number of recorded observations - where to learn from?" in *Int. Conf. on Discovery Science* (Thessaloniki), 659–673.

Vogel, C., Schobel, J., Schlee, W., Engelke, M., and Pryss, R. (2021). "Uniti mobile emi-apps for a large-scale european study on tinnitus," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (Mexico: IEEE), 2358–2362

Watson, H. A., Tribe, R. M., and Shennan, A. H. (2019). The role of medical smartphone apps in clinical decision-support: a literature review. *Artif. Intell. Med.* 100, 101707. doi: 10.1016/j.artmed.2019.101707

Conflict of Interest: RH was employed by Sivantos GmbH–WS Audiology during the development of the mHealth app software. In this article, he provided comments and feedback on the description of the purpose of the app.

The remaining 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.

frontiers | Frontiers in Neuroscience

ORIGINAL RESEARCH
published: 14 April 2022
doi: 10.3389/fnins.2022.831714

Deficits in Sense of Body Ownership, Sensory Processing, and Temporal Perception in Schizophrenia Patients With/Without Auditory Verbal Hallucinations

Jingqi He[1†], Honghong Ren[1†], Jinguang Li[1,2], Min Dong[3], Lulin Dai[4], Zhijun Li[1], Yating Miao[1], Yunjin Li[5], Peixuan Tan[6], Lin Gu[7,8], Xiaogang Chen[1] and Jinsong Tang[9,10*]*

[1] Department of Psychiatry, National Clinical Research Center for Mental Disorders, The Second Xiangya Hospital of Central South University, Changsha, China, [2] Affiliated Wuhan Mental Health Center, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China, [3] Guangdong Mental Health Center, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou, China, [4] Department of Neurosurgery, Center for Functional Neurosurgery, Ruijin Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China, [5] Department of Pathology, School of Basic Medical Sciences, Central South University, Changsha, China, [6] Department of Medical Psychology and Behavioral Medicine, School of Public Health, Guangxi Medical University, Nanning, China, [7] RIKEN Center for Advanced Intelligence Project, Tokyo, Japan, [8] Research Center for Advanced Science and Technology (RCAST), University of Tokyo, Tokyo, Japan, [9] Department of Psychiatry, Sir Run-Run Shaw Hospital, School of Medicine, Zhejiang University, Hangzhou, China, [10] Zigong Mental Health Center, Zigong, China

It has been claimed that individuals with schizophrenia have difficulty in self-recognition and, consequently, are unable to identify the sources of their sensory perceptions or thoughts, resulting in delusions, hallucinations, and unusual experiences of body ownership. The deficits also contribute to the enhanced rubber hand illusion (RHI; a body perception illusion, induced by synchronous visual and tactile stimulation). Evidence based on RHI paradigms is emerging that auditory information can make an impact on the sense of body ownership, which relies on the process of multisensory inputs and integration. Hence, we assumed that auditory verbal hallucinations (AVHs), as an abnormal auditory perception, could be linked with body ownership, and the RHI paradigm could be conducted in patients with AVHs to explore the underlying mechanisms. In this study, we investigated the performance of patients with/without AVHs in the RHI. We administered the RHI paradigm to 80 patients with schizophrenia (47 with AVHs and 33 without AVHs) and 36 healthy controls. We conducted the experiment under two conditions (synchronous and asynchronous) and evaluated the RHI effects by both objective and subjective measures. Both patient groups experienced the RHI more quickly and strongly than HCs. The RHI effects of patients with AVHs were significantly smaller than those of patients without AVHs. Another important finding was that patients with AVHs did not show a reduction in RHI under asynchronous conditions. These results emphasize the disturbances of the sense of body ownership in schizophrenia patients with/without AVHs and the associations with AVHs. Furthermore, it is suggested that patients with AVHs may have multisensory processing dysfunctions and internal timing deficits.

Keywords: rubber hand illusion, auditory verbal hallucinations, body ownership, sensory processing, temporal perception

OPEN ACCESS

Edited by:
Roland Schaette, University College London, United Kingdom

Reviewed by:
Jamie Ward, University of Sussex, United Kingdom
Mariano D'Angelo, Karolinska Institutet (KI), Sweden
Renzo Lanfranco, Karolinska Institutet (KI), Sweden

***Correspondence:**
Xiaogang Chen
chenxiaogang@csu.edu.cn
Jinsong Tang
tangjinsong@zju.edu.cn

[†] These authors have contributed equally to this work and share first authorship

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

Received: *08 December 2021*
Accepted: *15 March 2022*
Published: *14 April 2022*

Citation:
He J, Ren H, Li J, Dong M, Dai L, Li Z, Miao Y, Li Y, Tan P, Gu L, Chen X and Tang J (2022) Deficits in Sense of Body Ownership, Sensory Processing, and Temporal Perception in Schizophrenia Patients With/Without Auditory Verbal Hallucinations. Front. Neurosci. 16:831714. doi: 10.3389/fnins.2022.831714

INTRODUCTION

Schizophrenia is a severe mental disorder characterized by thought and perception disturbances. Individuals with schizophrenia usually have difficulties in distinguishing between reality and fantasy, which may cause unusual experiences of reality. Auditory verbal hallucinations (AVHs) are one of the most common presenting symptoms of schizophrenia, and are usually defined as the sensory perceptions of hearing voices that do not exist, occurs in approximately 74% of patients with schizophrenia (Dierks et al., 1999; Choong et al., 2007). Since the mechanisms underlying AVHs are unknown, researchers have proposed multiple theoretical models, such as unstable memories, source monitoring, top-down attention, and hybrid models of hallucinations (Ćurčić-Blake et al., 2017b).

From the view of cognition, most perception disturbances revolve around self-recognition failures, which means that patients tend to mistakenly attribute self-generated behaviors to an external source (Waters et al., 2010; Graham et al., 2014; Klaver and Dijkerman, 2016). Auditory verbal hallucinations are due to attribution errors, where internal mental events are mistaken as sensations originated from the surrounding environment (McGuire et al., 1995). Some researchers propose that AVHs result from an attentional bias toward internally generated information. This theory is supported by neuroimaging findings that suggest patients with AVHs also show deficits in processing exogenously presented sounds (David et al., 1996; Woodruff et al., 1997; Hugdahl et al., 2009; Kompus et al., 2011). Patients with AVHs can show an impaired performance on auditory attention tasks, such as the dichotic listening (DL) paradigm (Hugdahl et al., 2007). A meta-analysis found that activation in the left primary auditory cortex and the right rostral prefrontal cortex increased in the absence of external auditory stimuli but decreased in the presence of external auditory stimuli (Kompus et al., 2011).

In fact, numerous brain regions associated with audition are also involved in processing other sensory information and the integration of different types of sensory information (Caetano and Jousmäki, 2006; Butler et al., 2012; Knöpfel et al., 2019; Kassuba et al., 2020). Meanwhile, there are interactions between audition and other senses in some ways, such as touch and vision. For instance, long-term exposure to auditory noises can increase sensitivity to tactile frequency (Kassuba et al., 2020). By providing visual stimuli consistent with auditory stimuli, the representation of that sound can be strengthened in the auditory cortex (Atilgan et al., 2018). Moreover, multisensory integration effects can also act on some more advanced functions, such as emotional response and perception (Collignon et al., 2008; Pan et al., 2019), speech and language processing (Righi et al., 2018), and self-perception (Noel et al., 2018). According to the research based on the rubber hand illusion (RHI) paradigm, the sense of body ownership also relies on the reception and integration of self and externally generated multisensory information (Grechuta et al., 2021). Previous studies have shown that auditory cues can help to enhance the illusion, which suggests that auditory information plays a role in the framework of body ownership (Radziun and Ehrsson, 2018). In addition, a recent study has revealed that the sense of body ownership can be impacted by distal auditory signals that are independent of the action (Grechuta et al., 2021). We can speculate that, since AVHs involve abnormal auditory perceptions, they may have an impact on the processing of all types of sensory information, including the sense of body ownership. Although prior studies have established that deficits in multisensory integration are closely correlated with schizophrenia, the studies on the direct relationship between AVHs and integration processing are lacking, and most of them focus on audiovisual integration (Surguladze et al., 2001; Kim et al., 2003; Szycik et al., 2009).

The RHI paradigm is a classic and effective experimental method to induce body ownership illusions *via* the visual and tactile stimulations (Graham et al., 2014). The sense of body ownership relies on multisensory integration. The RHI paradigm induces a distorted sense of body ownership in an attempt to mediate conflicting visual and proprioceptive signals by integrating multisensory input to form a coherent representation of the body and the world (Tang et al., 2015; Klaver and Dijkerman, 2016). Simply put, the participants are more likely to report that the rubber hand is their own hand by hiding the real hand and, brushing the two hands synchronously while watching the fake hand, while these feelings can decrease under asynchronous conditions (Tsakiris and Haggard, 2005). In addition, the illusion can be reduced under circumstances where the rubber hand does not match the subject's visual body image, such as using a wooden block or changing the position of the subject's hand (Tsakiris and Haggard, 2005; Costantini and Haggard, 2007).

Due to their disturbed sense of self, patients with schizophrenia can experience abnormal body perception, including impairments in body ownership. Therefore, they are more susceptible to enhanced RHI (Thakkar et al., 2011; Zopf et al., 2021). The use of the RHI in patients with schizophrenia was first described by Peled et al. (2000), who found that patients have more profound experiences of the illusion than healthy controls (HCs), and subsequent studies have reported similar results (Peled et al., 2003; Thakkar et al., 2011; Germine et al., 2013). Furthermore, healthy individuals with psychosis-proneness also have more intensive experiences than those without schizotypal personality traits (Germine et al., 2013). A significant correlation between the severity of body-related perceptual symptoms and the intensity of RHI has been found in patients with schizophrenia (Germine et al., 2013; Zopf et al., 2021). In addition, researchers have observed that the sensitivity to detect temporal differences decreases in patients with schizophrenia by comparing the intensity of the illusion between synchronous and asynchronous conditions, which is related to their increased susceptibility to RHI (Zopf et al., 2021).

In view of the prevalence and specificity of AVH in psychosis, studying AVH is likely to provide a valuable perspective that is distinct from that generated by the study of schizophrenia alone. The RHI paradigm is an accessible way to observe an impaired sense of body ownership in individuals with schizophrenia. Unlike previous studies, it is a process of visual-tactile integration, so we can explore the performance of the patients in this special process and its relationship with specific

symptoms. Hence, in this study, we used the RHI paradigm to the schizophrenia patients with AVHs and compared their performance with HCs and the patients without AVHs. We aimed to observe the sense of body ownership and visual–tactile integration in schizophrenia patients with AVHs through an easy and intuitive method. We hypothesized that the illusion intensity of patients with AVHs might be greater than that of patients without AVHs and HCs due to the deficits in their sense of self.

MATERIALS AND METHODS

Participants

A total of 80 patients who met the DSM IV-TR criteria for schizophrenia were recruited from among outpatients attending the Second Xiangya Hospital of Central South University, China, from 2018 to 2019. A total of 36 healthy volunteers with no personal or family history of mental illness were recruited using poster advertisements and screened by matching their age, gender, and education years to the patients. All participants were native Chinese speakers and were interviewed by psychiatrists to evaluate their ability to understand and follow simple instructions. The inclusion criteria required that participants were right-handed, aged between 16 and 45, and were able to understand and cooperate with the research procedures. The exclusion criteria for all participants were (1) substance abuse, (2) a history of head trauma resulting in loss of consciousness, or (3) major medical or neurological illness. According to the presence of AVHs, patients were divided into two subgroups. Patients with AVHs were defined as patients who experienced AVHs at least once a day in the past year, while patients without AVHs were those who had never experienced AVHs since the onset of the disease. The subgroup of patients with AVHs contained 47 patients, while the subgroup of patients without AVHs contained 33 patients. Patients in each group were matched for age, gender, education years, medication, and duration of illness. This study was approved by the Second Xiangya Hospital Ethics Committee (No. S006, 2018), and all participants fully understood the research procedures and provided written informed consent.

The Rubber Hand Illusion Setup

Each participant sat down at the table opposite the experimenter. They placed their right hand in an open-side box to hide it from view. A life-sized rubber hand, wearing a blue latex glove, was placed to the left of the real right hand. The distance between the two hands was 15 cm. To maintain a consistent appearance, participants were required to wear a blue latex glove on their right hand. In addition, their right arm was covered by a long black cloak, which could prevent the participant from observing the location of their real hand. Two rotating brushes, driven by an electric motor (Xinda XD60D94-12Y-505, China), struck the right index finger of both the real and rubber hands at a constant rhythm of approximately 1 Hz. The experiment included two conditions, namely, synchronous and asynchronous. In the synchronous condition, the two brushes moved simultaneously without interval time. In the asynchronous condition, there was a 500-min delay in the brush that struck the real hand, which meant that tactile stimulation would be given to the participants after the visual stimulation of seeing the brush strike the rubber hand. Before the start of the experiment, the rubber hand was hidden from view.

Experimental Design

According to the condition of visuo-tactile stimulation, the experiment was divided into two blocks, and all participants experienced both blocks. The stimulation order was determined by coin flipping and was completely random and counterbalanced. Each block consisted of a 2-min induction period followed by a 3-min period during which the participants located their right index finger. Two measures were applied to evaluate the intensity of the illusion, namely, proprioceptive drift and subjective experience.

Proprioceptive Drift

At the beginning of the task, all participants were asked to make a judgment about the baseline position of their right index finger. Then, the participant was allowed to see the rubber hand and the electric motor started. After a 2-min induction period (Tang et al., 2015; Prikken et al., 2019; Zopf et al., 2021), the participant made a further judgment about their index finger every 1 min. Five judgments would be made in total, and each task lasted for 6 min. All judgments were documented using the readings observed in a ruler placed on top of the box. Drift was defined as the change in the perceived location from the baseline measurement, calculated by subtracting the estimates before the illusion from the estimates after the illusion. The average of the four measurements was used in the analysis.

Subjective Experience

After completing each task, participants were asked to complete a questionnaire in Chinese consisting of nine statements describing the perceptual experiences and sensations, such as "It seemed as if I were feeling the touch of the paintbrush in the location where I saw the rubber hand" and "I felt as if the rubber hand were my hand" (Botvinick and Cohen, 1998). Answers were scored on a scale ranging from 1 (completely disagree) to 5 (completely agree).

Clinical Evaluation

Clinical evaluation was performed independently by two experienced senior psychiatrists. The severity of illness was assessed by the 30-item Positive and Negative Syndrome Scale (PANSS) (Kay et al., 1987). For the subgroup of patients with AVHs, the 7-term Auditory Hallucinations Rating Scale (AHRS) was conducted to assess the severity of AVH (Hoffman et al., 2003).

Statistical Analysis

All statistical analyses were performed using SPSS (version 25, IBM Inc., New York, United States). All statistical tests were two-tailed, and the effects were considered significant if $p < 0.05$. Shapiro–Wilk's tests were used as normality tests. As the data did not fall into a normal distribution, we used non-parametric tests:

TABLE 1 | Demographic and clinical characteristics of patients with auditory verbal hallucinations (AVHs), patients with no AVHs and healthy controls.

	AVHs patients (n = 47)	Non-AVHs patients (n = 33)	HCs (n = 36)	Statistic test (df)	p value
Demographic					
Gender(M/F)	25/22	15/18	19/17	$\chi^2(2) = 0.541$	0.763
Age (years)	25.9 ± 5.3 (17–38)	27.0 ± 6.3 (18–42)	26.9 ± 5.7 (19–36)	$F(2,113) = 0.511$	0.601
Education(years)	12.6 ± 2.4 (9–17)	13.7 ± 2.7 (9–19)	13.8 ± 2.6 (6–17)	$F(2,113) = 2.933$	0.057
Clinical					
Duration of illness (months)	7.9 ± 5.0	6.0 ± 3.8	–	t(78) = 2.196	0.060
Medication (CPZE mg/day)	639.6 ± 273.3	566.7 ± 411.7	–	t(78) = 0.878	0.344
PANSS Total	58.6 ± 13.6	49.7 ± 15.7	–	t(78) = 2.692	0.009*
PANSS Positive	16.2 ± 4.1	10.4 ± 3.9	–	t(78) = 6.576	<0.001*
PANSS Negative	14.9 ± 5.5	13.1 ± 6.1	–	t(78) = 1.130	0.167
PANSS General	27.4 ± 6.6	26.2 ± 7.6	–	t(78) = 0.820	0.443
P3 item	4.8 ± 0.8	–	–	–	–
AHRS Score	25.9 ± 3.8	–	–	–	–

*Values are provided as mean ± SD unless otherwise stated. PANSS, Positive and Negative Syndrome Scale; AHRS, Auditory Hallucinations Rating Scale; CPZE, chlorpromazine equivalent dose. *p < 0.05.*

the Kruskal–Wallis H test was used in the intergroup comparison; Wilcoxon signed-rank test was used in analyses within the same group; and Spearman's rank correlation coefficient test was used in exploring the relationship between the severity of symptoms and the intensity of the illusion.

RESULTS

Participants

The demographic and clinical characteristics of the participants are summarized in **Table 1**. The study population consisted of 47 patients with AVHs, 33 patients without AVHs, and 36 HCs. No significant differences were found in gender, age, and education among the three groups. Compared to patients without AVHs, patients with AVHs had higher PANSS total and positive scores. However, no group differences were found for the negative or general scores between the two patient groups.

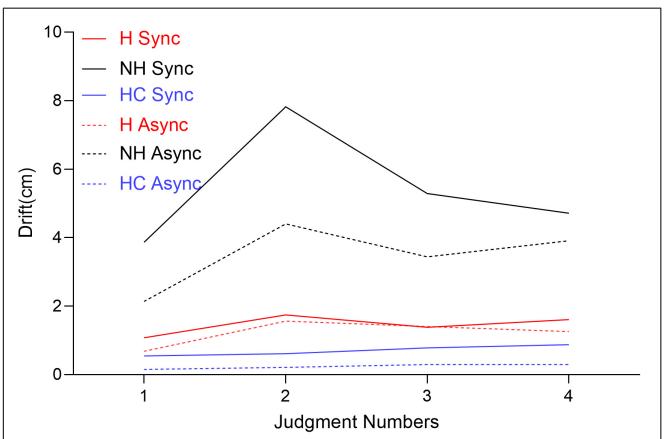


FIGURE 1 | Line graph to show the change in the proprioceptive drift of patients and healthy controls (HCs) over time. Judgments were taken at 1 min intervals.

Medication and the Illusion

Several studies show that the RHI might relate to dopaminergic pathways (Albrecht et al., 2011; Ding et al., 2017; Waldmann et al., 2020), so the medication was entered into the correlation analysis with the illusion. We explored their relationship using Spearman's rank correlation coefficient. Significance level was defined at $\alpha = 0.0125$ after Bonferroni correction. There was no significant correlation between medication and synchronous proprioceptive drift [$\rho(80) = -0.167$, $p = 0.138$], asynchronous proprioceptive drift [$\rho(80) = -0.087$, $p = 0.441$], synchronous subjective score [$\rho(80) = 0.121$, $p = 0.286$], or asynchronous subjective score [$\rho(80) = 0.240$, $p = 0.032$].

Preliminary Analysis: Baseline Accuracy of Hand Position Estimation

To test whether there were differences in localization at baseline status between the patient groups and the HCs, we first calculated the difference between the baseline readings of the experimenter and that of the participants. Then we conducted a Kruskal–Wallis H test, which showed that there were no significant differences among the three groups in either the synchronous condition [$H(2) = 5.314$, $p = 0.070$] or the asynchronous condition [$H(2) = 3.441$, $p = 0.179$]. Therefore, we believed that the patients had no difficulty in hand localization prior to the onset of the illusion-inducing stimulation.

Proprioceptive Drift

The perceived hand location varied over time (**Figure 1**). As shown in the figure, in both patient groups, the maximum proprioceptive drift appeared in the second measurement, while in HCs, it appeared in the last measurement.

The Kruskal–Wallis H test showed that whether in the synchronous condition [$H(2) = 23.834$, $p < 0.001$] or the

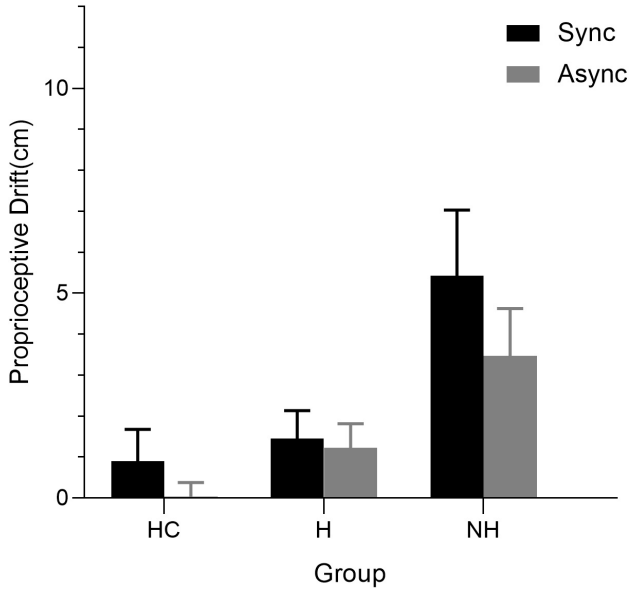


FIGURE 2 | Mean proprioceptive drift of patients with auditory verbal hallucinations (AVHs) (H), patients without AVHs (NH), and HCs under two stimulation conditions: synchronous (Sync) and asynchronous (Async). Error bars reflect 95% confidence intervals.

asynchronous condition [H(2) = 24.198, $p < 0.001$], there was a significant difference in proprioceptive drift among all groups (**Figure 2**). Pairwise multiple comparisons were performed using Dunn's pairwise tests with Bonferroni correction, which showed that, under the synchronous conditions, patients without AVHs had the largest mean drift, followed by patients with AVHs and HCs in succession, and the difference in patients without AVHs between patients with AVHs ($p < 0.001$) and HCs ($p < 0.001$) was significant; however, the difference between patients with AVHs and HCs was not significant ($p = 0.844$). Similar results were found in asynchronous conditions, and the pairwise comparisons between the patients without AVHs and the other two groups were significantly different (pH-NH = 0.015, pHC-NH < 0.001, pHC-H = 0.040). The Wilcoxon signed-ranks test showed that the total drift was greater after synchronous rather than asynchronous stimulation in patients without AVHs ($Z = 1.984$, $p = 0.047$) and HCs ($Z = 2.451$, $p = 0.014$), but no significant differences between the two conditions were observed among patients with AVHs ($Z = 0.321$, $p = 0.748$) (**Figure 2**).

We also performed a generalized linear mixed model (GLMM) to estimate the proprioceptive drift of the patients, and the PANSS score ($p = 0.515$), medication ($p = 0.743$), stimulation condition ($p < 0.001$), and group ($p < 0.001$) were involved in the analysis. The results confirmed that there were no associations between PANSS score or medication and proprioceptive drift.

Subjective Experience

The illusion scores were compared between the three groups using the Kruskal–Wallis H test followed by the *post hoc* Dunn's test with Bonferroni correction. The results showed that the illusion score was significantly different among the three groups in the two conditions [synchronous condition: H(2) = 19.159, $p < 0.001$; asynchronous condition: H(2) = 19.446, $p < 0.001$] (**Figure 3**). Under the synchronous conditions, consistent with the proprioceptive drift, the score of patients without AVHs was significantly higher than that of patients with AVHs ($p = 0.009$) and HCs ($p < 0.001$), while there was no significant difference between the other two groups ($p = 0.303$). Under the asynchronous conditions, each patient group had a higher score than HCs (pHC-NH = 0.002, pHC-H < 0.001), but no significant differences were found between the two patient groups (pH-NH = 1.000) (**Figure 3**). The synchronous illusion scores were also higher than the asynchronous illusion scores in HCs ($Z = 4.054$, $p < 0.001$) and patients without AVHs ($Z = 4.276$, $p < 0.001$), while no significant difference between the illusion scores under the two conditions was found among patients with AVHs ($Z = 0.138$, $p = 0.890$). Similarly, the results of GLMM showed that medication ($p = 0.091$) and PANSS score ($p = 0.523$) were irrelevant to the illusion score.

Relationship Between the Illusion and the Severity of Symptoms

We explored the relationship between the illusion intensity and the severity of symptoms using Spearman's rank correlation coefficient. Both proprioceptive drift and subjective experience under all conditions were entered in the correlation analysis. There was no significant correlation between the measures of the illusion for the two conditions and PANSS total scores or each subscale score in the two patient groups. In addition, for the subgroup of patients with AVHs, the total AHRS score and each item score were included in further analysis. However, no significant correlation was found either.

DISCUSSION

The aim of this study was to assess the deficits in sense of body ownership and multisensory integration in schizophrenia patients with/without AVHs. We used the RHI paradigm in HCs and patients with/without AVHs and compared their performances. The main findings were as follows: (1) compared to HCs, the illusion of schizophrenia patients with or without AVHs both showed up more quickly and was stronger. (2) Patients with AVHs experienced a weaker illusion than patients without AVHs. (3) Unlike HCs and patients without AVHs, patients with AVHs did not show a reduction in the RHI illusion under asynchronous conditions.

In agreement with previous studies, this study showed that the proprioceptive drift peak appeared earlier and was higher in both patient groups than in HCs. The strong and quick illusion of individuals with schizophrenia revealed that there was an impairment of body ownership, which is a component of self-awareness (Thakkar et al., 2011). In fact, the sense of body ownership involves two processes: the bottom-up integration of multisensory inputs and the top-down prediction (Grechuta et al., 2019). Although disturbances of body ownership in

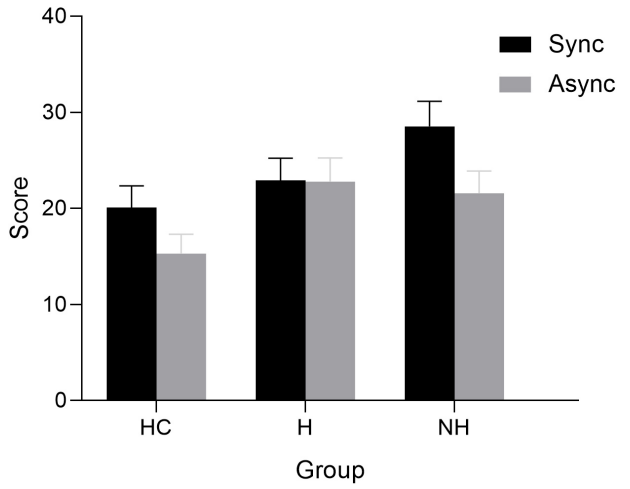


FIGURE 3 | Mean overall subjective experience score of patients with AVHs (H), patients without AVHs (NH), and healthy controls (HC) under two stimulation conditions: synchronous (Sync) and asynchronous (Async). Error bars reflect 95% confidence intervals.

individuals with schizophrenia are often linked with their deficits in multisensory integration, the role of top-down processes cannot be ignored. Ferri et al. designed a new induction procedure for the RHI in which the anticipation replaces the actual synchronized visuo-tactile stimulation and applied it to patients with schizophrenia (Ferri et al., 2013, 2014). The results showed that the patients with schizophrenia had a weaker illusion than the HCs in this paradigm and revealed that patients might have a predictive mechanism different from healthy individuals (Ferri et al., 2014). There have been RHI studies using an unpredictable delay in the asynchronous condition, and their results confirmed the presence of top-down differences between population groups (Botan et al., 2018, 2021). However, there have not been RHI studies on schizophrenia using unpredictable delays. Hence, further RHI studies on schizophrenia can take advantage of unpredictable delays in the asynchronous condition to better understand the predictive mechanisms of patients with schizophrenia.

It is well accepted that a combination of subjective and objective measures is conducive to the better evaluation of the RHI (Klaver and Dijkerman, 2016). In our study, the two measurements' results were consistent. Nevertheless, contradictory results have been reported in previous studies, which show dissociation between subjective ratings and drift (Shimada et al., 2009; Rohde et al., 2011; Romano et al., 2015). Additionally, studies on the associations between RHI and hypnotizability have suggested that the RHI processes contain implicit imaginative suggestion effects, which may have an impact on the reliability of so-called objective measures such as proprioceptive drift and skin conductance response (Lush et al., 2020, 2021). Lower self-awareness is related to higher hypnotizability (Cardeña and Terhune, 2019). Notably, higher hypnotizability enhances RHI effects and becomes a potential confounding factor (Lush et al., 2020, 2021). However, lower hypnotizability and enhanced RHI effects can both be observed in patients with schizophrenia (Frischholz et al., 1992). Hence, further discussions on hypnotizability and schizophrenia are needed, and relevant confounders should be identified and removed.

However, contrary to our expectation, patients with AVHs showed weaker illusion intensity than patients without AVHs, except that the two groups of patients had similar subjective experience scores under asynchronous conditions. We suggest that there can be several possible reasons to explain the result. The occurrence of RHI is a result of multisensory inputs and integration, so the reduction of the illusion may not mean a better sense of body ownership for patients with AVHs, but a reduction of the response to external stimulations and of multisensory integration. In fact, the development of AVHs is caused not only by the increased activation of endogenously evoked processing but also by the decreased activation of exogenously evoked processing (Kompus et al., 2011). Previous imaging studies have found that the same areas in the auditory cortex will be activated when an external auditory stimulus is not present; instead, when receiving the stimulus, the area will be deactivated (Binder et al., 1996; Hugdahl et al., 1999). There is an inhibitory dysfunction in the anterior cingulate cortex (ACC) of individuals with hallucinations, which leads to a deficient top-down inhibition of automatic activations. Therefore, the self-generated information can compete with the exogenous stimulations for attentional resources, resulting in a reduced response to the exogenous stimulations (Hunter et al., 2006; Hubl et al., 2007). We assume that competition may occur in the response to multisensory stimuli but is not limited to external auditory stimuli.

The RHI is a response to tactile and visual stimuli; in other words, the reduced illusion may represent a reduced response to external stimuli. In addition, the RHI is based on multisensory integration, while impaired multisensory processing but not unisensory processing by individuals with schizophrenia has been detected in some paradigms (de Gelder et al., 2005; Ross et al., 2007; Germine et al., 2013; Stevenson et al., 2017). Among all of the symptoms of schizophrenia, hallucinations are the most closely associated with multisensory integration, and previous studies have shown that the more types of hallucinations the patients experienced, the more severe their impairments were (Williams et al., 2010; Postmes et al., 2014).

A relationship between the audiovisual speech-perception network and hallucinations has been recognized, and this network is involved in the audiovisual integration. In addition, some regions in the impaired auditory association cortex, such as the superior temporal gyrus (STG) and the prefrontal cortex, of patients with AVHs may also play a role in tactile perception (Foxe et al., 2002; Schürmann et al., 2006; Bolognini et al., 2010; Spitzer and Blankenburg, 2012; Vergara et al., 2016; Ćurčić-Blake et al., 2017a; Hjelmervik et al., 2020). Existing studies have shown that external auditory signals can enhance the sensitivity to detect tactile frequency (Crommett et al., 2017). We deduced that the tactile frequency sensitivity in patients with AVHs might be reduced, which further caused an increased sensory asynchrony between tactile and visual stimuli and reduced the illusions. In short, there may be some remarkable overlaps between the

regions involved in hallucinations and multisensory integration (Surguladze et al., 2001; Kim et al., 2003; Szycik et al., 2009).

Accordingly, a deficit in visual-tactile integration may also exist in patients with AVHs, contributing to the reduced illusions of patients with AVHs. Another likely explanation is that hallucinations and delusions are two of the most common symptoms of schizophrenia, and delusions may be more prevalent in the group of patients without AVHs. The RHI can be seen as a false belief that the fake hand is one's own hand, similar to a delusion in a way, and previous studies have confirmed that the RHI effects are positively correlated with delusions in patients with schizophrenia (Germine et al., 2013; Crespi and Dinsdale, 2019; Prikken et al., 2019). It has been found that the experiences of delusions are associated with the structural changes within the insula, and dysfunctions in the region are critical to the increased RHI effects (Cascella et al., 2011; Crespi and Dinsdale, 2019).

An unexpected finding was that there was no difference in the illusion effects of patients with AVHs between the synchronous and asynchronous conditions. In the previous RHI studies on patients with schizophrenia, this phenomenon tended to be associated with their timing deficits (Zopf et al., 2021). It is commonly known that individuals with schizophrenia have a reduced sensitivity to asynchrony, and they usually need longer temporal intervals between stimulations to detect the asynchrony (Foucher et al., 2007; Arzy et al., 2011; Stevenson et al., 2017). This reduced sensitivity has been linked to body-related perceptual symptoms and passivity symptoms (Graham et al., 2014; Zopf et al., 2021). The result of a study using the RHI paradigm was in line with the findings, and it also indicated that the reduced sensitivity was irrelevant to the increased RHI susceptibility (Zopf et al., 2021). In this study, we set the delay time to 500 min, which was not long enough to avoid the normal reduction in RHI typically seen under asynchronous conditions (Graham et al., 2014). We found it in patients without AVHs but not in patients with AVHs, indirectly suggesting that the impairments in temporal processing in patients with AVHs were more severe than in patients without AVHs. Even so, the illusion effect did not become stronger along with it, which confirms the finding mentioned above (Zopf et al., 2021).

These findings are supported by relevant neuroimaging studies, showing that the functional and anatomical changes in the posterior temporal cortex not only participate in the formation of hallucinations (Barta et al., 1990; Levitan et al., 1999; Kim et al., 2003; Cachia et al., 2008), but are also involved in multisensory integration processes, including temporal processing (Miller and D'Esposito, 2005; Stevenson et al., 2011). Some researchers have proposed that there might not be specific brain networks for temporal perception, and instead they widely exist in each neural network (Stevenson et al., 2017). However, most neuroimaging studies on the relationship between multisensory integration, temporal processing in it and hallucinations are focused on audiovisual coupling, and the precise mechanisms and the regions involved with visual-tactile integration in patients suffering from hallucinations are still unknown. Hence, further studies should take more types of sensory integration into account.

CONCLUSION

In conclusion, there was a disturbance in the sense of body ownership in individuals with schizophrenia whether they experienced AVHs or not. In addition, those patients with AVHs seemed to show a reduced multisensory processing and integration. Meanwhile, deficits in temporal perception, which indicate impaired multisensory integration, were observed. The reason might be that there are some overlaps between the brain regions involved in multisensory integration and hallucinations. Notably, our findings of temporal perception deficits were in accordance with the lack of facilitation of RHI for the synchronous condition, so measures of time perception should be considered in future studies. The current findings were based on behavioral measures only, whereas past neurobiological studies about the integration process were mostly limited to audiovisual stimulation. Therefore, some neuroimaging tools, such as magnetic resonance or electroencephalogram, can be used in future studies to more deeply explore the underlying mechanisms.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Second Xiangya Hospital Ethics Committee. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

JT and XC designed the study. JH, HR, JL, MD, and LD collected the samples and clinical information. JH analyzed and discussed the experimental result. JH and HR wrote the first draft of the manuscript. ZL, YM, YL, and PT performed manuscript revision. LG, JT, and XC reviewed the draft and contributed to manuscript preparation. All authors contributed to and approved the final manuscript.

FUNDING

This research was supported by the National Natural Science Foundation of China (Nos. 81871057 and 82171495 to JT and No. 81871056 to XC) and the Fundamental Research Funds for the Central Universities of The Central South University (CX20190251).

ACKNOWLEDGMENTS

We are grateful to all the participants in this study.

REFERENCES

Albrecht, M. A., Martin-Iverson, M. T., Price, G., Lee, J., Iyyalol, R., and Waters, F. (2011). Dexamphetamine effects on separate constructs in the rubber hand illusion test. *Psychopharmacology* 217, 39–50. doi: 10.1007/s00213-011-2255-y

Arzy, S., Mohr, C., Molnar-Szakacs, I., and Blanke, O. (2011). Schizotypal perceptual aberrations of time: correlation between score, behavior and brain activity. *PLoS One* 6:e16154. doi: 10.1371/journal.pone.0016154

Atilgan, H., Town, S. M., Wood, K. C., Jones, G. P., Maddox, R. K., Lee, A. K. C., et al. (2018). Integration of visual information in auditory cortex promotes auditory scene analysis through multisensory binding. *Neuron* 97, 640–655.e4. doi: 10.1016/j.neuron.2017.12.034

Barta, P. E., Pearlson, G. D., Powers, R. E., Richards, S. S., and Tune, L. E. (1990). Auditory hallucinations and smaller superior temporal gyral volume in schizophrenia. *Am. J. Psychiatry* 147, 1457–1462. doi: 10.1176/ajp.147.11.1457

Binder, J. R., Frost, J. A., Hammeke, T. A., Rao, S. M., and Cox, R. W. (1996). Function of the left planum temporale in auditory and linguistic processing. *Brain J. Neurol.* 119(Pt 4), 1239–1247. doi: 10.1093/brain/119.4.1239

Bolognini, N., Papagno, C., Moroni, D., and Maravita, A. (2010). Tactile temporal processing in the auditory cortex. *J. Cogn. Neurosci.* 22, 1201–1211. doi: 10.1162/jocn.2009.21267

Botan, V., Fan, S., Critchley, H., and Ward, J. (2018). Atypical susceptibility to the rubber hand illusion linked to sensory-localised vicarious pain perception. *Conscious. Cogn.* 60, 62–71. doi: 10.1016/j.concog.2018.02.010

Botan, V., Salisbury, A., Critchley, H. D., and Ward, J. (2021). Vicarious pain is an outcome of atypical body ownership: evidence from the rubber hand illusion and enfacement illusion. *Q. J. Exp. Psychol.* 74, 1888–1899. doi: 10.1177/17470218211024822

Botvinick, M., and Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature* 391:756. doi: 10.1038/35784

Butler, J. S., Foxe, J. J., Fiebelkorn, I. C., Mercier, M. R., and Molholm, S. (2012). Multisensory representation of frequency across audition and touch: high density electrical mapping reveals early sensory-perceptual coupling. *J. Neurosci.* 32, 15338–15344. doi: 10.1523/jneurosci.1796-12.2012

Cachia, A., Paillère-Martinot, M. L., Galinowski, A., Januel, D., de Beaurepaire, R., Bellivier, F., et al. (2008). Cortical folding abnormalities in schizophrenia patients with resistant auditory hallucinations. *Neuroimage* 39, 927–935. doi: 10.1016/j.neuroimage.2007.08.049

Caetano, G., and Jousmäki, V. (2006). Evidence of vibrotactile input to human auditory cortex. *Neuroimage* 29, 15–28. doi: 10.1016/j.neuroimage.2005.07.023

Cardeña, E., and Terhune, D. B. (2019). The roles of response expectancies, baseline experiences, and hypnotizability in spontaneous hypnotic experiences. *Int. J. Clin. Exp. Hypn.* 67, 1–27. doi: 10.1080/00207144.2019.1553759

Cascella, N. G., Gerner, G. J., Fieldstone, S. C., Sawa, A., and Schretlen, D. J. (2011). The insula-claustrum region and delusions in schizophrenia. *Schizophr. Res.* 133, 77–81. doi: 10.1016/j.schres.2011.08.004

Choong, C., Hunter, M. D., and Woodruff, P. W. (2007). Auditory hallucinations in those populations that do not suffer from schizophrenia. *Curr. Psychiatry Rep.* 9, 206–212. doi: 10.1007/s11920-007-0020-z

Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., et al. (2008). Audio-visual integration of emotion expression. *Brain Res.* 1242, 126–135. doi: 10.1016/j.brainres.2008.04.023

Costantini, M., and Haggard, P. (2007). The rubber hand illusion: sensitivity and reference frame for body ownership. *Conscious. Cogn.* 16, 229–240. doi: 10.1016/j.concog.2007.01.001

Crespi, B., and Dinsdale, N. (2019). Autism and psychosis as diametrical disorders of embodiment. *Evol. Med. Public Health* 2019, 121–138. doi: 10.1093/emph/eoz021

Crommett, L. E., Pérez-Bellido, A., and Yau, J. M. (2017). Auditory adaptation improves tactile frequency perception. *J. Neurophysiol.* 117, 1352–1362. doi: 10.1152/jn.00783.2016

Ćurčić-Blake, B., Bais, L., Sibeijn-Kuiper, A., Pijnenborg, H. M., Knegtering, H., Liemburg, E., et al. (2017a). Glutamate in dorsolateral prefrontal cortex and auditory verbal hallucinations in patients with schizophrenia: a (1)H MRS study. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 78, 132–139. doi: 10.1016/j.pnpbp.2017.05.020

Ćurčić-Blake, B., Ford, J. M., Hubl, D., Orlov, N. D., Sommer, I. E., Waters, F., et al. (2017b). Interaction of language, auditory and memory brain networks in auditory verbal hallucinations. *Prog. Neurobiol.* 148, 1–20. doi: 10.1016/j.pneurobio.2016.11.002

David, A. S., Woodruff, P. W., Howard, R., Mellers, J. D., Brammer, M., Bullmore, E., et al. (1996). Auditory hallucinations inhibit exogenous activation of auditory association cortex. *Neuroreport* 7, 932–936. doi: 10.1097/00001756-199603220-00021

de Gelder, B., Vroomen, J., de Jong, S. J., Masthoff, E. D., Trompenaars, F. J., and Hodiamont, P. (2005). Multisensory integration of emotional faces and voices in schizophrenics. *Schizophr. Res.* 72, 195–203. doi: 10.1016/j.schres.2004.02.013

Dierks, T., Linden, D. E., Jandl, M., Formisano, E., Goebel, R., Lanfermann, H., et al. (1999). Activation of Heschl's gyrus during auditory hallucinations. *Neuron* 22, 615–621. doi: 10.1016/s0896-6273(00)80715-1

Ding, C., Palmer, C. J., Hohwy, J., Youssef, G. J., Paton, B., Tsuchiya, N., et al. (2017). Parkinson's disease alters multisensory perception: insights from the Rubber Hand Illusion. *Neuropsychologia* 97, 38–45. doi: 10.1016/j.neuropsychologia.2017.01.031

Ferri, F., Chiarelli, A. M., Merla, A., Gallese, V., and Costantini, M. (2013). The body beyond the body: expectation of a sensory event is enough to induce ownership over a fake hand. *Proc. Biol. Sci.* 280:20131140. doi: 10.1098/rspb.2013.1140

Ferri, F., Costantini, M., Salone, A., Di Iorio, G., Martinotti, G., Chiarelli, A., et al. (2014). Upcoming tactile events and body ownership in schizophrenia. *Schizophr. Res.* 152, 51–57. doi: 10.1016/j.schres.2013.06.026

Foucher, J. R., Lacambre, M., Pham, B. T., Giersch, A., and Elliott, M. A. (2007). Low time resolution in schizophrenia Lengthened windows of simultaneity for visual, auditory and bimodal stimuli. *Schizophr. Res.* 97, 118–127. doi: 10.1016/j.schres.2007.08.013

Foxe, J. J., Wylie, G. R., Martinez, A., Schroeder, C. E., Javitt, D. C., Guilfoyle, D., et al. (2002). Auditory-somatosensory multisensory processing in auditory association cortex: an fMRI study. *J. Neurophysiol.* 88, 540–543. doi: 10.1152/jn.2002.88.1.540

Frischholz, E. J., Lipman, L. S., Braun, B. G., and Sachs, R. G. (1992). Psychopathology, hypnotizability, and dissociation. *Am. J. Psychiatry* 149, 1521–1525. doi: 10.1176/ajp.149.11.1521

Germine, L., Benson, T. L., Cohen, F., and Hooker, C. I. (2013). Psychosis-proneness and the rubber hand illusion of body ownership. *Psychiatry Res.* 207, 45–52. doi: 10.1016/j.psychres.2012.11.022

Graham, K. T., Martin-Iverson, M. T., Holmes, N. P., Jablensky, A., and Waters, F. (2014). Deficits in agency in schizophrenia, and additional deficits in body image, body schema, and internal timing, in passivity symptoms. *Front. Psychiatry* 5:126. doi: 10.3389/fpsyt.2014.00126

Grechuta, K., De La Torre Costa, J., Ballester, B. R., and Verschure, P. (2021). Challenging the boundaries of the physical self: distal cues impact body ownership. *Front. Hum. Neurosci.* 15:704414. doi: 10.3389/fnhum.2021.704414

Grechuta, K., Ulysse, L., Rubio Ballester, B., and Verschure, P. (2019). Self beyond the body: action-driven and task-relevant purely distal cues modulate performance and body ownership. *Front. Hum. Neurosci.* 13:91. doi: 10.3389/fnhum.2019.00091

Hjelmervik, H., Craven, A. R., Sinceviciute, I., Johnsen, E., Kompus, K., Bless, J. J., et al. (2020). Intra-Regional Glu-GABA vs Inter-Regional Glu-Glu Imbalance: a 1H-MRS Study of the Neurochemistry of Auditory Verbal Hallucinations in Schizophrenia. *Schizophr. Bull.* 46, 633–642. doi: 10.1093/schbul/sbz099

Hoffman, R. E., Hawkins, K. A., Gueorguieva, R., Boutros, N. N., Rachid, F., Carroll, K., et al. (2003). Transcranial magnetic stimulation of left temporoparietal cortex and medication-resistant auditory hallucinations. *Arch. Gen. Psychiatry* 60, 49–56. doi: 10.1001/archpsyc.60.1.49

Hubl, D., Koenig, T., Strik, W. K., Garcia, L. M., and Dierks, T. (2007). Competition for neuronal resources: how hallucinations make themselves heard. *Br. J. Psychiatry* 190, 57–62. doi: 10.1192/bjp.bp.106.022954

Hugdahl, K., Brønnick, K., Kyllingsbaek, S., Law, I., Gade, A., and Paulson, O. B. (1999). Brain activation during dichotic presentations of consonant-vowel and musical instrument stimuli: a 15O-PET study. *Neuropsychologia* 37, 431–440. doi: 10.1016/s0028-3932(98)00101-8

Hugdahl, K., Løberg, E. M., and Nygård, M. (2009). Left temporal lobe structural and functional abnormality underlying auditory hallucinations in schizophrenia. *Front. Neurosci.* 3, 34–45. doi: 10.3389/neuro.01.001.2009

Hugdahl, K., Løberg, E. M., Specht, K., Steen, V. M., van Wageningen, H., and Jørgensen, H. A. (2007). Auditory hallucinations in schizophrenia: the role of cognitive, brain structural and genetic disturbances in the left temporal lobe. *Front. Hum. Neurosci.* 1:6. doi: 10.3389/neuro.09.006.2007

Hunter, M. D., Eickhoff, S. B., Miller, T. W., Farrow, T. F., Wilkinson, I. D., and Woodruff, P. W. (2006). Neural activity in speech-sensitive auditory cortex during silence. *Proc. Natl. Acad. Sci. U.S. A.* 103, 189–194. doi: 10.1073/pnas.0506268103

Kassuba, T., Pinsk, M. A., and Kastner, S. (2020). Distinct auditory and visual tool regions with multisensory response properties in human parietal cortex. *Prog. Neurobiol.* 195:101889. doi: 10.1016/j.pneurobio.2020.101889

Kay, S. R., Fiszbein, A., and Opler, L. A. (1987). The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophr. Bull.* 13, 261–276. doi: 10.1093/schbul/13.2.261

Kim, J. J., Crespo-Facorro, B., Andreasen, N. C., O'Leary, D. S., Magnotta, V., and Nopoulos, P. (2003). Morphology of the lateral superior temporal gyrus in neuroleptic nai;ve patients with schizophrenia: relationship to symptoms. *Schizophr. Res.* 60, 173–181. doi: 10.1016/s0920-9964(02)00299-2

Klaver, M., and Dijkerman, H. C. (2016). Bodily experience in schizophrenia: factors underlying a disturbed sense of body ownership. *Front. Hum. Neurosci.* 10:305. doi: 10.3389/fnhum.2016.00305

Knöpfel, T., Sweeney, Y., Radulescu, C. I., Zabouri, N., Doostdar, N., Clopath, C., et al. (2019). Audio-visual experience strengthens multisensory assemblies in adult mouse visual cortex. *Nat. Commun.* 10:5684. doi: 10.1038/s41467-019-13607-2

Kompus, K., Westerhausen, R., and Hugdahl, K. (2011). The "paradoxical" engagement of the primary auditory cortex in patients with auditory verbal hallucinations: a meta-analysis of functional neuroimaging studies. *Neuropsychologia* 49, 3361–3369. doi: 10.1016/j.neuropsychologia.2011.08.010

Levitan, C., Ward, P. B., and Catts, S. V. (1999). Superior temporal gyral volumes and laterality correlates of auditory hallucinations in schizophrenia. *Biol. Psychiatry* 46, 955–962. doi: 10.1016/s0006-3223(98)00373-4

Lush, P., Botan, V., Scott, R. B., Seth, A. K., Ward, J., and Dienes, Z. (2020). Trait phenomenological control predicts experience of mirror synaesthesia and the rubber hand illusion. *Nat. Commun.* 11:4853. doi: 10.1038/s41467-020-18591-6

Lush, P., Seth, A. K., and Dienes, Z. (2021). Hypothesis awareness confounds asynchronous control conditions in indirect measures of the rubber hand illusion. *R. Soc. Open Sci.* 8:210911. doi: 10.1098/rsos.210911

McGuire, P. K., Silbersweig, D. A., Wright, I., Murray, R. M., David, A. S., Frackowiak, R. S., et al. (1995). Abnormal monitoring of inner speech: a physiological basis for auditory hallucinations. *Lancet* 346, 596–600. doi: 10.1016/s0140-6736(95)91435-8

Miller, L. M., and D'Esposito, M. (2005). Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. *J. Neurosci.* 25, 5884–5893. doi: 10.1523/jneurosci.0896-05.2005

Noel, J. P., Park, H. D., Pasqualini, I., Lissek, H., Wallace, M., Blanke, O., et al. (2018). Audio-visual sensory deprivation degrades visuo-tactile peri-personal space. *Conscious. Cogn.* 61, 61–75. doi: 10.1016/j.concog.2018.04.001

Pan, F., Zhang, L., Ou, Y., and Zhang, X. (2019). The audio-visual integration effect on music emotion: behavioral and physiological evidence. *PLoS One* 14:e0217040. doi: 10.1371/journal.pone.0217040

Peled, A., Pressman, A., Geva, A. B., and Modai, I. (2003). Somatosensory evoked potentials during a rubber-hand illusion in schizophrenia. *Schizophr. Res.* 64, 157–163. doi: 10.1016/s0920-9964(03)00057-4

Peled, A., Ritsner, M., Hirschmann, S., Geva, A. B., and Modai, I. (2000). Touch feel illusion in schizophrenic patients. *Biol. Psychiatry* 48, 1105–1108. doi: 10.1016/s0006-3223(00)00947-1

Postmes, L., Sno, H. N., Goedhart, S., van der Stel, J., Heering, H. D., and de Haan, L. (2014). Schizophrenia as a self-disorder due to perceptual incoherence. *Schizophr. Res.* 152, 41–50. doi: 10.1016/j.schres.2013.07.027

Prikken, M., van der Weiden, A., Baalbergen, H., Hillegers, M. H., Kahn, R. S., Aarts, H., et al. (2019). Multisensory integration underlying body-ownership experiences in schizophrenia and offspring of patients: a study using the rubber hand illusion paradigm. *J. Psychiatry Neurosci.* 44, 177–184. doi: 10.1503/jpn.180049

Radziun, D., and Ehrsson, H. H. (2018). Auditory cues influence the rubber-hand illusion. *J. Exp. Psychol. Hum. Percept. Perform.* 44, 1012–1021. doi: 10.1037/xhp0000508

Righi, G., Tenenbaum, E. J., McCormick, C., Blossom, M., Amso, D., and Sheinkopf, S. J. (2018). Sensitivity to audio-visual synchrony and its relation to language abilities in children with and without ASD. *Autism Res.* 11, 645–653. doi: 10.1002/aur.1918

Rohde, M., Di Luca, M., and Ernst, M. O. (2011). The Rubber Hand Illusion: feeling of ownership and proprioceptive drift do not go hand in hand. *PLoS one* 6:e21659. doi: 10.1371/journal.pone.0021659

Romano, D., Caffa, E., Hernandez-Arieta, A., Brugger, P., and Maravita, A. (2015). The robot hand illusion: inducing proprioceptive drift through visuo-motor congruency. *Neuropsychologia* 70, 414–420. doi: 10.1016/j.neuropsychologia.2014.10.033

Ross, L. A., Saint-Amour, D., Leavitt, V. M., Molholm, S., Javitt, D. C., and Foxe, J. J. (2007). Impaired multisensory processing in schizophrenia: deficits in the visual enhancement of speech comprehension under noisy environmental conditions. *Schizophr. Res.* 97, 173–183. doi: 10.1016/j.schres.2007.08.008

Schürmann, M., Caetano, G., Hlushchuk, Y., Jousmäki, V., and Hari, R. (2006). Touch activates human auditory cortex. *Neuroimage* 30, 1325–1331. doi: 10.1016/j.neuroimage.2005.11.020

Shimada, S., Fukuda, K., and Hiraki, K. (2009). Rubber hand illusion under delayed visual feedback. *PLoS One* 4:e6185. doi: 10.1371/journal.pone.0006185

Spitzer, B., and Blankenburg, F. (2012). Supramodal parametric working memory processing in humans. *J. Neurosci.* 32, 3287–3295. doi: 10.1523/JNEUROSCI.5280-11.2012

Stevenson, R. A., Park, S., Cochran, C., McIntosh, L. G., Noel, J. P., Barense, M. D., et al. (2017). The associations between multisensory temporal processing and symptoms of schizophrenia. *Schizophr. Res.* 179, 97–103. doi: 10.1016/j.schres.2016.09.035

Stevenson, R. A., VanDerKlok, R. M., Pisoni, D. B., and James, T. W. (2011). Discrete neural substrates underlie complementary audiovisual speech integration processes. *Neuroimage* 55, 1339–1345. doi: 10.1016/j.neuroimage.2010.12.063

Surguladze, S. A., Calvert, G. A., Brammer, M. J., Campbell, R., Bullmore, E. T., Giampietro, V., et al. (2001). Audio-visual speech perception in schizophrenia: an fMRI study. *Psychiatry Res.* 106, 1–14. doi: 10.1016/s0925-4927(00)00081-0

Szycik, G. R., Münte, T. F., Dillo, W., Mohammadi, B., Samii, A., Emrich, H. M., et al. (2009). Audiovisual integration of speech is disturbed in schizophrenia: an fMRI study. *Schizophr. Res.* 110, 111–118. doi: 10.1016/j.schres.2009.03.003

Tang, J., Morgan, H. L., Liao, Y., Corlett, P. R., Wang, D., Li, H., et al. (2015). Chronic administration of ketamine mimics the perturbed sense of body ownership associated with schizophrenia. *Psychopharmacology* 232, 1515–1526. doi: 10.1007/s00213-014-3782-0

Thakkar, K. N., Nichols, H. S., McIntosh, L. G., and Park, S. (2011). Disturbances in body ownership in schizophrenia: evidence from the rubber hand illusion and case study of a spontaneous out-of-body experience. *PLoS One* 6:e27089. doi: 10.1371/journal.pone.0027089

Tsakiris, M., and Haggard, P. (2005). The rubber hand illusion revisited: visuotactile integration and self-attribution. *J. Exp. Psychol. Hum. Percept. Perform.* 31, 80–91. doi: 10.1037/0096-1523.31.1.80

Vergara, J., Rivera, N., Rossi-Pool, R., and Romo, R. A. (2016). Neural parametric code for storing information of more than one sensory modality in working memory. *Neuron* 89, 54–62. doi: 10.1016/j.neuron.2015.11.026

Waldmann, A., Volkmann, J., and Zeller, D. (2020). Parkinson's disease may reduce sensitivity to visual-tactile asynchrony irrespective of dopaminergic treatment: evidence from the rubber hand illusion. *Parkinsonism Relat. Disord.* 78, 100–104. doi: 10.1016/j.parkreldis.2020.07.016

Waters, F., Woodward, T., Allen, P., Aleman, A., and Sommer, I. (2010). Self-recognition deficits in schizophrenia patients with auditory hallucinations: a meta-analysis of the literature. *Schizophr. Bull.* 38, 741–750. doi: 10.1093/schbul/sbq144

Williams, L. E., Light, G. A., Braff, D. L., and Ramachandran, V. S. (2010). Reduced multisensory integration in patients with schizophrenia on a target detection

task. *Neuropsychologia* 48, 3128–3136. doi: 10.1016/j.neuropsychologia.2010.06.028

Woodruff, P. W., Wright, I. C., Bullmore, E. T., Brammer, M., Howard, R. J., Williams, S. C., et al. (1997). Auditory hallucinations and the temporal cortical response to speech in schizophrenia: a functional magnetic resonance imaging study. *Am. J. Psychiatry* 154, 1676–1682. doi: 10.1176/ajp.154.12.1676

Zopf, R., Boulton, K., Langdon, R., and Rich, A. N. (2021). Perception of visual-tactile asynchrony, bodily perceptual aberrations, and bodily illusions in schizophrenia. *Schizophr. Res.* 228, 534–540. doi: 10.1016/j.schres.2020.11.038

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Neuroscience

ORIGINAL RESEARCH
published: 25 April 2022
doi: 10.3389/fnins.2022.821078

Altered Coupling of Cerebral Blood Flow and Functional Connectivity Strength in First-Episode Schizophrenia Patients With Auditory Verbal Hallucinations

Jingli Chen, Kangkang Xue, Meng Yang, Kefan Wang, Yinhuan Xu, Baohong Wen, Jingliang Cheng, Shaoqiang Han* and Yarui Wei**

Department of Magnetic Resonance Imaging, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China

OPEN ACCESS

Edited by:
William Sedley, Newcastle University, United Kingdom

Reviewed by:
Valentina Ciullo, Santa Lucia Foundation (IRCCS), Italy
Matthew J. Hoptman, Nathan Kline Institute for Psychiatric Research, United States

***Correspondence:**
Jingliang Cheng
fccchengjl@zzu.edu.cn
Shaoqiang Han
shaoqianghan@163.com
Yarui Wei
yarui_wei@163.com

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

Received: *23 November 2021*
Accepted: *09 February 2022*
Published: *25 April 2022*

Citation:
Chen J, Xue K, Yang M, Wang K, Xu Y, Wen B, Cheng J, Han S and Wei Y (2022) Altered Coupling of Cerebral Blood Flow and Functional Connectivity Strength in First-Episode Schizophrenia Patients With Auditory Verbal Hallucinations. Front. Neurosci. 16:821078. doi: 10.3389/fnins.2022.821078

Objective: Auditory verbal hallucinations (AVHs) are a major symptom of schizophrenia and are connected with impairments in auditory and speech-related networks. In schizophrenia with AVHs, alterations in resting-state cerebral blood flow (CBF) and functional connectivity have been described. However, the neurovascular coupling alterations specific to first-episode drug-naïve schizophrenia (FES) patients with AVHs remain unknown.

Methods: Resting-state functional MRI and arterial spin labeling (ASL) was performed on 46 first-episode drug-naïve schizophrenia (FES) patients with AVHs (AVH), 39 FES drug-naïve schizophrenia patients without AVHs (NAVH), and 48 healthy controls (HC). Then we compared the correlation between the CBF and functional connection strength (FCS) of the entire gray matter between the three groups, as well as the CBF/FCS ratio of each voxel. Correlation analyses were performed on significant results between schizophrenia patients and clinical measures scale.

Results: The CBF/FCS ratio was reduced in the cognitive and emotional brain regions in both the AVH and NAVH groups, primarily in the crus I/II, vermis VI/VII, and cerebellum VI. In the AVH group compared with the HC group, the CBF/FCS ratio was higher in auditory perception and language-processing areas, primarily the left superior and middle temporal gyrus (STG/MTG). The CBF/FCS ratio in the left STG and left MTG positively correlates with the score of the Auditory Hallucination Rating Scale in AVH patients.

Conclusion: These findings point to the difference in neurovascular coupling failure between AVH and NAVH patients. The dysfunction of the forward model based on the predictive and computing role of the cerebellum may increase the excitability in the auditory cortex, which may help to understand the neuropathological mechanism of AVHs.

Keywords: cerebral blood flow, functional connectivity, neurovascular coupling, auditory verbal hallucination, forward model

INTRODUCTION

Auditory verbal hallucinations (AVHs) are cardinal symptoms in schizophrenia. Defined as the auditory experience of "hearing voices" in the absence of external stimuli that cause them, AVH is suffered by 60–80% of the patients and often produces distress, functional disability, and behavioral alterations (Jardri et al., 2011; Hjelmervik et al., 2020). Considering the severe cognitive problems, poor quality of life, and high morbidity, the physiological mechanism underlying AVHs should be fully understood to promote effective treatment.

Many models have been proposed to account for the different mechanisms of AVHs involving a wide range of brain regions far beyond the auditory cortex in schizophrenia, including into the thalamus (Ferri et al., 2018) and cerebellum (Pinheiro et al., 2020a). Especially, in recent decades, neuroimaging techniques have provided evidence for the central role of the cerebellar circuit in the forward model, which links AVH patients to impaired cerebellar function or structure by erratic prediction and imprecise computation of sensory consequences and also affects higher-level cognitive processes (Sokolov et al., 2017; Moberget and Ivry, 2019; Pinheiro et al., 2020b). Predictive timing disturbances in the forward model are a special marker of SZ and have been associated with other cognitive dysfunctions documented in prior studies (Ciullo et al., 2018). The forward model suggests that the cerebellum compares expected and actual sensory feedback (Martha et al., 2018). Sensory error messages are specifically encoded in the cerebellum's Purkinje cells' complicated spike discharges (Brooks et al., 2015). The cerebellar output is small if the entering stimulus matches the predicted one; if a discrepancy–error message is received, activity in the cerebellum increases, and a vast area of cerebral cortex is alerted by increasing excitability (Marco Molinari et al., 2008). In schizophrenia patients with AVHs, differences are sent to cortical regions such as the left superior and middle temporal gyrus (STG/MTG) via the thalamus (Tourville et al., 2008). One previous study reported that an acute brain disorder causes interruption of the excitatory projections from the lesioned brain area to the anatomically intact brain regions (Warren et al., 1958). Similar to the former, a unilateral cerebellar lesion decreased the contralateral cortical excitability and remained a baseline hemispheric CBF unchanging contralateral to a cerebellar lesion, which was suggested to have impaired neurovascular coupling between the cerebellum and cerebral cortex (Enager et al., 2004). Furthermore, a corticocerebellar–thalamic–cortical circuit connects the cerebellum to numerous areas of the cerebral cortex, and the cerebellum may play a key role in this circuit in psychosis by coordinating or modulating elements of cortical activity (Andreasen and Pierson, 2008; Cao et al., 2018). If this circuitry is disrupted, it will cause "cognitive dysmetria," which is difficult to prioritize, process, coordinate, and respond to information, eventually leading to function decoupling (Nancy et al., 1998). However, how the neurovascular coupling alteration involved in information processing in these brain regions within these circuits is disrupted remains unknown.

In recent years, to get a deeper understanding of the alteration of neurovascular relationships in neurological diseases, researchers began to use the method of neurovascular coupling to explore the pathogenesis of diseases, such as primary open-angle glaucoma (Wang Q. et al., 2021), bipolar disorder and major depressive disorder (He Z. et al., 2019), and Alzheimer's disease (Drzezga et al., 2011). The method of neurovascular coupling, reflected by cerebral blood flow (CBF), functional connectivity strength (FCS), and their relationship, showed a direct relationship among functional activity, metabolism, and neural activity, which demonstrated that brain regions with higher spontaneous neural activity tend to have more robust connectivity and increased perfusion (Liang et al., 2013). Regional CBF is tightly coupled with brain metabolism and can be measured utilizing functional neuroimaging techniques, such as arterial spin labeling (ASL), which has been widely used in schizophrenia (Vaishnavi et al., 2010; Zhuo et al., 2017; Jing et al., 2018). Increased CBF of the left STG was found in AVH patients accompanied by the left MTG by using this technique (Homan et al., 2012; Zhu et al., 2017; Zhuo et al., 2017). The whole-brain functional connectivity strength (FCS) highlights the involvement of each voxel in transmitting information in the overall brain network by depicting a specific voxel as well as all other voxels in the brain that surpassed a predetermined optimum threshold (Liang et al., 2013; van Lutterveld et al., 2014; Li et al., 2020). There are functional connectivity measures on how well a local activity is integrated across brain regions, which help researchers better understand the dysfunctions in integrated brain networks and the exact coordination of inter-regions in schizophrenia (Lui et al., 2010). In the graph theory, the FCS is also known as "degree centrality" of weighted networks (Wang Y. et al., 2021), and brain areas with a high FCS are regarded functional hubs that are well connected to the rest of the brain (Liang et al., 2013). The whole-brain functional connectivity approach solves some of the limitations of seed-based rsFC analysis and independent component analysis (ICA), all of which are approaches in quantifying rsFC alterations. For example, in the absence of the underlying pathophysiology for a disease, the analysis of seed-based rsFC approach may be difficult due to the requirement for *a priori* definition of seed regions (Nair et al., 2014). The ICA may face uncertainty about the optimal number of components, contentious criteria for discriminating between noise and signal, and interpretive complications brought by a sophisticated algorithm (Fox and Raichle, 2007). Increased FCS in the left crus I, bilateral crus II, left cerebellum VI, vermis VI, vermis VII, and decreased FCS in the left temporal cortex have been discovered using the FCS approach to explore connection alterations in schizophrenia (Wang et al., 2017; Zhu et al., 2017; Basavaraju et al., 2019; Ding et al., 2019).

The CBF–FCS correlation measures the spatial consistency of CBF and FCS across voxels over the entire gray matter (Zhu et al., 2017). The metabolic demand per unit of connection strength is measured by the CBF/FCS ratio, which indirectly indicates the neurovascular coupling of a single voxel or local region (Zhu et al., 2017). Cerebral volume reduction, neural loss, abnormal astrocytes, and white matter pathway interruptions that may contribute to the neurovascular decoupling have been reported in recent studies (Cocchi et al., 2014; Wang Q. et al., 2021). Voxel-wise entire brain studies of the CBF/FCS ratios and

CBF–FCS correlations can provide more precise and sensitive information on the alterations in brain functional regions than voxel-wise whole-brain analysis of CBF and FCS indices simply. Using the approach of neurovascular coupling in individuals with schizophrenia, Zhu et al. discovered reduced CBF/FCS ratios in higher-order brain systems related to cognitive control and affective regulation and elevated CBF/FCS ratios in lower-order brain systems, such as sensory processing (Zhu et al., 2017). Unfortunately, up until now, no study has investigated the alteration in neurovascular coupling specific to first-episode drug-naïve schizophrenia (FES) patients with AVHs (AVH).

Considering that patients with AVHs are characterized by impaired information processing related to auditory and speech-related networks in the forward model, which is believed to be associated with common pathological processes of AVHs, we hypothesized that neurovascular coupling alterations should be atypical in AVH patients. Furthermore, because brain regions with CBF and FCS changes are spatially inconsistent, with different effect sizes and directions in the AVH and first-episode drug-naïve schizophrenia (FES) patients without AVHs (NAVH) (**Figures 1**, **2** and **Supplementary Tables 1, 2**), we hypothesized that the AVH and NAVH groups would show a reduced and different CBF–FCS coupling, as well as an increased or decreased CBF/FCS ratio accompanied by different CBF and FCS changes. The voxel-based CBF and FCS analyses to detect abnormal perfusion and neural activity in AVH patients and first-episode drug-naïve schizophrenia (FES) patients without AVHs (NAVH) by using ASL and BOLD–fMRI were performed. Three groups were compared on the basis of CBF–FCS coupling in overall gray matter and CBF/FCS ratio voxel-by-voxel.

MATERIALS AND METHODS

Participants

This study recruited 50 FES AVH patients, 50 FES NAVH patients, and 50 age- and sex-matched HC. After preprocessing, we removed zero patients with AVH, five patients with NAVH, and zero patients with HC due to head movement parameters exceeding 3 mm displacement or 3° of rotation. Among the subjects left after preprocessed, we excluded four AVH patients, six NAVH patients, and two HC subjects because they were

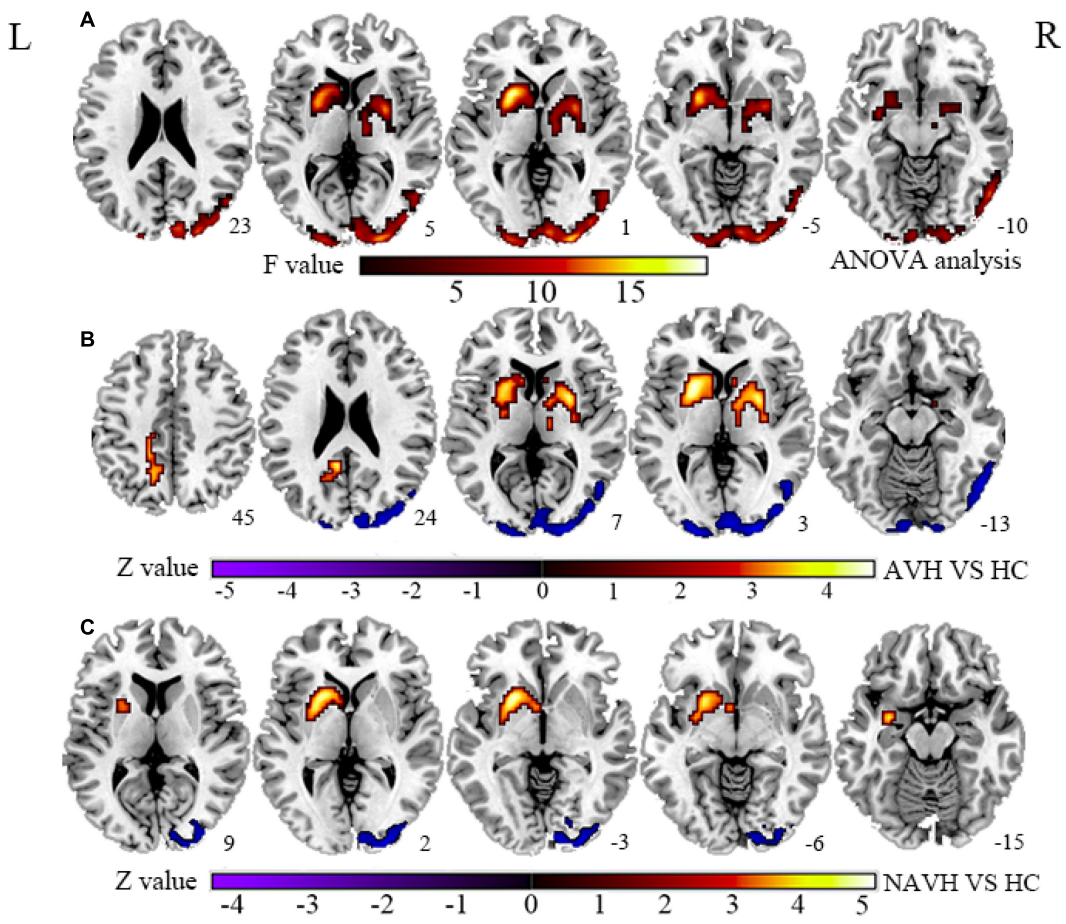


FIGURE 1 | (A) Brain regions with significant CBF changes between the AVH, NAVH, and HC groups. **(B)** Brain regions with significant CBF changes in the AVH group. **(C)** Brain regions with significant CBF changes in the NAVH group. CBF, cerebral blood flow; AVH, schizophrenia patients with AVHs; NAVH, schizophrenia patients without AVHs; HC, healthy control.

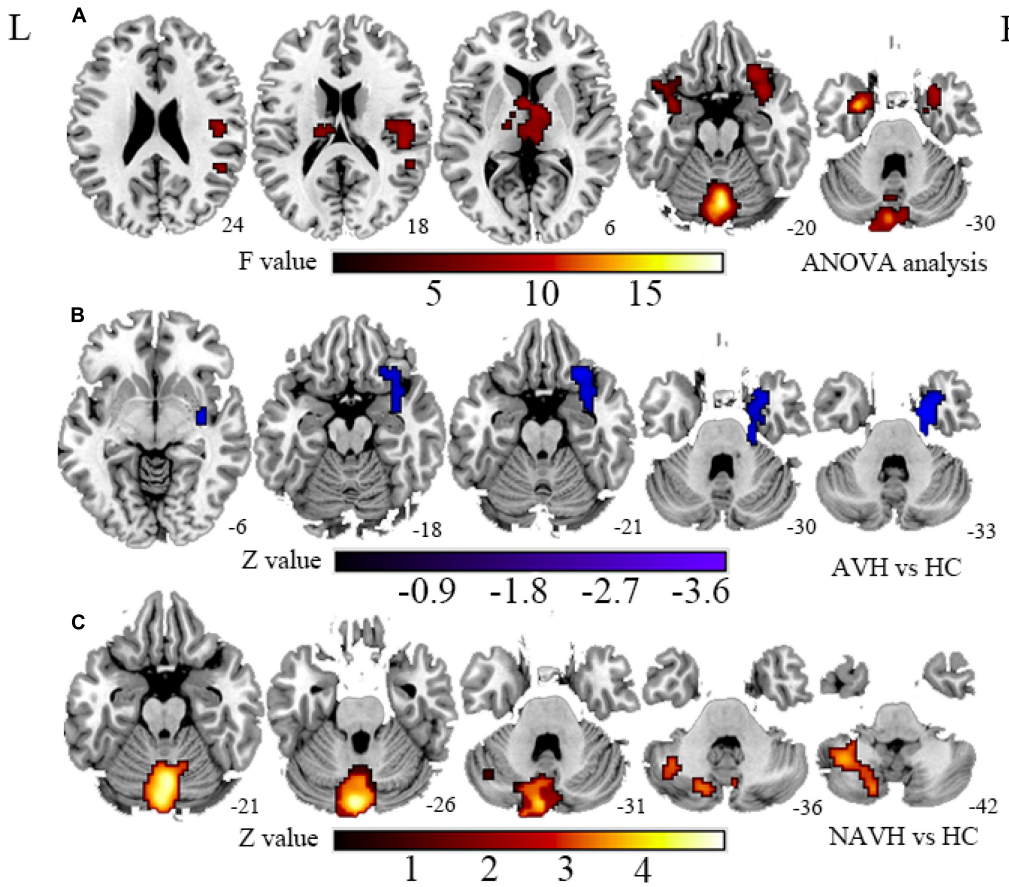


FIGURE 2 | (A) Brain regions with significant FCS changes between AVH, NAVH, and HC groups. **(B)** Brain regions with significant FCS changes in the AVH group. **(C)** Brain regions with significant FCS changes in the NAVH group. FCS, functional connectivity strength; AVH, schizophrenia patients with AVHs; NAVH, schizophrenia patients without AVHs.

missing data corresponding to CBF. Finally, included in the study were 46 AVH patients, 39 NAVH patients, and 48 HC. The detailed demographic and clinical data for these participants are shown in **Table 1**.

The diagnosis of schizophrenia is determined by a psychiatrist and evaluated and confirmed by an experienced psychologist using DSM-5 standards (Vahia, 2013). The Positive and Negative Symptom Scale (PANSS) was used to assess the severity of psychotic symptoms. A total of 46 patients reported experiencing AVHs within the past 4 weeks in the AVH group, most within the past week, while the other 39 patients reported no AVHs in their lifetime or in the past 4 months in the NAVH group. The Auditory Hallucination Rating Scale (AHRS) was used to assess the severity of AVHs. In the end, we collected PANSS data from 46 AVH patients and 39 NAVH patients and AHRS hallucination data from all AVH patients.

For the patient groups, the following are the exclusion criteria: (1) mental disorders caused by physical diseases other than schizophrenia, (2) alcohol addiction or a history of substance abuse, (3) contraindications to MRI, and (4) traumatic head injuries. The exclusion criteria for the HCs are any mental illness, neurological disease, and related family history. The age- and gender-matched HCs were recruited from the same geographic area. Moreover, the same exclusion criteria that were used for SZ patients were used for HCs. All subjects are right-handed. All subjects signed an informed consent form, and this study has been approved by the Ethics Committee of the First Affiliated Hospital of Zhengzhou University.

Data Acquisition

All subjects who met the enrollment conditions used the same eight-channel 3.0 Tesla magnetic resonance scanner (GE Discovery MR750, United States) to complete the MRI data collection. The collection location was located in the Magnetic Resonance Department of the First Affiliated Hospital of Zhengzhou University. Spatial 3D high-resolution T1-weighted images (3DT1) were acquired using a brain volume sequence with the following settings: repetition time (TR)/echo time (TE) = 8.2/3.2 ms, slice thickness = 1 mm, slice gap = 0 mm, flip angle = 12°, slice number = 1, field of view (FOV) = 25.6 × 25.6 cm^2, number of averages = 1, matrix size = 256 × 256, and voxel size = 1 × 1 × 1 mm^3. The resting-state perfusion imaging was performed using a pseudo-continuous ASL (pcASL) sequence with a 3D fast spin–echo acquisition

TABLE 1 | Demographic and clinical data of AVH patients, NAVH patients, and HC.

	AVH	NAVH	HC	$F/X^2/t$-values	p-Values
Number of subjects	46	39	48		
Age (SD)	21.7 (7.86)	20.2 (7.2)	21.9 (7.7)	0.659	0.519
Sex (M/F)	21/25	18/21	23/25	0.033	0.984
Hoffman hallucinations (SD)	23.8 (6.13)	–	–	–	–
PNASS (SD)					
Positive	20.1(5.6)	18.1 (6.1)	–	1.474	0.145
Negative	20.2 (5.1)	21.0 (5.8)	–	-0.630	0.531
General	41.6 (7.4)	42.6 (8.6)	–	-0.49	0.626
Total scores	82.3 (14.5)	81.6 (16.4)	–	0.189	0.851
PANSS hallucinations	4.1 (1.6)	2.3 (1.5)		4.125	0.000
PANSS delusions	4.7 (1.4)	3.8 (1.7)		2.107	0.040
FD (SD)	0.12 (0.08)	0.13 (0.10)	0.13 (0.07)	0.369	0.692

AVH, schizophrenia with AVH patients; NAVH, schizophrenia without AVH patients; HC, healthy control; F, female; M, male; FD, framewise displacement; PNASS, positive and negative syndrome scale.

and background suppression, and the parameters are as follows: TR/TE = 4,886/10.5 ms, slice thickness = 4.0 mm, slice gap = 0 mm, flip angle = 111°, slice number = 80, FOV = 24 × 24 cm^2, number of averages = 3, matrix size = 512 × 512, voxel size = 0.5 × 0.5 × 4 mm^3. A state of rest BOLD-fMRI data were collected using the following parameters in a gradient-echo single-shot echo-planar imaging (GRE-SS-EPI) sequence: TR/TE = 2,000/30 ms, slice thickness = 4 mm, slice gap = 0.5 mm, flip angle = 90°, slice number = 32, FOV = 22 × 22 cm^2, number of averages = 1, matrix size = 64 × 64, voxel size = 3.4375 × 3.4375 × 3.4375 mm^3. The duration of the resting state scan is 6 min.

fMRI Data Preprocessing

The BOLD-fMRI data are preprocessed by the Data Processing and Analysis of Brain Imaging (DPABI) toolbox[1], which is based on Statistical Parametric Mapping (SPM12[2]) and MATLAB (MathWorks). The following steps were performed: (1) removing the first five time points; (2) slice timing correction; (3) realigning; if a participant's maximum head motion was greater than 3 mm or 3° of rotation, they were excluded; (4) normalizing the BOLD-fMRI data space to the template of the Montreal Neurology Institute (resampled voxel size = 3 × 3 × 3 mm^3); (5) detrending; (6) filtering (0.01–0.08 Hz); (7) scrubbing the BOLD-fMRI data; and (8) regression of the Friston-24 motion parameters, cerebrospinal fluid signal, white matter signal.

Functional Connectivity Analysis

The FCS of the whole-brain gray matter is the average value of the functional connectivity strength between a given voxel X0 and all other voxels in the whole brain gray matter. Based on the gray matter template provided by the software, the Pearson correlation coefficient between each voxel and other voxel BOLD time series was calculated, the correlation threshold was set at 0.2 (Liu et al., 2015), and the complete gray matter function connection matrix of each subject was obtained. We used an isotropic Gaussian kernel [full width at half maximum (FWHM) = 6 mm] to spatially smooth the FCS map.

Cerebral Blood Flow Analysis

The CBF images were received from the ASL difference images by subtracting the label images from the control images. The CBF images were processed through a cloud platform (Beijing Intelligent Brain Cloud, Inc.[3]). (1) The CBF images were coregistered to the template of the Montreal Neurology Institute (resampled voxel size = 3 × 3 × 3 mm^3) and segmented into gray matter, white matter, and cerebrospinal fluid maps. (2) The images were spatially smoothened by using a Gaussian kernel with 6 mm fullwidth at half-maximum (FWHM).

Cerebral Blood Flow–Functional Connectivity Strength Correlation Analysis

We conducted correlation analyses across voxels for each participant to statistically analyze the correlation relation between CBF and FCS on the entire gray matter. First, the CBF and FCS maps were normalized into Z-scores for each participant by subtracting the mean and dividing by the SD of global values within the gray matter mask. The df_{eff} of across-voxel correlations was then calculated using the equation below:

$$df_{eff} = \frac{N}{(FWHMx \times FWHMy \times FWHMz)/v} - 2$$

where v is the volume of a voxel (3 × 3 × 3 mm^3), and N is the number of voxels (N = 66,817) used in the analyses. FWHMx × FWHMy × FWHMz were the average smoothness of the CBF and FCS maps (12.1 × 13.0 × 11.9 mm^3) estimated using Matlab's DPABI software (DPABI V3.0[4]). In our study, the df_{eff} of across-voxel correlations was 961. Finally, The CBF–FCS correlation coefficients were compared using one-way ANOVA.

[1]http://rfmri.org/dpabi

[2]http://www.fil.ion.ucl.ac.uk/spm

[3]www.humanbrain.cn

[4]http://rfmri.org/

Cerebral Blood Flow/Functional Connectivity Strength Ratio Analysis

To evaluate the amount of blood supply per unit of connectivity strength, we computed the CBF/FCS ratio of each voxel. Before calculating the CBF/FCS ratio, it is necessary to note that both CBF and FCS are the original values without *Z* transformation. After the ratio, perform *Z* transformation to improve normality. The operation steps are based on Matlab's DPABI software, using the whole-brain gray matter template as the mask, and calculate the CBF/FCS ratio for each subject.

Voxel-Wise Comparisons in Cerebral Blood Flow and Functional Connectivity Strength

To further understand what might be causing changes in the CBF/FCS ratio, we analyzed CBF and FCS changes between the three groups voxel-wise while controlling for age, gender, and GMV.

Clinical and Cognitive Associations

Based on the anatomical template, the average CBF/FCS ratio of the subregions with significant group differences on *F* map were extracted, and the non-parametric Spearman rank correlation analysis (Bonferroni corrected) was used to test the CBF/FCS ratio of each significant subregion and the clinical measures in the AVH and NAVH groups (PANSS positive, negative, general, total score, and AHRS).

Statistical Analysis

The intergroup differences of voxel-wise CBF, FCS, CBF–FCS correlation, and CBF/FCS ratio were tested by using analysis of variance (ANOVA) with age, gender, and GMV (Crow et al., 1980) of each subject as covariates. Multiple comparisons were corrected according to the Gaussian random field (GRF) theory (voxel-wise $p < 0.005$, cluster-wise $p < 0.05$, two-tailed) in the DPABI toolbox (see text footnote 1)[4].

Validation Analysis

The correlation threshold of $r = 0.2$ was applied in the FCS calculation (Liu et al., 2015). We repeatedly computed the whole-brain FCS with correlation thresholds of 0.1 and 0.3 to verify the stability of the results.

RESULTS

Spatial Distribution of the Functional Connectivity Strength, Cerebral Blood Flow, and Cerebral Blood Flow/Functional Connectivity Strength Ratio

The geographic distributions of FCS, CBF, and the CBF/FCS ratio were similar in the AVH, NAVH, and HC groups (**Supplementary Figure 1**). At the level of CBF index, the brain regions of the three groups of HC, AVH, and NAVH that showed similar CBF elevation were distributed in the medial/lateral prefrontal cortex, anterior/posterior cingulate cortex, precuneus, lateral temporal and parietal cortices, sensorimotor, and visual cortices. At the level of FCS index, the brain regions of the three groups of HC, AVH, and NAVH that showed similar FCS elevation were distributed in the lateral temporal cortex, prefrontal cortex, anterior and posterior cingulum, and visual cortex, which were all shown to have higher FCS. At the level of CBF/FCS ratios index, the brain regions of the three groups of HC, AVH, and NAVH that showed similar CBF/FCS ratio elevations were distributed in the medial prefrontal cortex, anterior cingulate cortex, sensorimotor cortex, and thalamus.

Cerebral Blood Flow–Functional Connectivity Strength Correlation

Although CBF was significantly correlated with FCS in both AVH, NAVH, and control groups (**Figure 3**), the three groups had no significant differences in CBF–FCS coupling (one-way ANOVA $F = 0.473$, $p = 0.624$; **Figure 3**). *Post-hoc* analysis: AVH group compared with HC (two-sample *t*-test, $t = -0.343$, $p = 0.732$), NAVH compared with HC (two-sample *t*-test, $t = 0.692$, $p = 0.491$), AVH compared with NAVH (two-sample *t*-test, $t = -1.017$, $p = 0.321$).

Cerebral Blood Flow/Functional Connectivity Strength Ratio

Compared with the HC group, both the AVH and NAVH groups exhibited decreased CBF/FCS ratio in left crus I/II, vermis VI/VII, as well as left cerebellum VI. Compared with the HC group, the AVH group showed increased CBF/FCS ratio in the left STG/MTG unparalleled and decreased CBF/FCS ratio in right cerebellum crus II (voxel level $p < 0.005$, cluster level $p < 0.05$, GRF-corrected, **Figure 4** and **Table 2**). Unfortunately, there was no difference in CBF/FCS ratio between the AVH and NAVH groups.

Cerebral Blood Flow and Functional Connectivity Strength

Compared with the HC group, the significant brain regions of CBF in AVH and NAVH patients are shown in **Figure 1** and **Supplementary Table 1** ($p < 0.05$, GRF-corrected), and the FCS is shown in **Figure 2** and **Supplementary Table 2** ($p < 0.05$, GRF-corrected). The AVH group did not show any difference in CBF compared with the NAVH group (voxel level $p < 0.005$, cluster level $p < 0.05$, GRF-corrected).

The Correlation Between Cerebral Blood Flow/Functional Connectivity Strength Ratio and Psychotic Symptoms

Supplementary Table 3 shows the associations of the PANSS positive, negative, and general subscores with the normalized CBF/FCS ratio of each significant subregion. **Figure 5** and **Supplementary Table 4** exhibit the relationships between the CBF/FCS ratio of each meaningful ROI and the AHRS. In the

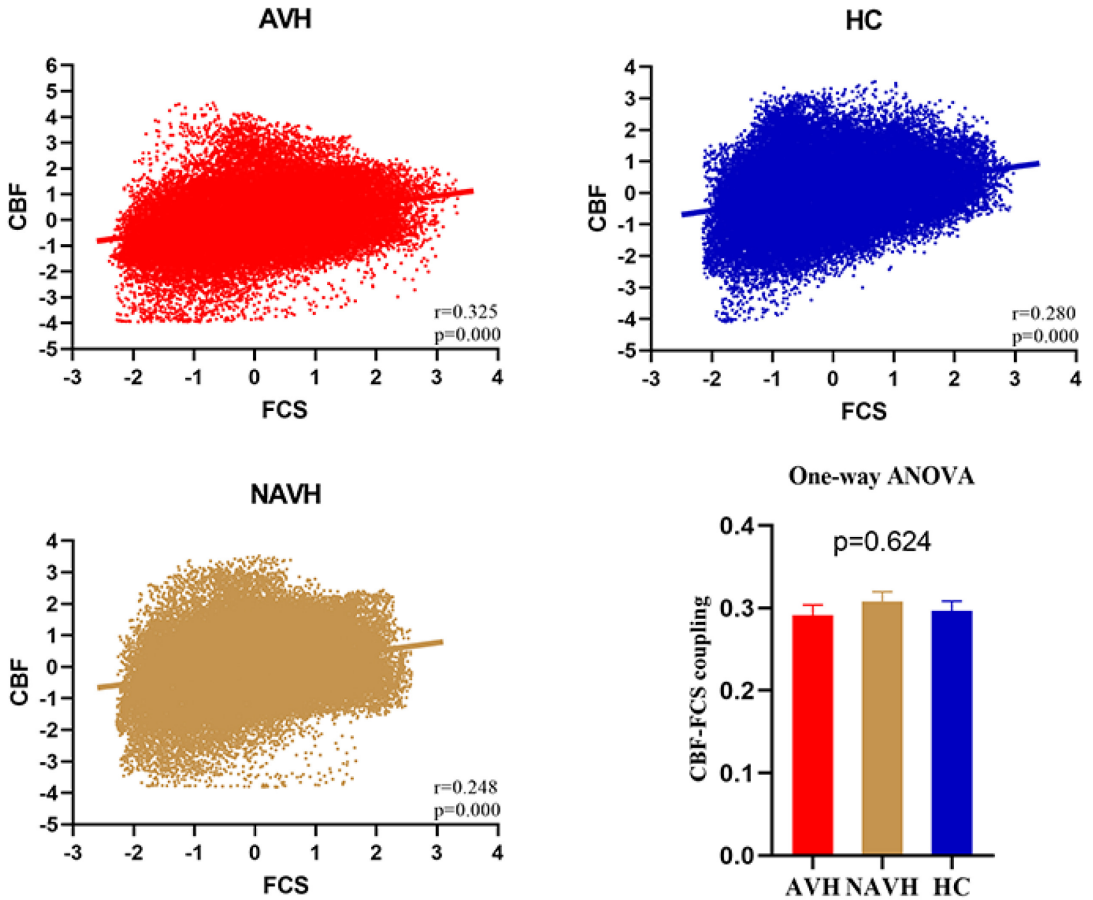


FIGURE 3 | Whole gray matter level cerebral blood flow–functional connectivity strength (CBF–FCS) coupling changes in schizophrenia. Scatter plots of the spatial correlations across voxels between CBF and FCS in an AVH (red), an NAVH patient (yellow), and an HC subject (blue), respectively. The mean whole gray matter level CBF–FCS coupling in AVH patients, NAVH patients, and HC. Although CBF is significantly correlated with FCS in both schizophrenia and control groups, there was no difference between the three groups. CBF, cerebral blood flow; FCS, functional connectivity strength; AVH, schizophrenia patients with AVHs; NAVH, schizophrenia patients without AVHs; HC, healthy control.

AVH group, we found a significance positive correlation between the CBF/FCS ratio in the left STG/MTG and the AHRS (left MTG: Spearman's $\rho = 0.343$, $p = 0.020$; left STG: Spearman's $\rho = 0.303$, $p = 0.041$). However, the significance did not survive the Bonferroni correction ($p < 0.05/51 = 0.001$).

Validation Analysis

We repeated our analysis using correlation thresholds of $r = 0.1$ and $r = 0.3$ to see if the correlation thresholds had any effect on our FCS calculation results. We found that the brain regions with significant CBF/FCS differences at $r = 0.1$ (**Supplementary Figure 2** and **Supplementary Table 5**) were consistent with $r = 0.2$ (**Figure 4** and **Table 2**). With $r = 0.3$ as the threshold (**Supplementary Figure 3** and **Supplementary Table 6**), the vermis VI/VII, left cerebellum crus I/II, left cerebellum VI, and right cerebellum crus II in the AVH and NAVH groups were preserved. The spatial distributions of CBF, FCS, and CBF/FCS ratio at $r = 0.1$ (**Supplementary Figure 4**) and $r = 0.3$ (**Supplementary Figure 5**) were very similar to those at $r = 0.2$ (**Figure 3**). The spatial correlations between CBF and FCS across voxels in anAVH patient (red), an NAVH patient (yellow), and an HC subject (blue) at $r = 0.1$ (**Supplementary Figure 6**) and $r = 0.3$ (**Supplementary Figure 7**) were comparable with $r = 0.2$ (**Figure 3**).

DISCUSSION

In this study, we adopted the method of neurovascular coupling in three groups (AVH patients, NAVH patients, and HC). Compared with HCs, the two schizophrenia groups showed widespread common decreased CBF/FCS ratios in the cerebellum regions, i.e., the left cerebellar crus I/II, left cerebellum VI, and vermis VI/VII. The AVH group exhibited additional alterations, including increased CBF/FCS ratio in the left STG/MTG and decreased CBF/FCS ratio in the right crus II. These converging results confirmed differences between patients with and without AVHs compared with HCs, respectively, suggesting neural mechanisms for hallucinations.

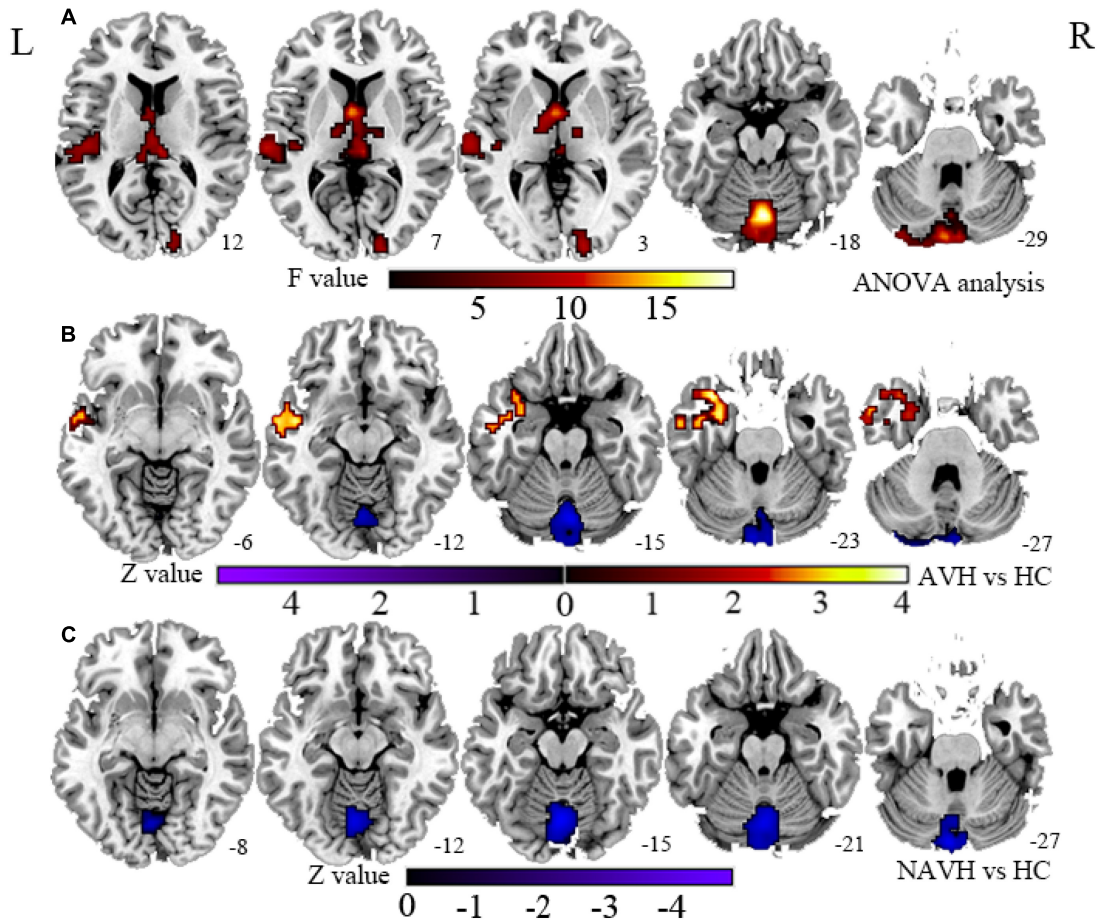


FIGURE 4 | (A) CBF/FCS ratio results for HC, AVH, and NAVH groups at a voxel-level height threshold of $p < 0.005$ (two sided) and cluster size GRF corrected threshold of $p < 0.05$. **(B)** Brain regions with significant CBF/FCS changes in the AVH group. **(C)** Brain regions with significant CBF/FCS changes in the NAVH group. CBF, cerebral blood flow; FCS, functional connectivity strength.

TABLE 2 | Brain regions with significant group differences in cerebral blood flow/functional connectivity strength (CBF/FCS) ratio.

Group differences	Regions	Cluster size (voxels)	Peak MNI coordinate			Peak *t*-values
			X	*Y*	*Z*	
AVH > HC	L_STG	98	−63	−24	6	10.91
	L_MTG	35	−45	−21	15	9.35
AVH/NAVH < HC	Vermis VI	71	3	−72	−18	19.27
	L_cerebellum crus II	57	−3	−87	−27	−15.00
	L_cerebellum VI	51	−9	−75	−18	−9.10
	L_cerebellum crus I	51	−33	−90	−30	−8.19
	Vermis VII	27	3	−72	−24	−10.30
AVH < HC	R_cerebellum crus II	22	6	−87	−27	−10.45

AVH, schizophrenia with AVHs patients; NAVH, schizophrenia without AVHs patients; HC, healthy control; MNI, Montreal Neurological Institute; L_MTG, left middle temporal gyrus; L_STG, left superior temporal gyrus.

Increased FCS may reflect a plastic or compensatory response mechanism to structural abnormalities or an undifferentiated state of brain activity characterized by a disruption of usually separated neural activity in schizophrenia (Fornito and Bullmore, 2015). Previous research has found that in schizophrenia, lower structural integrity is accompanied by better functional connectivity, implying that structural impairments can be mitigated by improved functional integration and functional plasticity (Cocchi et al., 2014; Straathof et al., 2018) or excessive attention to internal stimuli (Narr and Leaver, 2015).

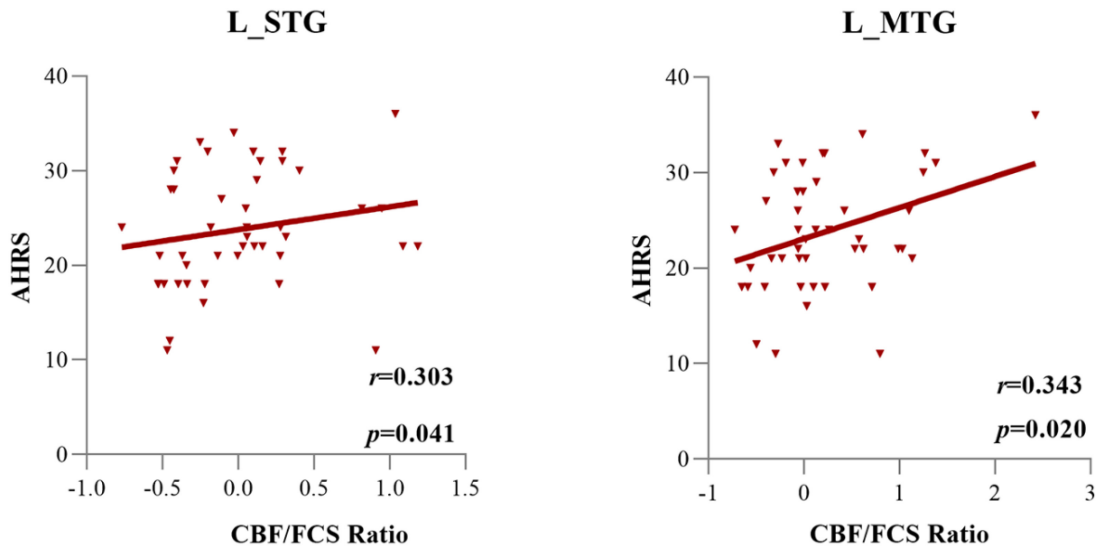


FIGURE 5 | The cerebral blood flow/functional connectivity strength (CBF/FCS) ratio of the left STG and left MTG was positively correlated with the severity of auditory verbal hallucinations, as assessed by AHRS scores in AVH patients.

Cerebellar morphology studies reported reduced GMV in the left cerebellum VI (He H. et al., 2019), left crus I/II (Ding et al., 2019), and cerebellum vermis (Zhong et al., 2018) accompanied by increased FCS between the cerebellum and cortical/subcortical networks or brain regions. Furthermore, structural damage, such as GMV reduction, diminished cortical thickness, and white matter alteration, may impede the exact coordination of inter-regional functional synchronization and reduce information transmission, resulting in decreased FCS in schizophrenia (Kelly et al., 2018; Pan et al., 2019). Previous studies with schizophrenia, for example, have demonstrated a disruption or deviations within the white matter interconnecting left hemisphere language regions (IFG/STG/MTG) accompanied by a low degree of functional connectivity between them (Jeong et al., 2009; de Weijer et al., 2013). The CBF changes in schizophrenia may reflect aberrant neural activity, changes in neurotransmitters, and microvasculature alterations related to neuroinflammation (Kirkpatrick and Miller, 2013; Muller, 2018; Sukumar et al., 2020). Homan et al. has reported that repetitive transcranial magnetic stimulation (TMS) treatment, associated with decreased neural activity, can reduce resting-brain perfusion of the left STG (Homan et al., 2012). AVH patients with persistent regional high perfusion of the left STG indicated that neuronal activity in the left STG was a characteristic biomarker in AVHs (Homan et al., 2013). Groen et al. (2011) showed spontaneous reactivation of memory traces together with increased CBF in the left MTG.

Based on the previous research, any single or a combination of neurovascular unit component impairments and abnormalities (significantly, the astrocytes), neural activity anomaly, or structural impairments can lead to abnormal neurovascular coupling (Howarth, 2014; Stobart and Anderson, 2013), precise coordination and integration of inter-regional functional synchronization is impaired (Pan et al., 2019), and information transmission is decreased eventually (Kelly et al., 2018). Any cause of abnormality in CBF or FCS might lead to a change in the ratio of CBF/FCS. Absolutely, many other factors could also result in neurovascular decoupling, but it is not the focus of our article (Girouard and Iadecola, 2006). The across-voxel correlation between CBF and FCS between the three groups showed no statistical difference on the whole gray matter level in our results. We speculate that the result may be due to the mutual cancelation of many influencing factors as above precisely. However, the CBF/FCS ratio, another coupling index, could show similar discrepancy. The CBF/FCS ratio directly indicated neurovascular coupling of a given voxel or local region and maintained balance in healthy brains (Liang et al., 2013; Zhu et al., 2017). The CBF/FCS ratio equilibrium can be altered in schizophrenia and may differ between AVH and NAVH individuals, with CBF and FCS altering in different directions.

Compared with HC, in the NAVH patients, the left crus I/II, vermis VI/VII, and left cerebellum VI showed increased FCS and normal CBF, suggesting that the decreased CBF/FCS ratio in these regions is predominantly driven by the increased FCS. Compared with HC, in the AVH patients, the decreased CBF/FCS ratio in the left crus I/II, vermis VI/VII, left cerebellum VI, and right crus II was predominantly driven by both CBF and FCS. Under these circumstances, the CBF/FCS ratio decrease was based on ROI analyses in the AVH group (**Supplementary Figure 8**). Interactions between the posterior lateral cerebellum (e.g., crus I/II) and the prefrontal cortex underlie the engagement of the cerebrum in higher-level functions like cognitive control and processing (Buckner et al., 2011; Carass et al., 2018). Many studies on the cerebellum structure have proven that the reduction in cerebellar gray matter in the left crus I and left crus II was related to the decrease in cognitive function (Kuhn et al., 2012; Kim et al., 2018; Moberget et al., 2018; Uwisengeyimana et al., 2020).

In a functional imaging study, the cerebellar VI, which is coupled to subcortical limbic regions and covered in the salience network of the intrinsic connection networks in the cerebral cortex, prioritizes processing of emotionally significant stimuli in a context-dependent way (Seeley et al., 2007; Habas et al., 2009). Moreover, anatomically, the cerebellar vermis VI and VII are located in the posterior section of the cerebellar vermis (Carass et al., 2018), The vermis situated in the midline of the cerebellum is equivalent to the limbic cerebellum, connected to the thalamus and the limbic system (Ichimiya et al., 2001), and plays a role in higher-level functions, such as affection or emotional regulation and cognitive processing (Schmahmann et al., 2007; Yucel et al., 2013; Womer et al., 2016). Therefore, cerebellar VI may have a role in determining the valence of important emotional cues and choosing suitable behavioral responses (Habas et al., 2009). In other words, these regions, which are associated with one of the cerebellum's major tasks in emotional processing, were activated by computing cues, resulting in diminished consolidation of emotional signals and delivery to the cerebral cortex (Adamaszek et al., 2017). Cerebellum VI is the part of the upper cerebellum that is functionally connected to sensory–motor-related areas. It is also projected to the sensorimotor brain network and is considered to be an important component of motor control and coordination (Grodd et al., 2001; Buckner et al., 2011; Ding et al., 2019). In cognitive processing, the sensorimotor network is involved in perceiving face expressions, emotions, and personal desires (Watanuki et al., 2016). Small cerebellum VI clusters were found in salience and sensorimotor networks. The overlap suggests an intracerebellar connection between them that might relate to limbic control of the motor system (Habas et al., 2009). Therefore, the motor learning of the cerebellum in the forward model may straightly exemplify the intimate relationship between the cerebellum's motor and cognitive domains (Jacobi et al., 2021).

Pastor et al. discovered that frequency-specific coupling between STG and crus II in the auditory cortical–cerebellar–thalamic loop regulates auditory cortex oscillatory activity in schizophrenia AVHs (Pastor et al., 2008). Furthermore, Coull and Nobre (2008) have demonstrated that cerebellar areas are almost always engaged by explicit time prediction, depending on the individual task setting. Previous research has suggested that the left crus I and bilateral crus II may have a role in AVHs by altering sensory feedback and, as a result, unpredictable prediction, as shown by the forward model (Runnqvist et al., 2016; Pinheiro et al., 2020b). In addition, the posterior vermis has shown involvement in a cerebello–thalamo–cortical circuit for error-related cognitive control in healthy adults (Womer et al., 2016) and auditory prediction in AVHs (Pinheiro et al., 2020a). Given the cerebellum's role in the forward model and its function as a comparator, the cerebellum can compare actual input with previous stimuli and test for discrepancies. If a discrepancy–error signal is discovered, cerebellar activity increases, and a vast portion of the cerebral cortex is alerted by increasing excitability (Marco Molinari et al., 2008). Our findings of decreased CBF/FCS ratio in cerebellum regions could implicate that cerebellum cognitive dysmetria is linked to gray matter structural damage or disruption of the cerebello-cortical circuit, resulting in patients with AVHs having difficulty synchronizing and integrating neuronal computations and processing to generate ordered and meaningful motor and cognitive activities (Nancy et al., 1998).

Compared with HC, the left STG/MTG showed a significantly increased CBF/FCS ratio in AVH patients. Further analysis based on ROI showed that these areas of AVH patients have higher CBF and lower FCS than HC (**Supplementary Figure 8**). Many structural (Onitsuka et al., 2004; Cui et al., 2018; Curtis et al., 2021), functional neuroimaging (Allen et al., 2008), and circuit studies (Benetti et al., 2015; Huang et al., 2019; Ren et al., 2021) have indicated that the left STG, particularly the primary and association auditory cortex, and the left MTG, play a key role in the etiology of AVHs. The STG on the left mainly deals with the perception of "speech," that is, understanding the phonetic and semantic features of the speech content (Modinos et al., 2013) and auditory feedback processing originating from the cerebellum (Christoffels et al., 2007). As we all know, the left MTG is known to be especially vital for the semantic processing of speech and mapping sound to meaning (Clos et al., 2014; Liu et al., 2016). It may be related to the internal attribution of the event. In this case, the self-stored semantic memory is considered an active and intentional agent (Blackwood et al., 2000). Reduced left STG/MTG gray matter volumes are linked to higher AVH severity (Allen et al., 2008). A recent research suggests that AVHs are caused by abnormally high resting-state activity in the auditory cortex (Kuehn and Gallinat, 2012; Hugdahl and Sommer, 2018). AVHs are caused by abnormal height or abnormal static activity in the left STG/MTG, which causes spontaneous internal signals to be misunderstood as external (Cho and Wu, 2013; Alderson-Day et al., 2015). Internal speech being mistakenly attributed to external or non-self sources could be the result of atypical self or reality monitoring, which is caused by the failure of the internal forward model (Cho and Wu, 2013; Moseley et al., 2013). Evidence from neuroimaging suggests that monitoring of one's own speech, overt or covert, is related to activity in auditory cortical regions such as the left STG (Allen et al., 2007; Moseley et al., 2013). In addition, a greater CBF/FCS ratio in the left STG/MTG was positively linked with hallucination severity measured by AHRS in patients with AVHs, probably reflecting these regions based on trait study engaging in more rapid AVH processing. So, our findings of increased CBF/FCS in the left STG/MTG may implicate that the spontaneous auditory activation of auditory representation information emerged, the coordination and integration of that local activity across brain regions were impaired, and the event to another person was misattributed eventually.

In summary, we use the combination of BOLD and ASL technology to reveal the disordered coupling of resting CBF and functional connectivity between AVH and NAVH patients. In addition, our results revealed that schizophrenic patients had widespread deficits in both low-level sensorimotor and higher-order cognitive networks of the cerebellum, which suggest potential impairment affection, emotion, and cognitive functions. Specifically, our findings may possibly implicate that the typical symptom of AVHs in schizophrenia might arise from the failure of a forward model originating from functional synchronization abnormality among networks in cerebellar

regions, which in turn might contribute to increase the activity of the cerebellum and alert the left STG/MTG by enhancing its excitability and, eventually, not recognizing that the experience is internally produced.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author/s.

ETHICS STATEMENT

This study has been reviewed and approved by the Ethics Committee of The First Affiliated Hospital of Zhengzhou University. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

YW, JCheng, and SH designed the study and wrote the protocol. JChen performed the data processing and statistical analyses and wrote the first draft of the manuscript. All authors contributed to and have approved the final manuscript.

FUNDING

This work was supported by the Natural Science Foundation of China (81601467 and 81871327) and the Medical Science and Technology Research Project of Henan Province (201701011).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.821078/full#supplementary-material

REFERENCES

Adamaszek, M., D'Agata, F., Ferrucci, R., Habas, C., Keulen, S., Kirkby, K. C., et al. (2017). Consensus paper: cerebellum and emotion. *Cerebellum* 16, 552–576. doi: 10.1007/s12311-016-0815-8

Alderson-Day, B., McCarthy-Jones, S., and Fernyhough, C. (2015). Hearing voices in the resting brain: a review of intrinsic functional connectivity research on auditory verbal hallucinations. *Neurosci. Biobehav. Rev.* 55, 78–87. doi: 10.1016/j.neubiorev.2015.04.016

Allen, P., Amaro, E., Fu, C. H. Y., Williams, S. C. R., Brammer, M. J., Johns, L. C., et al. (2007). Neural correlates of the misattribution of speech in schizophrenia. *Br. J. Psychiatry* 190, 162–169. doi: 10.1192/bjp.bp.106.025700

Allen, P., Laroi, F., McGuire, P., and Aleman, A. (2008). The hallucinating brain: a review of structural and functional neuroimaging studies of hallucinations. *Neurosci. Biobehav. Rev.* 32, 175–191. doi: 10.1016/j.neubiorev.2007.07.012

Andreasen, N. C., and Pierson, R. (2008). The role of the cerebellum in schizophrenia. *Biol. Psychiatry* 64, 81–88. doi: 10.1016/j.biopsych.2008.01.003

Basavaraju, R., Ithal, D., Thanki, M., Hr, A., Thirthalli, J., Pascual-Leone, L., et al. (2019). Intermittent theta burst stimulation of cerebellar vermis in schizophrenia: impact on negative symptoms and brain connectivity. *Schizophr. Bull* 45, S234.

Benetti, S., Pettersson-Yeo, W., Allen, P., Catani, M., Williams, S., Barsaglini, A., et al. (2015). Auditory verbal hallucinations and brain dysconnectivity in the perisylvian language network: a multimodal investigation. *Schizophr. Bull* 41, 192–200. doi: 10.1093/schbul/sbt172

Blackwood, N. J., Howard, R., JffYTCHE, D. H., Simmons, A., Bentall, R. P., and Murray, R. M. (2000). Imaging attentional and attributional bias: an fMRI approach to the paranoid delusion. *Psychol. Med.* 30, 873–883. doi: 10.1017/s0033291799002421

Brooks, J. X., Carriot, J., and Cullen, K. E. (2015). Learning to expect the unexpected: rapid updating in primate cerebellum during voluntary self-motion. *Nat. Neurosci.* 18, 1310–1317. doi: 10.1038/nn.4077

Buckner, R. L., Krienen, F. M., Castellanos, A., Diaz, J. C., and Yeo, B. T. (2011). The organization of the human cerebellum estimated by intrinsic functional connectivity. *J. Neurophysiol.* 106, 2322–2345. doi: 10.1152/jn.00339.2011

Cao, H., Chen, O. Y., Chung, Y., Forsyth, J. K., McEwen, S. C., Gee, D. G., et al. (2018). Cerebello-thalamo-cortical hyperconnectivity as a state-independent functional neural signature for psychosis prediction and characterization. *Nat. Commun.* 9:3836. doi: 10.1038/s41467-018-06350-7

Carass, A., Cuzzocreo, J. L., Han, S., Hernandez-Castillo, C. R., Rasser, P. E., Ganz, M., et al. (2018). Comparing fully automated state-of-the-art cerebellum parcellation from magnetic resonance images. *Neuroimage* 183, 150–172. doi: 10.1016/j.neuroimage.2018.08.003

Cho, R., and Wu, W. (2013). Mechanisms of auditory verbal hallucination in schizophrenia. *Front. Psychiatry* 4:155. doi: 10.3389/fpsyt.2013.00155

Christoffels, I. K., Formisano, E., and Schiller, N. O. (2007). Neural correlates of verbal feedback processing: an fMRI study employing overt speech. *Hum. Brain Mapp.* 28, 868–879. doi: 10.1002/hbm.20315

Ciullo, V., Piras, F., Vecchio, D., Banaj, N., Coull, J. T., and Spalletta, G. (2018). Predictive timing disturbance is a precise marker of schizophrenia. *Schizophr. Res. Cogn.* 12, 42–49. doi: 10.1016/j.scog.2018.04.001

Clos, M., Diederen, K. M., Meijering, A. L., Sommer, I. E., and Eickhoff, S. B. (2014). Aberrant connectivity of areas for decoding degraded speech in patients with auditory verbal hallucinations. *Brain Struct. Funct.* 219, 581–594. doi: 10.1007/s00429-013-0519-5

Cocchi, L., Harding, I. H., Lord, A., Pantelis, C., Yucel, M., and Zalesky, A. (2014). Disruption of structure-function coupling in the schizophrenia connectome. *Neuroimage Clin.* 4, 779–787. doi: 10.1016/j.nicl.2014.05.004

Coull, J., and Nobre, A. (2008). Dissociating explicit timing from temporal expectation with fMRI. *Curr. Opin. Neurobiol.* 18, 137–144. doi: 10.1016/j.conb.2008.07.011

Cui, Y., Liu, B., Song, M., Lipnicki, D. M., Li, J., Xie, S., et al. (2018). Auditory verbal hallucinations are related to cortical thinning in the left middle temporal gyrus of patients with schizophrenia. *Psychological. Med.* 48, 115–122. doi: 10.1017/S0033291717001520

Crow,T. J., Frith, C. D., Johnstone, E. C., and Owen, D. G. C. (1980). Schizophrenia and cerebral atrophy. *Lancet* 24, 1129–1130.

Curtis, M. T., Coffman, B. A., and Salisbury, D. F. (2021). Parahippocampal area three gray matter is reduced in first-episode schizophrenia spectrum: discovery and replication samples. *Hum. Brain Mapp.* 42, 724–736. doi: 10.1002/hbm.25256

de Weijer, A. D., Neggers, S. F., Diederen, K. M., Mandl, R. C., Kahn, R. S., Hulshoff Pol, H. E., et al. (2013). Aberrations in the arcuate fasciculus are associated with auditory verbal hallucinations in psychotic and in non-psychotic individuals. *Hum. Brain Mapp.* 34, 626–634. doi: 10.1002/hbm.21463

Ding, Y., Ou, Y., Pan, P., Shan, X., Chen, J., Liu, F., et al. (2019). Cerebellar structural and functional abnormalities in first-episode and drug-naive patients with schizophrenia: a meta-analysis. *Psychiatry Res. Neuroimaging* 283, 24–33. doi: 10.1016/j.pscychresns.2018.11.009

Drzezga, A., Becker, J. A., Van Dijk, K. R., Sreenivasan, A., Talukdar, T., Sullivan, C., et al. (2011). Neuronal dysfunction and disconnection of cortical hubs in non-demented subjects with elevated amyloid burden. *Brain* 134, 1635–1646. doi: 10.1093/brain/awr066

Enager, P., Gold, L., and Lauritzen, M. (2004). Impaired neurovascular coupling by transhemispheric diaschisis in rat cerebral cortex. *J. Cereb. Blood Flow Metab.* 24, 713–719. doi: 10.1097/01.WCB.0000121233.63924.41

Ferri, J., Ford, J. M., Roach, B. J., Turner, J. A., van Erp, T. G., Voyvodic, J., et al. (2018). Resting-state thalamic dysconnectivity in schizophrenia and relationships with symptoms. *Psychological. Med.* 48, 2492–2499. doi: 10.1017/S003329171800003X

Fornito, A., and Bullmore, E. T. (2015). Reconciling abnormalities of brain network structure and function in schizophrenia. *Curr. Opin. Neurobiol.* 30, 44–50. doi: 10.1016/j.conb.2014.08.006

Fox, M. D., and Raichle, M. E. (2007). Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. *Nat. Rev. Neurosci.* 8, 700–711. doi: 10.1038/nrn2201

Girouard, H., and Iadecola, C. (2006). Neurovascular coupling in the normal brain and in hypertension, stroke, and alzheimer disease. *J. Appl. Physiol.* 100, 328–335. doi: 10.1152/japplphysiol.00966.2005

Grodd, W., Hülsmann, E., Lotze, M., Wildgruber, D., and Erb, M. (2001). Sensorimotor mapping of the human cerebellum: fMRI evidence of somatotopic organization. *Hum. Brain Mapp.* 13, 55–73. doi: 10.1002/hbm.1025

Groen, G., Sokolov, A. N., Jonas, C., Roebling, R., and Spitzer, M. (2011). Increased resting-state perfusion after repeated encoding is related to later retrieval of declarative associative memories. *PLoS One* 6:e19985. doi: 10.1371/journal.pone.0019985

Habas, C., Kamdar, N., Nguyen, D., Prater, K., Beckmann, C. F., Menon, V., et al. (2009). Distinct cerebellar contributions to intrinsic connectivity networks. *J. Neurosci.* 29, 8586–8594. doi: 10.1523/JNEUROSCI.1868-09.2009

He, H., Luo, C., Luo, Y., Duan, M., Yi, Q., Biswal, B. B., et al. (2019). Reduction in gray matter of cerebellum in schizophrenia and its influence on static and dynamic connectivity. *Hum. Brain. Mapp.* 40, 517–528. doi: 10.1002/hbm.24391

He, Z., Sheng, W., Lu, F., Long, Z., Han, S., Pang, Y., et al. (2019). Altered resting-state cerebral blood flow and functional connectivity of striatum in bipolar disorder and major depressive disorder. *Prog. Neuropsychopharmacol. Biol. Psychiatry* 90, 177–185. doi: 10.1016/j.pnpbp.2018.11.009

Hjelmervik, H., Craven, A. R., Sinceviciute, I., Johnsen, E., Kompus, K., Bless, J. J., et al. (2020). Intra-regional glu-GABA vs inter-regional glu-glu imbalance: a 1H-MRS study of the neurochemistry of auditory verbal hallucinations in schizophrenia. *Schizophr. Bull.* 46, 633–642. doi: 10.1093/schbul/sbz099

Homan, P., Kindler, J., Hauf, M., Hubl, D., and Dierks, T. (2012). Cerebral blood flow identifies responders to transcranial magnetic stimulation in auditory verbal hallucinations. *Transl. Psychiatry* 2:e189. doi: 10.1038/tp.2012.114

Homan, P., Kindler, J., Hauf, M., Walther, S., Hubl, D., and Dierks, T. (2013). Repeated measurements of cerebral blood flow in the left superior temporal gyrus reveal tonic hyperactivity in patients with auditory verbal hallucinations: a possible trait marker. *Front. Hum. Neurosci.* 7:304. doi: 10.3389/fnhum.2013.00304

Howarth, C. (2014). The contribution of astrocytes to the regulation of cerebral blood flow. *Front. Neurosci.* 8:103. doi: 10.3389/fnins.2014.00103

Huang, J., Zhuo, C., Xu, Y., and Lin, X. (2019). Auditory verbal hallucination and the auditory network: From molecules to connectivity. *Neuroscience* 410, 59–67. doi: 10.1016/j.neuroscience.2019.04.051

Hugdahl, K., and Sommer, I. E. (2018). Auditory verbal hallucinations in schizophrenia from a levels of explanation perspective. *Schizophr. Bull.* 44, 234–241. doi: 10.1093/schbul/sbx142

Ichimiya, T., Okubo, Y., Suhara, T., and Sudo, Y. (2001). Reduced volume of the cerebellar vermis in neuroleptic-naive schizophrenia. *Soc. Biol. Psychiatry* 49, 20–27. doi: 10.1016/s0006-3223(00)01081-7

Jacobi, H., Faber, J., Timmann, D., and Klockgether, T. (2021). Update cerebellum and cognition. *J. Neurol.* 268, 3921–3925. doi: 10.1007/s00415-021-10486-w

Jardri, R., Pouchet, A., Pins, D., and Thomas, P. (2011). Cortical activations during auditory verbal hallucinations in schizophrenia: a coordinate-based meta-analysis. *Am. J. Psychiatry* 168, 73–81. doi: 10.1176/appi.ajp.2010.09101522

Jeong, B., Wible, C. G., Hashimoto, R.-I., and Kubicki, M. (2009). Functional and anatomical connectivity abnormalities in left inferior frontal gyrus in schizophrenia. *Hum. Brain Mapp.* 30, 4138–4151. doi: 10.1002/hbm.20835

Jing, R., Huang, J., Jiang, D., Tian, H., Li, J., and Zhuo, C. (2018). Distinct pattern of cerebral blood flow alterations specific to schizophrenics experiencing auditory verbal hallucinations with and without insight: a pilot study. *Oncotarget* 9, 6763–6770. doi: 10.18632/oncotarget.23631

Kelly, S., Jahanshad, N., Zalesky, A., Kochunov, P., Agartz, I., Alloza, C., et al. (2018). Widespread white matter microstructural differences in schizophrenia across 4322 individuals: results from the ENIGMA Schizophrenia DTI Working Group. *Mol. Psychiatry* 23, 1261–1269. doi: 10.1038/mp.2017.170

Kim, T., Lee, K. H., Oh, H., Lee, T. Y., Cho, K. I. K., Lee, J., et al. (2018). Cerebellar structural abnormalities associated with cognitive function in patients with first-episode psychosis. *Front. Psychiatry* 9:286. doi: 10.3389/fpsyt.2018.00286

Kirkpatrick, B., and Miller, B. J. (2013). Inflammation and schizophrenia. *Schizophr. Bull.* 39, 1174–1179. doi: 10.1093/schbul/sbt141

Kuehn, S., and Gallinat, J. (2012). Quantitative meta-analysis on state and trait aspects of auditory verbal hallucinations in schizophrenia. *Schizophr. Bull.* 38, 779–786. doi: 10.1093/schbul/sbq152

Kuhn, S., Romanowski, A., Schubert, F., and Gallinat, J. (2012). Reduction of cerebellar grey matter in Crus I and II in schizophrenia. *Brain Struct. Funct.* 217, 523–529. doi: 10.1007/s00429-011-0365-2

Li, Q., Dong, C., Liu, T., Chen, X., Perry, A., Jiang, J., et al. (2020). Longitudinal changes in whole-brain functional connectivity strength patterns and the relationship with the global cognitive decline in older adults. *Front. Aging Neurosci.* 12:71. doi: 10.3389/fnagi.2020.00071

Liang, X., Zou, Q., He, Y., and Yang, Y. (2013). Coupling of functional connectivity and regional cerebral blood flow reveals a physiological basis for network hubs of the human brain. *Proc. Natl. Acad. Sci. U.S.A* 110, 1929–1934. doi: 10.1073/pnas.1214900110

Liu, C., Xue, Z., Palaniyappan, L., Zhou, L., Liu, H., Qi, C., et al. (2016). Abnormally increased and incoherent resting-state activity is shared between patients with schizophrenia and their unaffected siblings. *Schizophr. Res.* 171, 158–165. doi: 10.1016/j.schres.2016.01.022

Liu, F., Zhu, C., Wang, Y., Guo, W., Li, M., Wang, W., et al. (2015). Disrupted cortical hubs in functional brain networks in social anxiety disorder. *Clin. Neurophysiol.* 126, 1711–1716. doi: 10.1016/j.clinph.2014.11.014

Lui, S., Li, T., Deng, W., Jiang, L., Wu, Q., Tang, H., et al. (2010). Short-term effects of antipsychotic treatment on cerebral function in drug-naive first-episode schizophrenia revealed by "resting state" functional magnetic resonance imaging. *Arch. Gen. Psychiatry* 67, 783–792. doi: 10.1001/archgenpsychiatry.2010.84

Marco Molinari, R., Chiricozzi, F. R., Clausi, S., Tedesco, A. M., De Lisa, M., and Leggio, M. G. (2008). Cerebellum and detection of sequences, from perception to cognition. *Cerebellum* 7, 611–615. doi: 10.1007/s12311-008-0060-x

Moberget, T., Doan, N. T., Alnaes, D., Kaufmann, T., Cordova-Palomera, A., Lagerberg, T. V., et al. (2018). Cerebellar volume and cerebellocerebral structural covariance in schizophrenia: a multisite mega-analysis of 983 patients and 1349 healthy controls. *Mol. Psychiatry* 23, 1512–1520. doi: 10.1038/mp.2017.106

Moberget, T., and Ivry, R. B. (2019). Prediction, Psychosis, and the Cerebellum. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* 4, 820–831. doi: 10.1016/j.bpsc.2019.06.001

Modinos, G., Costafreda, S. G., van Tol, M.-J., McGuire, P. K., Aleman, A., and Allen, P. (2013). Neuroanatomy of auditory verbal hallucinations in schizophrenia: a quantitative meta-analysis of voxel-based morphometry studies. *Cortex* 49, 1046–1055. doi: 10.1016/j.cortex.2012.01.009

Moseley, P., Fernyhough, C., and Ellison, A. (2013). Auditory verbal hallucinations as atypical inner speech monitoring, and the potential of neurostimulation as a treatment option. *Neurosci. Biobehav. Rev.* 37, 2794–2805. doi: 10.1016/j.neubiorev.2013.10.001

Muller, N. (2018). Inflammation in schizophrenia: pathogenetic aspects and therapeutic considerations. *Schizophr. Bull.* 44, 973–982. doi: 10.1093/schbul/sby024

Nancy C., Andreasen, N. C., Paradiso, S., and O'Leary, D. S. (1998). Cognitive dysmetria" as an integrative theory of schizophrenia: a dysfunction in corticalsubcortical-cerebellar circuitry? *Schizophr. Bull.* 24, 203–218. doi: 10.1093/oxfordjournals.schbul.a033321

Nair, A., Keown, C. L., Datko, M., Shih, P., Keehn, B., and Muller, R. A. (2014). Impact of methodological variables on functional connectivity findings in autism spectrum disorders. *Hum. Brain Mapp.* 35, 4035–4048. doi: 10.1002/hbm.22456

Narr, K. L., and Leaver, A. M. (2015). Connectome and schizophrenia. *Curr. Opin. Psychiatry* 28, 229–235. doi: 10.1097/yco.0000000000000157

Onitsuka, T., Shenton, M. E., Salisbury, D. F., Dickey, C. C., Kasai, K., Toner, S. K., et al. (2004). Middle and inferior temporal gyrus gray matter volume abnormalities in chronic schizophrenia: an MRI study. *Am. J. Psychiatry* 161, 1603–1611. doi: 10.1176/appi.ajp.161.9.1603

Pan, Y., Das, T., Khan, A., Dempster, K., Mackinley, M., and Palaniyappan, L. (2019). T17 Acute conceptual disorganization in untreated firstepisode psychosis: a 7t dti study of cingulum trac. *Schizophr. Bull.* 45, S209–S210.

Pastor, M. A., Vidaurre, C., Fernandez-Seara, M. A., Villanueva, A., and Friston, K. J. (2008). Frequency-specific coupling in the cortico-cerebellar auditory system. *J Neurophysiol.* 100, 1699–1705. doi: 10.1152/jn.01156.2007

Pinheiro, A., Bouix, S., Makris, N., Schwartze, M., Shenton, M., and Kotz, S. (2020a). T163 structural and connectivity changes in the cerebellum contribute to experiencing auditory verbal hallucinations. *Schizophr. Bull.* 46, S293.

Pinheiro, A. P., Schwartze, M., and Kotz, S. A. (2020b). Cerebellar circuitry and auditory verbal hallucinations: an integrative synthesis and perspective. *Neurosci. Biobehav. Rev.* 118, 485–503. doi: 10.1016/j.neubiorev.2020.08.004

Ren, J., Hubbard, C. S., Ahveninen, J., Cui, W., Li, M., Peng, X., et al. (2021). Dissociable auditory cortico-cerebellar pathways in the human brain estimated by intrinsic functional connectivity. *Cereb. Cortex* 31, 2898–2912. doi: 10.1093/cercor/bhaa398

Runnqvist, E., Bonnard, M., Gauvin, H. S., Attarian, S., Trebuchon, A., Hartsuiker, R. J., et al. (2016). Internal modeling of upcoming speech: a causal role of the right posterior cerebellum in non-motor aspects of language production. *Cortex* 81, 203–214. doi: 10.1016/j.cortex.2016.05.008

Schmahmann, J. D., Weilburg, J. B., and Sherman, J. C. (2007). The neuropsychiatry of the cerebellum - insights from the clinic. *Cerebellum* 6, 254–267. doi: 10.1080/14734220701490995

Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., et al. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *J. Neurosci.* 27, 2349–2356. doi: 10.1523/JNEUROSCI.5587-06.2007

Sokolov, A. A., Miall, R. C., and Ivry, R. B. (2017). The cerebellum: adaptive prediction for movement and cognition. *Trends Cogn. Sci.* 21, 313–332. doi: 10.1016/j.tics.2017.02.005

Stobart, J. L., and Anderson, C. M. (2013). Multifunctional role of astrocytes as gatekeepers of neuronal energy supply. *Front. Cell Neurosci.* 7:38. doi: 10.3389/fncel.2013.00038

Straathof, M., Sinke, M. R. T., Dijkhuizen, R. M., and Otte, W. M. (2018). A systematic review on the quantitative relationship between structural and functional network connectivity strength in mammalian brains. *J. Cereb. Blood Flow Metab.* 39, 189–209. doi: 10.1177/0271678X18809547

Martha, L., Streng, M. L., Popa, L. S., and Ebner, T. J. (2018). Modulation of sensory prediction error in Purkinje cells during visual feedback manipulations. *Nat. Commun.* 9, 1099. doi: 10.1038/s41467-018-03541-0

Sukumar, N., Sabesan, P., Anazodo, U., and Palaniyappan, L. (2020). Neurovascular uncoupling in schizophrenia: a bimodal meta-analysis of brain perfusion and glucose metabolism. *Front. Psychiatry* 11:754. doi: 10.3389/fpsyt.2020.00754

Tourville, J. A., Reilly, K. J., and Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *Neuroimage* 39, 1429–1443. doi: 10.1016/j.neuroimage.2007.09.054

Uwisengeyimana, J. D., Nguchu, B. A., Wang, Y., Zhang, D., Liu, Y., Qiu, B., et al. (2020). Cognitive function and cerebellar morphometric changes relate to abnormal intra-cerebellar and cerebro-cerebellum functional connectivity in old adults. *Exp. Gerontol.* 140, 111060. doi: 10.1016/j.exger.2020.111060

Vahia, V. N. (2013). "Cautionary Statement for Forensic Use of DSM-5," in *Diagnostic and Statistical Manual of Mental Disorders*, ed. American Psychiatric Association (Virginia, US: American Psychiatric Association).

Vaishnavi, S. N., Vlassenko, A. G., Rundle, M. M., Snyder, A. Z., Mintun, M. A., and Raichle, M. E. (2010). Regional aerobic glycolysis in the human brain. *Proc. Natl. Acad. Sci. U.S.A* 107, 17757–17762. doi: 10.1073/pnas.1010459107

van Lutterveld, R., Diederen, K. M., Otte, W. M., and Sommer, I. E. (2014). Network analysis of auditory hallucinations in nonpsychotic individuals. *Hum. Brain Mapp.* 35, 1436–1445. doi: 10.1002/hbm.22264

Wang, Q., Qu, X., Chen, W., Wang, H., Huang, C., Li, T., et al. (2021). Altered coupling of cerebral blood flow and functional connectivity strength in visual and higher order cognitive cortices in primary open angle glaucoma. *J. Cereb. Blood Flow Metab.* 41, 901–913. doi: 10.1177/0271678X20935274

Wang, S., Zhan, Y., Zhang, Y., Lv, L., Wu, R., Zhao, J., et al. (2017). Abnormal functional connectivity strength in patients with adolescent-onset schizophrenia: a resting-state fMRI study. *Eur. Child. Adolesc. Psychiatry* 26, 839–845. doi: 10.1007/s00787-017-0958-2

Wang, Y., Jiang, Y., Su, W., Xu, L., Wei, Y., Tang, Y., et al. (2021). Temporal dynamics in degree centrality of brain functional connectome in first-episode schizophrenia with different short-term treatment responses: a longitudinal study. *Neuropsychiatr. Dis. Treat* 17, 1505–1516. doi: 10.2147/NDT.S305117

Warren, H., Kempinsky, M. D., and Louis, S. T. (1958). Experimental study of distant effects of acute focal brain injury a study of diaschisis. *AMA Arch. NeurPsych.* 79, 376–389. doi: 10.1001/archneurpsyc.1958.02340040020002

Watanuki, T., Matsuo, K., Egashira, K., Nakashima, M., Harada, K., Nakano, M., et al. (2016). Precentral and inferior prefrontal hypoactivation during facial emotion recognition in patients with schizophrenia: A functional near-infrared spectroscopy study. *Schizophr. Res.* 170, 109–114. doi: 10.1016/j.schres.2015.11.012

Womer, F. Y., Tang, Y., Harms, M. P., Bai, C., Chang, M., Jiang, X., et al. (2016). Sexual dimorphism of the cerebellar vermis in schizophrenia. *Schizophr. Res.* 176, 164–170. doi: 10.1016/j.schres.2016.06.028

Yucel, K., Nazarov, A., Taylor, V. H., Macdonald, K., Hall, G. B., and Macqueen, G. M. (2013). Cerebellar vermis volume in major depressive disorder. *Brain Struct. Funct.* 218, 851–858. doi: 10.1007/s00429-012-0433-2

Zhong, C., Liu, R., Luo, C., Jiang, S., Dong, L., Peng, R., et al. (2018). Altered structural and functional connectivity of juvenile myoclonic epilepsy: an fMRI study. *Neural. Plast* 2018, 7392187. doi: 10.1155/2018/7392187

Zhu, J., Zhuo, C., Xu, L., Liu, F., Qin, W., and Yu, C. (2017). Altered coupling between resting-state cerebral blood flow and functional connectivity in schizophrenia. *Schizophr. Bull.* 43, 1363–1374. doi: 10.1093/schbul/sbx051

Zhuo, C., Zhu, J., Qin, W., Qu, H., Ma, X., and Yu, C. (2017). Cerebral blood flow alterations specific to auditory verbal hallucinations in schizophrenia. *Br. J. Psychiatry* 210, 209–215. doi: 10.1192/bjp.bp.115.174961

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Neuroscience

ORIGINAL RESEARCH
published: 03 June 2022
doi: 10.3389/fnins.2022.830703

Circadian Sensitivity of Noise Trauma-Induced Hearing Loss and Tinnitus in Mongolian Gerbils

*Jannik Grimm, Holger Schulze and Konstantin Tziridis**

Experimental Otolaryngology, University of Erlangen-Nuremberg, Erlangen, Germany

Noise-induced hearing loss (HL) has a circadian component: In nocturnal mice, hearing thresholds (HT) have a significantly stronger effect to acoustic trauma when induced during the night compared to rather mild effects on hearing when induced during daytime. Here, we investigate whether such effects are also present in diurnal Mongolian gerbils and determined whether trauma-induced HL correlated with the development of a tinnitus percept in these animals. In particular, we investigated the effects of acoustic trauma (2 kHz, 115 dB SPL, 75 min) on HT and tinnitus development in 34 male gerbils exposed either at 9 AM, 1 PM, 5 PM, or 12 PM. HT was measured by acoustic brainstem response audiometry at defined times 1 day before and 1 week after the trauma. Possible tinnitus percepts were assessed behaviorally by the gap prepulse inhibition of the acoustic startle response at defined times 1 day before and 1 week after the trauma. We found daytime-dependent changes due to trauma in mean HT in a frequency-dependent manner comparable to the results in mice, but the results temporally shifted according to respective activity profiles. Additionally, we found linear correlations of these threshold changes with the strength of the tinnitus percept, with the most prominent correlations in the 5 PM trauma group. Taken together, circadian sensitivity of the HT to noise trauma can also be found in gerbils, and tinnitus strength correlates most strongly with HL only when the trauma is applied at the most sensitive times, which seem to be the evening.

Keywords: rodents, tinnitus, chronobiology, acoustic trauma, gap prepulse inhibition of the acoustic startle response, auditory brainstem response audiometry

OPEN ACCESS

Edited by:
Richard Carl Gerum,
York University, Canada

Reviewed by:
Alexander Galazyuk,
Northeast Ohio Medical University, United States
Prashanth Prabhu,
All India Institute of Speech and Hearing (AIISH), India

****Correspondence:***
Konstantin Tziridis
konstantin.tziridis@uk-erlangen.de

Specialty section:
This article was submitted to Auditory Cognitive Neuroscience, a section of the journal Frontiers in Neuroscience

Received: *07 December 2021*
Accepted: *09 May 2022*
Published: *03 June 2022*

Citation:
Grimm J, Schulze H and Tziridis K (2022) Circadian Sensitivity of Noise Trauma-Induced Hearing Loss and Tinnitus in Mongolian Gerbils. Front. Neurosci. 16:830703. doi: 10.3389/fnins.2022.830703

INTRODUCTION

Circadian rhythms substantially affect the behavior and physiology of animals (Konopka and Benzer, 1971; Gachon et al., 2004; Richards and Gumz, 2013). Underlying mechanisms are centrally controlled by the suprachiasmatic nucleus in the hypothalamus (Moore and Eichler, 1972) and synchronized to light emission received *via* the eyes. In addition, sound can entrain circadian rhythms by modulating melatonin production or body temperature (Goel, 2005), and circadian clocks were found in the cochlea and the inferior colliculus (Meltser et al., 2014; Park et al., 2016). This is based on variable mRNA expression and protein production, possibly affecting neuronal processing within the auditory system over the course of the day (Park et al., 2016).

Effects of such circadian clocks on the auditory system are evident in auditory-related behavior such as the startle response to a loud sound. In mice, a nocturnal species (Ripperger et al., 2011), it has been shown that the amplitude of the acoustic startle response (ASR) is larger during daytime

(their resting or inactive phase) compared to the night (active phase) (Meltser et al., 2014). In other nocturnal species such as rats (Stephan, 1983) or humans as a diurnal species (Miller and Gronfier, 2006), this effect has been shown to be reversed (Frankland and Ralph, 1995). Interestingly, such internal clocks in mice not only affect auditory processing but also modulate the effects of noise trauma on the hearing thresholds (HT). Then, 2 weeks after acoustic trauma, the trauma-elevated thresholds completely recovered in CBA/Ca/Sca male mice, which received trauma (6–12 kHz bandpass noise, 100 dB SPL, 60 min) during the inactive phase. On the other hand, animals that received trauma during the active phase showed a permanent high-frequency hearing loss (HL) (Meltser et al., 2014). Furthermore, the brain-derived neurotrophic factor (BDNF) mRNA and protein expression-related responses of the internal clocks, e.g., in the cochlea and inferior colliculus in the latter animals were significantly reduced, indicating a lack of BDNF-related protective response ability after active phase trauma (Fontana et al., 2019). Nevertheless, the results can vary dependent on the species or even strain used for the animal experiments (Meltser et al., 2014; Harrison, 2019; Sheppard et al., 2019) and might not be usable for a direct comparison with human physiology. We here attempted to reduce at least one of the constrains with rodents, namely, being mostly nocturnal animals, using a diurnal rodent species that in our view showed to be a good model for human HL and tinnitus development after an acoustic trauma.

In this study, we investigated whether an acoustic trauma applied at different times of the day in diurnal male Mongolian gerbils (Roper, 1976) leads to variable changes in HT. If so, we were furthermore interested in the question, of whether different trauma effects on HT also lead to a different rate and/or strength of tinnitus development in these animals. Our hypotheses were, first, that gerbils show a higher sensitivity to acoustic trauma during their active phase and, second, that tinnitus development is different at different times based on the differences in HL. The fact that tinnitus severity and the amount of hearing loss are correlated in humans is well described [e.g., (Searchfield et al., 2007)], but tinnitus strength/loudness and severity are not necessarily identical (Hiller and Goebel, 2007), which is a constrain of our animal model as we are not able to tell, if animals with a loud percept do also suffer more. Nevertheless, we do see tinnitus intensity and HL correlations in other studies with this animal model [e.g., (Lanaia et al., 2021)], which shows the principle approach to be valid. In this context, we recently proposed a new mechanism of tinnitus development based on neuronal stochastic resonance (SR) (Krauss et al., 2016a). It was proposed that, following noise trauma-induced HL, the SR adds neuronal noise to weak auditory signals at the level of the dorsal cochlear nucleus, thereby lifting otherwise sub-threshold signals above the response threshold of dorsal cochlear nucleus neurons so that the auditory information can be further propagated upstream the auditory pathway. In other words, the added noise would be "tuned up" in a frequency-specific manner to reduce noise trauma-induced HL. Simultaneously, the added noise signal is also propagated upstream the auditory pathway to the auditory cortex where it is finally perceived as tinnitus. In other words, tinnitus is the cost of the rescued hearing threshold (Krauss et al., 2018, 2019). This model of tinnitus development led to a number of predictions that, in the meantime, were demonstrated to be true: temporary auditory information depletion leads to partially reversible tinnitus (Krauss and Tziridis, 2021), chronic auditory information depletion leads to permanent tinnitus percepts (Tziridis et al., 2021), and patients with tinnitus have less HL compared to those without tinnitus (Gollnast et al., 2017). Interestingly, a linear correlation between HL and tinnitus strength has been described, indicating a direct link between HL and tinnitus development (Lanaia et al., 2021). However, to date, it is still unclear whether circadian effects (e.g., activity or hormonal levels) also play a role in either the tinnitus development itself or whether the tinnitus strength that develops after a HL is induced. Therefore, we aimed to investigate with this study if a noise trauma applied at different times of the day leads to hearing loss of different strengths in a diurnal species and if a thereby induced tinnitus development differs depending on the trauma induction time.

MATERIALS AND METHODS

Ethical Statement

Mongolian gerbils (*Meriones unguiculatus*) were housed in type IV cages in a UniProtect Air Flow Cabinet (Zoonlab, Castrop-Rauxel, Germany, ambient noise level 55 dB SPL) in groups of 3 to 4 animals with free access to water and food (complete food for Gerbils, V1644-000, ssniff Spezialdiäten GmbH, Soest, Germany) at 24°C room temperature and 50% relative air humidity under 12-/12-h dark/light cycle (6 AM to 6 PM light). Animals had access to nesting material (paper nestlings) and other enrichments (cardboard tubes). The use and care of animals were approved by the state of Bavaria (Regierungspräsidium Unterfranken, Würzburg, Germany, no. 54.2.2-2532-2-540). A total of thirty-four 10–12-week-old (65 to 77 g) male gerbils purchased from Janvier (Saint Berthevin, France) were used in this study.

Time Regime of Behavioral Testing, Auditory Brainstem Responses, and Acoustic Trauma

All animals were handled before the beginning of the experiments and accustomed to the setup environment to minimize stress. Animals were separated into four groups with 8 to 10 animals each, animals were randomly allocated to the groups by the experimenter, and therefore, he was not blinded. The only difference between the groups was the time of the acoustic trauma (2 kHz, 115 dB SPL, 75 min) which either started at 9 AM, 1 PM, 5 PM, or 12 PM, no control group was included in this study. Relative to the Zeitgeber time (ZT), i.e., the onset of light was at 6 AM, this is ZT3, ZT7, ZT11, and ZT18. These daylight times were selected to span the light part of the activity cycle roughly equal, and the 12 PM time point was chosen as it was the center time of the dark cycle. All other measurements were taken at the identical time of the day. This means that the gap prepulse inhibition of the ASR gap-prepulse inhibition of the acoustic startle reflex

(GPIAS) measurement for behavioral tinnitus assessment always started at 9 AM with a duration of 60 min, and the auditory brainstem response (ABR) measurements started at 10 AM with a duration of 30 min per ear and a randomized order of measurements for both ears. An overview of the timing is depicted in **Figure 1**, all behavioral and electrophysiological methods used have previously been described in detail (Ahlf et al., 2012; Tziridis et al., 2014, 2015; Krauss et al., 2016b; Schilling et al., 2017, 2019) and are therefore only described briefly here.

Behavioral Testing (GPIAS)

The ASR is a widely used behavioral approach that does not require any training prior to testing. The original prepulse inhibition of the ASR has been modified into GPIAS to assess a possible tinnitus percept (Turner et al., 2006; Turner, 2007). In this study, animals were tested first with a gap–no-gap paradigm to measure the prepulse inhibition (PPI) of the startle response amplitudes usually 1 to 2 days before the acoustic trauma. These data were used as the healthy baseline condition for the calculation of the effect size of the change of PPI (cf. Statistical Analyses). For ASR measurement, animals were placed in a 15-cm-long, 4.2-cm-wide acrylic tube on a 3D-acceleration sensor platform in front of two speakers (one for the background stimulus and one for the startle pulse) inside a dark acoustic chamber (Schilling et al., 2017; Gerum et al., 2019). For adaptation, the animals were given 15-min rest, following which the stimulation started with five adaptation stimuli that were discarded from further analysis. The proper GPIAS stimuli consisted of 60 stimuli of band-pass filtered (digital 4^{th}-order Butterworth filter, 48 dB/octave) background noise (10 s, 60 dB SPL) with center frequencies of 1, 2, 4, and 8 kHz and half an octave width each. They were presented with (30 stimuli) or without (30 stimuli) a 50-ms gap (2-ms sin^2-ramps) 100 ms before a startle stimulus (white noise burst, 20-ms duration, 2-ms sin^2-ramps, 105 dB SPL) in a pseudorandomized manner. A complete behavioral measurement of 240 stimuli had a duration of roughly 45 min. Additionally, 5 habituation stimuli at the beginning of the whole procedure were presented but not recorded. The exact same measurements were repeated 7 days after the acoustic trauma at the same time of the day (cf. **Figure 1**).

Hearing Thresholds (ABR)

After the GPIAS measurement, the animals were anesthetized with 0.3 to 0.4 ml of a mixture of ketamine and xylazine (ketamine 500 mg/kg, xylazine 25 mg/kg, s.c.) and placed on a remote-controlled heating pad to obtain individual baseline audiograms (i.e., ABR) of each healthy animal (**Supplementary Figure S1**). The stimulation frequencies between 1 and 8 kHz in octave steps for stimulation intensities ranging from 0 to 90 dB SPL in 5 dB steps were presented free-field *via* a loudspeaker placed 3 cm in front of the pinna of the measured ear. For each ear, stimulus and intensity, 300 repetitions (each) of 6-ms-long phase-inverted double stimuli (100-ms intrastimulus interval, 500-ms interstimulus interval) were presented. The complete measurement of one ear took less than 30 min. For the ABR measurements, three silver electrodes were placed subcutaneously, one retro-aural above the bulla of the tested ear (recording electrode), another one central between both ears (reference electrode), and a last one at the basis of the tail (ground electrode). The signal was recorded and filtered (bandpass filter 400 to 2,000 Hz) *via* a Neuroamp 401 amplifier (JHM, Mainaschaff, Germany). As for the GPIAS, the exact same measurements were repeated 7 days after the acoustic trauma at the same time of the day.

Acoustic Trauma

Then, 1 to 2 days after the baseline measurements of GPIAS and ABR, the animals were again anesthetized with the same ketamine–xylazine mixture and placed on a remote-controlled heating pad, 10 cm in front of a speaker inside a dark acoustic chamber. A mild binaural acoustic trauma was induced using a 2 kHz pure tone at 115 dB SPL lasting for 75 min. These adjustments were selected as they provided the highest rates of tinnitus development in our laboratory, with the trauma frequency being slightly below the best hearing range of the animals (Tziridis et al., 2014). The start of the sound exposure was either at 9 AM, 1 PM, 5 PM, or 12 PM to sample the two peaked circadian cycles of a male gerbil (Roper, 1976). In the 12 PM condition, animals were handled only under red light to prevent resetting of the circadian system, as mice and gerbils share the lack of red receptor cones (Rocha et al., 2016). The animals were usually sleeping for 120 min, and after fully waking up, they were returned to their home cage into the normal 12-/12-h light (6 AM to 6 PM)/dark (6 PM to 6 AM) rhythm.

Statistical Analyses

The obtained data of the GPIAS and ABR measurements were evaluated objectively (i.e., automatically) by custom-made Python programs (Ahlf et al., 2012; Gollnast et al., 2017). The threshold determination was performed by an automated approach using the root mean square (RMS) values of the ABR amplitudes fitted with a hard-sigmoid function utilizing the background activity as offset (Schilling et al., 2019). The mean threshold was set at the level of slope change of this hard-sigmoid fit independent for each frequency and for each time point. HL was calculated by subtracting the post-trauma from the pre-trauma thresholds and tested by parametric tests. Positive values indicate worse hearing after the trauma, and negative values indicate better hearing. The HL was calculated separately for each ear, and the ear with less HL was called the less affected ear and the other the stronger affected ear. Mean HL is the mean binaural HL for each animal. For statistical evaluation, two-factorial ANOVAs with Tukey's *post hoc* tests were performed. Additionally, we investigated as a kind of *post hoc* test one-factorial ANOVAs at different stimulation frequencies to "zoom in" on possible temporal threshold effects of the trauma. Significant hearing loss was also tested by Bonferroni-corrected single sample *t*-tests vs. zero.

Tinnitus development was tested for each animal and frequency individually by *t*-tests of the log-normalized PPI: The log-normalization of the response amplitudes of gap (A_{gap}) and no-gap (A_{no-gap}) by $\log(A_{gap}/A_{no-gap})$ is necessary as it has been shown that, only after this calculation, parametrical testing (e.g.,

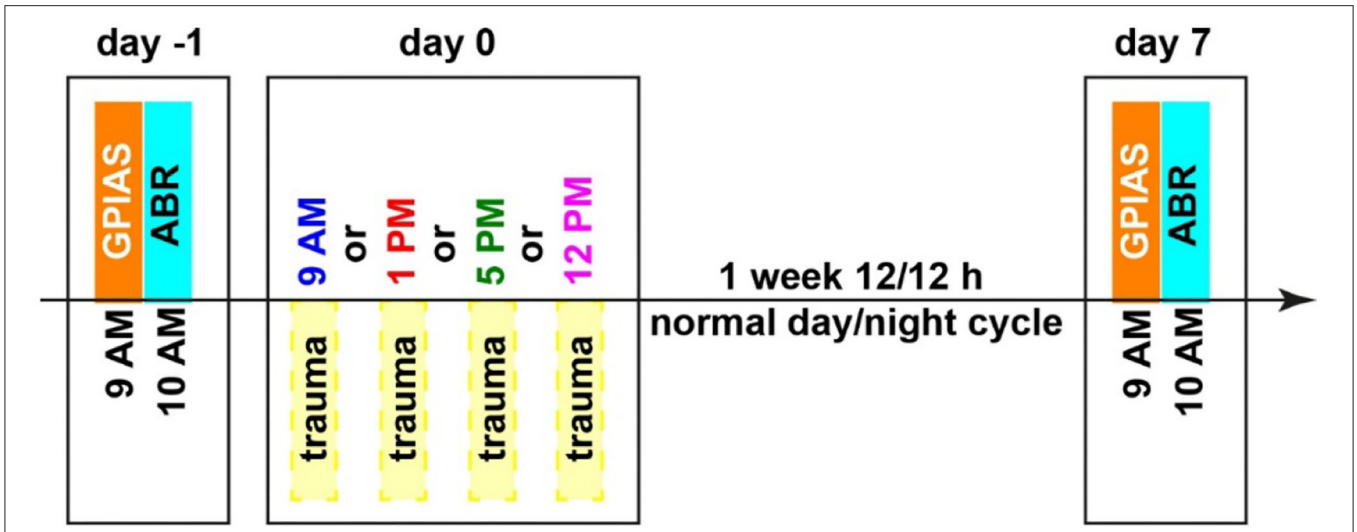


FIGURE 1 | Time regime of behavioral testing for tinnitus assessment (GPIAS), hearing threshold measurements (ABR), and mild acoustic trauma (2 kHz, 115 dB SPL, 75 min). Different trauma times are depicted in different colors, which will also be used in the following figures. Zeitgeber times (ZT, time of light on) correspond as follows: 9 AM = ZT 3, 10 AM = ZT 4, 1 PM = ZT 7, 5 PM = ZT 11 and 12 PM = ZT 18.

by *t*-tests) is allowed (Schilling et al., 2017). A significant decrease ($p < 0.05$, Bonferroni-corrected) of the GPIAS-induced change of the response to the ASR after the trauma ($GPIAS_{post}$) relative to conditions before the trauma ($GPIAS_{pre}$) in that test was rated as an indication of a tinnitus percept at that specific frequency. The effect size of the log-normalized PPI change can be interpreted as a correlate of the subjective tinnitus loudness. With this behavioral approach, we can group animals independent of any other measurements into those with behavioral signs of a tinnitus percept and those without (Schilling et al., 2017). The resulting distributions were tested by a chi-squared test for multiple groups.

Gap-prepulse inhibition of the acoustic startle reflex effect size was assessed by the Kruskal–Wallis ANOVAs independent for each time of trauma induction. ABR threshold changes (less affected ear, stronger affected ear, and mean HL) were assessed independently from the GPIAS analyses by two-factorial ANOVAs (factors *stimulation frequency* and *time of trauma*). Correlations of behavior and HL were calculated with multiple linear regression analyses for the four different time points separately. Additionally, the behavioral data were correlated with the mean HL and the individual HL of the two differently affected ears.

RESULTS

Hearing Loss Induced by Noise Trauma at Different Times of the Day

First, we investigated the mean binaural HL induced by the trauma at the four different times of the day by a two-factorial ANOVA with the factors *time of day* and *frequency*. **Figure 2A** depicts the results of this analysis. It shows no significant *time* effect averaged across all frequencies (also cf. **Table 1**) on the mean HL of both ears (right panel: $F_{(3,104)} = 0.50$, $p = 0.68$) but a significant *frequency* effect averaged across all times (center panel: $F_{(3,104)} = 3.42$, $p = 0.02$), indicating that the pure tone trauma induced a frequency-specific HL. The *interaction* of both factors revealed no significant differences in the mean binaural HL over the four times of the day and the four frequencies ($F_{(9,104)} = 0.77$, $p = 0.65$). In other words, binaural HL was comparable in all four groups, resulting in a frequency-specific HL around 4 kHz with the biggest variances seemingly around the trauma and the 4 kHz frequencies. This is also the case when grouping the data in only two-time groups, "early" (9 AM and 1 PM) and "late" (5 PM and 12 PM), for a higher statistical power as can be seen in **Supplementary Figure S2A**. In **Figure 2B,** the one-factorial ANOVA of the mean binaural HL at the trauma frequency over the different *times of day* of the trauma is given. No significant effect could be found ($F_{(3,24)} = 2.13$, $p = 0.12$), indicating that the mean binaural HL at the trauma frequency was comparable between the four trauma groups. Nevertheless, we found a significant HL (single sample *t*-test vs. 0) in the 5 PM trauma group only (mean ± standard deviation: 14.7 ± 15.3 dB, $p = 0.03$). For the most affected frequency of 4 kHz, we also did not see any time dependency of the mean binaural HL on the *time* of trauma by a 1-factorial ANOVA ($F_{(3,24)} = 0.41$, $p = 0.75$), and the single sample *t*-test vs. Also, 0 only revealed a significant HL at 5 PM (17.61 ± 9.71 dB, $p = 0.001$). In conclusion, there might be a trend for higher HL values in binaural HL analysis when the trauma was applied at 5 PM, but the effect is not very prominent.

Next, we focused our analyses on the two individual ears of each animal separately. In all animals, one ear showed a lower mean HL (less affected ear) compared to the other (stronger affected ear). In the less affected ear (**Figure 2C**), we found a similar result for the two-factorial ANOVA of the mean HL as in the binaural analysis, with no significant effect of the *time of day* the trauma was applied ($F_{(3,104)} = 0.49$, $p = 0.69$), a significant *frequency* effect ($F_{(3,104)} = 3.61$, $p = 0.016$), and no significant

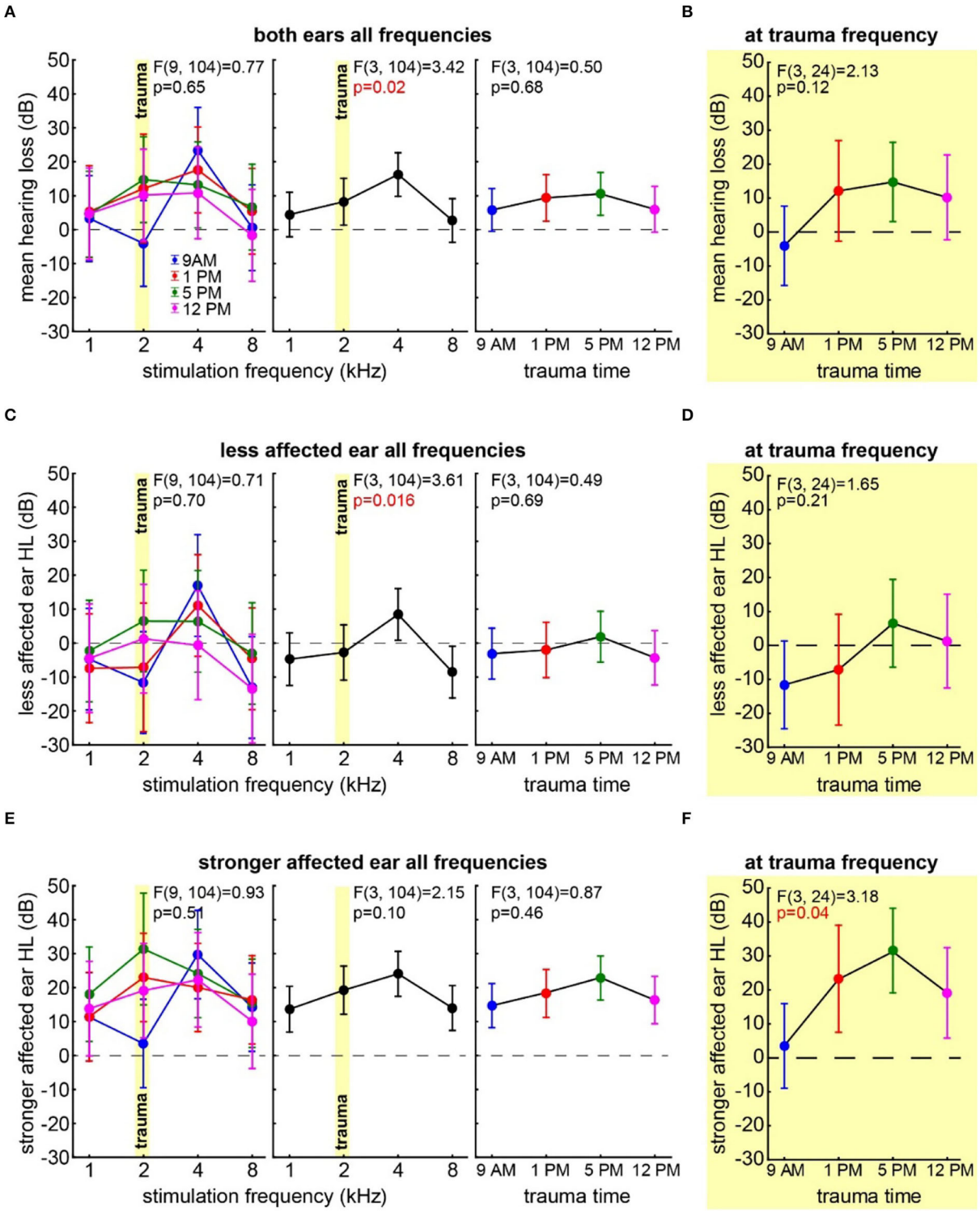


FIGURE 2 | Hearing loss one week after trauma at different times of the day. **(A)** Mean HL across both ears analyzed by a two-factorial ANOVA with factors *time of day* (right panel) and *frequency* (center panel), the interaction of both factors is depicted in the left panel, *time of day* is color-coded, the trauma frequency is marked by a yellow bar, and whiskers give the 95% confidence intervals, significant *p*-values of the F-statistics are red. **(B)** One-factorial ANOVA of the mean HL of both ears at the trauma frequency over the *time of day*. **(C)** Two-factorial ANOVA of the HL of the less affected ear with factors *time of day* and *frequency*. **(D)** One-factorial ANOVA of the HL of the less affected ear at the trauma frequency over the *time of day*. **(E)** Two-factorial ANOVA of the HL of the stronger affected ear with factors *time of day* and *frequency*. **(F)** One-factorial ANOVA of the HL of the stronger affected ear at the trauma frequency over the *time of day*.

TABLE 1 | Statistics of one-factorial ANOVAs of HL over the different *frequencies* and single sample *t*-tests vs. 0 in the four trauma time conditions.

Trauma condition	Frequency-dependent F statistics	Frequency-dependent *p*-value	General effect: mean ± standard deviation	General effect: single-sample *t*-test vs. 0 *p*-value
9 AM	$F_{(3,28)} = 3.21$	0.038	14.70 ± 21.00	0.14
1 PM	$F_{(3,28)} = 0.40$	0.228	17.71 ± 21.11	0.07
5 PM	$F_{(3,24)} = 1{,}55$	0.754	21.41 ± 14.22	0.01
12 PM	$F_{(3,24)} = 1.00$	0.409	16.33 ± 17.15	0.06

interaction of both factors ($F_{(9,104)} = 0.71$, $p = 0.70$). Also, in the one-factorial ANOVA of the mean HL of the less affected ear at the trauma frequency (**Figure 2D**), no significant differences between the HL induced by the trauma at different *times of the day* could be found ($F_{(3,24)} = 1.65$, $p = 0.21$). Note that the less affected ear shows hardly any change relative to the pre-trauma condition, and no mean HL value at any point in time was significantly different from zero. In the strongest affected frequency of 4 kHz (cf. **Figure 2C**, center panel), we did not see any *time* dependency in the HL of the less affected ear (one-factorial ANOVA: $F_{(3,24)} = 0.61$, $p = 0.62$), and the single sample *t*-tests vs. 0 only showed a significant HL at 5 PM (11.07 ± 10.57 dB, $p = 0.03$). Taken together, the less affected ear did show a frequency-dependent HL around 4 kHz, but also one could only speak of a trend that the HL in the 5 PM trauma group might be stronger compared to the other groups.

The stronger affected ear on the other hand showed a somewhat different pattern (**Figure 2E**), as neither factor *time of day* ($F_{(3,104)} = 0.87$, $p = 0.46$) nor *frequency* ($F_{(3,104)} = 2.15$, $p = 0.10$) nor *interaction* ($F_{(9,104)} = 0.93$, $p = 0.51$) showed any kind of significant effect on the mean HL in that ear. Note that in the factor *frequency,* a weak tendency could be found. Nevertheless, contrary to the other two analyses, we did not see strong differences in the HL in the factor *frequency* in these ears (cf. **Figures 2A,E**, center panels). For comparison with the less affected ear and the mean binaural HL (i.e., the two analyses above), we compared the effect of the *time* of the trauma also at the trauma frequency of 2 and at 4 kHz by two one-factorial ANOVAs. In the first one-factorial ANOVA of the mean HL at the trauma frequency of the stronger affected ear (**Figure 2F**), we found a significant effect of the *time of the day* of the trauma induction with $F_{(3,24)} = 3.18$, $p = 0.04$ and significant mean HL values (single sample *t*-test vs. 0) at 1 PM ($p = 0.013$), 5 PM ($p = 0.005$), and 12 PM ($p = 0.036$). In the second one-factorial ANOVA at 4 kHz, we found no time dependency of the HL of the stronger affected ear ($F_{(3,24)} = 0.24$, $p = 0.87$), but except for the 1 PM trauma, we found time significant hearing losses with single sample *t*-tests vs. 0 (9 AM: 29.73 ± 28.23 dB, $p = 0.02$, 5 PM: 24.15 ± 10.84 dB, $p < 0.001$, 12 PM: 22.29 ± 19.48 dB, $p = 0.02$). In other words, the HL in the stronger affected ear was greater in the quality of a general threshold increase, but in a specific frequency, the ranges significantly depended on the trauma time.

Effect Size of Behavior Change by Trauma at Different Times of the Day

In the case of the tinnitus assessment, we could not disentangle the effects of the single ears, as we measured the behavior of the awake animal. First, we analyzed the median effect size of the response change at the different frequencies of the four different trauma time groups independently by the Kruskal–Wallis ANOVAs (**Figure 3**). The median effect size represents the overall response of all animals and gives a general overview of the change of the PPI responses post- vs. pre-trauma but not the individual percept of each animal. Note that only negative values indicate a possible tinnitus percept at a given frequency. In this general overview, we found no significant differences in the responses to the different frequencies at 9 AM ($H_{(3,32)} = 1.64$, $p = 0.65$), 1 PM ($H_{(3,28)} = 4.31$, $p = 0.23$), or 5 PM ($H_{(3,32)} = 2.46$, $p = 0.48$), but at least a strong tendency for such a frequency-specific effect on the effect size of the PPI difference at 12 PM with $H_{(3,28)} = 7.57$ and $p = 0.056$. A comparable result was found, when the data were grouped into only two trauma time groups, i.e., early and late (**Supplementary Figure S2B**). No significant difference emerged in the early group ($H_{(3,N=60)} = 4.31$, $p = 0.23$) whereas the late statistics showed a tendency for a frequency dependency of the effect size ($H_{(3,N=60)} = 6.85$, $p = 0.077$). To estimate the frequency of individual appearance of tinnitus percepts across animals in the four tested frequency ranges, we counted the number of significant negative effect sizes at each frequency and trauma time condition and plotted them as a percentage of complete responses in **Figure 4**. In only 26% of all responses, a tinnitus percept was found with a peak of 10% of all responses after trauma at 5 PM. Note that a noise trauma at 5 PM also indicated the strongest HL. Nevertheless, a chi-squared test for multiple groups revealed no significant differences ($X^2_{(3,104)} = 4.31$, $p = 0.23$) between the group response distributions. In other words, when only looking at the behavioral changes related to possible tinnitus percepts, we could not identify any significant dependencies on the time of trauma.

Linear Regression Analyses of Hearing Loss and Effect Size

With the idea of the model of SR for tinnitus development (cf. Introduction) in mind, we finally investigated the relationship between hearing loss (which is believed to

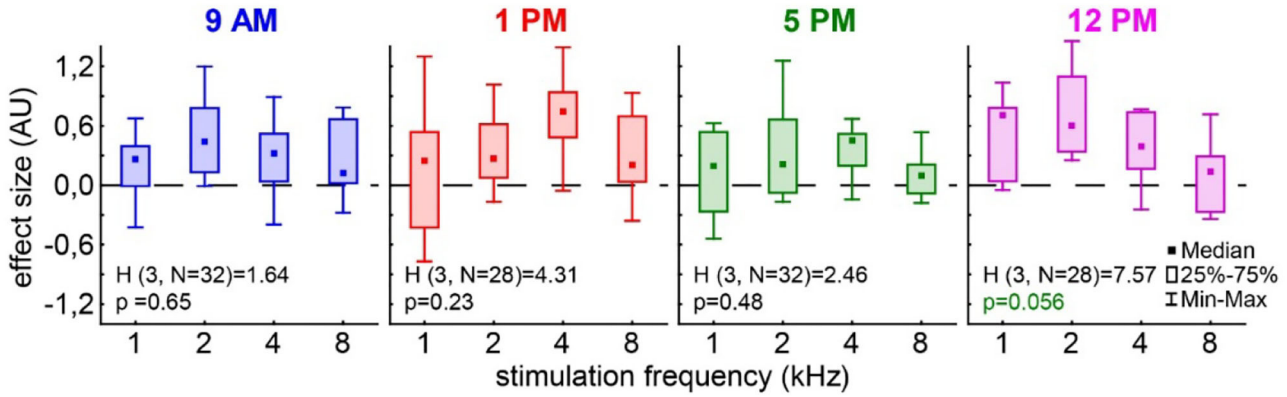


FIGURE 3 | Effect size of the PPI change post vs. pre-trauma. The Kruskal–Wallis ANOVAs (H-statistics) were calculated independently for each trauma time group. Boxes depict the interquartile range, and whiskers give the range of the complete distribution.

trigger SR) and the tinnitus percept (a consequence of SR) at the given frequencies by multiple linear regression analyses (**Figure 5**). First, we looked at the relationship between the frequency-matched values of the mean binaural hearing loss and the behavioral effect size in the four different trauma time groups (**Figure 5A**). We found a significant linear regression in the 5 PM trauma group with $r^2 = 0.22$ and $p = 0.01$. In other words, higher HL values led to more values with positive effect sizes in that group, while hearing threshold gain correlates with negative effect size values, indicating a tinnitus percept. Additionally, in the 1 PM as well as in the 12 PM trauma group, we see tendencies for such a linear correlation, both with r^2-values around 0.11 and $p = 0.09$. When grouping the data into only two trauma time groups (early and late), the results were even more clear with the early trauma animals showing no significant correlation between HL and effect size with $r^2 = 0.02$ and $p = 0.26$ and the late animals showing such a correlation with $r^2 = 0.12$ and $p = 0.007$ (**Supplementary Figure S2C**). In the less affected ears (**Figure 5B**), we found an even stronger correlation between HL and effect size (r^2-value of 0.34 and a $p = 0.001$) in the 5 PM trauma group but only a tendency for a linear correlation of both variables in the 1 PM trauma group ($r^2 = 0.11$, $p = 0.09$). Finally, in the stronger affected ears (**Figure 5C**), we did not observe any significant correlation between the two variables. Note that most of the HL values in these ears were positive, indicating a real hearing loss.

Concluding the results, we were able to prove that only the combinatory view of hearing loss and the behavioral correlates of a tinnitus percept showed the circadian effects of acoustic trauma in detail, whereas each variable alone did show only minor temporal dependencies. However, contrary to our hypothesis, the strongest and most general effects found were in the afternoon, when the activity of the animals was minimal (cf. **Figure 6**).

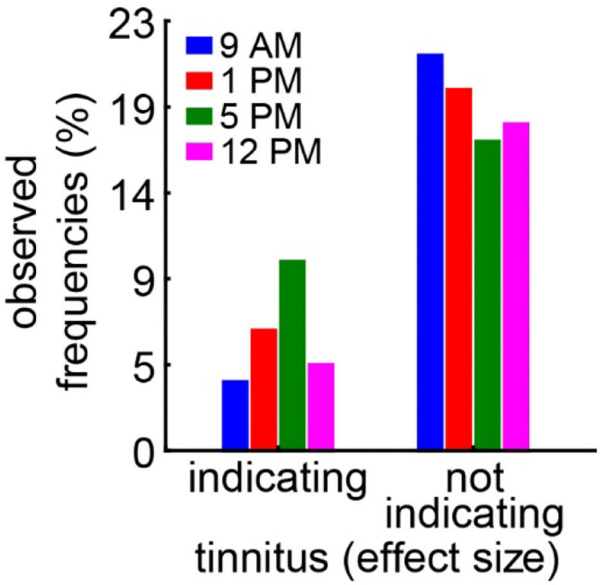


FIGURE 4 | Relative number of frequencies with or without behavioral signs of a tinnitus percept. A significantly negative effect size indicates a possible tinnitus percept and is counted as such. A chi-squared test for multiple groups ($p = 0.23$) reveals no significant difference in the response distributions.

DISCUSSION

In this study, we aimed to investigate, first, whether noise trauma applied at different times of the day leads to hearing loss of different strengths in a diurnal species, the Mongolian gerbil, and, second, if tinnitus development induced by such noise trauma differs depending on trauma induction time. We could show only minor effects in the single variables, but in the combination, we found clear correlations between HL and behavioral correlates of tinnitus at different time points. When analyzing the variables independently, we found that, especially around the trauma frequency, the stronger affected ear showed a significant dependency of the HL on the time of the trauma with the amount of HL peaking around 5 PM – corresponding to the end of the activity cycle in these

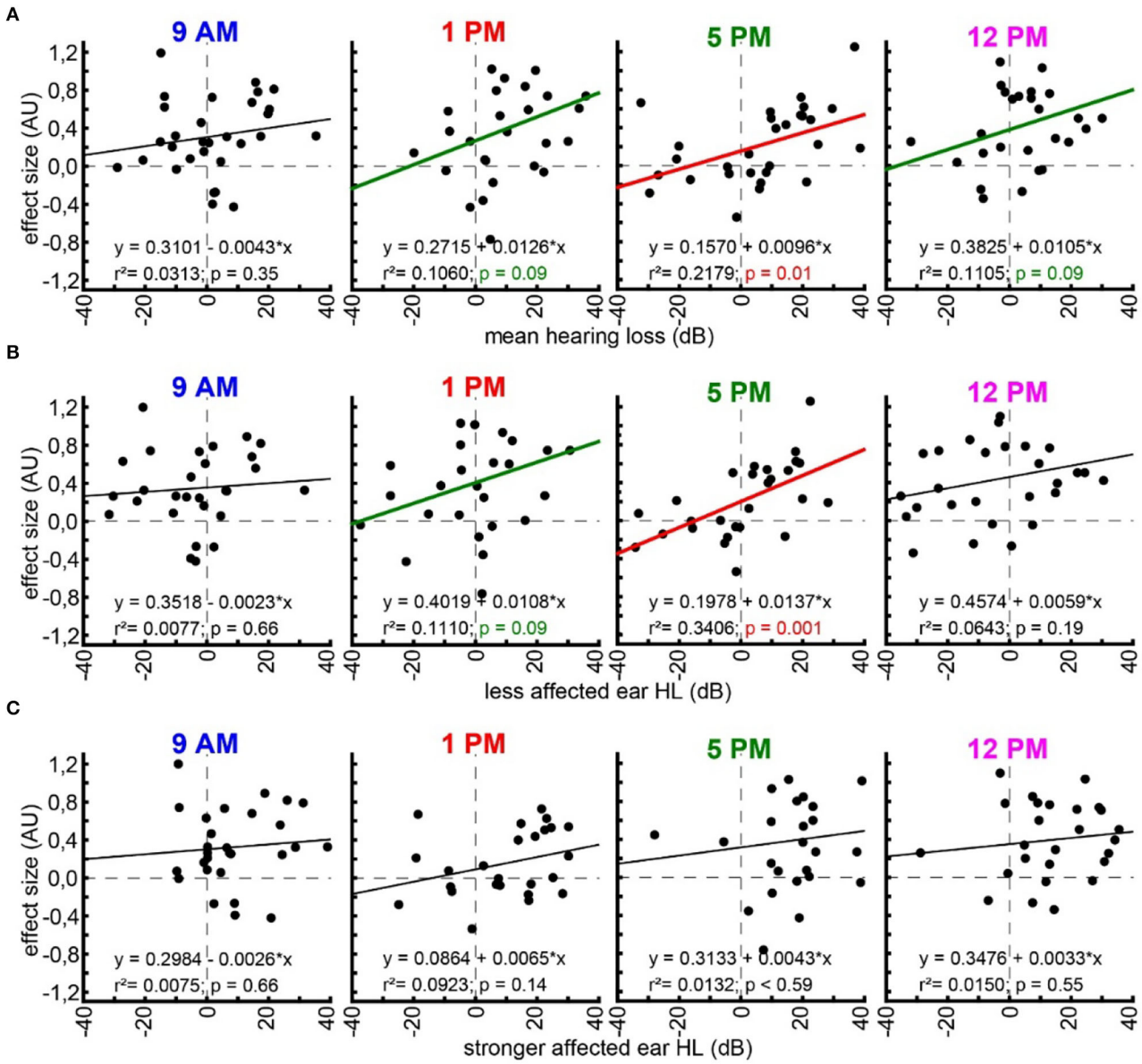


FIGURE 5 | Linear regression analyses of hearing loss and effect size. **(A)** Multiple linear regressions for effect size and mean binaural hearing loss for each trauma time group (panels from left to right). Linear regression formula, r^2-value, and *p*-value are given for each case. Black regression lines and *p*-values indicate no significant regression, green indicated a tendency, and red indicates a significant regression. **(B)** Multiple linear regressions for effect size and less affected ear hearing loss. **(C)** Multiple linear regressions for effect size and stronger affected ear hearing loss.

animals. The rate of observed tinnitus frequencies also tended to be highest around this time of the day (i.e., 5 PM) even though we were not able to see any significant differences in the effect sizes. When correlating the individual tinnitus-related behavioral changes with the individual changes in the hearing thresholds, we could show a clear correlation between tinnitus strength and hearing loss. This correlation was not seen for all trauma induction times but was peaked at 5 PM, i.e., the time of minimal activity of the animals (cf. below). Surprisingly, this correlation seems to be present not only in the less affected ears but also in mean binaural hearing loss values.

Before going into a detailed discussion, we would like to state that the study has several limitations. First of all, we did not use sham trauma groups for control, as this would need a sham group for each time point. Therefore, we only compared the effects of the trauma at different time points. Second, no histology was performed to investigate the effects of the trauma on the cochleae, as this was not the focus of this study. Nevertheless, we could already observe that tinnitus is most probably based

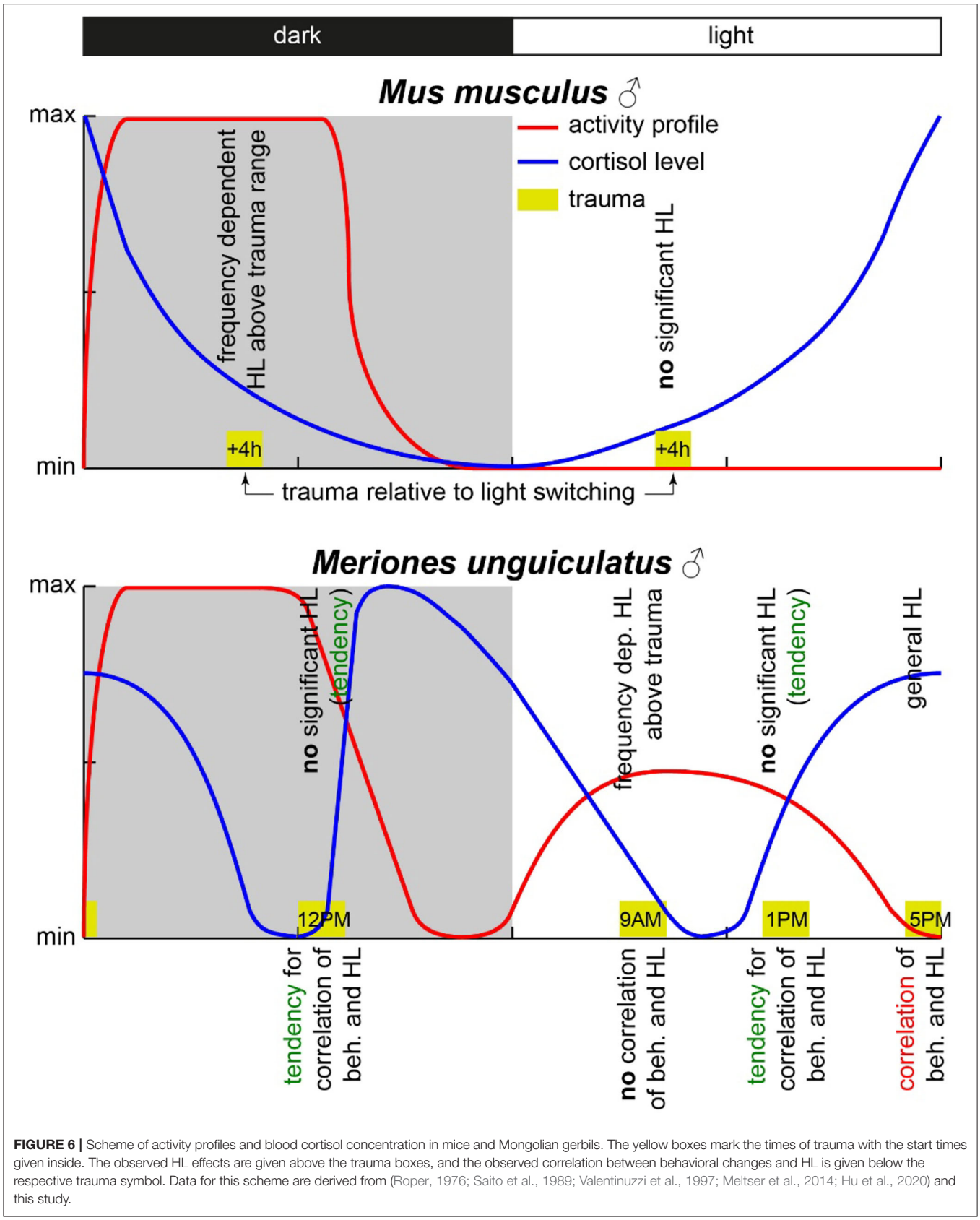


FIGURE 6 | Scheme of activity profiles and blood cortisol concentration in mice and Mongolian gerbils. The yellow boxes mark the times of trauma with the start times given inside. The observed HL effects are given above the trauma boxes, and the observed correlation between behavioral changes and HL is given below the respective trauma symbol. Data for this scheme are derived from (Roper, 1976; Saito et al., 1989; Valentinuzzi et al., 1997; Meltser et al., 2014; Hu et al., 2020) and this study.

on a synaptopathy of the ribbon synapses of the inner hair cells of the cochlea (Tziridis et al., 2021) and have no doubt that we would find similar results also in these animals. Third, we always used mild traumata to induce a maximum amount of tinnitus frequencies in our studies [e.g., (Ahlf et al., 2012; Tziridis et al., 2014, 2015; Krauss et al., 2016b)]. As it has been speculated for other rodents, that mild (or too strong) traumata might be not optimal for circadian studies, as the effects might be too small to be detected or might be dampened by "noise" (Fontana et al., 2019; Sheppard et al., 2019). We agree with this view on the level of single variables, but as we could see in this study, by combining the relevant variables, a strong effect can be shown even in mild acoustic trauma data. Finally, the circadian behavior of gerbils is clearly different from other rodents (cf. below and **Figure 6**). For example, the foraging behavior of animals is usually spread over the activity period, i.e., for mice, this is the night butthis is the whole day for gerbils. Eating has a clear impact on clock rhythms in several organs (Gachon et al., 2004; Mohawk et al., 2012) and might therefore be one of the uncontrolled factors that affect the outcome of our but also other studies.

In nocturnal mice, the most prominent effect on HT after noise trauma was found when broadband noise trauma was applied during their active phase (Meltser et al., 2014), but during their resting period during the day, the animals showed hardly any impairment in hearing after 2 weeks of recovery [for other acoustic trauma-related circadian effects and overview cf. also (Kiefer et al., 2022)]. In this study with Mongolian gerbils, we found the most prominent effects after 1 week after a 2 kHz pure tone trauma around the end of their activity cycle, i.e., near the end of the light phase. For an easier comparison of the two animal models, we present a synopsis of the activity profiles and corresponding blood cortisol levels at the respective trauma times in **Figure 6**. The circadian rhythm within the cochlea is, among others, controlled *via* the suprachiasmatic nucleus and glucocorticoids (Lightman and Conway-Campbell, 2010). The rough scheme presented here is based on data from five different studies in the literature (Roper, 1976; Saito et al., 1989; Valentinuzzi et al., 1997; Meltser et al., 2014; Hu et al., 2020). The peak activity of mice starts with the beginning of the dark cycle and lasts roughly 7 to 8 h (Siepka and Takahashi, 2005), whereas in male gerbils, an activity cycle with two peaks has been described (Roper, 1976). One activity phase started roughly 1 h before the beginning of the light phase (in our case around 5 AM), peaking around 5 h into the light phase (here: 11 AM) and having a trough at the beginning of the dark cycle (here: 6 PM). After that, the activity peaked again 4 h into darkness (here: 10 PM) and then went back to the second trough 1 h before the light cycle starts again. In our approach, we applied the trauma either 2 h before (9 AM) or 2 h after (1 PM) the daytime peak activity as well as near the activity trough at the end of the light cycle (5 PM) and mid-night activity at 12 PM. In contrast to mice, the strongest trauma effects were found around the activity trough at 5 PM (**Table 1**) with the stronger affected ear showing a very broad HL and explicitly loosing at 2 kHz around 27 dB. Also, the averaged hearing impairment of both ears still reached 14 dB at 5 PM, whereas the values of the 9 AM trauma showed an extremely focused hearing impairment, only affecting the 4 kHz frequency range (mean binaural HL (9 AM) one-factorial ANOVA: $F_{(3,28)} = 3.21$, $p = 0.038$, Tukey's *post hoc* 2 kHz vs. 4 KHz: $p = 0.037$). In other words, we saw different HL patterns when applying an acoustic trauma at different times of the day. These patterns are more complex than those in mice, where trauma during the peak activity leads to strong hearing impairment. This higher complexity in the gerbil compared to mice may be because the activity cycle in gerbils with its two activity peaks is more complex than the cycle in mice that shows only one activity peak. In general, the different trauma effects in gerbils could be due to several different reasons as the circadian rhythms affect the different organ systems differently. One probable candidate for mediating the described effect is cortisol (**Figure 6**). In mice, the peak cortisol levels in the blood are found at the beginning of the dark cycle (Gong et al., 2015) corresponding to the strong effect the trauma has on the animals during that time. In gerbils, the cortisol levels were at a minimum at the beginning of the light cycle and then were roughly stable over the day (Liu et al., 2019) well into the dark phase. When looking at the described activity patterns and the blood cortisol levels of mice and gerbils at the different trauma times, a pattern emerges. We found a frequency-specific effect of the trauma on the hearing thresholds of the two animal species at times when activity is high and cortisol levels are dropping. We found no significant general HL (even though we see tendencies for this in Mongolian gerbils) when the activity is low or at least dropping (this could be a possible reason for described tendencies in gerbils) and blood cortisol levels are rising. Finally, in gerbils, we found a general, not frequency-specific, HL when activity is low and blood cortisol levels are high (not tested in mice).

The second question we asked is whether the different trauma times also lead to the differences in tinnitus development following the trauma. It is well accepted that, in humans, tinnitus can be a consequence of hearing loss [e.g., (Heller, 2003; Pan et al., 2009; Boussaty et al., 2021; Tziridis et al., 2021)] and that tinnitus severity and hearing loss intensity in patients with noise-induced hearing loss correlate (Searchfield et al., 2007; Mazurek et al., 2010), but it is unknown whether different trauma times also play a role in its development. Our approach here was the analysis of the effect size of the PPI change post- vs. pre-trauma (Schilling et al., 2017). It enabled us not only to detect a possible tinnitus percept at a given frequency range (± ½ octave around a center frequency) but also calculate a relative tinnitus strength, with the more negative values representing a louder percept (Krauss and Tziridis, 2021; Lanaia et al., 2021). Positive values are most likely the result of cortical processing and learning (Moreno-Paublete et al., 2017) that counteract any negative value of the effect size. In other words, we most probably underestimate the average number of tinnitus percepts, which may also explain the strong positive offsets of the calculated linear regressions shown in **Figure 5** and **Supplementary Figure S2C**. Independent of this possible offset, the method has the advantage of being very fast as it does not require any learning but is hard to cross-validate and is therefore debatable (Eggermont, 2013). Nevertheless, in the effect size data (**Figure 3**), we found no strong differences in the PPI response changes to the different frequencies relative to the different times of the pure tone trauma

applied. The tendency found for such a difference was in the effect size data for the trauma at 12 PM. When looking into the details of the analysis, it became clear that this possible difference was a more positive value at 2 kHz, indicating even less tinnitus/more cortical learning under that specific condition. In addition, the counts of the significant tinnitus frequencies (**Figure 4**) did not show any significant differences between the tinnitus/non-tinnitus frequency ratios. Therefore, in the tinnitus strength or numbers, we could not find any trauma time dependency.

This changed when correlating both variables with each other based on our hypothesis of SR-dependent tinnitus development. This hypothesis assumes that the information flow from the auditory periphery to the central nervous system is constantly optimized by the well-described neurobiological mechanism of SR (Shannon, 1948; Gammaitoni et al., 1998) (cf. Introduction). The data presented here provide an additional aspect of the correlation of both variables (**Figure 5**) as it became clear that the correlation between tinnitus loudness and hearing loss is most prominent when the trauma is applied at the end of the light cycle (i.e., end of activity phase) at 5 PM and the behavior is correlated with the less affected ear. This could indicate that the perceived tinnitus may be monaural at least in the majority of the investigated animals and/or only smaller hearing losses can be rescued by the proposed mechanism as suggested by our study and others (Zeng et al., 2000; Schilling et al., 2020). The fact that the correlation between HL and tinnitus-related behavior was dependent on the trauma induction time indicates that HL is not the only factor that determines tinnitus development but that the tinnitus developing mechanism itself is influenced by the time of trauma induction. In other words, tinnitus development has a circadian component irrespective of noise trauma severity.

CONCLUSIONS

We observed that the time of acoustic trauma plays a significant role not only in the amount of hearing loss it inflicts but also in the strength of a subsequent development of a tinnitus percept. Diurnal as well as nocturnal animals seem to show a similar pattern as the vulnerability for a frequency-dependent hearing loss seems to be highest around their peak activity phase, while at lower activities, there is either no significant impairment or a more generalized effect on all frequencies. The development of tinnitus follows this pattern, in line with the hypothesis that the impairment is the source of the development of the percept. For future investigations – also in humans – it might be crucial to take these circadian effects into account.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The animal study was reviewed and approved by Regierungspräsidium Unterfranken, Würzburg, Germany, No. 54.2.2-2532-2-540.

AUTHOR CONTRIBUTIONS

KT and HS planned the experiments and wrote the manuscript. JG performed all experiments. JG and KT performed the statistical analyses. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We are grateful to Mrs. Jwan Rasheed for her excellent technical support.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2022.830703/full#supplementary-material

REFERENCES

Ahlf, S., Tziridis, K., Korn, S., Strohmeyer, I., and Schulze, H. (2012). Predisposition for and prevention of subjective tinnitus development. *PLoS ONE.* 7, e44519. doi: 10.1371/journal.pone.0044519

Boussaty, E. C., Friedman, R. A., and Clifford, R. E. (2021). Hearing loss and tinnitus: association studies for complex-hearing disorders in mouse and man. *Human Genet.* 141, 1–10. doi: 10.1007/s00439-021-02317-9

Eggermont, J. J. (2013). Hearing loss, hyperacusis, or tinnitus, what is modeled in animal research? *Hear. Res.* 295, 140–149. doi: 10.1016/j.heares.2012.01.005

Fontana, J. M., Tserga, E., Sarlus, H., Canlon, B., and Cederroth, C. (2019). Impact of noise exposure on the circadian clock in the auditory system. *J. Acoust. Soc. Am.* 146, 3960–3966. doi: 10.1121/1.5132290

Frankland, P. W., and Ralph, M. R. (1995). Circadian modulation in the rat acoustic startle circuit. *Behav. Neurosci.* 109, 43. doi: 10.1037/0735-7044.109.1.43

Gachon, F., Nagoshi, E., Brown, S. A., Ripperger, J., and Schibler, U. (2004). The mammalian circadian timing system, from gene expression to physiology. *Chromosoma.* 113, 103–112. doi: 10.1007/s00412-004-0296-2

Gammaitoni, L., Hänggi, P., Jung, P., and Marchesoni, F. (1998). Stochastic resonance. *Rev. Mod. Phys.* 70, 223. doi: 10.1103/RevModPhys.70.223

Gerum, R. C., Rahlfs, H., Streb, M., Krauss, P., Grimm, J., Metzner, C., 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. *Front. Behav. Neurosci.* 13, 140. doi: 10.3389/fnbeh.2019.00140

Goel, N. (2005). Late-night presentation of an auditory stimulus phase delays human circadian rhythms. *Am. J. Physiol. Regul. Integr. Compar. Physiol.* 289, R209–R216. doi: 10.1152/ajpregu.00754.2004

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. *Front. Neurol.* 8, 31. doi: 10.3389/fneur.2017.00031

Gong, S., Miao, Y. L., Jiao, G. Z., Sun, M. J., Li, H., Lin, J., et al. (2015). Dynamics and correlation of serum cortisol and corticosterone under different physiological or stressful conditions in mice. *PLoS ONE.* 10, e0117503. doi: 10.1371/journal.pone.0117503

Harrison, R. T. (2019). *Effect of Changes to the Circadian Rhythm on Susceptibility to Noise-and Drug-Induced Hearing Losses.* Columbus, OH: The Ohio State University.

Heller, A. J. (2003). Classification and epidemiology of tinnitus. *Otolaryngol. Clin. North Am.* 36, 239–248. doi: 10.1016/S0030-6665(02)00160-3

Hiller, W., and Goebel, G. (2007). When tinnitus loudness and annoyance are discrepant, audiological characteristics and psychological profile. *Audiol. Neurotol.* 12, 391–400. doi: 10.1159/000106482

Hu, H., Kang, C., Hou, X., Zhang, Q., Meng, Q., Jiang, J., et al. (2020). Blue light deprivation produces depression-like responses in mongolian gerbils. *Front. Psychiatry.* 11, 233. doi: 10.3389/fpsyt.2020.00233

Kiefer, L., Koch, L., Merdan-Desik, M., Gaese, B. H., and Nowotny, M. (2022). Comparing the electrophysiological effects of traumatic noise exposure between rodents. *J. Neurophysiol.* doi: 10.1152/jn.00081.2021

Konopka, R. J., and Benzer, S. (1971). Clock mutants of Drosophila melanogaster. *Proc Nat Acad Sci.* 68, 2112–2116. doi: 10.1073/pnas.68.9.2112

Krauss, P., Schilling, A., Tziridis, K., and Schulze, H. (2019). Models of tinnitus development: from cochlea to cortex. *HNO* 67, 172–177. doi: 10.1007/s00106-019-0612-z

Krauss, P., and Tziridis, K. (2021). Simulated transient hearing loss improves auditory sensitivity. *Sci. Rep.* 11, 14791. doi: 10.1038/s41598-021-94429-5

Krauss, P., Tziridis, K., Buerbank, S., Schilling, A., and Schulze, H. (2016b). Therapeutic value of ginkgo biloba extract EGb 761(R) in an animal model (Meriones unguiculatus) for noise trauma induced hearing loss and tinnitus. *PLoS ONE.* 11, e0157574. doi: 10.1371/journal.pone.0157574

Krauss, P., Tziridis, K., Metzner, C., Schilling, A., Hoppe, U., and Schulze, H. (2016a). Stochastic resonance controlled upregulation of internal noise after hearing loss as a putative cause of tinnitus-related neuronal hyperactivity. *Front. Neurosci.* 10, 597. doi: 10.3389/fnins.2016.00597

Krauss, P., Tziridis, K., Schilling, A., and Schulze, H. (2018). Cross-modal stochastic resonance as a universal principle to enhance sensory processing. *Front. Neurosci.* 12, 578. doi: 10.3389/fnins.2018.00578

Lanaia, V., Tziridis, K., and 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. doi: 10.3389/fnbeh.2021.698516

Lightman, S. L., and Conway-Campbell, B. L. (2010). The crucial role of pulsatile activity of the HPA axis for continuous dynamic equilibration. *Nat. Rev. Neurosci.* 11, 710–718. doi: 10.1038/nrn2914

Liu, X., Zheng, X., Liu, Y., Du, X., and Chen, Z. (2019). Effects of adaptation to handling on the circadian rhythmicity of blood solutes in Mongolian gerbils. *Animal Models Exper. Med.* 2, 127–131. doi: 10.1002/ame2.12068

Mazurek, B., Olze, H., Haupt, H., and Szczepek, A. J. (2010). The more the worse, the grade of noise-induced hearing loss associates with the severity of tinnitus. *Int. J. Environ. Res. Public Health.* 7, 3071–3079. doi: 10.3390/ijerph7083071

Meltser, I., Cederroth, C. R., Basinou, V., Savelyev, S., Lundkvist, G. S., and Canlon, B. (2014). TrkB-mediated protection against circadian sensitivity to noise trauma in the murine cochlea. *Curr. Biol.* 24, 658–663. doi: 10.1016/j.cub.2014.01.047

Miller, M. W., and Gronfier, C. (2006). Diurnal variation of the startle reflex in relation to HPA-axis activity in humans. *Psychophysiology.* 43, 297–301. doi: 10.1111/j.1469-8986.2006.00400.x

Mohawk, J. A., Green, C. B., and Takahashi, J. S. (2012). Central and peripheral circadian clocks in mammals. *Annu. Rev. Neurosci.* 35, 445–462. doi: 10.1146/annurev-neuro-060909-153128

Moore, R. Y., and Eichler, V. B. (1972). Loss of a circadian adrenal corticosterone rhythm following suprachiasmatic lesions in the rat. *Brain Res.* 42. doi: 10.1016/0006-8993(72)90054-6

Moreno-Paublete, R., Canlon, B., and Cederroth, C. R. (2017). Differential neural responses underlying the inhibition of the startle response by pre-pulses or gaps in mice. *Front. Cell Neurosci.* 11, 19. doi: 10.3389/fncel.2017.00019

Pan, T., Tyler, R. S., Ji, H., Coelho, C., Gehringer, A. K., and Gogel, S. A. (2009). The relationship between tinnitus pitch and the audiogram. *Int. J. Audiol.* 48, 277–294. doi: 10.1080/14992020802581974

Park, J. S., Cederroth, C. R., Basinou, V., Meltser, I., Lundkvist, G., and Canlon, B. (2016). Canlon, Identification of a circadian clock in the inferior colliculus and its dysregulation by noise exposure. *J. Neurosci.* 36, 5509–5519. doi: 10.1523/JNEUROSCI.3616-15.2016

Richards, J., and Gumz, M. L. (2013). Mechanism of the circadian clock in physiology. *Am. J. Physiol. Regul. Integr. Compar. Physiol.* 304, R1053–R1064. doi: 10.1152/ajpregu.00066.2013

Ripperger, J. A., Jud, C., and Albrecht, U. (2011). The daily rhythm of mice. *FEBS Lett.* 585, 1384–1392. doi: 10.1016/j.febslet.2011.02.027

Rocha, F. A. D. F., Gomes, B. D., Silveira, L. C. D. L., Martins, S. L., Aguiar, R. G., and de Souza, J. M., et al. (2016). Spectral sensitivity measured with electroretinogram using a constant response method. *PLoS ONE.* 11, e0147318. doi: 10.1371/journal.pone.0147318

Roper, T. J. (1976). Sex differences in circadian wheel running rhythms in the Mongolian gerbil. *Physiol. Behav.* 17, 549–551. doi: 10.1016/0031-9384(76)90121-9

Saito, M., Nishimura, K., and Kato, H. (1989). Modifications of circadian cortisol rhythm by cyclic and continuous total enteral nutrition. *J. Nutr. Sci. Vitaminol.* 35, 639–647. doi: 10.3177/jnsv.35.639

Schilling, A., Gerum, R. C., Krauss, P., Metzner, C., Tziridis, K., and Schulze, H. (2019). Objective estimation of sensory thresholds based on neurophysiological parameters. *Front. Neurosci.* 13, 481. doi: 10.3389/fnins.2019.00481

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. *Front. Behav. Neurosci.* 11, 198. doi: 10.3389/fnbeh.2017.00198

Schilling, A., Krauss, P., Hannemann, R., Schulze, H., and Tziridis, K. (2020). Reduktion der Tinnituslautstärke. *HNO.* 69, 891–898. doi: 10.1007/s00106-020-00963-5

Searchfield, G., Jerram, C., Wise, K., and Raymond, S. (2007). The impact of hearing loss on tinnitus severity. *Austr. New Zealand J. Audiol.* 29, 67–76. doi: 10.1375/audi.29.2.67

Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Techn. J.* 27, 379–423. doi: 10.1002/j.1538-7305.1948.tb01338.x

Sheppard, A., Liu, X., Alkharabsheh, A., Chen, G.-D., and Salvi, R. (2019). Intermittent low-level noise causes negative neural gain in the inferior colliculus. *Neuroscience.* 407, 135–145. doi: 10.1016/j.neuroscience.2018.11.013

Siepka, S. M., and Takahashi, J. S. (2005). Methods to record circadian rhythm wheel running activity in mice. *Methods Enzymol.* 393, 230–239. doi: 10.1016/S0076-6879(05)93008-5

Stephan, F. K. (1983). Circadian rhythms in the rat, Constant darkness, entrainment to T cycles and to skeleton photoperiods. *Physiol. Behav.* 30, 451–462. doi: 10.1016/0031-9384(83)90152-X

Turner, J. G. (2007). Behavioral measures of tinnitus in laboratory animals. *Prog. Brain Res.* 166, 147–156. doi: 10.1016/S0079-6123(07)66013-0

Turner, J. G., Brozoski, T. J., Bauer, C. A., Parrish, J. L., Myers, K., Hughes, L. F., et al. (2006). Gap detection deficits in rats with tinnitus, a potential novel screening tool. *Behav. Neurosci.* 120, 188–195. doi: 10.1037/0735-7044.120.1.188

Tziridis, K., Ahlf, S., Jeschke, M., Happel, M. F., Ohl, F. W., and 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. doi: 10.3389/fneur.2015.00022

Tziridis, K., Forster, J., Buchheidt-Dorfler, I., Krauss, P., Schilling, A., Wendler, O., et al. (2021). Tinnitus development is associated with synaptopathy of inner hair cells in Mongolian gerbils. *Eur. J. Neurosci.* 54, 4768–4780. doi: 10.1111/ejn.15334

Tziridis, K., Korn, S., Ahlf, S., and Schulze, H. (2014). Protective effects of Ginkgo biloba extract EGb 761 against noise trauma-induced hearing loss and tinnitus development. *Neural. Plast.* 2014, 427298. doi: 10.1155/2014/427298

Valentinuzzi, V. S., Scarbrough, K., Takahashi, J. S., and Turek, F. W. (1997). Effects of aging on the circadian rhythm of wheel-running activity in C57BL/6 mice. *Am. J. Physiol.* 273, R1957–R1964. doi: 10.1152/ajpregu.1997.273.6.R1957

Zeng, F.-G., Fu, Q.-J., and Morse, R. (2000). Human hearing enhanced by noise. *Brain Res.* 869, 251–255. doi: 10.1016/S0006-8993(00)02475-6

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.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

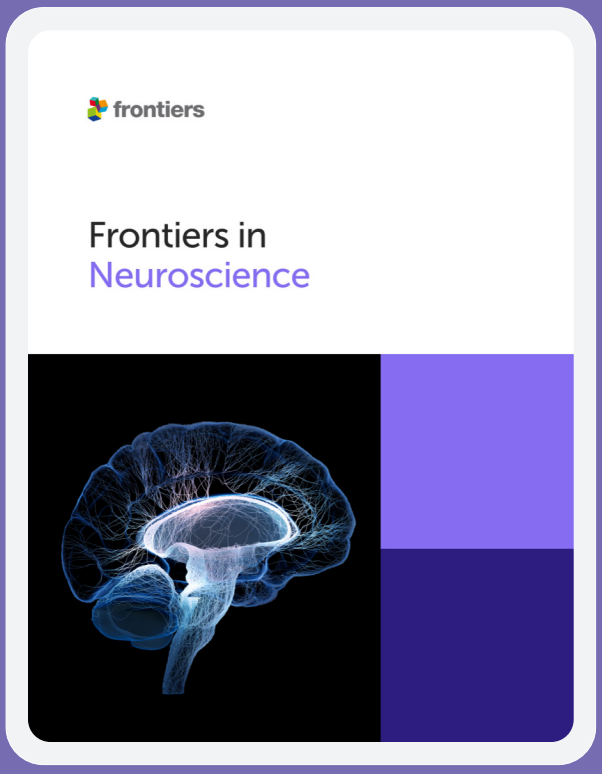
Frontiers in Neuroscience

Provides a holistic understanding of brain function from genes to behavior

Part of the most cited neuroscience journal series which explores the brain - from the new eras of causation and anatomical neurosciences to neuroeconomics and neuroenergetics.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact