Deciphering Real-World Listening: Auditory Cognition in Complex Acoustic Scenes

Xiaoxuan Guo

郭晓轩

Biosciences Institute, Faculty of Medical Sciences

Newcastle University

2025

Declaration

I, Xiaoxuan Guo, confirm that the work presented in this thesis is my own. Information from other sources has been referenced and acknowledged appropriately. The experimental design and data collection presented in Chapter 3 were carried out in collaboration with other lab members in the Auditory Cognition Lab; study design and data collection of Section 5.1 were carried out by a previous lab member with undergraduate students; data analysis and visualisation of both studies were performed independently.

Acknowledgements

The past four years have been an incredible experience for me. I am deeply grateful to the many people who have supported me along the way. Firstly, I would like to thank my supervisor Prof Tim Griffiths, who gave me the opportunity to start the PhD journey, trusted me to explore and learn, and fostered my academic and career growth. I have to thank Dr Will Sedley, whose steady guidance, support, and encouragement have been invaluable throughout this journey. I would not be able to complete my PhD without Dr Ester Benzaquén, who has been like a third supervisor to me and a very good friend. She has taught me many essential skills that enabled me to complete this thesis. I know I can always count on her for help, inspiration, and advice. Dr Ekaterina Yukhnovich has been a great PhD buddy, always ready to lend a helping hand. She has also taught me important lab skills and helped me cope with my PhD. I want to thank Hasan Çolak, Dr Meher Lad, Dr Pradeep Dheerendra, Dr Emma Holmes, and Dr Joel Berger for generously providing valuable materials to support my projects and offering insightful advice along the way. I am appreciative of Dr Quoc Vuong, Dr Colline Poirier, Dr Kai Alter, and Prof Steven Rushton. Their thoughtful feedback and recommendations have significantly enhanced this work. I am grateful to Dr Kai Alter and Prof Adam Tierney for agreeing to review this work.

The greatest gift of my life is my loving family and friends. I owe all my achievements to my parents for their selfless love and unwavering support. They have given me the courage to explore life and the means to support my whims. My grandma is one of those life's true heroines who has an astonishing capacity for love, care and endurance. Listening to her voice has always been a great cure for stress. Watching Inuyasha wag his tail was another. I would like to thank my friends, Dr-to-be Huayu and Hosanna, who have walked with me throughout this journey despite both being hundreds of miles away. I want to thank my partner Matt, who is the most patient person I have ever met. I would not be able to complete my PhD feeling happy and content without his love and support.

Abstract

Speech-in-noise (SIN) difficulty can be explained by a variety of auditory cognitive factors, and has been linked to general cognitive performance. This work summarises the mechanisms of the auditory system supporting speech perception in noise and reviews the commonly used hearing tests that predict SIN ability. The first two chapters identify the outstanding questions in the field which led to the two main objectives of this thesis: exploring the inter-relations of the auditory cognitive predictors of SIN, and developing new measures of SIN perception that can better assess real-life listening and facilitate research into the link between listening and cognition. Experiments were carried out exploring the interactions of the auditory cognitive predictors of SIN perception using multivariate analyses, including auditory peripheral sensitivity, age, central hearing, auditory short-term memory, phonological working memory, and general intelligence. New listening tests were developed to better predict SIN processing. I designed a dynamic auditory figure-ground paradigm to measure an important aspect of central sound processing: the ability to segregate an auditory figure consisting of roving pure-tone segments following the pitch trajectory of natural speech from a random-frequency tone cloud. Neural responses to the dynamic figure-ground and SIN were investigated to reveal the underlying neural mechanisms of sound segregation and sustained tracking of the target stream.

I present evidence showing that central sound segregation, auditory-specific short-term memory, pure-tone audiogram, and age can explain 47% of the variance in SIN perception. Dynamic auditory figure-ground can predict both sentence- and word-level SIN better than fixed-frequency figure-ground and can be used to elicit neural entrainment generated by high-level cortical regions and the medial temporal lobe, consistent with previous literature. The peak amplitude of the entrainment response to the dynamic figure-ground correlated with SIN performance, suggesting that the neural entrainment to the dynamic figure-ground can be a biomarker for SIN ability. This work reveals important interactions between the auditory cognitive mechanisms contributing to SIN perception. It suggests new measures to predict real-life listening in both hearing disorders and brain disorders.

Table of Contents

List of Figures

List of Tables

1. Chapter 1: General Introduction

1.1 Critical determinants of speech-in-noise perception

Speech perception is often challenged with competing speech sounds (multiple speakers talking concurrently with the target speaker) or environmental sounds (air conditioning system, traffic noise, etc.). Such listening conditions are often described as “speech-in-noise”, or more colloquially, the “cocktail party problem” (Cherry, 1953). Speech-in-noise (SIN) is essential for people to perform their daily social and occupational commitments, and as with other types of hearing impairment, difficulty understanding speech could also lead to isolation or psychiatric disorders such as depression and anxiety (Rutherford et al., 2018; Scinicariello et al., 2019; Blazer & Tucci, 2019). The underlying mechanisms for SIN could also potentially explain the association between hearing loss and later-life cognitive decline (Griffiths et al., 2020).

SIN processing relies on the cooperation between the peripheral and the central auditory systems and multiple cognitive mechanisms. Sound waves are filtered and converted to electrical signals by the cochlea, a spiral-shaped cavity in the inner ear (Casale et al., 2024). The cochlea encodes sound based on the place code for frequency (tonotopic representation where the apex represents the lowest frequencies and the base the highest), the rate code for sound level, and the temporal code for accurate pitch discrimination and melody perception (Oxenham, 2013; Rutherford et al., 2021; Ehret, 2009). A healthy cochlea can detect a wide range of frequencies at a very low sound level (clinical normal threshold below 20 dB HL). Hearing thresholds, typically measured by a pure-tone audiometer, are used as a gold standard to determine hearing in clinics. However, many studies have shown that the audiogram does not tell the whole story when describing a person’s real-life listening ability (Zadeh et al., 2021; Anderson et al., 2013; Vermiglio et al., 2012; George et al., 2007). This is because the foreground and background voices often overlap over the frequencies that are represented in the cochlea and the auditory nerve. To perform sound segregation, in addition to the initial sound processing in the cochlea, the acoustic signals go through a series of segmentation and integration processes based on spectrotemporal information. These are eventually transformed into separate perceptual objects in the auditory cortex. At the cortical level, processing the SIN sounds requires the support of a complex auditory cognitive network involving memory and attention to segregate the target sound from a noisy background when they share very similar acoustic properties. This process is called auditory streaming, or “auditory scene analysis”. In the cocktail party paradigm, being able to parse complex acoustic information into meaningful auditory objects is crucial for successful speech comprehension.

Due to the variety of mechanisms involved, it is difficult to determine the root cause or causes of SIN difficulty in a patient. There is a wide range of methods available for audiologists to use in examining the auditory periphery, but resources are limited for testing high-level auditory functions. To tease apart the roles of different levels of central processing, one of the most prominent issues in the field is the need to develop measures for central auditory processing relevant to SIN perception that can be used easily in clinics and research. The primary objectives of this work is to investigate the auditory cognitive mechanisms involved in SIN processing and devise a new set of tests that could be used in clinical practice to examine real-life listening ability, and in research to investigate SIN perception and its auditory cognitive predictors. I will begin by reviewing the key concepts related to the auditory system and SIN perception, and then discuss the high-level mechanisms for SIN processing that might be affected by cognitive decline or brain disorders such as dementia.

1.2 Speech-in-noise difficulty and the auditory system

1.2.1 The peripheral auditory system

Speech perception begins with auditory signals being picked up by the peripheral auditory system (PAS). The PAS is responsible for capturing and converting sound signals into interpretable signals, it then feeds these signals into the central auditory system for further analysis. The human PAS is structured with the outer ear for sound wave collection, the middle ear for the transmission of acoustic vibration and impedance matching when air contacts the cochlear fluid (Bruns, 2021), and the inner ear that contains the cochlea for the transformation of acoustic waves into neural signals. The cochlea has around 3500 inner hair cells (IHCs) and over 12000 outer hair cells (OHCs) that respond selectively to different frequencies (NIH, 2019). The IHCs transduce sound vibrations into electrical signals for further processing, and the OHCs, powered by motor protein prestin (Zheng et al., 2000), mechanically amplify sound levels (Liberman, 2017). Dysfunctions of the PAS could lead to hearing loss which leads to difficulty understanding speech in a noisy environment. Hearing loss is often categorised into three essential types: conductive disorders caused by failure to conduct sound waves to the inner ear or resonance of the cochlear duct, sensorineural disorders caused by damaged sensory cells or cochlear neurons, or a mixture of the above which is more commonly seen (Howarth & Shone, 2006; Kelly, 2009). The causes of hearing difficulty can be genetic, age-related, noise-induced, or caused by infections, vascular insults, ototoxic drugs, etc., and they can correspond to different pathologies in the peripheral system.

Hearing loss is considered highly heritable. Pichora-Fuller & MacDonald reported that the heritability coefficients in humans range from 0.22 to 0.55 (Pichora-Fuller & MacDonald, 2009), which means genetic factors play an important role in the high prevalence of hearing loss. Due to the inevitability of environmental effects on a person's hearing, it is difficult to extract the genetic mechanism when designing a study, but mutations in the mitochondrial DNA (mtDNA) have been proposed to be a common genetic risk factor for both age-related and noise-induced damage (Ensink et al., 1998; Winston & Lei, 2023). A systematic review identified that hearing impairment caused by mtDNA deficits was mostly sensorineural hearing loss (40.8%) (Fancello et al., 2023). Sensorineural hearing loss often leads to SIN problems even when the periphery sensitivity is restored due to peripheral distortion and central temporal processing deficits (Decruy et al., 2020).

Age-related hearing loss can be attributed to several cochlear changes. Degeneration or loss of hair cells starting at the basal end of the cochlea, causing high-frequency hearing loss (Slade et al., 2020), and loss of cochlear nerve axons can reduce speech discrimination regardless of hearing sensitivity (Howarth & Shone, 2006; Peelle & Wingfield, 2016). Symptoms of age-related hearing loss are often more prominent in adverse listening conditions such as SIN conditions or rapid speech presentation. Such deficits usually affect auditory temporal processing more than spectral processing, and the effect of temporal-processing degeneration can manifest in all levels of speech processing including prosodic patterns, gap-detection, and acoustic cues such as harmonicity that contribute to periodicity (Pichora-Fuller & MacDonald, 2009). Difficulty in speech perception under temporally complex conditions could explain why some older people with normal audiograms still perform suboptimally on SIN tasks compared to their younger counterparts.

Noise-induced neural degeneration can be caused by mechanical, metabolic, or immune damage after intensive noise exposure, often eventually leading to progressive hearing loss (Natarajan et al., 2023). Threshold shifts can be temporary with effective intervention, such as halting harmful noise exposure. However, extreme acoustic intensity could lead to immediate, permanent threshold elevation, usually causing cochlear damage. This can include dysfunctions of the hair cells or their separation from the cilia, so no effective vibration is received by the hair cells. Alternatively, the basilar membrane may be separated from the hair cells, resulting in damaged sound encoding and difficulty understanding speech in challenging auditory environments (Ding et al., 2019).

Hidden Hearing Loss

Damage to synaptic connections among hair cells or cochlear neurons can happen before cell damage, which is more than often not detected by pure tone audiometry (Liberman, 2017). The study by Tremblay et al. (2015) found that people who participate in loud hobbies are more likely to have cochlear damage, such as loss of synaptic connections, without necessarily exhibiting permanent threshold elevation. This type of subclinical functional impairment is called “hidden hearing loss” (HHL). HHL is usually used to define damage to the synapses between inner hair cells and the auditory nerve fibres. SIN deficits are prominent in most accounts of HHL (Tremblay et al., 2015). While PTA is not sensitive to HHL, other measures have been used, including auditory brainstem responses (ABRs), which are responsive to dysfunctions in intensity coding, and sound-evoked auditory nerve compound action potential that can capture amplitude reduction (Kujawa & Liberman, 2009; Furman et al., 2013; Tremblay et al., 2015; Kohrman et al., 2020).

Animal studies have shown that noise overexposure or ageing primarily affects cochlear neurons, within which synaptic connections are the most susceptible to damage (Liberman & Kujawa, 2017). This is termed cochlear synaptopathy. It results from damaged presynaptic ribbons and postsynaptic nerve terminals, which lead to the disconnection of IHCs from the auditory nerve fibres (ANFs). Moreover, studies have consistently reported that noise-induced cochlear synaptopathy often selectively impacts ANFs with low to medium spontaneous rates (Hoben et al., 2017; Smith et al., 2019). Low spontaneous rate ANFs correspond to higher thresholds and a wider dynamic range (Shi et al., 2016), which are essential for parsing complex acoustic inputs. However, there is evidence supporting that synaptic damage could be reversible, but the level of recovery is still under debate (Kujawa & Liberman, 2009; Lin et al., 2011; Shi et al., 2013). While HHL was thought to be caused by loss of low spontaneous rate ANFs, more recent data suggest that recovered ANFs continue to exhibit changed functionality (Shi et al., 2016; Song et al., 2016).

1.2.2 Subcortical Neural Circuits and Acoustic Features Extraction

SIN signals must undergo processing in the ascending auditory pathway from the cochlea to the auditory cortex. Subcortical pathways play an important role in extracting sound features encoded by the PAS (Figure 1.1 for an illustration by Davies & Sugano (2020). They process spatial information (sound source) and spectrotemporal features (envelope, periodicity, fundamental frequency).

The cochlear nucleus receives signals from the cochlear nerves and is the first relay station of auditory information (Mendoza, 2011). The tonotopic organisation is retained in the cochlear nuclei to transmit the frequency information from the cochlea (Malmierca & Smith, 2009). Some specialised cells in the cochlear nucleus preserve the timing information and others encode the intensity (Winter, 2015). Regarding the processing of spectrotemporal features, ANFs first send periodicity information that evokes synchronous neural activity in the anteroventral cochlear nucleus, and this information forms the basis of pitch perception that is crucial to SIN (Anderson et al., 2010).

The superior olivary complex (SOC) is where the first major stage of binaural processing takes place and the auditory information from the cochlear nucleus converges (Walton & Burkard, 2001). SOC consists of the lateral superior olive encoding the interaural level difference, the medial superior olive encoding the interaural time differences, and the medial nucleus of the trapezoid body providing inhibitory input (Winter, 2015). The SOC is also implicated in enhancing SIN ability by reducing the disturbance of noisy signals (Sardone et al., 2019), but the effect has not been consistently demonstrated and some believe that it might be task-dependent (Mishra & Lutman, 2014; Felix et al., 2018).

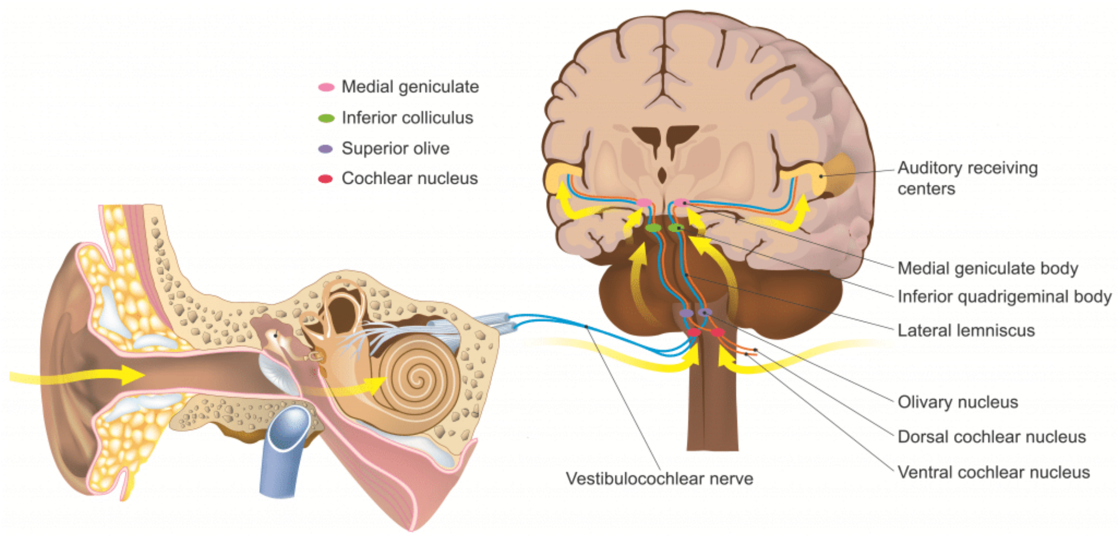

Figure 1.1 An illustration of the auditory system from the cochlea to the auditory nerve fibres, the subcortical structures and the auditory cortex. © PATTARAWIT/ Adobe Stock (#: 145672550).

SOC joins projections from the cochlear nucleus to form the lateral lemniscus tract that carries acoustic information to the inferior colliculus (Winter, 2015). The inferior colliculus is the relay centre for most ascending auditory tracts from the auditory brainstem and descending tracts from the cortex, which is critical in representing spectrotemporal features of sounds and sound localisation (Yang et al., 2020; Malmierca, 2015). In addition, the inferior colliculus has also been found to play a role in sensory prediction, decision-making, and reward prediction, which could aid SIN processing (Du et al., 2024). The inferior colliculus projects to the medial geniculate body (MGB) located in the thalamus (Alagramam & Weisz, 2023). From the inferior colliculus, the auditory pathways are either lemniscal projections with tonotopic organisations and a high-fidelity representation of acoustic features or non-lemniscal projections which have less sharp tonotopic organisation, but supply more context-dependent information (Malmierca, 2015; Anderson & Linden, 2011). The thalamocortical axons from the MGB relay information to the primary auditory cortex (Hain, 2007). The MGB actively shapes the neural representations of spectrotemporal features of sounds and changes in its structure and function could lead to various neurological disorders. For example, SIN difficulty can be caused by changes in the MGB, which have been found to precede standard neuropathological markers of Alzheimer's disease presenting SIN as one of the early symptoms (Bartlett, 2013).

1.2.3 Cortical Processing of SIN

Compared to the role of the auditory periphery and subcortical pathways in SIN, the central auditory system has been more comprehensively investigated due to its role in processing more detailed time-frequency features. However, isolating deficits caused by purely central auditory processing damage is difficult, as degraded peripheral inputs can inflict significant changes to cortical functions. Other top-down mechanisms such as executive functions and working memory also modulate the speech-tracking process, especially when the target speech is masked by noise (Pichora-Fuller et al., 2016; Alain et al., 2018). SIN processing recruits a complex neural network. In addition to the superior temporal gyrus (STG) and Broca's area, activations have also been observed in the prefrontal cortex (PFC), the left inferior frontal gyrus (IFG), and the parietal cortex (Alain et al., 2018). The human STG is involved in speech feature extraction and multisensory integration. This region is crucial for speech sound identification. It is the site where finer spectral and temporal features are encoded and integrated over a longer time frame to establish perceptual sequences (Yi et al., 2019). A meta-analysis carried out by Alain & colleagues demonstrated heightened activities in the left STG (specifically the planum temporale) elicited by SIN (Alain et al., 2018).

In addition to the conventional auditory cortex, the prefrontal cortex also aids the process of speech under adverse listening conditions. PFC is generally believed to be involved in executive functions, attention and working memory (Friedman & Robbins, 2022). When the STG is not sufficient for processing complex auditory signals, neurons of other regions such as PFC are activated to enhance speech tracking. Prefrontal activation during SIN could reflect the engagement of auditory working memory in processing detailed linguistic patterns or enhanced attention to the target sound (Alain et al., 2018). The left fronto-parietal network (IFG and inferior parietal lobe (IPL) has been shown to play an important role in effortful listening as well (Alain et al., 2018). The IFG is often indicated in predictive coding, in which the medial prefrontal cortex has been suggested to play a role in computing error signals that are subsequently passed on to the lateral prefrontal region for the generation and temporary maintenance of predictions in the dorsal-lateral prefrontal cortex (Alexander & Brown, 2018). The predictions work to constrain perception, which could benefit the process of accurately and effectively forming auditory percepts from degraded or distorted sound sources. The IPL has been associated with the cognitive processing of language (within the angular gyrus), phonological and semantic processing, and speech production (Coslett & Schwartz, 2018; Deroche et al., 2017; Brownsett & Wise, 2010).

1.2.4 Hippocampus in SIN perception

Outside the classic auditory pathways, studies have increasingly shown that the hippocampus is also involved in speech processing. The hippocampus is a key part of the medial temporal lobe (MTL) that supports learning, memory formation (Whitlock et al., 2006; Squire & Zola-Morgan, 1991), and spatial memory (Eichenbaum, 2017). A recent review identified a wide array of hippocampal functions in processing auditory stimuli, including passive listening of simple auditory stimuli, associating sound with a reward or punishment, auditory working memory, consolidation of auditory episodic memory, auditory sequence learning and prediction, pattern separation (storage of distinct activity patterns) and completion (memory retrieval based on a partial cue) in auditory scene analysis, speech perception and memory, etc. (Billig et al., 2022). Many of these functions are involved in analysing complex auditory scenes such as SIN.

The human hippocampus aids the acoustic pattern recognition, which is important for sound segregation (Kumar et al., 2016). In the visual domain, its role in pattern recognition is not only for memory but also for online perception (Mitchnick et al., 2022; Kragel et al., 2021), and the same could be true in the auditory domain as well (Billig et al., 2022). While it is unclear if the hippocampus can guide the instantaneous pattern analysis in auditory scene analysis, it certainly helps with sound grouping over time. Researchers examined the activities in the MTL with electrocorticography and found low-frequency increases in the hippocampus and parahippocampus during auditory working-memory maintenance (Kumar et al., 2021), and hippocampal involvement in the encoding, maintenance, and retrieval of auditory working memory (Kumar et al., 2016). In addition, the hippocampus is also believed to play a role in generating predictive processing of auditory sequences (Bonetti et al., 2024; Stachenfeld et al., 2017). This ability could aid in predicting incoming speech stimuli based on the structure of speech and detect deviations from expected sequences so the auditory system can adjust to dynamic changes in SIN patterns.

1.3 Theories accounting for auditory scene analysis

Auditory scene analysis is a fundamental skill of the auditory system from the auditory periphery to the cortex that groups a complex auditory signal into perceptually meaningful objects. When the incoming auditory stimuli are speech, this process is called SIN perception. Auditory scene analysis engages not only a bottom-up process that encodes the acoustic features of incoming sounds (frequency regularities, temporal synchrony, timbre, harmonicity, etc.) but also top-down mechanisms including pattern recognition facilitated by learning as well as attention (Bregman, 1994). The principles of grouping in Gestalt psychology (Köhler, 1967) have been mainly used to account for visual perceptual grouping, but they can also be applied to auditory perceptual organisation (Chakrabarty & Elhilali, 2019; Chen, 2005). Some of the principles that can be used to explain auditory grouping include figure-ground articulation, proximity, common fate, similarity, and continuity. Traditionally, auditory segregation has been categorised into two types. For auditory cues that occur at the same time, auditory segregation is based on the commonality of the sound stream's onset/offset, which exploits a grouping mechanism termed "common fate" by Gestalt laws. For auditory cues that start from a different time but remain constant over time, for example, a sequential sound with consistent frequency, segregation is based on the commonality of acoustic features, or "proximity" in Gestalt laws. Much work has been carried out to unify the two types of processes and develop a general model (see review Gutschalk & Dykstra, 2014; Kwak & Han, 2020). Here, existing models of the auditory periphery and high-level grouping mechanisms that best accommodate current psychophysical and neurophysiological data are discussed.

A classic paradigm used to probe auditory streaming is a sequence of alternating high- and low-frequency tones that can be perceived as two streams (Figure 1.2(a)) (Bregman, 1994). Fishman and colleagues proposed that adaptation within frequency bands could explain auditory stream segregation (Fishman et al., 2014, 2001). They found that increased frequency separation, presentation rate, and duration of the tones enhanced spatial differentiation of the neural responses to the tones along the tonotopic map in the primary auditory cortex (A1) (Fishman et al., 2014). Their model of streaming can account for whether Streams A and B (Figure 1.2(a)) are perceived as one or two separate streams. When the frequency differences between A and B are wide, the wide spatial separation along the tonotopic map means that these sounds do not overlap regardless of their presentation rate or duration. When the frequency differences between the two alternating sounds are moderate, the neural responses overlap in places. Neural adaptation due to fast presentation rates can aid sound segregation by generating distinct activity patterns for both sounds. When the frequency differences between the two sounds are small, they are always perceived as a single stream. However, researchers discovered that the tonotopic representation in the A1 alone was not sufficient to account for auditory streaming especially when the frequencies overlap (Elhilali & Shamma, 2008). Psychophysical data of human participants showed a significant perceptual difference between synchronous and asynchronous sounds, while animal work showed that neural responses of A1 were independent of such discrepancy. This means that in addition to the tonotopic representation of sounds, the streaming process must rely on other mechanisms. Researchers thus proposed an auditory streaming model that incorporates temporal coherence analysis within each individual stream as a central step for auditory scene analysis (Elhilali & Shamma, 2008; Shamma et al., 2013; Shamma et al., 2011). Coherence was defined as the "average similarity or coincidence of their (different channels) responses measured over a given time window" (Shamma et al., 2011, p.10). The temporal coherence theory suggests a two-stage process in the auditory system: feature analysis and coherence analysis. First, the auditory periphery picks up sound waves formed by auditory signals from various sources, and then the cochlea filters the sound waveforms and converts them into firing patterns across a range of neurons representing different spectral frequencies. After the extraction of the basic acoustic features (such as pitch, timbre, and loudness), the central auditory system in the second stage computes the correlation between feature-selective neurons and groups the neurons with similar temporal firing patterns together (Shamma et al., 2011). In addition to the feed-forward processing, temporal coherence analysis also suggests that selective attention modulates stream formation via tuning neural responses to certain acoustic features or enhancing neural synchrony of different neural populations (Niebur et al., 2002).

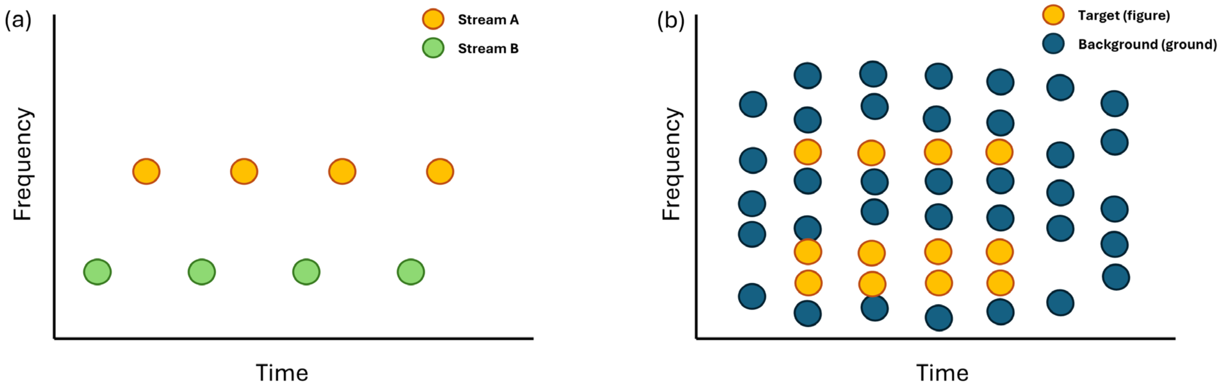

Figure 1.2 A schematic representation of different streaming paradigms. Figure 1.2(a) illustrates the alternating-tone task used to probe streaming based on neural adaptation. Figure 1.2(b) illustrates the auditory figure-ground paradigm used to probe streaming based on temporal coherence.

temporal coherence model was called “stochastic figure-ground (SFG)” (Teki et al., 2013). This paradigm consists of a figure made of repeating pure-tone elements and a ground of randomised frequency elements (Figure 1.2(b)). The temporal coherence analysis based on the SFG paradigm postulates that a coherence matrix is generated across all channels of the spectrogram of each stimulus, and the stimulus containing a higher coherence level (number of elements in a single time frame) would also present higher cross-correlation values compared to a background of randomised channels. The detection of the auditory figure was shown to recruit the temporal coherence mechanism (Teki et al., 2013). The researchers modelled the responses at the auditory cortex and calculated the temporal coherence as the difference between the mean of the maximum cross-correlation for the target and the background. They found that the temporal coherence increased with an increasing length of figure duration (Figure 1.3, Teki et al., 2013). This is very similar to figure-detection performance in human participants, whose detection sensitivity increased with increasing coherence level and chord length (Teki et al., 2013).

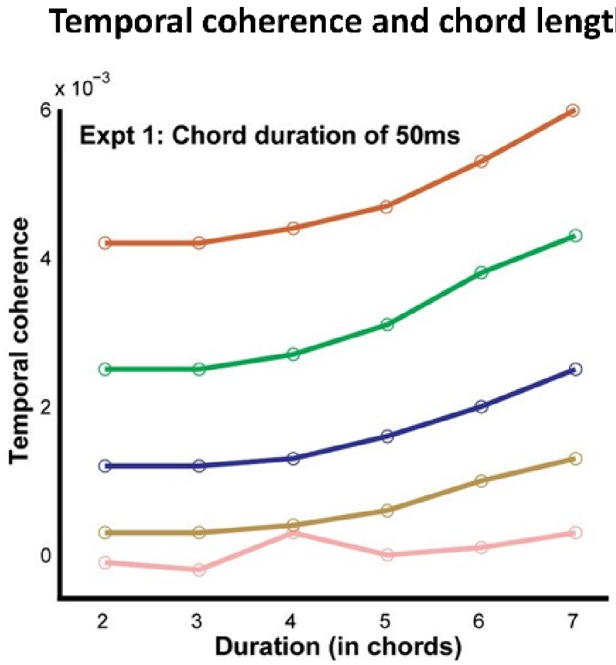

Figure 1.3 Temporal coherence and SFG (the images were taken from Teki et al., 2013)The line graph shows the relationship between temporal coherence and duration in chords (the number of pure-tone elements over time). The different line colours were used to create a higher visual contrast. Each line represented a different combination of temporal coherence, as marked by the y-axis, and duration, as marked by the x-axis.

In terms of the neural correlates of auditory scene analysis, Gutschalk & Dykstra (2014) reviewed human neurophysiological data acquired from a variety of pure-tone elements masked by random multi-tone masker paradigms (Gutschalk & Dykstra, 2014). The results were consistent with the multi-layered processing proposed by the temporal coherence analysis model. The initial 80 ms of brain oscillations were shown to be associated with the processing of spectral features, which corresponds to the first stage of temporal coherence analysis. Cortical-evoked potentials, including the negativity at 75-200 ms evoked by identifying auditory targets, the enhanced N1 component, or the object-related negativity during sound segregation, have been reported to be less attributable to specific acoustic features but more to the perception of the target stream (Gutschalk & Dykstra, 2014). The auditory-evoked N1 responses have been thought to originate from the non-primary auditory cortex, which has been reported to support auditory streaming in the cortex (Snyder & Alain, 2007; Gutschalk et al., 2005). A recent fMRI study reported activities in the early auditory cortex (A1), which were modulated by task difficulty in a complex auditory figure-ground task (temporally consistent pure-tone elements masked by a tone cloud) (Holmes et al., 2021). Neuroimaging findings using a similar paradigm showed significant bilateral activations in the superior temporal sulcus (STS) and intraparietal sulcus (IPS) as an effect of increased temporal coherence (Teki et al., 2011). High-level auditory cortical responses consistent with Teki's finding were also found in an experiment with non-human primates, where three macaque monkeys were trained to detect the target sound (Schneider et al., 2018). The researchers found similar perceptual patterns in monkeys as in human subjects, and they also found activation in the rostral belt and parabelt (like the high-level auditory cortex in humans) in functional imaging. The IPS is another crucial region in auditory scene analysis, which is often associated with sensory integration and top-down attention modulation in auditory streaming (Cusack, 2005). While it is outside the conventional auditory area, a strong activation was shown during the segregation of sound streams, which was considered to reflect strong perceptual "pop-out" during auditory scene analysis (Teki et al., 2011).

In conclusion, auditory scene analysis has been hypothesised to rely on the analysis of acoustic features based on adaptation within frequency bands in the A1. Grouping more complex auditory stimuli could employ temporal coherence analysis engaging high-level cortical mechanisms located in the non-primary auditory cortex and the parietal lobe.

1.4 Cognitive mechanisms related to SIN: hearing and cognition

The intricate relationship between auditory processing and cognition has drawn attention to the question of a potential causal link between hearing and cognition. Epidemiological studies have long reported that hearing loss is associated with cognitive decline. A longitudinal study exploring the connection between peripheral hearing and cognition found that hearing loss was independently correlated with cognitive decline in older adults (Lin et al., 2012, 2013). The researchers recruited 1984 participants in total who were followed for 6 years. The results showed that people with hearing loss not only had a 41% greater annual rate of cognitive decline, but the severity of hearing loss at baseline was also positively correlated with the acceleration of the cognitive decline. A meta-analysis found significant associations between age-related hearing loss and dementia, as well as a small association between hearing loss and cognitive functions (global cognition, executive function, episodic memory, processing speed, semantic memory, and visuospatial skills) (Loughrey et al., 2018).

They hypothesised that the link between age-related hearing loss and cognitive decline was most likely due to a common aetiology, such as vascular disease. They also reported increased short-term memory and executive function recruitment in hearing-impaired individuals. The change in cognitive functions could be a form of compensation for hearing loss as the reallocation of cognitive resources can negatively influence general cognitive ability, causing a decline in processing speed and memory. Consistent with this finding, a longitudinal study (Merten et al., 2020), assessing hearing sensitivity (measured by PTA), SIN perception (measured by Word Recognition in Competing Message), and cognition (Trail-Making Test (TMT) among 1274 middle-aged adults, found a small effect of SIN on TMT scores and a non-directional association between PTA and TMT scores.

One prominent issue with the study of the relationship between cognition and hearing is that it is difficult to eliminate confounding factors. In addition to general cognitive abilities, there are other associated variables relevant to speech processing in noise. With regard to demographic information, for example, studies found that hearing loss seems to be more prevalent among populations that are older, white, male, smokers, or diagnosed with depression (Lin et al., 2013; Tremblay et al., 2015). Other risk factors such as dysfunctions in the metabolic system (Sun et al., 2015) or the immune system (Chaitidis et al., 2020), lower education level, high LDL cholesterol, physical inactivity, air pollution, visual loss, etc. (Livingston et al., 2024), have also been identified. A large-scale experiment carried out by Tremblay et al. (2015) looked into the risk factors for SIN difficulty in an adult population of 686 aged from 21 to 84, and discovered that self-reported listening difficulty was related to mental health status such as depression and medication history, clinical consultations regarding ear infections, as well as neuropathy-type symptoms. Participants who reported listening difficulty also shared a higher likelihood of reporting symptoms of peripheral neuropathy such as numbness, imbalance, and temporary loss of sensation and/or depression, especially among people with impaired visual function. While cross-modality studies have provided evidence for audiovisual interactions for age-related hearing loss (Bishop & Miller, 2009), mental health and peripheral neuropathy fall outside the scope of audiology and have not been rigorously studied with auditory processing. It is therefore important for researchers to be aware of such potential confounds when conducting experiments.

In addition to age-related cognitive decline, hearing loss has also been associated with dementia. Dementia is not the same as age-related cognitive decline, but a group of pathological brain conditions affecting cognition that include Alzheimer's disease and vascular pathology. Studies of the association between hearing loss and dementia have mainly examined 'all-cause dementia'. Two recent Lancet reviews reported that hearing loss in mid-life could potentially account for approximately 8% - 9% of total dementia cases and is likely to be the largest modifiable risk factor (about 9.1%) of dementia (Livingston et al., 2024, 2017). Longer exposure to hearing loss was found to relate to an increased risk of dementia (Ford et al., 2018), in which men with hearing loss had a 69% higher hazard of developing dementia than those without. This could indicate that there is a potential causal link between hearing loss and dementia. However, the use of hearing aids did not reduce cognitive decline over a 3-year period (Lin et al., 2023), suggesting that the association might rest somewhere higher than the auditory periphery. SIN impairment has been suggested to predict a 61% increased risk of dementia in a large-scale UK Biobank study (n=82039; followed up for a median of 10 years) (Stevenson et al., 2021). In a cross-sectional study, SIN was found to have a stronger association with cognitive function than PTA with cognition (Hoff et al., 2023). Moreover, when comparing both age and hearing-matched participants with or without mild cognitive decline, a significant difference was found in their ability to process SIN (Mamo & Helfer, 2021). These studies suggest that the association between hearing loss and dementia might not be attributed to peripheral hearing threshold elevation alone, and central involvement as measured with SIN tasks could be more predictive. However, research on SIN and dementia is extremely limited. This research gap was also flagged by the most recent Lancet review on the risk factors of dementia (Livingston et al., 2024).

Hypotheses linking hearing and cognitive decline

Different theories have been proposed in an attempt to explain the relationship between age-related hearing loss and cognitive decline or dementia development, among which the most prominent ones are: a. the common cause hypothesis, b. the sensory deprivation hypothesis, and c. the information degradation hypothesis (Merten et al., 2020; Pronk et al., 2019; Roberts & Allen, 2016). There are also evolving theories discussing genetic connections between hearing loss and dementia or links between SIN-specific hearing loss and dementia (Griffiths et al., 2020).

Firstly, the common cause hypothesis suggests that both age-related perceptual failure and cognitive deterioration are driven by a third common factor (Lindenberger & Baltes, 1994). This cause could be the functional decline of the brain or a common pathology affecting both the peripheral auditory system and cortical regions related to auditory processing and general cognition. Christensen and colleagues used factor analysis modelled by 10 perceptual and cognitive variables and identified a common cause factor that could reflect "conscious understanding" or some form of ageing (Christensen et al., 2001). Regarding research on Alzheimer's disease (AD), an association between AD and cochlear pathology seems to exist in early-onset hearing loss, and pathological changes of AD also exist in the auditory pathway and auditory cortex (Griffiths et al., 2020).

The sensory deprivation hypothesis assumes that a lack of stimulus input caused by deficits in the peripheral system precedes and causes cognitive decline. The hypothesis suggests that people with hearing loss are more likely to live in social isolation, which would in turn greatly reduce auditory input as well as other forms of sensory input. Although Stevenson et al., (2021) found that depression and social isolation alone did not mediate the relationship between hearing loss and dementia in a large-scale Biobank study. Researchers believe that it takes a prolonged span of time for sensory deprivation to cause salient structural changes to the brain and its cognitive functions but the deterioration in cognition is inevitable (Uchida et al., 2019). Research on sensory deficits (mainly early-life or congenital sensory deficits) has demonstrated the possibility of cortical structural changes, for instance in white matter integrity or connectivity, or functional rearrangements due to early-life sensory deprivation. A study on patients with congenital olfactory deprivation showed alterations in the secondary but not the primary olfactory cortex (Peter et al., 2020). Studies on the auditory system (Hribar et al., 2014; Lazard et al., 2014) discovered functional rearrangements and multiple structural changes in the superior temporal gyrus, Heschl's gyrus, and the planum temporale for post-lingual deafness as well as congenital deafness. Fine and Park also pointed out in their review of visual studies that while the occipital lobe showed very subtle reorganisations in people with early blindness, novel functional responses such as tactile, auditory, working memory, language, and mathematics were prominent (Fine & Park, 2018).

The information degradation hypothesis is similar to the sensory deprivation hypothesis in the sense that they both assume degraded inputs (caused by compromised peripheral processing or auditory masking) would lead to perceptual failure which then affects the cognitive system, except the latter supports a relatively immediate effect of degraded input on cognitive processing (Monge & Madden, 2016). Compromised sensory input requires more resources such as attention, working memory, and executive function to be allocated to perceptual processing as a form of compensation, which could subsequently cause suboptimal cognitive performance due to a diversion of these resources from other roles. A study by Gilmore et al. (2006) supported the information degradation hypothesis, in which even young adults were significantly influenced by degraded visual stimuli instantaneously. However, when comparing young adults with an equivalent degree of hearing loss with their older counterparts, younger people still performed better in cognitive tasks, suggesting that degraded speech input could not be the sole factor responsible for poor cognitive outcomes (Gordon & Fitzgibbons, 1997).

A model focusing on the role of the MTL proposed an interaction between brain activity related to auditory cognition and dementia pathology (Griffiths et al. 2020). This model stems from a similar idea as the information degradation theory but assumes that the altered cortical activity interacts with AD pathology. As described in section 1.2.2, the hippocampus is involved in auditory pattern analysis, especially for novel sounds or SIN stimuli (Billig et al., 2022; Griffiths et al., 2020). The MTL is also the region where the earliest neurofibrillary changes in typical AD show (Teipel et al., 2013; Xie et al., 2018). Under this hypothesis, there are two possible types of interaction that could explain the link between hearing loss and cognitive decline: 1. heightened engagement of MTL during effortful listening increases AD pathology; 2. AD pathology leads to altered neural activity in MTL, which causes excitotoxic neuronal degeneration (Griffiths et al., 2020). Further animal work is needed to determine the direction of causation. While this theory mainly accounts for AD development, it could also explain links between hearing loss and general cognitive decline as the MTL is important for both SIN perception and general cognition (Section 1.2.4).

Recently, many groups have looked into the genetic relationship between hearing loss and cognitive decline (Sarant et al., 2020). Sarant & colleagues investigated how genetic risk factors for hearing impairment and cognitive disorders such as AD might interact or influence each other. While no causal links have been found yet, they did find genetic correlations between hearing loss and AD, and that genetic risk factors for AD also influence speech-in-noise perception (Brenowitz et al., 2020; Mitchell et al., 2020).

While these hypotheses are by no means mutually exclusive, they could lead to different intervention strategies. For instance, if it were indeed genetic reasons that establish the link, hearing loss itself would no longer be considered a modifiable factor for dementia. As mentioned previously, the most salient issue is that most studies working on the interactions of sensory system and cognition are correlational and cross-sectional. More longitudinal studies with larger sample sizes and long-term interventional studies might be conducive to uncovering the potential causal connection. Most of the above-mentioned models involve an interaction between SIN pathology and AD pathology. This is in line with the findings of many of the studies cited in this section, in which SIN was found to be an independent predictor for dementia. It is essential to break down hearing into stages such as peripheral hearing loss and SIN hearing loss to better understand the main predictor of the relationship between hearing loss and cognitive decline or dementia. Research into fundamental determinants of SIN could also potentially aid dementia treatment.

2. Chapter 2: Evaluation of Hearing Tests that Predict Real-life Listening

2.1 Introduction

As I have established through the previous chapter, the auditory pathways form a highly intricate system, where dysfunction can appear at various levels while presenting very similar symptoms. This means that hearing problems can be difficult to test. A wide range of hearing tests that can be indicative of real-life listening are available now. Some commonly used tests include pure-tone audiometry for peripheral hearing sensitivity, tympanometry and stapedial reflexes for the middle ear function, and otoacoustic emissions (OAES) for the hair cell functions. Tests used to evaluate neural transmission from the cochlea to the brainstem and the primary auditory cortex are more often used now too, including the auditory brainstem response (ABR), auditory middle latency response (MLR), and frequency following response (FFR). In recent years, speech audiometry is also used to assess a person's real-world listening ability. These are predominantly sentence-in-babble tests such as the QuickSIN (Killion et al., 2004), LiSN-S (Cameron & Dillon, 2007), BKB-SIN (Etymotic Research, 2005), Hearing in Noise Test (HINT, Nilsson et al., 1994), AzBio Sentences in Noise (Spahr et al., 2012), but can also be word-in-noise tests such as the WIN test (Wilson, 2003). To assess hearing sensitivity based on neural responses, auditory steady-state response (ASSR) can be used; cortical auditory evoked potentials (CAEPs) are also available for more specialised testing. Outside of clinical practice, numerous tests for central hearing and non-speech measures for central sound processing have been devised to assess real-world listening ability as well such as auditory figure-ground tests (Guo et al., 2022; Teki et al., 2013), auditory short-term memory tests (Lad et al., 2020a), gap detection tests (Phillips et al., 1997) and various self-assessed measures such as the Speech, Spatial and Qualities of Hearing Scale (SSQ, Gatehouse & Noble, 2004).

All of the above-mentioned measures can be roughly categorised into objective verbal tests, non-verbal tests, and subjective tests. While most of the behavioural tests such as the PTA and sentence-in-babble tests, can be conceived as being subjective in the sense that they rely on the personal response/assessment of hearing ability, most studies categorise these as objective measures. In this Chapter, subjective and objective measures refer to self-evaluated and performance-based/physiological measures, respectively. The objective verbal SIN tests focus on speech recognition and comprehension, which rely on both the language and cognitive domains. Non-verbal measures assess the functionality of the auditory pathways important for sound processing that is independent of language ability but crucial for language learning, comprehension, and production. They interact with cognitive abilities, the most important of which include auditory working memory, general intelligence, and attention. Subjective measures refer to patients' self-evaluation of their real-life listening experience assessed through questionnaires. For diagnostic purposes and obtaining reliable performance in research, objective measures are generally preferred.

Recent reviews have identified most of the commonly used verbal SIN tests in clinics for paediatric practices (Sanchez et al., 2022) and for French speakers (Reynard et al., 2022), but tests of hearing and listening functions that predict SIN perception are yet to be reviewed. In this review, I will evaluate the tests for real-life listening ability, focusing on their ability to predict SIN perception while addressing the issues in application or task development. The aim is to provide a comprehensive perspective of SIN testing including methods that are not often used and provide considerations for the further development of relevant hearing tests.

2.2 Behavioural methods: Non-verbal Measures of Speech-in-noise

As I have discussed previously, a lot of behavioural measures of SIN perception are non-verbal, which have fewer restrictions on the patient's age, language, and educational background. This section reviews the behavioural hearing tests with non-verbal stimuli where a relationship with SIN performance has been established or investigated.

2.2.1 Pure-tone audiogram and speech-in-noise

Pure-tone audiogram (PTA) has been used as the most common test for audiological practice. It reflects the perceptual sensitivity at 0.25-8 kHz and has been used extensively both in clinics and research as the primary hearing screening tool. However, PTA does not fully explain SIN perception. While responses to pure tones travel up to the primary auditory cortex, the audiogram can only accurately provide information on peripheral sensitivity (Musiek et al., 2017) and does not necessarily predict SIN listening. Füllgrabe et al. (2015) reported that real-life listening deteriorates with ageing regardless of hearing sensitivity, suggesting that the two aspects of listening ability — periphery and central, might not fully align. More and more studies are reporting the essential role extended high-frequency (EHF) plays in predicting SIN function in some circumstances (Polspoel et al., 2022; Zadeh et al., 2019). However, similarly to the normal-range audiogram, EHF is not always found to be predictive of SIN. To further explore the strength of the relationship between PTA, either in standard frequency or extended high-frequency, and SIN performance, I have conducted a review of the literature on the effect of PTA on SIN ability. The review focused on obtaining a group estimator of the strength of the PTA-SIN relationship and evaluated the effect of age, hearing, and sample size on this relationship.

Methods

A database search with the PubMed default timescale setting (last update before August 2024) with the search terms "*(("PTA") OR ("pure tone audiogram")) AND (speech in noise) NOT (review)*" revealed 218 independent studies on PubMed looking into the link between standard-frequency PTA (SF-PTA) and different speech measures of real-world listening. Ninety papers were selected for full-text screening after the title and abstract screening, 14 of which were eventually deemed relevant for the topic with most of the required information reported. This includes sample size, correlation coefficient or other comparable metrics (standardised effect or r squared), and relevant demographic features (PTA and age), which were extracted for data analysis. The extracted data are summarised in Table 2.1.

The same selection procedure was used to examine extended high-frequency audiometry (EHF-PTA) and SIN (Table 2.1), with search words: (("extended high-frequency PTA") OR ("extended high-frequency pure tone audiogram")) AND (speech in noise) NOT (review). This added another 40 papers to the previous search on standard PTA, totalling 258 papers for screening. Nine papers were identified out of 40 that focused on extended high-frequency PTA and reported the necessary data for this review.

The information retrieved from relevant publications includes the correlation coefficient or r-squared that quantifies the relationship between PTA and SIN, age, hearing ability, and sample size. The coefficient, age and hearing ability were visualised in the 3-D scatter plots in Figure 2.1 by taking the averaged age and the maximum PTA threshold of the inclusion criteria as the x- and y- axes and the correlation coefficients of the SIN-PTA relationship as the z-axis. For studies that did not report criteria on hearing sensitivity, 15 dB and 50 dB are plotted as 'normal hearing' and 'normal to severe hearing loss'. The two numbers were chosen randomly to represent the averaged audiogram of the populations that could be characterised as having 'normal hearing' or 'normal to severe hearing loss' based on the guidelines of British Society of Audiology, which defined normal hearing as below 21 dB HL, mild hearing loss as 21- 40 dB HL, moderate hearing loss as 41 – 70 dB HL, and severe hearing loss as 71-95 dB HL (British Society of Audiology, 2018). For studies that did not report an average age, the mean of the reported range was used to plot the bubble plot. The size of the bubble is scaled by the effect size. For studies that reported only r-squared values, they were transformed into r-values by taking the square roots. It is important to note that Figure 2.1 is intended to provide an intuitive illustration of the data only and it does not reflect real data accurately due to the lack of descriptive data on hearing sensitivity and age.

To evaluate the effect on a group level, the total score was calculated as the mean of the coefficients on the studies identified. Confidence intervals (95%) were calculated based on the sample size and the absolute effect size of the studies using the metafor package in R version 4.4.1. The result was plotted as forest plots for PTA and EHF-PTA respectively in Figure 2.2. The impact of age was further investigated with a post-hoc meta-regression analysis using the restricted maximum likelihood method. The dependent variable was the effect size, the moderator was age, and variance estimates were used to weight the studies. The analysis tested whether age significantly influenced the effect sizes of the relationship between PTA or EHF-PTA and SIN.

The type of speech materials could also impact the association between PTA and speech recognition in noise. A cross-sectional study by Wilson et al. (2007) compared some of the most frequently used SIN tests: BKB-SIN, HINT, QuickSIN, and WIN and found that tests with lower semantic context showed a stronger association with the pure-tone thresholds, e.g. the WIN test showed the strongest correlation with PTA. Therefore, an independent sample t-test was conducted to examine if the type of speech materials can influence the strength of the relationship between SF-PTA/EHF-PTA and SIN measures. This test was carried out with the entire dataset including both the SF and EHF reports. The single-word tests and DiN tests are categorised into one group, and the sentence tests are into another group.

Results and discussion

The overall effect size for PTA and SIN was numerically larger than EHF-PTA but there was no significant difference between the strength of the relationship between SF-PTA and SIN compared to EHF-PTA and SIN (p = 0.391). The result suggested that EHF-PTA might not predict SIN better than standard PTA. However, the sample of the EHF studies consisted mainly of younger people below 50 years old with good hearing. As shown in Table 2.1, hearing ability seems an essential factor that modulates the relationship between SIN perception and hearing sensitivity for the standard frequencies. People with normal hearing tend to show non-significant correlations or relatively small to medium significant effects between PTA and SIN. It is possible that if the impact of hearing sensitivity were removed from the analysis, EHF might show better predictive power than SF PTA. However, this analysis was impossible as descriptive results of PTA were not reported.

Article	Correlation Coefficient	Hearing sensitivity	Age	Sample Size	Speech materials
Moore et al., 2020	0.188***	Mostly normal	R: 6~11	1457	VCV pseudoword in speech-modulated noise
Jansen et al., 2014	0.670*	Normal to severe HL	R: 22~59	118	CVC in speech-shaped noise
Wong et al., 2008	0.770**	Normal to profound HL	M: 44.7 (SD: 13.5)	30	HINT
George et al., 2007	0.710***	<60 dB (HI)	M: 65.5 (R: 46~81)	21	Sentence in stationary noise
George et al., 2007	0.39(ns)	<15 dB (NH)	M: 63.5 (R: 53~78)	13	Sentence in stationary noise
Merten et al., 2022	0.250*	M: 13.9 (SD: 9.3)	M: 55 (SD: 14)	2585	SiB
Borch Petersen et al., 2016	R^2=0.101**	M: 65.3 (SD: 12.2)	M: 52.6 (SD:11.4)	283	Sentence in noise
Wilson, 2011	0.750*	<20 dB	M: 62.3 (R: 20~89)	3143	WIN
Bochner et al., 2015	-0.593***	M: 41.47 dB (SD: 21.22), 0.5-2 kHz	M: 62.4 (SD: 20.8)	70	SiB
Bochner et al., 2015	-0.600***	M: 60.14 dB HL (SD = 19.99), 2-8 kHz	M: 62.4 (SD: 20.8)	70	SiB
Bochner et al., 2015	0.633***	M: 41.47 dB (SD: 21.22), 0.5-2 kHz	M: 62.4 (SD: 20.8)	70	QuickSIN
Bochner et al., 2015	0.768***	M: 60.14 dB HL (SD = 19.99), 2-8 kHz	M: 62.4 (SD: 20.8)	70	QuickSIN
Anderson et al., 2013	0.118(ns),	≤ 45 dB	M: 63.89 (SD: 4.83)	120	QuickSIN, WIN, HINT

	0.103(ns), 0.112(ns)				
Besser et al., 2015	0.39*	<25 dB	M: 72.0 (SD: 4.3)	26	LiSN-S
Besser et al., 2015	−0.53**	<25 dB	M:21.7 (SD: 2.6)	26	LiSN-S
Diedesch et al., 2021	R^2 = 0.467*	<20 dB	M: 21.3 (SD: 2.5)	16	QuickSIN
Vermiglio et al., 2012	Ns (0.5_2 kHz), 0.41* (3-6kHz), 0.37*(5-6kHz)	<25 dB	M: 32.78 (R: 10.71)	215	HINT
Vermiglio & Fang, 2021	0.002(ns)	<20 dB	M: 31.82 (SD:10.16)	325	HINT
Zadeh et al., 2021	0.49/0.50***	>20 (HI)	M: 54.2 (SD: 9.2)	40	DIN
Zadeh et al., 2021	ns	≤20 (NH)	M: 29.4 (SD: 10.2)	70	DIN
Extended High-Frequency Audiometry					
Trine & Monson, 2020	0.320*	<25 dB	M: 21.3 (R:19 - 25)	41	SiB
Smith et al., 2019	R^2 = 0.013(ns)	0.25–8 kHz; ≤ 20 dB HL ;>8kHz at <10dB	M: 22.56 (R: 18 – 30)	194	QuickSIN
Besser et al., 2015	0.72**	<25 dB	M: 72.0 (SD: 4.3)	26	LiSN-S
Besser et al., 2015	0.09(ns)	<25 dB	M: 21.7 (SD: 2.6)	26	LiSN-S
Çolak et al., 2024	0.634***	≤ 20 dB	M:24.44 (R:19 -34)	32	SIB
Ananthanarayana et al., 2024	0.39*	≤25 dB	M: 21.1 (R:18 - 33)	37	SiB
Drennan, 2022	0.30(significant)	Normal hearing	R:18 - 72	119	WIN
Zadeh et al., 2019	0.38**	≤20 dB	M: 29.5 (SD = 9.1)	116	DIN (broadband noise)
Zadeh et al., 2019	0.17(ns)	≤20 dB	M: 29.5 (SD = 9.1)	116	DIN (lowpass filter)
Zadeh et al., 2021	0.50***	> 20 HL (HI)	M: 54.2 (SD: 9.2)	40	DIN
Polspoel et al., 2022	-0.48*, -0.51*	≤20 dB	R:20-26	24	CVC, SiB

Table 2.1 Relationship between PTA and SIN. The negative correlations are from studies using adaptive SIN tests and PTA, where a higher score indicated lower performance. Three asterisks (***) denote the significance level at p<0.001, two represent p<0.01, and one asterisk represents p<0.05. The PTA results of each study are extracted from participants' inclusion criteria. Age is reported either as range (R), mean (M), or standard deviation (SD). The correlation coefficients marked as ‘ns’ are nonsignificant. HI: hearing impaired. NH: normal hearing. SiB: sentence-in-babble test. HINT: Hearing in Noise Test. LiSN-S: Listen in Spatialized Noise. QuickSIN: Quick Speech-in-Noise. DIN: digit-in-noise. VCV: vowel-consonant-vowel. CVC: consonant-vowel-consonant.

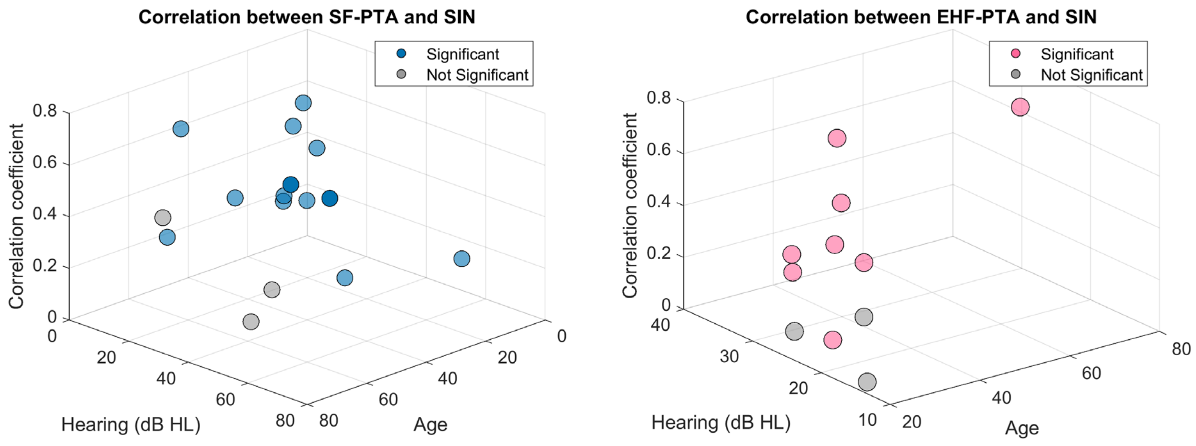

Figure 2.1 The impact of age and hearing sensitivity on the relationship between audiogram results and SIN. The x-axis plots the average age, and the y-axis shows the upper limit of the PTA inclusion criteria. The z-axis plots the effect size of the PTA-SIN relationship. The bright blue dots on the left report the correlation between SF-PTA and SIN. The dark grey bubble reports the nonsignificant correlation coefficients. The bright pink on the right shows the significant correlation coefficients between EHF-PTA and SIN, and the dark grey shows the nonsignificant ones.

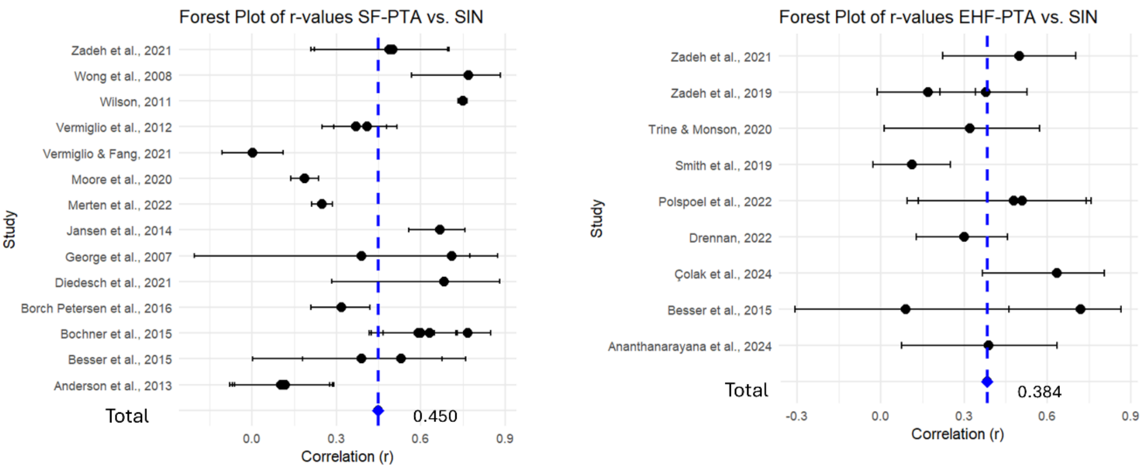

Figure 2.2 Forest plot of PTA or EHF-PTA and SIN. The absolute values of the correlations are plotted here. The individual data are marked by black dots with confidence intervals in black bars. The average result is marked by a blue diamond with the exact number marked next to it. The dotted vertical line marks the average correlation coefficient. Studies reporting multiple r values are plotted within the same line.

On the other hand, age significantly affected the association between EHF-PTA and SIN. The regression results revealed a small but significant effect of age on the EHF-SIN relationship (β= 0.008, p = 0.031), but not of age on the standard PTA-SIN relationship (β= 0.001, p = 0.792). This suggests that EHF is more sensitive to age-related changes in the PTA-SIN relationship. As EHF captures the variation of hearing better for healthy young people who do not have age-related high-frequency loss, this result aligns with the expectation.

Another factor that impacted the coefficient size reported is the sample size. A correlation test between the correlation coefficient and sample size revealed a negative correlation between the size of the coefficients of the PTA-SIN relationship (both SF and EHF combined) and the corresponding sample size (rho = -0.431, p = 0.018). In other words, datasets with larger sample sizes, such as Moore et al., 2020, tended to report a weaker relationship between PTA and SIN. This raises a serious issue with data validity. Certain studies could be underpowered to provide scientifically reliable results, which drove the overall effect size to higher when the actual strength between PTA and SIN could still be lower than what I synthesised in this report (around 0.4).

Finally, I found that the type of stimuli had no impact on the PTA vs. SIN relationship. The independent sample t-test showed a non-significant mean difference between the word measures and the sentence measures (p = 0.787).

Conclusion

Overall, PTA results can be used as an indicator of speech-in-noise performance, but this predictive relationship tends to be weaker in younger people when testing for EHF hearing. Having a large sample size does not always guarantee a strong association in this instance but too small a sample size could lead to potentially spurious results or insufficient power to detect any relationship between PTA and SIN. Single-word or sentence tests did not differ in their association with PTA.

This review also identifies an important research problem in reporting non-significant results. Some studies did not report non-significant results, and those that did (as shown in Table 2.1) did not always provide specific coefficients. Hence these non-significant findings could not be synthesised with others for a systematic comparison. Research on the extended-high frequency PTA is also very limited to young people with normal hearing and more studies need to be carried out with a wider range of populations. All of the above factors would influence the meta-analysis greatly, rendering the results of the analysis less reliable.

2.2.2 Temporal processing

In addition to losing hearing sensitivity, compromised frequency or temporal resolution could also lead to poor listening ability (Bramhall et al., 2019). Temporal processing is an important part of SIN processing. Indeed, it is thought to underlie most auditory processing capacities (Shinn, 2003). Temporal processing is broadly defined as the perception of time-related aspects of sound, including temporal resolution, sequencing, integration and masking. The behavioural measures used most in relation to SIN performance are those measuring temporal resolution and temporal ordering.

Temporal resolution

Temporal resolution has been commonly measured with temporal modulation transfer function (TMTF) and gap detection (GDT). Gap detection measures the shortest time possible to discriminate between two sounds, and has been shown to reflect real-world listening abilities (Blankenship et al., 2022; Heeke et al., 2018). However, researchers found low behavioural performance consistency between the two tests, which means that TMTF and GDT could potentially measure different processes (Shen, 2014; Shen & Richards, 2013). GDT stimulus is usually comprised of pure tones or broadband noises. The two sounds can have the same frequency ranges to form a within-channel gap detection task or different frequency ranges to form a between-channel GDT task. Phillips et al. (1997) demonstrated that the two tasks reflected different processes. The within-channel detection was considered to be discontinuity detection, which could be performed at the peripheral level by the same set of perceptual channels activated by the stimulus. Whereas the between-channel was theorised to engage more complex central sound processing when the underlying perceptual timing operation required cross-channel comparison. The between-channel gap detection task was hence hypothesised to be more relevant to speech perception (Phillips & Smith, 2004). To investigate if the two paradigms engaged different mechanisms, an EEG study investigated them with an event-related design and found a significantly higher amplitude for between-channel detection compared to the within-channel detection (Lister et al., 2007). The result was replicated with a larger sample of both older and younger participants (Lister et al., 2011). However, Heinrich et al. (2004) demonstrated comparable mismatch negativity (MMN) responses (which is a negative cortical-evoked potential in response to the detection of an oddball in a series of repeating sounds) to the two types of gap detection tasks in both amplitude and latency, which was source localised to the primary auditory cortex.

The current evidence seems to lean towards the theory that regards the between-channel and within-channel gap-detection tasks to tap into different perceptual domains, but it is difficult to conclude if between-channel can better predict SIN performance. Blankenship et al. (2022), for instance, reported a better correlation between CAEP elicited by within-channel GDT with speech perception in noise (both word and sentence perception) compared to between-channel detection in cochlear implant (CI) users. However, this correlation was only based on electrophysiological responses, and no performance-level association was reported. Another study with CI patients reported that within-channel gap-detection tasks significantly predicted SIN perception (Xie et al., 2022a).

Studies comparing the two types of GDT are limited. In clinics, within-channel GDT tests have been more commonly used, such as the Gap-in-Noise test (GIN, Musiek et al., 2005), the Random Gap Detection Test (RGDT, Keith, 2000), the Adaptive Test of Temporal Resolution (this test has a component of between-channel detection, Lister et al., 2006), Auditory Fusion Test-Revised (McCroskey & Keith, 1997). They are used as a way of assessing temporal resolution that can inform the diagnosis of auditory processing disorder. However, in terms of the predictability of SIN listening, the literature suggests that within-channel gap detection does not consistently reflect the performance of commonly used SIN performance. An early psychoacoustic experiment looking at temporal acuity, sentence-in-noise, and reverberation found a strong correlation between the two measures (Irwin & McAuley, 1987). However, the study was conducted with a very small sample (8 participants) so the results might not be reliable. Similarly, a behavioural correlation was established by Feng et al. (2010) with native Mandarin speakers with high-frequency hearing loss. However, the researchers cautioned that the data showed large individual variations and needed to be validated. A significant correlation was found between RGDT and sentence-in-babble measures in CI patients (Blankenship et al., 2016). Heeke et al., (2018), on the other hand, found a negative correlation between RGDT and HINT threshold measures and the correlation was not significant after correction for multiple comparisons. While older people showed lower gap-detection ability, their SIN ability still held up, and no significant correlation was found between

GIN and R-SPIN (DeMetropolis et al., 2021). This result was consistent with the study conducted by Hoover et al., (2015), who found non-significant correlations between GIN and SIN perception in normal hearing people, and Cesur & Derinsu (2020) in CI users. Snell & Frisina (2000) proposed that the relationship between the SIN measures and gap detection might be modulated by age. They investigated younger and older age groups and found a significant correlation between SIN perception and gap thresholds in younger participants but no association in the older age group. However, the opposite results were found by the same group when they attempted to replicate the result (Snell et al., 2002). Overall, there is little consistency on the literature reporting an association between SIN and gap detection measures.

TMTF, on the other hand, has been reported to have a more stable relationship with SIN performance. TMTF measures the smallest sinusoidal-modulation depth a person can use to discriminate an amplitude-modulated tone or noise from a sound that is not modulated (Eggerrmont, 2015). Studies (George et al., 2006, 2007) have shown that temporal acuity measured by detection of the amplitude-modulated noise explained a large variance of speech intelligibility in modulated noise. A similar result was found later (Narne, 2013), in which the TMTF was found to be a significant predictor of speech in speech spectrum-shaped noise.

In summary, there is consistent evidence of a correlation between temporal resolution measured by TMTF and the detection of speech in speech-shaped noise. Gap-detection tests, on the other hand, while having wide clinical application in assessing hearing disorders, do not show consistent results in predicting speech perception in babble noise. Research evidence supports a link between within-channel gap detection and speech perception in noise with CI patients, but the clinical measure of GIN was reported to differ from the traditional psychophysical gap-detection paradigms and could not predict SIN perception (Hoover et al., 2015). The between-channel gap-detection paradigm is not well-researched and could potentially be relevant to SIN processing.

Temporal ordering

Temporal ordering is often measured with frequency and duration pattern tests. The frequency and duration pattern tests were the most widely available clinical tests for temporal processing (Shinn, 2003), which measure the ability to distinguish the tone of a different frequency (high vs low sound) or length (short vs long sound) out of three tones, respectively. However, research is limited on these tests regarding their relationship with SIN perception. A recent study reported no correlation between duration patterns and SIN tests in children with central auditory processing disorder (Spandita & Jain, 2024). A systematic review on temporal ordering tests in Brazil reported the use of frequency and duration patterns tests in diagnosing some speech-related disorders such as dyslexia, developmental language disorder, autism spectrum disorder, reading and writing disorders etc., but no studies reported a specific relationship between speech recognition in noise and temporal ordering (Delecrode et al., 2014).

2.2.3 Measures of auditory stream segregation

Auditory streaming can be elicited by all kinds of mixtures of sound, with the target and the background sound ranging from speech (conversations, words, numbers) to degraded speech (vocoded or sine-wave speech) to non-linguistic stimuli (polyphonic music, pure tones, stochastic-figure-ground). While SIN paradigms are more ecological in terms of simulating real-life conversations, the speech stimuli used are complex, conveying not only acoustic information (timbre, pitch, harmonicity, etc.), but also linguistic (semantics, syntax, pragmatics, etc.) and social cues (age, sex, familiarity, etc.). Studies have found that increased linguistic complexity can directly lower performance on SIN tasks within and across participants (Warzybok et al., 2015; Coene et al., 2016). Similarly, increased familiarity with the sound input (speech of a close family member) or of the speech content (recently read passages) could improve SIN performance (Holmes, To, et al., 2021). It is possible for people to exploit linguistic or social cues to generate expectations and compensate for compromised auditory grouping mechanisms. To assess or detect potential damage to the central mechanisms, therefore, researchers have attempted to remove the linguistic and social contents from SIN tests and created auditory stream segregation tests.

As mentioned in Section 1.3, a classic paradigm is segregation based on rapidly alternating tones of two frequencies, following the method of Bregman & Campbell (1971). Two streams consisting of the lower-frequency tones and the higher-frequency ones respectively could be formed under certain presentation rates and frequency differences. Significant correlations were found between these tasks with speech in steady-state speech-shaped noise and babble noise in cochlear implant users (Hong & Turner, 2006), but research is limited.

Auditory figure-ground was developed with a temporally coherent figure with repeating frequencies masked by a tone cloud with randomised frequency. This was named *stochastic figure-ground (SFG),* or *fixed-frequency auditory figure-ground (AFG-Fixed).* The stimulus was first tested in humans in a psychophysical and fMRI study (Teki et al., 2011), where participants listened to SFG and were instructed to detect the figure while ignoring the ground. The results demonstrated perceptual sensitivity to the presence of a figure. However, the prototype SFG detection performance did not correlate with SIN performance (Holmes & Griffiths, 2019). An attempt to make SFG more speech-like was made by Holmes & Griffiths in a more recent study (Holmes & Griffiths, 2019), where they added a gap discrimination task and complex frequency patterns similar to the formants in natural speech (roving) to the SFG. The researchers correlated the performance of the new versions of SFG with SIN tasks and found that the gap discrimination task correlated with SIN performance significantly ($r = 0.45$), and independently of PTA prediction in a stepwise regression model (r^2 change = 0.05). Similarly, figure discrimination with coherent roving patterns showed a significant correlation with SIN ($r = 0.44$), which was also independent of PTA (r^2 change = 0.04).

Measures of stream segregation are not widely used in clinics or research. However, the relationship between SFG and SIN suggests the potential for using it as a complementary test for real-life listening.

2.2.4 Measures of short-term memory and working memory

In addition to auditory processing abilities, cognitive performance can also predict SIN perception. Akeroyd reviewed the relationship between SIN and aspects of cognition and reported working memory as the most effective measure of SIN perception (Akeroyd, 2008). The working memory here was verbal working memory as measured by the reading span test. Working memory measured by the reading span has been well-researched with SIN perception. A systematic review identified the relationship between 5 domains of cognition and SIN performance: “processing speed ($r = .39$), inhibitory control ($r = .34$), working memory ($r = .28$), episodic memory ($r = .26$), and crystallised IQ ($r = .18$)” (Dryden et al., 2017). Füllgrabe & Rosen (2016) found that the association between working memory and SIN could be age specific. The researchers reviewed 41 datasets and found that the strength of the association between working memory and SIN is very weak ($r=0.18$, $p = 0.162$) for young listeners (aged 18-39), whereas the association was stronger ($r \geq 0.44$, $p \leq 0.011$) for older age groups (aged 40-59 and 70-91) (Füllgrabe & Rosen, 2016).

For more specific auditory short-term memory (non-speech), Lad et al. (2024) proposed a new paradigm examining auditory memory precision for frequency and amplitude modulation rate (AM), which differed from the classic frequency/amplitude detection tasks that compare the frequency or modulation rates of two-sound presentations. The auditory short-term memory tests of frequency and amplitude discrimination tasks implemented a delay of up to 4 seconds after the first sound and required the participants to match the frequency or AM rate to the first sound using a slider (Lad et al., 2024). The researchers found a significant correlation between sentence-in-babble perception and memory for frequency precision ($\rho = -0.36$) but not for amplitude precision (Lad et al., 2020a). In a more recent study, however, they found that memory for AM precision ($r^2 = 0.24$) was more important than that of frequency (Lad et al., 2024). Due to this inconsistency, more studies are needed to validate this paradigm.

2.3 Behavioural methods: verbal measures of SIN perception

2.3.1 Verbal objective measures of speech-in-noise

When it comes to assessing a person's real-life listening ability unaccounted for by the pure-tone thresholds, speech-based tests are the most used form of testing due to their high ecological validity. There are many considerations for the application or development of such tests. First, the test stimuli can have a range of variations: the target stimuli can be sentences of different phonetic, syntactic, or semantic complexity, or can be formed of words or syllables. The background noise can be stationary, degraded speech or speech-shaped noise, or babble noise with varying numbers of speakers. The outcome measures of a SIN test can be active, where participants' response is required, or passive. Commonly used response modes are verbal or nonverbal, and open-set or close-set, depending on the purpose of the test. Task accuracy (percentage of correct responses) and signal-to-noise ratio (SNR) are often used to quantify participant's performance. Other outcome measures include speech-based frequency following response (FFR), speech-evoked auditory brainstem response (ABR), speech-evoked pupillometry, etc., where the evaluation of a person's SIN ability does not necessarily depend on their performance.

Sentences or words

In terms of providing an ecologically rich form of testing for real-word listening, sentence-in-babble is widely considered the most suitable type of assessment. This is evidenced by a recent survey of British Audiologists and ENT surgeons (Bernard et al., 2024), which reported that the most commonly used SIN tests in the clinics for adults were QuickSIN (Killion et al., 2004) and LiSN-S (Cameron & Dillon, 2007), and for children was LiSN-S, both are sentence-in-babble tasks. Using sentences has benefits that go beyond ecological validity. Sentence tests present more words than single-word tests when controlling for the test duration, which can give a more accurate description of SIN ability (Weißgerber et al., 2013; Wilson et al., 2007). People are also more sensitive to detecting minor stimulus degradation with longer speech stimuli (Antons et al., 2012). However, as sentences need to be formed in a set structure, the choice of word categories is less flexible, and people are more likely to form predictions of the upcoming words based on the syntactic features of the sentences. It is also difficult to balance the phonemes. In addition, sentence tests tend to have more semantic context and rely heavily on working memory as well as language competence, which can make the interpretation of the test results more ambiguous. For example, language ability may decline with age even in normal-hearing individuals (Colby & McMurray, 2023; Payne et al., 2014; Waters & Caplan, 2001).

On the other hand, word-based tests have the advantage of flexibility: the materials used are likely to be phonemically balanced and tailored to different levels of literacy. They reflect more purely on SIN perception instead of language competence. Wilson et al. (2007) also found that the word-in-noise (WIN) test as well as a low-contextual sentence-in-babble test (QuickSIN) provided more separation in recognition performance between the normal-hearing and hearing-impaired groups, making it a good tool for hearing diagnosis. The drawback of the word tests is that they are not as ecologically valid as using sentences. Not providing a language context means that the task can be too challenging to do for people with substantial hearing loss, CI users, or for children.

Response mode

Task design has a crucial impact on what is measured and the outcome of the test. The most commonly used clinical tests tend to employ open-set responses (in which a participant repeats back what is heard). As previously mentioned, tests that ask for verbal responses are arguably the most ecological form of test, and the best form for people who are unable to give accurate responses with a keyboard or mouse. The drawbacks of open-set tests are the potential confounds involved in the tasks. Firstly, having to give verbal responses can present a challenge for certain populations such as post-stroke patients with speech production difficulty. The process itself poses demands not only on speech perception in noise, but also on word recognition, language processing, lexical access, language production, and working memory (Klem et al., 2015). This is the reason that sentence repetition is often used as a measure of Developmental Language Disorder (Wang et al., 2022). On the other hand, close-set tasks require computer literacy and are less ecological. However, they do not involve language production and are a purer measure of perception.

Other issues for SIN test application

Test results can be skewed by participant accent and dialect, vocabulary size, cognition, and attention, as well as factors involved in test administration such as testing environment and equipment.

When speech is heard in an unfamiliar dialect or accent in a noisy environment, this can disproportionately impact people’s speech processing. This problem affects not only non-native speakers but also native speakers who are unfamiliar with different dialects and accents. For example, adult speakers of the Southern Standard British English have been found to show slower processing speed when listening to Glaswegian English, especially in adverse listening conditions (Adank et al., 2009). Similarly, Bent et al. (2021) showed a decrement in a variety of native accents in young adults with normal hearing, and this effect was more pronounced in children: even as adults did not perform differently in British vs. American accented speech, children can still struggle.

Aside from word recognition accuracy, other aspects of speech processing are affected by accent. For example, The LiSN-S has similar normative data for British and American children but the talker advantage measure requires a corrective factor

(Murphy et al., 2019). Speech processing takes more effort when people are confronted with a less familiar accent (Van Engen & Peelle, 2014), suggesting that accented speech (to a given listener) may engage a somewhat distinct set of cognitive and perceptual mechanisms than non-accented speech. Research also showed that older adults might have different cognitive strategies to younger adults when processing accented speech, modulated by cognitive flexibility and inhibitory control (Ingvalson et al., 2017). These findings highlight a potential problem with the implementation of hearing assessments both in research and clinics, where practitioners are often limited by the materials available to them, and the materials might not be suitable for the population that they test. Such is the case in UK audiology practice. Parmar et al., (2022) reported that only 20.4% of publicly funded audiology practices give speech tests in the UK. This is partly due to limited clinical resources but also because of the lack of widespread availability of materials geared towards British English. Many commonly used speech tests for hearing impairments used in the UK are not available in British English or validated with British populations. As previously mentioned, the most commonly used speech-based screening tools were QuickSIN and LiSN-S for adults, and LiSN-S for children. Both of the tests were recorded in American or Australian English only. McLaughlin et al. (2018) found that people's relative skill at processing SIN did not even correlate with their skill at processing accented speech. This means that when people are tested with a SIN test in a less familiar accent, they could show lower performance leading to misdiagnosis of hearing problems simply due to their lower ability to process accents and not due to their SIN ability. Consequently, speech-based tests can easily misidentify hearing problems by using a uniform standard (Dawes, 2011; Dawes & Bishop, 2007).

2.3.2 Verbal subjective measures of speech-in-noise

From the patient's perspective, hearing difficulties might be best defined by experience. Many people suffer from effortful listening when talking to others, especially in noisy environments and the most useful way to quantify this experience is arguably self-evaluation such as a questionnaire and/or interview. Table 2.2 summarises some of the most frequently used questionnaires that have been designed to capture a systematic picture of a person's hearing profile.

Test	Scope	Considerations
SSQ (Gatehouse & Noble, 2004)	Speech understanding in competing contexts, spatial hearing, and qualities of hearing experience (listening effort and naturalness, clarity).	It provides a comprehensive evaluation of real-world hearing, but the full version is lengthy. Some of the questions are complex and subject to individual interpretation.
Hearing Handicap Inventory for Adults (Newman et al., 1990)	Assesses the emotional and social/situational impact of hearing loss.	The whole test is short and provides a comprehensive emotional and social evaluation. However, it is not focused on identifying SIN problems and is used only for people under 65, and it has a weak association with PTA and word recognition.
Abbreviated Profile of Hearing Aid Benefit (Cox & Alexander, 1995)	Assesses the outcome of a hearing aid fitting on SIN perception and aversiveness of sounds.	It has a focus on background noise in one subscale, but it is limited to hearing aid users.
Glasgow Hearing Aid Benefit Profile (Gatehouse, 2000)	Assesses the benefit of hearing aids in SIN settings and various listening environments	Focuses on real-world hearing aid benefits but is again limited to hearing aid users.
Listening in Daily Life Questionnaire (Anderson & Smaldino, 1999)	Assesses real-world listening difficulties in education.	Demonstrated efficacy in evaluating how classroom acoustics and background noise affect students with hearing loss, but it is not applicable to the wider population.

Table 2.2 Commonly used subjective measures of SIN with testing scopes and key considerations for implementation.

Self-assessment provides important information that influences the diagnosis of hearing disorders, strategies for fitting hearing aids, and the setups of educational environments. However, a key issue with the self-assessed tests is the correspondence with the objective measures. There is a discrepancy between the subjective and objective tools of real-life hearing assessment (Choi et al., 2019; Pedersen & Rosenhall, 1991; Matthews et al., 1990), but literature on this topic is not consistent as some studies also found a significant association between self-assessment and speech audiometry (Eckert et al., 2017; Mendel, 2007). An important factor is that self-assessments rely on factors that are not related to hearing thresholds or SIN perception such as mental status and personality. For example, Wöstmann et al.(2021) reported the significant effect extraversion has on the subjective (but not objective) hearing-in-noise tests. This discrepancy is also due to self-assessed tests tending to focus more on personal experience of hearing quality, whereas objective measures tend to test hearing acuity, speech recognition and comprehension. It is therefore important to account for both measures especially when assessing hearing aid performance.

2.4 Physiology: biomarkers of speech-in-noise processing

In addition to behavioural methods, there is a wide range of physiological methods available to assess the function of the auditory system from the auditory periphery to the high-level cortices involved in processing complex SIN signals. While many of the behavioural methods discussed in the previous sections, such as PTA and gap detection, HINT, are commonly used in clinics and research, they are not always the best assessment to choose. For patients not able to give reliable responses, such as infants or people who suffer from language production disorders (e.g. dysarthria, expressive aphasia), physiological responses would more accurately reflect a person's auditory processing abilities. This section reviews some of the most used tools for assessing the auditory system that can predict SIN behavioural performance.

2.4.1 Auditory periphery

The otoacoustic emissions (OAEs) test is a commonly used tool to examine the cochlear function that could predict SIN perception. Specifically, it measures the outer hair cell function via the echo sound travelling back to the middle ear produced by the vibration of the OHCs when stimulated by clicks. Due to the link between OAE responses and hearing sensitivity, the test (commonly transient evoked OAE and distortion product OAE) is often used for new-born hearing screening as it does not require behavioural responses (Smith & Cone, 2021). Studies have shown that not only can OAEs be used to measure cochlear health, but they can also be used as an indicator for central auditory processing disorder, which could have a large effect on SIN perception (Iliadou et al., 2018). OHCs are innervated by cholinergic efferent fibres of the medial olivocochlear (MOC) system (Fuchs & Lauer, 2019), which has been identified as an important system that benefits signal processing in noise (Chintanpalli et al., 2012; de Boer et al., 2012). The absence of acoustic reflexes and OAE suppression was also shown to be related to self-reported speech processing in noise (Lautenschlager et al., 2015). However, recent research on the medial olivocochlear reflex measured by transient evoked OAEs reported no modulatory influence on a SIN task (Gafoor & Uppunda, 2023). The group continued to review the research on the role of MOC in SIN perception using meta-analysis and found that MOC reflex measured by OAE accounts for less than 1% of the variations in SIN (Gafoor & Uppunda, 2024). Although this does not provide strong evidence for the lack of relationship between MOC and SIN perception itself as OAE only indirectly measures MOC reflex (Lichtenhan et al., 2015), OAE suppression has been shown not to predict SIN. In conclusion, while OAE can be used as a reliable measure for hearing, it cannot provide sufficient insight into real-world listening.

2.4.2 Ascending pathways

The subcortical and brainstem structures are critical for early auditory processing, especially in encoding the temporal and spectral features of speech. Auditory brainstem responses (ABR) are characterised by a series of waves that represent different levels of neural activity from the auditory nerve to the inferior colliculus (Parkkonen et al., 2009). Not all components have been well researched in association with SIN and some have been shown to predict SIN ability poorly. ABR wave I amplitude, for example, has been used in clinics for decades and has been associated with cochlear synaptic integrity (Bramhall, 2021). In a 2019 study, researchers found no significant correlations between wave I amplitude and an objective SIN test QuickSIN ($r = -0.05$), and a self-reported SIN ability measured by SSQ ($r = 0.31$) (Bhatt & Wang, 2019).

Auditory brainstem responses to speech and other complex stimuli (cABRs) on the other hand, seem to consistently show good predictability of SIN perception (Anderson & Kraus, 2010). Speech-ABR consists of both a transient response to the speech onset and a sustained response also known as frequency following response (FFR) (Sinha & Basavaraj, 2011). The researchers found that ABR responses elicited by speech correlated with SIN significantly, such as consonant differentiation (/da/, /ba/, /ga/) with HINT ($r = 0.492$) (Hornickel et al., 2009), the encoding of fundamental frequency (F0) with QuickSIN (r = 0.523) (Anderson & Kraus, 2010), and the second harmonics (H2) with HINT (r=0.486) (Chandrasekaran et al., 2009). However, a more recent finding in hearing aid users revealed that the relationship between speech-ABRs and sentence/word-in-noise or subjective reports did not hold up after considering hearing thresholds (BinKhamis et al., 2019). This suggests that speech-ABR might not predict independent variance of SIN in addition to PTA. More studies are required to examine ABR and SIN perception with PTA as a potential confound.

On the other hand, the frequency or envelope following responses demonstrated that it explained a significant variance in SIN independently of PTA (Mepani et al., 2021). Thompson et al. (2019) also found a similar result after accounting for age. FFR has been used to describe a broad range of brainstem responses including speech envelopes and has been differentiated sometimes by terms such as “spectral FFR” and “envelope FFR” (Aiken & Picton, 2008). While FFR has been seen as a measure of brainstem activity, more and more evidence has emerged to support the hypothesis of FFR having more central involvement (Gnanateja et al., 2021; Coffey et al., 2019, 2016). Nonetheless, EEG-recorded FFR found that the subcortical sources dominated the electrical FFR, as well as a link between FFR and SIN (Bidelman & Momtaz, 2021). However, the study had a small sample (n = 12), and EEG source reconstruction does not have the same level of spatial resolution as MEG.

Overall, subcortical temporal processing measured electrically generally exhibited a weak to moderate correlation with behavioural SIN thresholds. However, ABR shares a large variance with peripheral hearing sensitivity and might not independently predict real-life listening. The body of literature investigating FFR in relation to SIN is relatively limited and further investigation is needed for validation, ideally providing clearer quantification of its contribution after accounting for PTA and age.

2.4.3 Cortical recordings that predict SIN performance

The cortex is the final important stop for processing complex sounds. Auditory evoked potentials such as middle latency response (MLR), auditory steady-state response (ASSR), and cortical auditory evoked potentials (CAEPs) have been widely applied in research and clinics to measure a person’s hearing. Neural entrainment to continuous speech is another new area of research that could potentially be used as a biomarker for SIN ability.

MLR is thought to be generated by the auditory cortex primarily and provides information on the neural integrity of the central auditory system (Musiek & Nagle, 2018). However, many studies showed non-significant associations between MLR characteristics and speech perception in noise (Alemi & Lehmann, 2019; Purdy & Kelly, 2016). ASSR is used to determine hearing thresholds for people who are unable to give responses in traditional behavioural tests. It records bioelectric activities which are phase-locked to the presentation rate of a click train, or the modulation frequency of amplitude-modulated sounds, with the main generators for the most commonly used ASSR of 40 Hz modulation rate located at the primary auditory cortex (Manting et al., 2021). The recording can be performed with a simple montage; usually with one active electrode, two reference channels, and one ground. As discussed in the previous section (2.1.4), behavioural tasks using amplitude-modulated sounds show a significant correlation with speech perception. Studies recording ASSR for amplitude modulation sweeps have also demonstrated a significant correlation between the amplitude of ASSR and speech recognition threshold in noise at 30 – 40 Hz (r = 0.61) but not beyond 40 Hz, which (higher frequencies) yields predominant responses in subcortical locations (Manju et al., 2014). Similarly, a strong correlation (r = 0.89) was found in CI patients with 40 Hz ASSR. A comparison between younger and older age groups also found that ASSR responses predicted SIN performance independent of age (McClaskey et al., 2019).

Finally, cortical auditory evoked potentials (CAEPs) are very often used in research into speech and are available for more specialised testing in clinics. The CAEPs are recorded responses to auditory stimuli (such as syllables in noise), with a classic P1 response at around 50 ms, an N1-P2 response at 100 ms and 180 ms, followed by P3 at 300 ms (Martin et al., 2007). Auditory evoked potentials have also been proposed as a measure for SIN perception. A study using a simple tone-in-noise paradigm found that the signal-to-noise ratio can affect the amplitude and latency of N1, P2, and N2, but not P1 (Billings et al., 2009). The same group later found that the N1 amplitude and latency were a strong predictor ($r_{amplitude}$ = 0.72, $r_{latency}$ = 0.77) of SIN perception using monosyllabic sounds (/ba/) masked by speech spectrum continuous noise (Billings et al., 2013). P1 amplitude was also reported to predict syllable identification in Gaussian noise (Dias et al., 2021). A significant P2 increase was reported to be associated with auditory training, suggesting a link between neural plasticity to speech processing as well (Tremblay & Kraus, 2002). More recent developments in the field have moved towards more ecologically valid stimuli, with real words or sentences masked by babble noise. Researchers investigated the relationship between N1-P2 peak-to-peak amplitude of a word-in-noise task and word-in-noise and sentence-in-noise performance and found a significant correlation (r = 0.30) with word perception in noise in CI users (Berger et al., 2023). However, a correlation was not found with the sentence-in-noise measure (Berger et al., 2023).

Detection of auditory changes has recently been proposed to predict SIN perception. Acoustic change complex (ACC) and mismatch negativity (MMN) are two types of CAEP that reflect the automatic sensory processing of stimulus change (e.g. frequency and intensity) (Velluti, 2018). ACC detects changes in a continuous auditory stimulus. The task typically involves detecting a shift in intensity, frequency, or other acoustic features within a sound sequence. ACC reflects the ability to detect changes in the auditory cortex and exhibits a classic peak pattern similar to the N1-P2 complex as evoked by simpler paradigms. ACC can be recorded both with and without an active task, making it a useful tool for assessing auditory perception in clinics (Sanju et al., 2023; Kim, 2015). MMN detects deviations in repetitive regular sounds. A common paradigm used is the oddball paradigm, where a series of standard tones is interrupted by a deviant tone. ACC has a higher signal-to-noise ratio and needs fewer stimulus presentations to reach sufficient power (Kim, 2015). In terms of their relevance to SIN perception, the latency of ACC was shown to predict SIN independent of PTA (R^2=0.36) (Vonck et al., 2022). In CI patients, ACC N1 latency was also shown to correlate with the Consonant-Nucleus-Consonant (CNC) word perception test, but similar to the Berger et al. (2023) study, it did not correlate with the sentence-in-noise measure (McGuire et al., 2021). P2 latency only correlated with the digit-in-noise score but not WIN, and N1- P2 amplitude here did not correlate with any SIN measures (McGuire et al., 2021). MMN, on the other hand, shows little evidence of its ability to predict SIN. A study reported a significant correlation between MMN amplitude elicited by the syllable /bu/ and sentence-in-noise perception (Koerner et al., 2016), but with only a small sample (n = 15) with no reported effect size. A comprehensive assessment of MMN central sound processing in older adults over 60 years old of a much larger sample (n

= 56) found no correlation between MMN latency or amplitude and syllable-in-white-noise perception (Brückmann et al., 2021). However, a comparison between speech-in-babble and speech-in-quiet showed a significant difference in MMN latencies, where the noise condition had earlier MMN peaks compared to the quiet condition (Kozou et al., 2005). Due to limited literature, it is impossible to conclude a relationship between MMN and SIN. It is conceivable that as MMN has a relatively low SNR and is not always present in normal-hearing individuals despite good behavioural performance (Bishop & Hardiman, 2010), the effect of any relationship between MMN and behavioural SIN thresholds is difficult to find.

Finally, for a more ecological recording, researchers have used longer continuous speech materials (such as audiobooks) and investigated the linear transformation of the target speech to EEG signals. The prediction accuracy of these linear models or waveform morphology has been reported to indicate speech perception in noise (Brodbeck & Simon, 2022; Ding & Simon, 2014; Kegler et al., 2022; Panela et al., 2024). Tracking of the speech envelope in particular has been shown to predict speech intelligibility in stationary speech-weighted noise ($r = 0.51$) (Van Hirtum et al., 2023), however, this positive correlation was not always found (Kösem et al., 2023). Other acoustic features have been investigated too, such as pitch and temporal coherence in speech tracking (Bachmann et al., 2021; Teoh et al., 2019; O'Sullivan et al., 2015), but they have not shown a direct correlation with SIN on the behavioural level.

To summarise, a wide variety of auditory evoked potentials have been used to study listening to speech in a challenging environment. The long-established components, such as ASSR, N1-P2, and ACC, elicited by simple speech sounds (syllables or words) exhibit the most stable relationships with SIN perception.

2.4.4 Other physiological measures

In addition to EEG/MEG recordings of speech stimuli, there are a few emerging tools for studying the physiological correlates of SIN perception in research. Pupillometry for example has been increasingly used in SIN research. During effortful listening, pupil dilation increases and eye movement decreases, and these can be used as indicators of cognitive load and listening effort during SIN perception (Cui & Herrmann, 2023; Koelewijn et al., 2012; Zekveld et al., 2010). The dilation responses vary in different phases: adult pupils were shown to dilate during auditory processing, while dilation decreases during retention (post-stimulus-onset) (Trau-Margalit et al., 2023). Children respond differently to adults, and show consistent increases in dilation, suggesting more effortful listening for children when challenged by SIN listening (Trau-Margalit et al., 2023). While no evidence has been found that pupil characteristics can predict SIN perception, a combination approach can prove useful, where pupillometry is used together with EEG recordings to help interpret EEG responses to SIN (Ershaid et al., 2024; Kılıç et al., 2024). For instance, Ershaid et al. (2024) found a significant increase in speech tracking in response to more challenging listening conditions. As they also found larger pupil dilation, the researchers concluded that the effect of challenging SIN perception on EEG speech tracking reflected resource allocation and listening effort.

Facial expressions have also been proposed to be indicative of effortful listening (See Venkitakrishnan & Wu, 2023 for a review on the topic), as well as heart rate and skin conductance (Andersson et al., 2023; Christensen et al., 2021; Shoushtarian et al., 2019; Mackersie et al., 2015). However, they generally lack sensitivity to SIN SNR changes and are not feasible to be used as reliable measures of SIN perception (Cvijanović et al., 2017).

To summarise, among the physiological responses to SIN stimuli, speech-evoked brainstem and cortical-evoked potentials are by far the most reliable measures of SIN recognition or detection performance. Pupillometry is a useful tool to gain insight into cognitive resource allocation and can be used in combination with electrical recordings to provide a more detailed picture of the neural encoding of speech.

2.5 Conclusion

This review has summarised most of the behavioural and physiological responses that predict (or not predict) SIN perception and the strength of the relationship between them. The most commonly used behavioural measure that predicts SIN is PTA, with an average correlation coefficient of around 0.472 (Figure 2.1). However, the effect size of the association between PTA and SIN might not reflect the effect due to the issues with data reporting (lacking nonsignificant results, underpowered studies, for example). Importantly, younger people with intact hearing do not tend to show this correlation. Tests for temporal acuity are also very common and explain a significant variance in SIN perception. Auditory streaming tested by stochastic figure-ground measures an independent variance of SIN in addition to PTA ($r = 0.441$). This shows a great potential of using the figure-ground paradigm for SIN assessments. Such non-verbal measures benefit from having no linguistic and socioeconomic confounds and can provide a 'pure' measure of central auditory processing. Future studies should focus on validating the results in different populations and improving the paradigm so they can provide a reliable assessment of central hearing. Speech-based tests and subjective questionnaires are becoming more popular for both clinical assessment and research as a measure of real-life listening, but the discrepancy between the two types of tests means that the choice of tests should be more cautious. For the patient's comfort (such as during the fitting of hearing aids), questionnaires are preferred, but for accurate SIN recognition assessment, sentence- or word-in-noise tests are preferred. Finally, physiological measures can be used when participants are not able to respond as instructed, or as complementary methods of SIN assessments. Electrical measurements, especially ABR, ASSR, N1-P2, and ACC, provide stable biomarkers for SIN perception with a moderate to strong effect size. These should be considered for SIN testing when peripheral measures are insufficient to explain real-life listening.

3. Chapter 3: Exploring the auditory cognitive mechanisms of speech-in-noise perception

In the previous chapters, I have reviewed the essential mechanisms of natural listening and the most commonly-used measures to assess SIN ability. In this chapter, I move forward to bring them together and investigate the interactions among the important predictors of SIN, aiming to establish a clearer portrayal of how different components of sound analysis and cognition contribute to real-world listening. SIN is a complicated process that can be predicted by many auditory and cognitive factors. In this study, I roughly categorised these factors into the auditory peripheral functions, short central auditory processing (CPS), long central auditory processing (auditory-specific memory, CPL), verbal short-term and working memory, fluid intelligence, reading ability/crystallised intelligence, and musical sophistication. The main aim of this chapter is to explore the variance of SIN perception that these auditory cognitive predictors can explain, while accounting for the interactions among themselves.

In addition to examining how auditory cognitive functions predict SIN, another direction of the relationship can be explored, which relates to the hypothesis explaining the link between listening difficulty and cognitive decline (Section 1.2.3). General cognition was used as the outcome measure to explore this question. The goal was to further detail the hypothesis of hearing loss causing cognitive decline and specify what aspects of listening (e.g. peripheral hearing, central auditory processing, verbal SIN processing) contribute to cognitive changes while accounting for age.

To account for the interactions of a large number of variables, I used the structural equation modelling (or structural equation models, SEM), which models the relationship between different types of variables based on prior expectations from literature that yields the relationships between the different variables. These variables can be indicator variables, latent variables, endogenous and exogenous variables, moderating variables, mediating variables, etc. An observed variable is one that is measured directly, and a latent variable is a factor unmeasured but indicated by other observed variables (these are called indicator variables). An endogenous variable is a variable affected by other variables within the model, whereas an exogenous variable is unaffected by other variables in the model. A moderator variable is a variable that moderates or affects the relationship between two variables, while a mediator variable explains this relationship directly. Structural equation modelling is an attempt to construct a 'complete' model that reflects the direct effects among these variables while accounting for the relative importance of indirect effects, such as the interaction between covariates on outcomes.

Modelling the interactions among the proposed variables requires a large sample to achieve enough power. SEM has a variety of standards to decide an appropriate sample size based on the number of observations (N) per statistical estimates (q), which range from 20:1 to 5:1 (Bentler & Chou, 1987; Kline, 2015) or based on the absolute sample size of 250 if using the Satorra-Bentler scaled method (Hu & Bentler, 1999). Considering the large sample, I considered online testing first as the best way of data collection. An online testing platform coded with JavaScript was developed. To ensure the reliability of this platform in collecting behavioural performance of auditory tasks, I conducted a test-retest reliability check. The study was therefore carried out in two steps: online-testing validation and main experiment.

3.1 Online validation (home-testing)

Research on human behaviour and cognition has traditionally been conducted in laboratory settings, where environmental factors are stable and can be controlled. However, research/data collection using online methods has been gaining popularity and has experienced almost exponential growth since the COVID-19 pandemic, including online auditory testing. A quick database (PubMed) search revealed that around 40% of the online studies starting from 1972 were carried out between 2020-2023.

Online data collection has been widely used for various research areas due to its advantages over lab testing: lower cost, easier recruitment process, larger sample size, faster data acquisition, and possibly more ecological validity. The use of online methods for survey and questionnaire data has been successfully implemented for over a decade. However, unlike lab-based testing, behavioural online studies introduce specific issues that need to be considered. For example, the timing of events and recorded reaction times may vary between participants (Bridges et al., 2020). Online testing poses requirements for participants that are not present in lab settings. This is particularly important for auditory research: participants must own an appropriate set of equipment suitable for hearing tests (a computer or laptop with a good soundcard, high-quality headphones without noise cancellation features, internet service), they must have digital literacy, and they need to be motivated to complete the tasks as instructed without being monitored. In addition, the home testing environment is not ideal. It is difficult to control for environmental confounds. As a result, online data tends to be noisier and harder to interpret. These problems can be mitigated by a large and representative sample size and careful validity checks to ensure that the paradigm is robust under the home-testing condition.

A validation study should be an essential step when developing an online hearing test. However, reviews on auditory online testing (Bright & Pallawela, 2016; Irace et al., 2021) indicate that most home-based hearing tests are not validated – the majority of studies make inferences based on online data alone without validating the paradigm in lab settings. Even when a paradigm has been validated in some way, their method of validation might not be reliable. Some studies carried out both self-testing and guided-testing in the lab to ensure consistent performance with different response modes (Corona et al., 2020), discounting the environmental factors in online home testing. Other validation studies on peripheral hearing screening only compared online results against participants' PTA thresholds, but not compared lab performance on the same task with PTA (Jansen et al., 2010).

The current study examined the validity of an online battery of auditory cognitive tests by comparing lab-testing performance with online performance. The battery includes the pure-tone audiogram (PTA, lab only), the antiphasic digit-in-noise test (DiN) (De Sousa et al., 2020), the sentence-in-babble (SiB) test based on the English Oldenburg sentences, the auditory figure-ground test (AFG) (Holmes & Griffiths, 2019; Teki et al., 2011), and a matrix reasoning task (lab only) (Chierchia et al., 2019). I also collected some demographic information of all participants, including their musical experience using the Goldsmith Musical Sophistication Index (MSI) (Müllensiefen et al., 2014). This battery covers the major online auditory tests: tests for peripheral auditory functions (PTA, DiN) and central sound processing (SiB, AFG). All the individual tests have already been shown to be effective in lab-based research. PTA has long been used as a measure for hearing sensitivity, and while it does not fully determine real-life listening, the test has shown a significant correlation with speech-in-noise tasks and is potentially the predictor that explains the most variance of SiN perception as demonstrated by the previous review (Section 2.2.1). DiN was developed as an online tool for hearing screening and has been widely used as a substitute for audiogram in various regions speaking different languages when audiogram is not an option (Smits & Houtgast, 2005; Potgieter et al., 2018; Ceccato et al., 2021). Based on the previous studies exploring the relationship between DiN and SIN perception (Kaandorp et al., 2015; Smits et al., 2013), I expected a strong correlation between DiN and SiB performance in this study as well. The SiB and AFG have been used for lab testing but have not been validated online (Holmes & Griffiths, 2019). The AFG test was shown to be a predictor of speech-in-noise ability, which was independent of hearing sensitivity (Section 2.2.3). A similar result should be found with the online test.

3.1.1 Methods

Participants

A total of 41 English native speakers were recruited for the experiment, one of whom only participated in the lab-testing session. Forty participants (15 male) were included in the data analysis, aged 19 to 67 (mean=32.90; SD=15.18). Participants had a range of peripheral hearing thresholds measured by pure-tone audiometry in decibels hearing level (dB HL) (see Figure 3.1), but had no history of neurological disorders, brain injuries, speech and language disorders, or hearing impairment. This study was approved by the research ethics committee of Newcastle University and written informed consent was obtained from all participants.

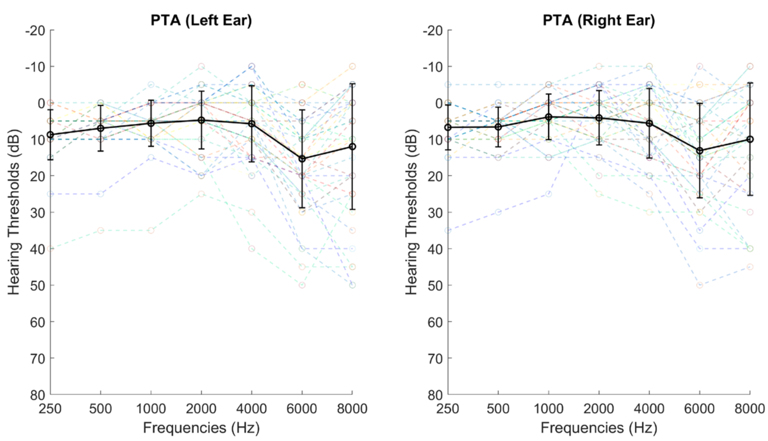

Figure 3.1 Pure-tone audiograms of the participants. The coloured dashed lines plot the individual PTA thresholds and the black lines with circles plot the average PTA. The error bars show the standard deviation. The x-axis represents the frequencies in Hz and the y-axis is the hearing thresholds in dB HL.

Materials

A headphone check was implemented using the dichotic Huggins Pitch (HP) as described in Milne et al. (2021), to ensure headphone use at home. The test stimuli consisted of three intervals of white noise (1000 ms), with one of the intervals containing a Huggins Pitch stimulus (Cramer and Huggins, 1958), where the same white noise is presented but in one of the ears, which has a 180° phase shift over a narrow-band (± 6%) centred at 600 Hz, giving a perception of pitch when presented binaurally (Chait et al., 2006; Yost and Watson, 1987). Participants performed 6 trials of the HP, where the HP percept was randomly presented in one of the three intervals.

The antiphasic DiN task (De Sousa et al., 2020) is a test of peripheral hearing thresholds using three digits that are presented with an inverted phase between two ears masked by speech-weighted noise presented in-phase. The task was implemented as a one-up one-down adaptive paradigm starting at 0 dB SNR and ended after 11 reversals. SNR changes started at 10 dB, followed by 5 dB after 3 reversals, and proceeded to 2 dB and 1 dB steps after 5 and 7 reversals respectively. Participants were instructed to select the corresponding digits they heard from a 3x3 number pad (numbered 0-9) presented on the screen.

The SiB task was adapted from the English Oldenburg matrix set, read by a male speaker with a British accent (Holmes & Griffiths, 2019). All sentences had the same structure [<name> <verb> <number> <adjective> <noun/object>] and were formed by a random combination of close-set options. The masking noise was a 16-talker babble presented in a changing SNR ratio using a one-up one-down adaptive procedure that terminated after 10 reversals. SNR steps started at 5 dB and were lowered to 2 and 1 dB after 2 and 5 reversals. Participants were asked to choose individual words from a 5x10 matrix on the screen.

The prototype stochastic figure-ground (SFG, also referred to as AFG to avoid confusion caused by different acronyms) was created using frequency bursts (called chords) and is formed by two separate elements termed “figure” and “ground”. An auditory ground was composed of random frequency components, while a figure was composed of frequency components repeating over time. Each chord lasted 50 ms and the background spanned 70 chords, which were formed by 5-15 frequency components randomly selected from a log-spaced frequency pool (180 – 7246 Hz). The figure was formed by 3 frequency components repeating over 42 chords from the same pool. Two stimuli were presented per trial with ground, and one of them contained a 6-chord gap within the figure. Participants were asked to decide which stimulus contained the gap. The target-to-masker ratio (TMR) changed in a one-down one-up adaptive procedure, starting with 4 dB TMR, followed by 2- and 1-dB steps after 1 and 4 reversals, respectively. The task terminated with a maximum of 10 reversals.

Procedure

The experiment included two sessions: lab testing and online (home) testing. All participants took part in both sessions in a counterbalanced order. For the lab testing session, testing took place in a soundproof booth, using an external soundcard (RME FireFace UC) and Sennheiser HD 380 pro headphones. First, PTA was measured for both ears across six frequencies (0.25 to 8 kHz) with an interacoustics diagnostic audiometer AD226. Then, participants performed the matrix reasoning task, after which they performed the main experimental tasks. During the main experimental protocol, which was the same both in-lab and online, participants were first presented with a 350 Hz continuous tone and asked to adjust their volume settings to a comfortable level. After this, participants were presented with a 100 ms 350 Hz tone in each ear separately to ensure sound was presented dichotically. Then, participants performed the headphone check followed by the three auditory tasks (DiN, SiB, and AFG) in a randomised order.

For the online testing session, participants were instructed to sit in a quiet place and ensure they had access to a computer/laptop, the internet, Google Chrome, and headphones (over-the-ear headphones preferred, in-ear headphones were also acceptable). Participants using Windows systems were instructed to turn off enhanced sound settings. After completing the main experiment, participants were asked to send information on the model of the headphones they used as well as general feedback if they had any.

All tasks were coded in JavaScript and Chrome was used as the browser for task presentation. Participants received a £10 voucher after the completion of the second session.

Data analysis

I conducted the following data analysis. The scores of HP were calculated by summing the answers (1=correct detection, 0=incorrect detection) of each item. Thresholds for SiN and AFG were calculated using the median of the last 5 reversals, and the last 4 reversals were taken for DiN. The descriptive statistics were calculated with SPSS Statistics 29. To examine the mean performance difference between the home and lab sessions, paired-sample t-tests were performed on the auditory measures. The overall datasets were not normally distributed, and the Spearman correlation coefficient was used to determine the correlations between the lab and home results. The intraclass correlation coefficient (ICC) with a two-way mixed effects model with absolute agreement was used to measure the test-retest reliability between the lab and home session. ICC is commonly used to estimate the association between variables similar to a correlation, but it considers both correlation and bias when assessing reproducibility (Liu et al., 2016). The absolute agreement measures are used to determine the level of agreement of raters, in this instance the scores of two testing sessions.

3.1.2 Results

Headphone check

Only one person failed to achieve the maximum score in the HP test (1/6) in the lab despite having their confirmed use of headphones. Online, a total of 35 (87.5%) participants scored 6/6 in the headphone check, similar to what was reported previously (Milne et al., 2021).

Descriptive statistics comparison of mean performance

The means and standard deviations are reported in Table 3.1. The t-test showed a significant mean difference between the home and lab measures (t_{SiB} (39) = -3.667, p < 0.001; t_{DIN} (39) = -2.116, p = 0.041; t_{AFG} (39) = -2.176, p = 0.036). All lab performance was better than the home session.

Tests	Lab (mean/SD (unit))	Home (mean/SD (unit))
PTA	7.813 / 6.662 (dB HL)	
SiB	-5.725 / 2.539 (dB)	-4.413 / 2.428 (dB)
DiN	-18.750 / 2.069 (dB)	-17.675 / 2.709 (dB)
AFG	-32.050 / 8.108 (dB)	-28.463 / 7.759 (dN)

Table 3.1 Descriptive statistics of the auditory measures tested in the lab and at home.

Test-retest reliability: comparing home and lab testing results

The performance of home and lab testing for all the tests is shown in Figure 3.2. The ICC scores are shown in Table 3.2. The SiB test was the only test that showed consistency between lab and online testing performance (R_{ICC}= 0.682, p<0.001). The DiN test home-testing results demonstrated a nonsignificant ICC with the lab-testing score (p=0.244). Similarly, the stochastic figure-ground performance at home and in the lab did not correlate (p=0.197). The correlation showed similar results (Figure 3.2).

Test	R_{ICC} (p)	CI Lower	CI Upper
SiB	0.682 (p < 0.001)	0.326	0.842
DiN	0.187 (p = 0.244)	-0.446	0.556
AFG	0.225 (p = 0.197)	-0.377	0.576

Table 3.2 ICC scores of the auditory measures.

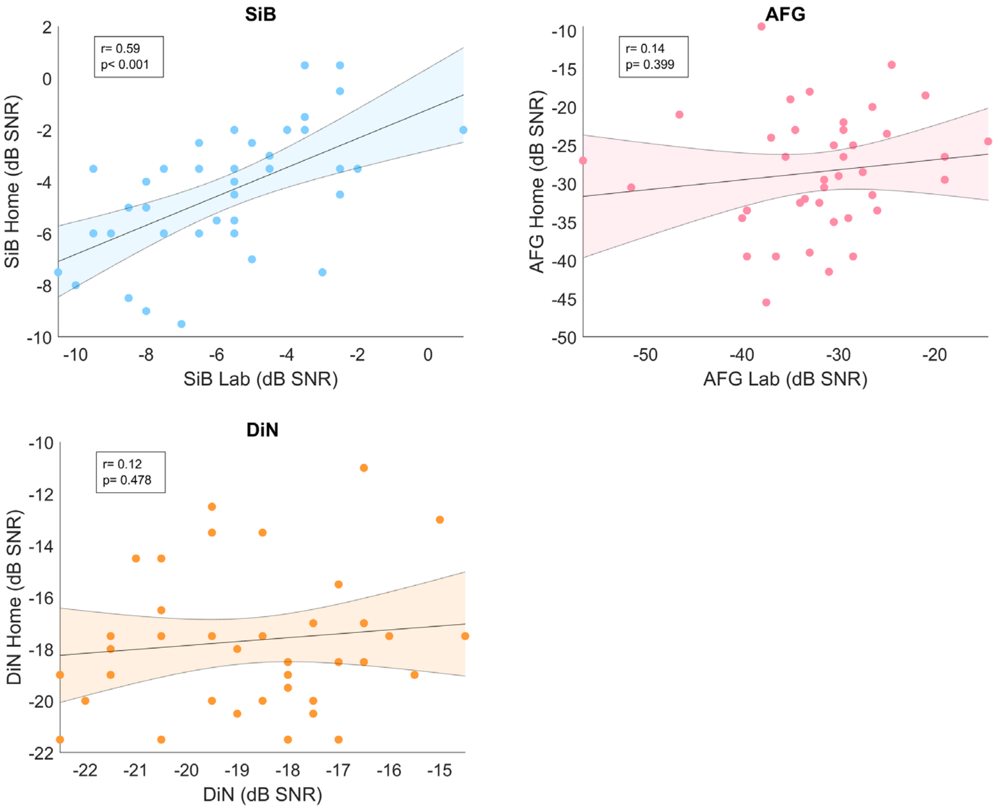

Figure 3.2 Correlation between home and lab performance. SNR thresholds calculated in-lab are shown in the x-axis while thresholds calculated at-home are shown along the y-axis.

Relationships between auditory tests at home and in the lab

The PTA thresholds did not correlate with the DiN measure in the lab or at home. However, it correlated significantly with both the SiB and AFG measures in the lab (rho$_{PTA\text{-}SIN}$ = 0.505, p < 0.001, rho$_{PTA\text{-}AFG}$ = 0.351, p = 0.024) and at home (rho$_{PTA\text{-}SIN}$ = 0.446, p = 0.004, rho$_{PTA\text{-}AFG}$ = 0.344, p = 0.032).

As is shown in Figure 3.3.3, DiN and AFG scores were compared with the SiB task scores. The DiN and SiB thresholds showed a non-significant association in the lab, but the home testing scores showed a stronger correlation (rho = 0.58). Similarly, AFG did not significantly correlate with SiB in the lab, but the online AFG measure correlated significantly with online SiB (p = 0.002). Finally, when comparing the DiN performance with AFG (see Figure 3.3), home testing results showed a modest association (rho = 0.36) which was not found in lab testing (p = 0.553).

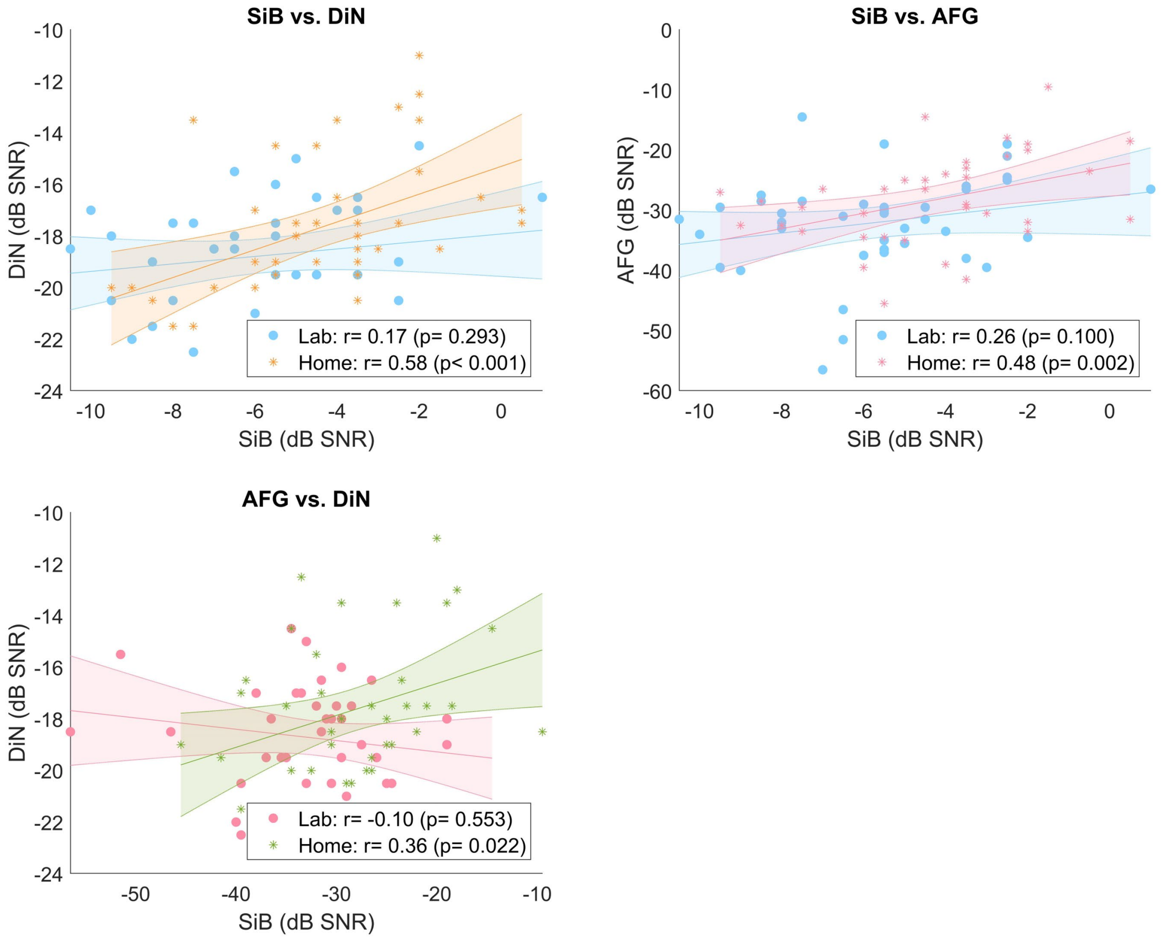

Figure 3.3 Correlation of the auditory measures tested in the lab and at home.

3.1.3 Discussion

Based on the outcome of the current experiment, the testing performance in different environments was highly inconsistent. The results at home are more likely to reflect an external effect like the attentional effect (based on post-hoc speculation) instead of the true auditory processes relevant to the tasks. The problems with online auditory experiments: various testing equipment, environments with distractions, and low motivations, are difficult to overcome.

The effectiveness of Huggins pitch as a headphone check

Huggins pitch was used as a headphone check task. However, our results showed that this task might not just correspond to headphone usage. One participant with normal hearing sensitivity could not hear the pitch sounds despite proper headphone usage and normal hearing. Different types of headphones (over-ear/in-ear, open-back/close-back) and the quality of their make could also make a difference, but this requires further research to uncover the specific impact and the level of the impact. Some participants also reported that Huggins pitch was easier to hear with a speaker instead of with headphones. The scoring method of this task is also problematic, as this should be an all or none task and the variety of scores is more likely to be an attentional effect or chance performance. For studies that require strict headphone-wearing, monitoring by the researcher might still be the best way.

Home testing reliability

The results of the test-retest reliability check showed little consistency in performance in different testing environments. The lab testing session for the DiN, SiB, and AFG resulted in significantly higher performance compared to home testing. Both SiB and AFG showed significant correlations with PTA. Consistency with the hearing thresholds has been used to validate certain online auditory measures. However, the inconsistent correlations between other auditory measures at home compared to in the lab showed that testing against PTA alone might not be enough to validate a hearing test. DiN test, for example, showed a strong association with the SiB thresholds at home despite lacking association with peripheral hearing when the test itself was developed as a tool to substitute audiogram for online testing. Similarly, the AFG task was developed as a measure for central sound segregation and was found to correlate moderately with the SiB test (Holmes & Griffiths, 2019). The original study required around 90 participants to bring out the statistical significance, however, with the online testing, only half of the sample size was needed to achieve a similar effect size.

These results raise a major issue of online data validity. As shown by the online results, it was easier to obtain significant correlations and thus possibly ‘desirable’ results for researchers. While the relationships found between online hearing measures could be attributed to a general effect of external factors such as attention. These confounds could easily be ignored without validating the paradigm with lab-based tests. This highlights the importance of task validations, but as mentioned previously, thorough test-retest reliability checks for online testing platforms both at different time points and in different testing locations were rarely performed.

Finally, in the case of unstable performance across testing environments, adjustments are possible for the online paradigm. To improve the reliability of an online auditory task, it could be helpful to make the task more engaging or implement an attention checker and give rewards for better focus. It is also important to have priors from literature, so the findings of an online experiment can be evaluated in the context of other studies. In summary, validation study is essential for such online batteries, as online results might be "too good to be true". As the findings of this study suggested that the online paradigms might not be robust, the main experiment was conducted in the lab.

3.2 Main experiments (laboratory-testing)

While the predictive relationship between SIN and the numerous auditory cognitive factors has been well established in literature, how they function as an integral system is yet to be investigated. In this study, I aim to test multiple auditory and cognitive indicators and examine how they interact with each other and with SIN perception using multivariate analysis. To model these complex interactions between the auditory cognitive predictors, both hierarchical regression and SEM were used.

Firstly, objective SIN perception can be measured by verbal sentence-in-babble and word-in-babble tests. In this experiment, both tests were used to better quantify verbal SIN perception. Based on the review of Chapter 2, the most important auditory cognitive predictors for SIN perception are the auditory peripheral functions, short central auditory processing (CPS), long central auditory processing (auditory-specific memory, CPL), verbal short-term and working memory, and general intelligence. The pure-tone audiogram (PTA) can be used to measure peripheral hearing sensitivity, which has shown a strong correlation with the verbal SIN measures (Chapter 2.1.3). The transient central auditory processing involves spectrotemporal analysis of the auditory information, which can be assessed with the auditory figure-ground (AFG) tasks (Holmes & Griffiths, 2019; Teki et al., 2011). Temporal acuity can be tested by the between-channel gap detection to see if the between-channel task can better predict SIN (GAP-Det, Phillips et al., 1997).

In addition to the transient sound analysis, central processing also involves retaining and manipulating the incoming auditory signals, which requires auditory memory. The CPL latent structure was therefore indicated by the auditory memory task for frequency precision (AUM-Freq) and precision for amplitude modulation rate (AUM-Amp) (Lad et al., 2020b, 2024). Digit span backward (DS-backward) from the Wechsler Adult Intelligence Scale (WAIS) was also used to test phonological working memory, as well as the transformation of information and mental manipulation of working memory (Wechsler, 1955). All three measures have been shown to have a strong association with SIN perception (Section 2.2.4). General intelligence, which has been shown to contribute to SIN perception (Dryden et al., 2017; Akeroyd, 2008), was tested by a matrix reasoning task (MTX) (Chierchia et al., 2019).

Additionally, I also collected some demographic information that might impact SIN perception, including age and musical sophistication. Age has been reported to be one of the most important predictors of SIN ability (Billings & Madsen, 2018). The musical sophistication of the participants was assessed with the Goldsmith Musical Sophistication Index (MSI) (Müllensiefen et al., 2014). Music training has been shown to improve SIN perception (Maillard et al., 2023; Parbery-Clark et al., 2009) although this result has not been consistently found (McKay, 2021; MacCutcheon et al., 2020); music training has also been reported to reduce the impact of ageing on central sound processing but not the auditory periphery (Zendel et al., 2019). However, the effect of musicality on speech perception is not always shown at the cortical level (Jasmin et al., 2024). The Wechsler Test of Adult Reading (WTAR) was also used to collect the reading ability of irregular words, which reflects crystallised intelligence and premorbid intelligence (Venegas & Clark, 2011). Here, we use WTAR primarily as a test of literacy. However, for potential future studies on patients, this is an important measure to differentiate participants with no cognitive impairments from potential mild cognitive impairment.

I hypothesise that all the above-mentioned auditory cognitive predictors are relevant to SIN recognition, and different domains of auditory cognition (periphery, central, general cognition) can explain independent variances in SIN perception.

3.2.1 Methods

Participants

A total of 177 datasets were included in this analysis with 115 female and 62 male participants aged 48.56 on average (SD = 15.15). They had a wide range of hearing abilities (see Figure 3.4).

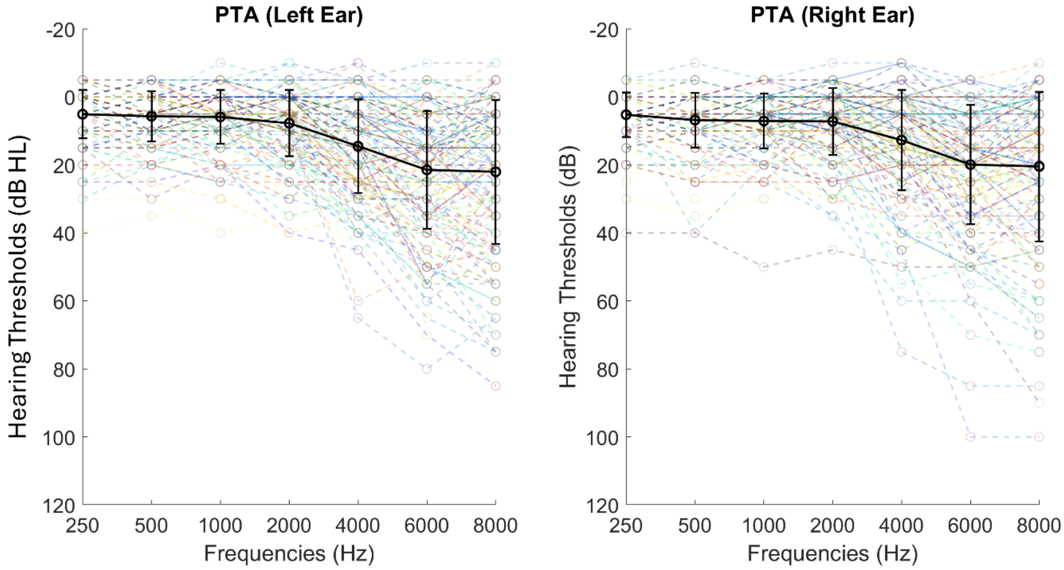

Figure 3.4 PTA results. The individual PTA thresholds are plotted in coloured dashed lines with circles. The group average is plotted in black lines with circles. The error bars are the standard deviations. The x-axis represents the frequencies tested in Hz, and the y-axis represents the hearing thresholds in dB HL.

The same set of tasks was presented to participants using a computer monitor (Dell Inc.) in a soundproof booth. The auditory stimuli were played through a sound card (RME FireFace UC) connected with headphones (Sennheiser HD 380 Pro).

Materials

The SiB were the same as used in the online validation study. See Section 3.1.1 for more details. In addition to the SiB test, a word-in-noise (WiN) was added to capture a different aspect of SIN perception (Guo, et al., 2024). I developed this task in collaboration with colleagues at Newcastle University, Iowa University, and UCL based on the ITCP test (Geller et al., 2021). The details of the test development are described in Section 4.1. The task had 120 trials with balanced female and male speaker sounds. The target words were common monosyllabic words, and the babble noise was a 8-talker babble. Participants were presented with 4 alternatives per trial and asked to choose the one that corresponded to the target word. The SNR was 2 dB, and the outcome was measured as the proportion of correct answers.

The AFG gap discrimination (AFG-Gap) task was the same as the one used for the online study. See Section 3.1.1 for more details. In addition, I used the AFG figure detection task (AFG-Det), which had a similar stimulus configuration as AFG-Gap. The figure was made of 3 components that repeated for 6 chords long. The ground was composed of randomised frequency over 40 chords. Two sounds per trial were presented to the participants, one of which contained a figure. The adaptive procedure was the same as the AFG-Gap task, with a one-down one-up adaptive design, starting with 4 dB TMR, followed by 2- and 1-dB steps after 1 and 4 reversals.

The GAP-Det was a between-channel gap detection task based on Phillips et al. (1997). The GAP-Det stimulus consisted of two narrow-band noises with a bandwidth of 0.25 octaves with a 0.5 ms ramp separated by a silent interval. The first noise was a 10 ms sound centred at 4 KHz. The second noise was a 1 KHz tone of 300 ms. Participants were presented with a pair of these sounds, where one sound contained a gap, and the other one had no gap (the “no-gap” sound had 1 ms between sounds). The inter-stimulus interval was 600 ms long. The task was assessed with a 1-up 2-down staircase procedure with the duration of the gap changing based on performance. The starting duration was 200 ms. The test terminated after 19 reversals which allowed most participants to reach a stable performance. The outcome was the median of the last 6 reversals.

The two AUM tasks included a frequency and AM rate discrimination task described in Lad et al. (2022). The two tests shared the same paradigm but different auditory stimuli. The stimuli were pure tones from 440 to 880 Hz for the frequency discrimination task and white noise modulated with a sine wave (100% depth) from 5 to 20 Hz for the AM-precision task. Participants were asked to keep a sound in mind and ‘find’ the corresponding sound with the same frequency or modulation rate on a fixed horizontal scale that they could interact with after a delay. After showing a fixation cross at the centre of the screen, the initial stimulus would be played for 1 s. After a 1-4 second delay, a slider would appear with a movable marker for participants to click to match with the first sound they heard. Each click would generate a corresponding sound for the participants to match with the experimental sound. There were no limitations on the number of clicks participants were allowed to do. The performance of the two tasks was quantified by ‘precisions’, scored using a Gaussian function that estimated the inverse of the standard deviation of the errors in each trial across the whole experiment (Lad et al., 2024).

The Goldsmiths Musical Sophistication Index questionnaire was used to assess participant's general music competence of both musicians and non-musicians. The details were described by Müllensiefen et al. (2014). The questionnaire measures different aspects of musical sophistication such as active engagement (listening or practising), perceptual abilities, and emotional responses to music. The maximum score is 126.

The Matrix Reasoning task was adapted from Chierchia et al. (2019). The test item consisted of a 3 x 3 matrix containing abstract shapes. One cell of the matrix was empty and needed to be completed by the participant within 30 seconds. 26 items were selected from Test Form 1. The first five items were used for practice. Starting from five, the main test items were chosen sequentially from 6 to 25 with one extra harder item added (item 47) to avoid ceiling performance. All participants were shown the same items in the same presentation order.

The raw score of the DS-backward task was used for data analysis, which was calculated as the correct responses scored as 1 summed together. The correct responses of WTAR were also scored as 1 and summed in the end. I used the standardised scores based on the WAIS manual for data analysis (Wechsler, 1955).

Data Analysis

To analyse the relationship between variables, I conducted bivariate correlation tests between the two speech measures and other auditory cognitive predictors included in this study (Spearman). To correct for multiple comparisons, the Holm-Bonferroni correction was used. Stepwise linear regression was carried out with the SiB and WiN scores as the outcome variables and PTA, age, the two AFG measures, gap detection, the two AUM measures, digit-span test, Gold-MSI and matrix reasoning as the predictors. To test if age played an important role in the data, a post-hoc correlation analysis was carried out separating participants into a younger and older group splitting from the median age (50.28). These analyses were conducted in MATLAB R2021a.

To better understand the inter-relationships among the auditory cognitive predictors and account for different aspects of SIN perception, SEMs were also constructed using the lavaan package (version 0.6-15) in R (version 4.2.1). SEM is a multivariate analysis method that allows the modelling of multiple observed variables to indicate a latent variable, which is a hypothetical construct that is not directly measured but can be inferred by their observed variables. Maximum likelihood estimation was used with nonnormality correction based on the Satorra-Bentler scaled test statistic. Robust measures were reported in this study (Brosseau-Liard et al., 2012; Brosseau-Liard & Savalei, 2014). The models were evaluated by a set of criteria (Hu & Bentler, 1999; Kline, 2015). These included the Bentler comparative fit index (CFI) and Tucker-Lewis Index (TLI), the root-mean-square error of approximation (RMSEA), the standardised root mean squared residual (SRMR), and the chi-square test. Both RMSEA and SRMR are absolute measures of the estimated discrepancy between the predicted and observed models. The SRMR is a measure of the mean absolute correlation residual measuring the differences between the original correlations (observed) and the implied correlations by the model. RMSEA ≤ 0.06 and SRMR ≤ 0.08 have been suggested to indicate a close model fit (Hu & Bentler, 1999). RMSEA up to 0.10 is considered a fair fit, but above 0.10 is generally unacceptable (Browne & Cudeck, 1992). CFI and TLI, on the other hand, are incremental indices that reflect the relative improvement of the model fit compared to a baseline model (Kline, 2015). TLI is non-normed so it can fall outside the 0-1 range whereas CFI is normed, but the cutoff for both of them is above 0.95 for a good fit (Hu & Bentler, 1999). The chi-square (χ^2) result was also reported (Kline, 2015). The null hypothesis for the chi-square test was that the predicted model perfectly reflects the true data. Thus, a nonsignificant chi-square would indicate a good model fit. These criteria are summarised in Table 3.3.

Fit Index	
χ^2 (p)	≥ 0.05
RMSEA	< 0.100
CFI	> 0.90
TLI	> 0.90
SRMR	≤0.08

Table 3.3 Criteria for acceptable model fit.

To confirm the choice of the scaling variable of the two latent structures, confirmatory factor analysis (CFA) was performed on the latent structures (Figure 3.5). Scaling variables are used to assign scales to latent variables, which is essential when identifying a model. The method used in lavaan is the Fixed Marker (FM) scaling that fixes the loading of the chosen scaling variable to 1 (Lavaan.Org - Model Syntax 2, n.d.). CFA could only be carried out on CPS and CPL, as only two indicators were available in the SIN latent structure and the model could not be identified. The choice of scaling variable for SIN was thus based on theory alone. The structural equation models constructed in this study would account for sound segregation on a short timescale (CPS) and longer timescale (CPL), as well as complex cognitive processes measured by matrix reasoning, musicality and reading abilities. As sentence-level perception would capture these processes better than word-level perception, the SiB test was chosen to be the scaling variable for SIN.

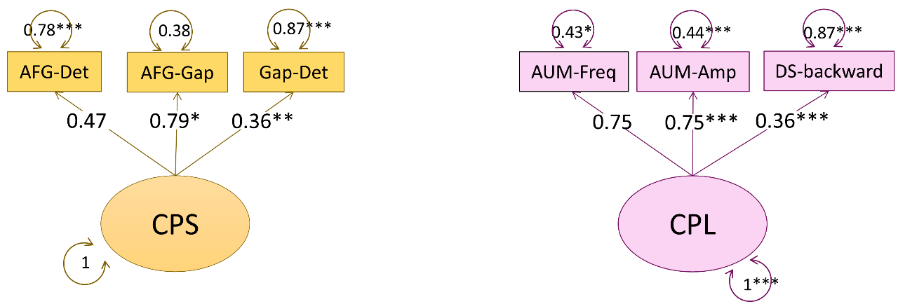

Figure 3.5 Confirmatory factor analysis. The oval shape represents the latent variable. The rectangles are the indicators. The arrowed circles are the error terms representing measurement errors not captured by the indicator or variance unexplained for the latent variable by the indicators. One asterisk represents $p < 0.05$, two represent $p < 0.01$, three represent $p < 0.001$.

Two structural equation models were constructed based on the CFA (Figure 3.6). Model I theorises that both PTA and Age predict SIN performance. The relationship between PTA, Age, and SIN was discussed in more detail in Section 2.2.1 and Section 4.2. The CPL was shown to predict SIN perception (Lad et al., 2020b, 2024) and hence was constrained to predict SIN in the SEM. The DS-backward test was found to be associated with SIN perception on a sentence level (Shokuhifar et al., 2024) and training on DS-backward ability was shown to improve SIN perception (Ingvalson et al., 2015). Music (Hennessy et al., 2022; Zendel et al., 2019), general intelligence as measured by the MTX test (Akeroyd, 2008), and reading ability have also been associated with SIN perception. They were thus all configured to predict SIN in Model I. Model II had the same latent constructs as Model I but with the MSI, MTX, and WTAR removed to simply the overall model. These exogenous variables reduce the degrees of freedom thus having an impact on the model fit. For example, chi-square tends to fit the data better for models with higher complexity whereas RMSEA incorporates the degrees of freedom and can lead to overfitting with more complex models. Under-identification due to insufficient sample size compared to the number of estimates can be a problem as in Model I. For Model I, the N:q was around 9:1, and for Model II it was 11:1, both under the ideal N:q ratio of 20:1 but over the acceptable limit (5:1).

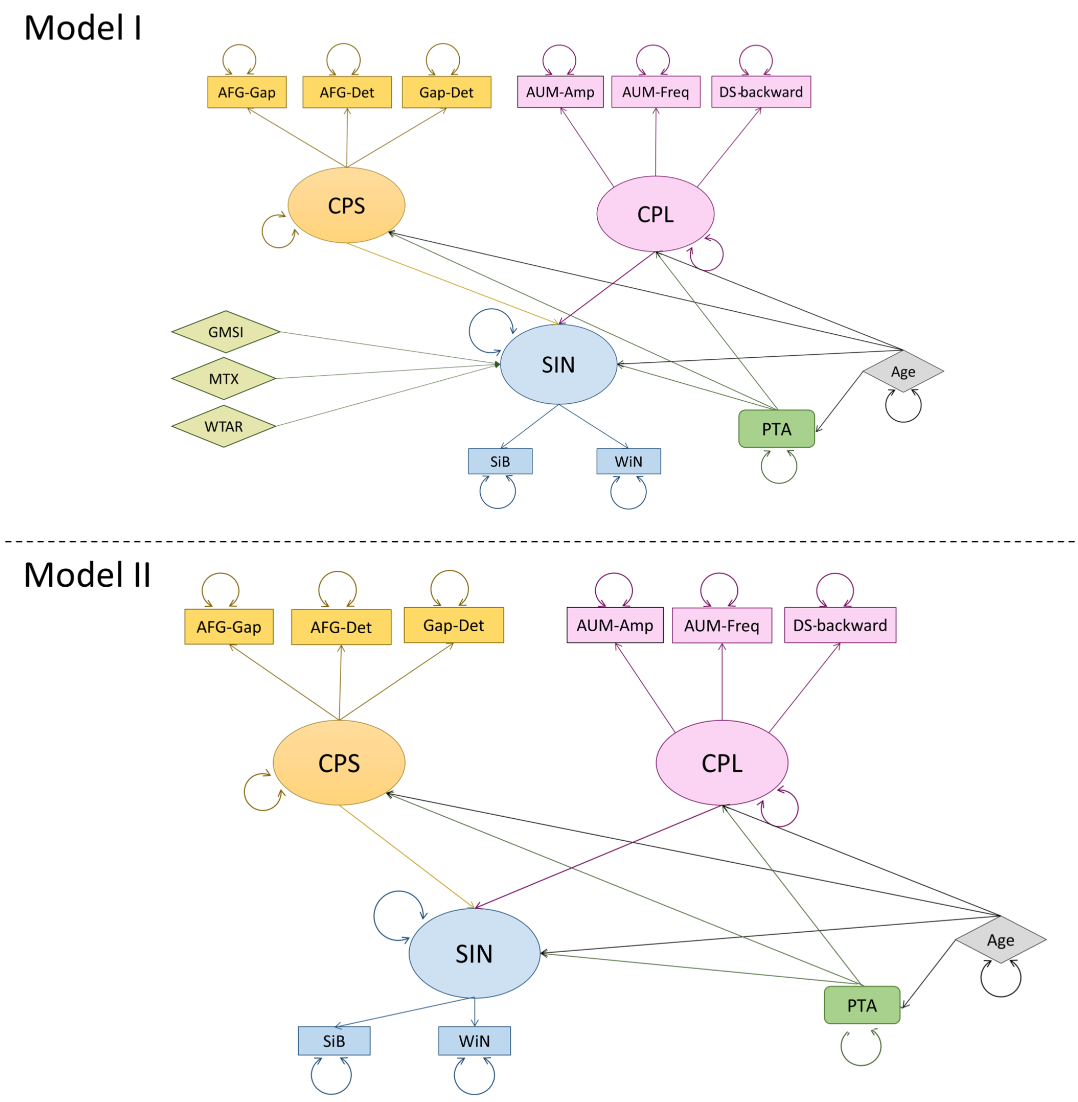

Figure 3.6 Conceptual SEM models. The latent variables are plotted in oval, with arrows pointing at their indicators plotted in rectangles. The exogenous variable is plotted in a diamond. The observed variable not included in any latent structure is plotted in a rectangle with rounded edges. The arrowed half circles are the error terms. All latent constructs are colour-coded: e.g., the CPS latent variable and its indicators as well as the arrowed lines are all plotted in orange.

3.2.2 Results

The correlation coefficients are summarised in Table 3.4. Both SiB and WiN had significant associations with the auditory cognitive predictors included in this study.

	Age	PTA	AFG-Gap	AFG-Det	Gap-Det	AUM-Freq	AUM-Amp	DS	MTX	MSI	WTAR
WiN	-.667***	-.616***	-.508**	-.325**	-.458**	.427***	.424***	.226*	.416***	.258*	.234**
SiB	.587***	.529***	.429***	.231**	.320***	-.323***	-.349***	-.209**	-.310***	-.168**	-.275**

Table 3.4 Correlation coefficients with Holm-Bonferroni corrected alpha thresholds. Three asterisks indicate p< 0.001, two indicate p < 0.01, one asterisk indicates p< 0.05.

The linear regression results are shown in Table 3.5. Age was the most important predictor for both SIN measures with the largest variance. For SiB, reading ability and DS-backward were also important predictors. Together the model accounted for 43.86% of the SiB variance ($F(4,172) = 35.374$, $p < 0.001$). For the word-level SIN measure, however, the AUM frequency discrimination was more important. The model including 6 variables accounted for 53.75% of the variance of WiN ($F(6,169) = 34.879$, $p < 0.001$).

Predictors	WiN (Adjusted r^2 change)	Predictors	SiB (Adjusted r^2 change)
Age	0.423***	Age	0.341***
+AUM-Freq	0.059*	+WTAR	0.059*
+WTAR	0.021*	+DS-backward	0.022*
+AFG-Gap	0.016 (ns)	+AFG-Gap	0.017*
+DS-backward	0.010*		
+AFG-Det	0.008*		

Table 3.5 The standardised coefficient beta of SiB and WiN. Three asterisks represent p level smaller than 0.001, two asterisks represent p<0.01, and one asterisk represents p<0.05, ns represents non-significant result.

The correlation analysis based on two different age groups is shown in Figure 3.7. The younger group showed a significant correlation between the SIN measures and both AUM and gap detection measures. The older group, however, showed a significant correlation in the WiN condition but a weak or nonsignificant correlation in the SiB condition.

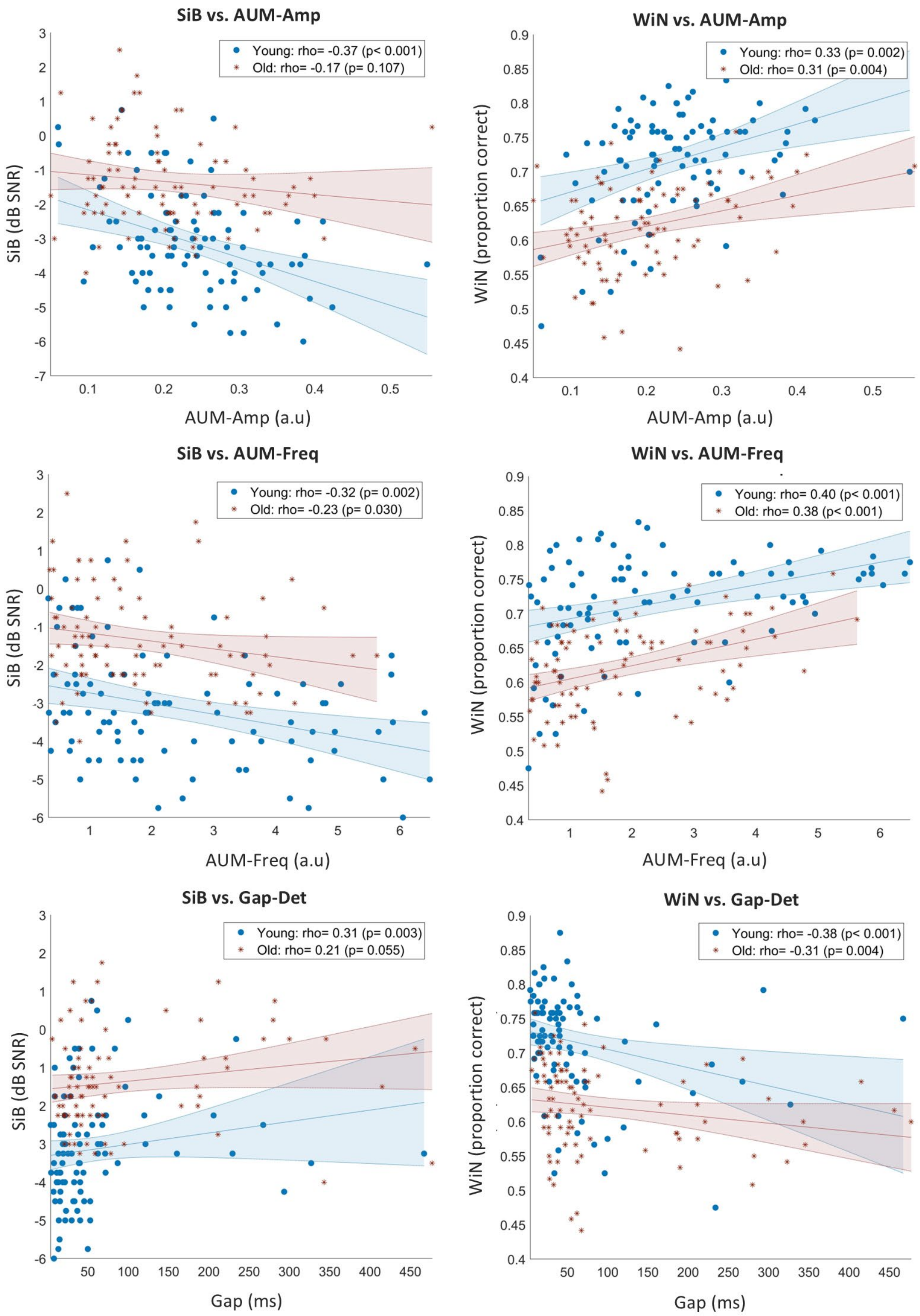

Figure 3.7 Correlation between SiB/WiN with the two AUM measures and gap detection based on the age group. Blue dots plot the data of the younger group and the red dots the older group. Shaded areas plot the error with the best line of fit plotted in the middle. Correlation coefficients and p values are shown in the legend.

The fit indices of Models I & II are presented in Table 3.6. The fit indices for Model I indicated an unacceptable fit and was rejected, but most of the fit indices for

Model II were within our criteria for a close fit. Chi-square was significant, which usually suggested a poor model fit. However, chi-square is a measure that is heavily influenced by the sample size. With a larger sample size chi-square very often shows a significant result regardless of the actual model fit (Bentler & Bonett, 1980). With all the fit indices and their theoretical validity taken into consideration, Model II was accepted. In Model II, the path connecting AUM and AFG was significant. PTA did not predict SIN significantly (Figure 3.8). The model explained 47% of the variance in SIN (adjusted $R^2 = 0.469$).

Fit Index	Model I	Model II
χ^2	181.629 (p<0.001)	58.203 (p<0.001)
RMSEA	0.116	0.078
CFI	0.824	0.951
TLI	0.756	0.921
SRMR	0.126	0.085

Table 3.6 Fit indices of SEM Model I and Model II.

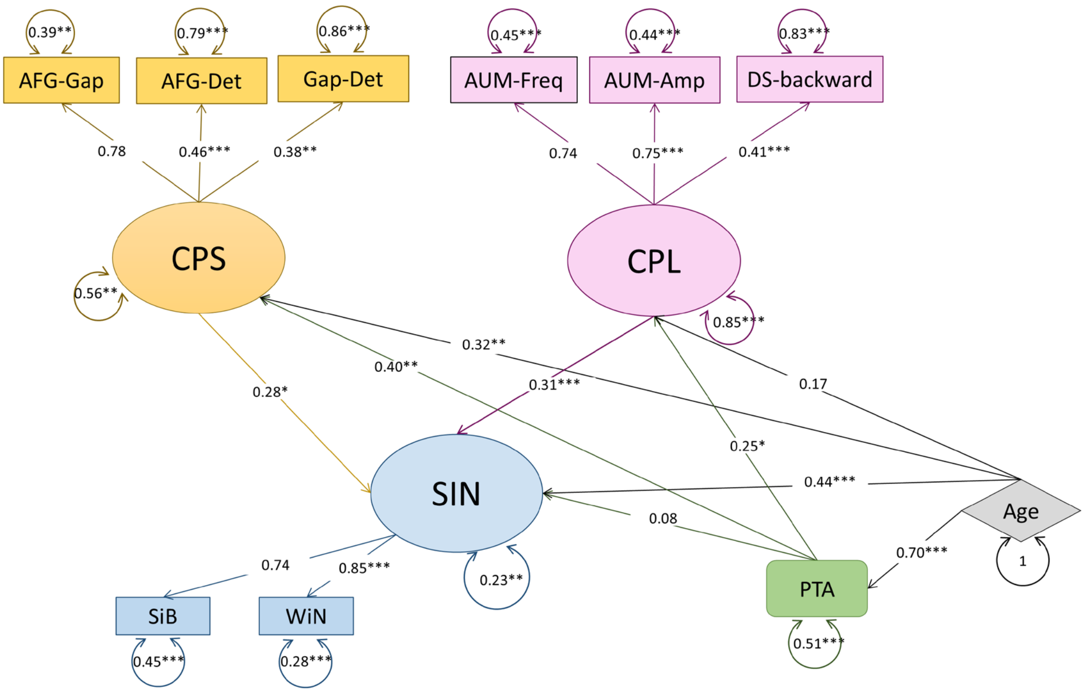

Figure 3.8 Model II with path estimates. The significance level is marked with an asterisk: three asterisks reflect p<0.001, two reflect p<0.01, and one reflects p<0.05. The scaling variables do not have an estimate hence no significance level is marked. The latent variables are plotted in oval, with arrows pointing at their indicators plotted in rectangular. The exogenous variables are plotted in diamond. The arrows pointing from the exogenous variables to the latent variables indicate a predictive relationship. The arrowed half circles are the error terms.

3.2.3 Discussion

Different Auditory Processing Mechanisms for Sentence vs. Word Level SIN Perception

While the bivariate correlation demonstrated significant correlations between all the auditory cognitive variables included in this test battery and the two SIN measures, the hierarchical regression models revealed the roles played by particular predictors in different aspects of real-world listening. The WiN test is a task focusing on consonant perception and relies heavily on the fundamental grouping or sound segregation. Age, as expected, came out as the most important predictor explaining the highest variance of WiN perception. PTA was not a significant predictor in this data likely because the sample was relatively young and had mostly normal-hearing participants. Based on the literature review presented in Chapter 2, a stronger association between PTA and SIN tends to be found in older samples with elevated thresholds.

The AUM-Freq task combined both frequency discrimination and holding a certain frequency in mind over time, which is particularly important for the consonant perception task. AUM-Amp did not predict either SiB or WiN significantly in the regression models. The result was congruent with the initial study by Lad et al. (2020), which found a significant association with the AUM frequency precision task but not the AM precision task. However, a recent study with a larger sample revealed that AUM-Amp was an important predictor of SIN as well (Lad et al., 2024). The sample characteristics were a major difference between those studies. In addition to the larger sample size 153 (Lad et al., 2024) vs. 44 (Lad et al., 2020), the more recent study also had an older sample (average 67 vs 30). The current study had a large (n = 177) sample size and a relatively young sample (averaged age = 49). It could potentially explain the consistency with results reported by Lad et al. (2020).

From the age-split correlation results, a significant effect of age can be found in affecting the relationship between the SIN perception scores and AUM scores. The importance of precision for both frequency and amplitude rate on the WiN task for both groups was expected, as consonant perception relies on both mechanisms. In this case, the correlation reflected more of the perceptual not the memory aspect of the AUM tasks: the ability to perceive and distinguish the specific frequency or amplitude rate. The SiB task, however, provided more linguistic content and the task can be approached differently based on individual abilities. For the younger population, I found a stable moderate association between SIN and AUM. The older population, on the other hand, revealed non-significant SiB to AUM-Amp and Gap-Det relationships. The younger group (with their mostly intact peripheral and central hearing ability) likely used acoustic cues to segregate the sentence from noise, while retaining the information for the duration of the task. For the older group with lower perceptual acuity and potentially deteriorated central hearing, the reliance on either was weaker. The regression model on SiB showed that WTAR and DS-backward were more important. Both tasks assessed something less specific to the acoustic features and more generally related to memory, language, and cognition. The older group could rely on their working memory and reading ability to perform the SiB task.

AFG-Gap was another predictor of both SIN measures. The AFG detection task, however, did not explain a significant variance in SiB. This result was congruent to what was found by Holmes, et al. 2019, which demonstrated a small added variance of the AFG-Gap to a sentence-in-babble task after accounting for PTA but not after accounting for AFG-Det as well. Here, I added a word-level test and found that only AFG detection task predicted WIN. The two paradigms differed only in their tasks: the AFG-Gap task was more related to figure-tracking over time and the AFG-Det was more related to instantaneous segregation as participants only needed several chords to perform the detection task (Teki et al., 2013). It is possible that this more transient form of sound segregation matters more to word recognition than sentence recognition in noise.

Both the previous literature (Holmes & Griffiths, 2019) and the current study have consistently found that the AFG gap discrimination task outperformed the AFG figure detection task in predicting either sentence or word perception in noise. This could be because detecting figures does not necessarily require figure-tracking over time. Participants could potentially exploit other mechanisms to achieve successful figure-ground segregation, such as differences in the acoustic energy level. The fixed-frequency figure would always generate higher energy at certain frequencies compared to the random-frequency ground (Figure 3.9). This is particularly a problem with the figure detection task. Since participants did not have to track the whole figure to perform the task, the increased power of the figure could be used as a strong cue to detect the presence of a target sound instantaneously so segregation based on temporal coherence might not be used in this instance. The problem was somewhat mitigated in the gap discrimination condition, where both sounds contained a figure per trial and participants were asked to keep track of the figure over time in order to hear the gap. This could explain why the gap discrimination task showed a stronger association with SIN perception and a higher contribution to the AFG latent structure.

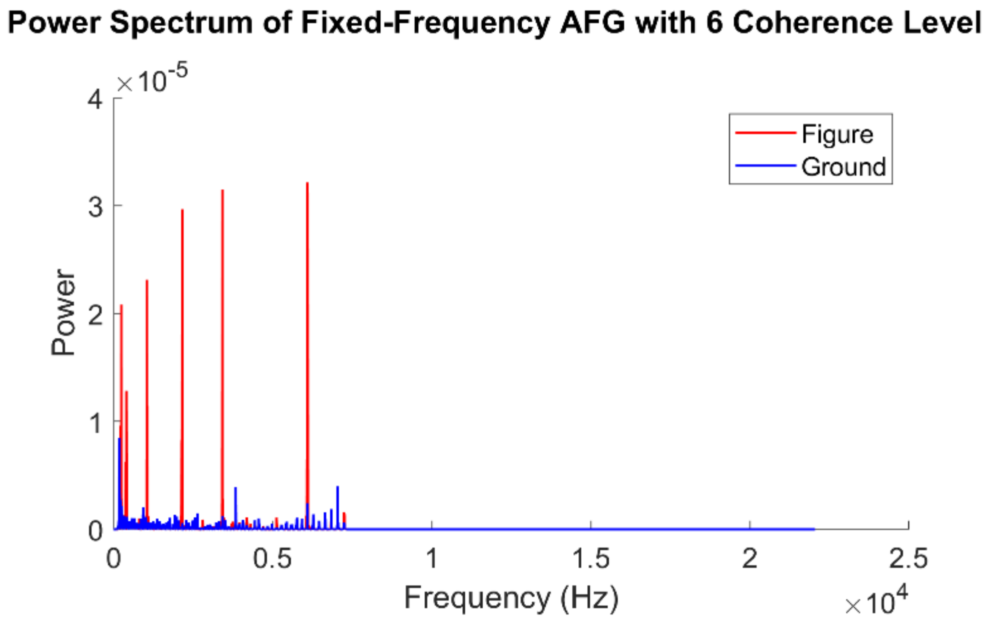

Figure 3.9 Power spectrum of a fixed-frequency figure-ground stimulus at 0 target-to-masker ratio.

Understanding Auditory Cognitive Predictors of Speech-in-Noise Perception in a Structural Equation Model

Model II revealed noteworthy inter-relationships between the peripheral, and central auditory processing, cognition, age, and real-world listening. Age was the dominant predictor of SIN in this model. This impact was not only direct but also mediated through CPS with a significant path coefficient of 0.32 to CPS. Age also modified PTA significantly through a high coefficient. However, age did not modify CPL significantly. This was probably due to the dominant effect of AUM-Freq on the CPL construct. As shown in Figure 3.7, the two age groups had very similar effect sizes in their respective correlations between AUM-Freq and SIN. This suggested that the frequency-domain AUM task was relatively robust against ageing. The sample was relatively young as well, which might have limited the range of data acquired. DS-backward also did not correlate with age significantly (r = -0.01).

Consistent with the regression results, PTA did not predict SIN directly. However, the model revealed an important mediation effect of PTA modifying SIN through a large impact on CPS and CPL. The significant modifying effect on CPL was driven by the two AUM measures. In the current study, PTA did not correlate with DS-backward significantly ($p = 0.217$). However, as the two AUM measures assess both the perception and retention of frequency and AM rate, PTA should directly modify the AUM measure.

The two latent structures modified SIN significantly with a similar effect size. The current model suggests that after accounting for age and PTA, the transient central sound processing had a similar level of impact on SIN perception as central processing involving short-term and working memory. Together with the regression results, they suggested an intriguing direction for the development of new hearing measures that can combine both processes in order to best predict SIN perception. The test should include the sound segregation aspect of AFG and involve working memory in the task.

3.2.4 Limitations and future direction

The study was initially designed to be an online experiment with the expectation of obtaining a much larger sample size. However, the online study validation showed poor test-retest reliability and data collection was hence carried out in the lab only. This has greatly limited our speed of data collection, resulting in a poor fit for Model I. A larger sample should be able to power the more complex interactions in Model I. However, the sample size was sufficient for a less complex model (Model II).

The AFG design can be improved to eliminate the power difference between the figure and the ground. Besides using a discrimination task where both stimuli have a figure, one way of removing the power accumulation due to frequency repetition is the introduction of frequency-variant figure components. Having a changing frequency contour would also make the auditory figure more like natural speech, which carries dynamic frequency contours such as the fundamental frequency and formants. In addition, I have demonstrated that the current AFG was more important for word than sentence perception. It is possible that additional frequency variation would improve its predictability of sentence-level perception as well. Incorporating a memory task similar to the AUM tasks into the AFG design could also improve its ability to predict sentence-level perception.

3.3 Explore the potential link between SIN listening and cognitive functions

The first two sections of Chapter 3 are based on strong priors in literature. This chapter, however, will dive into the realm of more speculative investigations of a potential causal influence of SIN hearing on cognitive decline based on cross-sectional priors only. I will use the same method and dataset as Section 3.2 and study the possible contributions of peripheral and central auditory processing and SIN ability in cognitive functions.

3.3.1 Modelling the link between hearing loss and cognitive decline

As reviewed in Section 1.4, a potential link between hearing loss and dementia or cognitive decline has been both proposed and was evidenced in cross-sectional studies (Loughrey et al., 2018) as well as longitudinal studies (Lin et al., 2012, 2013; Merten et al., 2020). Upon detailed examination, researchers found that this relationship might have strong central auditory involvement as measured by SIN tests; SIN impairment was independently associated with incident dementia with a 61% increased risk (Stevenson et al., 2021), SIN performance had a stronger association with cognitive function than PTA (Hoff et al., 2023), and age- and hearing-matched participants with or without mild cognitive decline showed a significant difference in SIN performance (Mamo & Helfer, 2021). In this section, I aim to explore the data from Section 3.2 further to examine the link between SIN perception and cognition with structural equation modelling.

Similar to the analysis carried out in the previous section, I used the SEMs consisting of a central auditory processing (CAP) latent variable indicated by AUM-Amp, AFG-Gap and Gap-Det, and a SIN latent variable indicated by WiN and SiB. Having more than three indicators is not recommended for SEM (Hayduk & Littvay, 2012). I chose the AUM-Amp over AUM-Freq as the scaling variable based on literature, where AUM-Amp was found to be a better predictor of ACE-3 measured cognition (Lad et al., 2024). AFG-Gap was shown in Section 3.2 to have the strongest correlation with SIN. I also chose the gap detection task to group under the CAP latent structure as it was the only measure of temporal acuity in the test battery. The outcome latent variable here was a general cognition variable (GCog), indicated by the cognitive measure in the study: DS-backward, MTX, and WTAR. These cognitive tests measure different domains of cognition including working memory (DS), fluid intelligence (MTX), crystallised intelligence and reading ability or literacy (WTAR). Age was configured to predict not only SIN and CAP as in the previous section but also GCog. Here, all cognitive measures used the non-age adjusted raw scores. The conceptual model with GCog as the outcome variable is shown in Figure 3.10. The order of the indicators was decided based on the strength of the correlations. The hypothesis tested was PTA, CAP, and SIN could modify GCog independently. The choice of the scaling indicators can impact the magnitude of the unstandardised regression path estimates (Klopp & Klößner, 2021) but it would not affect the model fits based on the maximum likelihood estimation (Bollen et al., 2022).

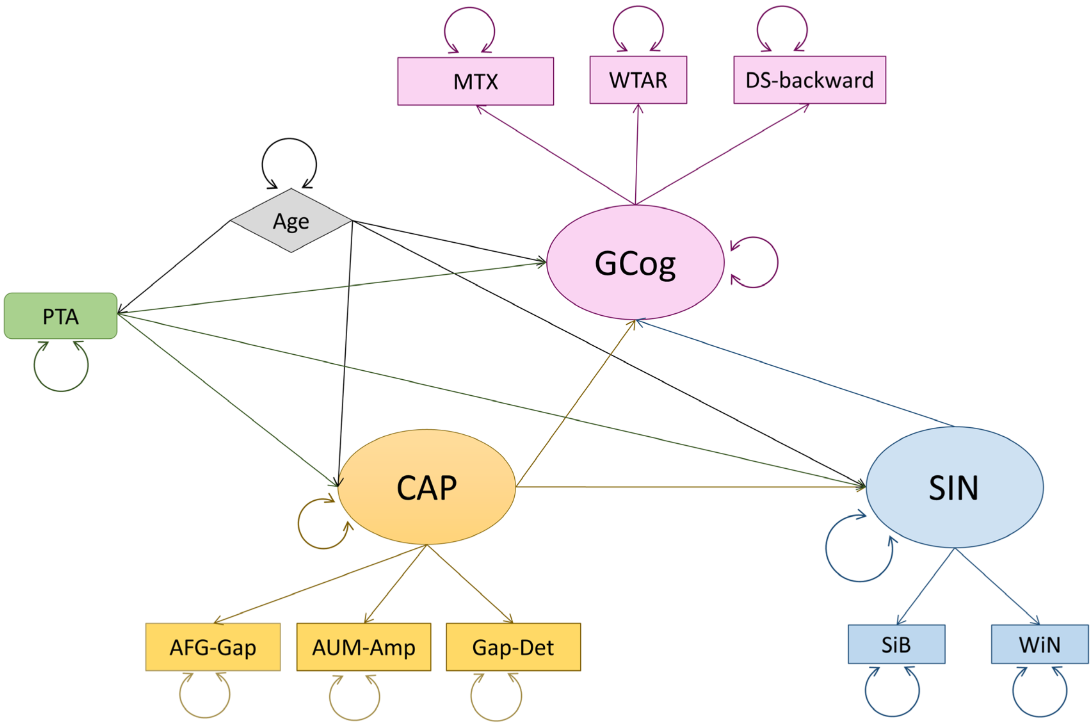

Figure 3.10 Conceptual GCog model. The latent variables are plotted as ovals, with arrows pointing at their indicators plotted in rectangles. The exogenous variable is plotted as a diamond. The observed variable under no latent structure is plotted in a rectangle with rounded edges. The arrowed half circles are the error terms (or residuals, defined as variance unexplained by the measure due to score unreliability). All latent constructs are colour-coded.

3.3.2 Results

Correlations among the cognitive variables and age and auditory variables are tabulated in Table 3.7. The table summarises Spearman's rank correlation coefficients among the variables. To correct for multiple comparisons, Holm-Bonferroni correction was used based on 3*9 pairs of comparisons. The adjusted p-values are shown in the table. I consistently found a moderate correlation between MTX and age as well as all the auditory predictors. WTAR and DS-Backward correlated with WIN, SIB, gap detection, and AUM-Freq.

	Age	PTA	WIN	SIB	AFG-Gap	AFG-Det	Gap-Det	AUM-Freq	AUM-Amp
MTX	-0.39***	-0.36***	0.42***	-0.33***	-0.33***	-0.28***	-0.38***	0.49***	0.44***
WTAR	0.10	-0.02	0.10	-0.17	-0.05	-0.09	-0.24*	0.41***	0.17
DS-backward	-0.02	-0.08	0.23*	-0.22*	-0.15	-0.02	-0.31***	0.33***	0.36***

Table 3.7 Correlation coefficients between cognitive measures with auditory measures. This table summarises the correlation coefficients among the cognitive variables and auditory variables with statistical significance at $p < 0.05$ marked with one asterisk, $p < 0.01$ with two asterisks, and $p < 0.001$ with three asterisks.

Structural model

The cognitive model (Figure 3.11) showed an acceptable fit for all the fit indices (Table 3.8). The model was accepted. All the paths showed significant path coefficients except for PTA to SIN, SIN to GCog, and Age to GCog. The model explained 48% of the variance in cognition (adjusted $R^2 = 0.484$).

Fit Index	
RMSEA	0.082
CFI	0.945
TLI	0.909
SRMR	0.072

Table 3.8 Fit indices of the cognitive model.

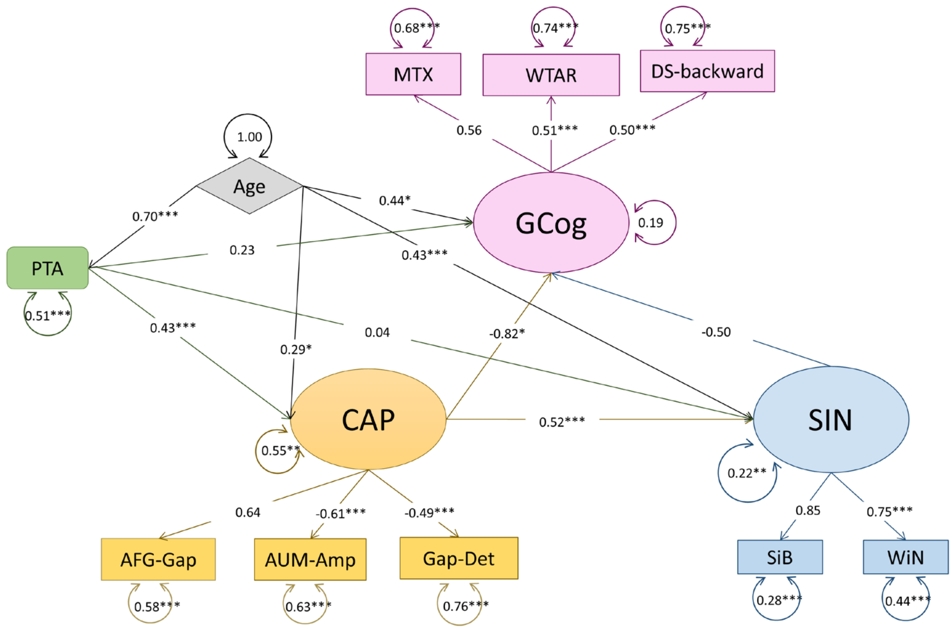

Figure 3.11 Cognition model with path estimates. The significance level is marked with asterisks: *** suggested p<0.001, ** suggested p<0.01, * is p<0.05.

3.3.3 Discussion

The results demonstrated important associations between cognition and auditory predictors including peripheral hearing and central hearing. The model explained 48% of the variance in the general cognition latent variable. Consistent with previous literature, SIN measures correlated significantly with measures of fluid intelligence, working memory and crystallised intelligence. However, PTA only correlated with the measure of WTAR. This finding suggests, in terms of the association between hearing and cognition, reading ability and the cognitive control and executive aspect of working memory are not as relevant as fluid intelligence. Reading ability retains relatively well over ageing so it is not surprising to see a nonsignificant correlation here despite using the WTAR raw score (Dykiert & Deary, 2013). However, age has been established to have a negative influence on digit span performance (GrÉGoire & Van Der Linden, 1997). However, I did not find this correlation in the current sample (see Figure 3.12 for a detailed visual illustration). Figure 3.12 shows that different age groups scored almost exactly the same on the DS-backward test. It is unclear if this is due to our sample characteristics but there was a decent range of age, and the sample size should be enough for the correlation test. More investigations are needed to uncover the reason for the lack of association between DS and age.

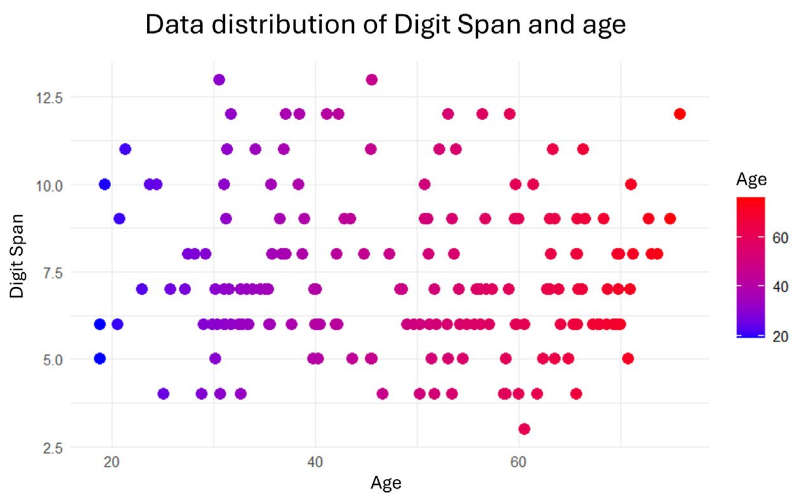

Figure 3.12 Scatterplot of DS-backward and age. The x-axis plots the age, and the y-axis plots the SD-backward score. The plot was colour-scaled based on age.

Central sound segregation as measured by the two AFG tests was also shown to be associated with the matrix reasoning task, but not with the WTAR or DS. Temporal acuity measured by the gap detection task and the two auditory memory tasks, however, exhibited low to moderate associations with both MTX and DS-backward. The perceptual precision and the memory of frequency were important for all three domains of cognition.

To understand the interactions among these complicated pairs of relationships, the SEM provided an exploratory solution. The CAP variable had the most significant contribution from AFG-Gap, followed by AUM-Amp and Gap-Det. Consistent with the findings of the previous section, age and CAP predicted SIN significantly. PTA predicted CAP significantly and influenced SIN indirectly through CAP. To explore the question of what aspects of hearing can predict cognition, age, PTA, CAP and SIN were configured to predict GCog. The current data demonstrated that CAP was the most important factor in modifying changes in GCog. SIN also had a high coefficient but was not significant after accounting for CAP, PTA, and age in this model. It is possible that the correlations found in the previous literature between the general cognition measures and SIN measures (Hoff et al., 2023; Stevenson et al., 2022; Merten et al., 2020) reflected the central auditory processing aspect of SIN perception instead of the linguistic or social cues. PTA did not change GCog directly but might have modified GCog through its impact on CAP. Age modified all latent constructs and

PTA significantly, highlighting the importance of accounting for age as a general factor affecting both listening and cognition in experimental designs.

While the current study tested a statistical causal model linking SIN perception, hearing, age, and central sound processing with cognition, SEM models are not designed to derive causal relations (Bollen & Pearl, 2013). The model construction was based on mainly previous observational studies (Section 1.4). To validate the findings, longitudinal studies and patient studies are needed with more controlled manipulations of the experimental condition. Although no strong causal link between central auditory processing and cognitive functions can be concluded here, the current study provides a direction for future patient work: most of the studies in the field focus on peripheral hearing and one or two aspects of cognition such as executive function or working memory. I demonstrated the importance of considering a more comprehensive picture of hearing and cognition in multivariate models including central hearing measures, SIN tests, and age to delineate the specific roles of listening in different domains of cognitive functions.

In conclusion, the exploratory analysis here showed evidence supporting a potentially important role of central sound processing in general cognition consisting of fluid intelligence, phonological working memory, and crystallised intelligence. SIN measures have a small to moderate correlation with all domains of cognition measured in this study, but the structural model showed that SIN did not directly modify cognition. Instead, central sound processing, which explained a larger variance in SIN than age, predicted general cognition in SEM. The results suggest the potential predictive power of central sound processing involved in SIN perception on cognitive changes. While the pure-tone thresholds failed to predict cognition in this study, they influenced cognition indirectly through central auditory processing.

Future directions

The present findings point to a potentially important role played by central auditory processing measured by sound segregation, auditory short-term memory and temporal acuity in performing cognitive tasks. Further research could investigate this link further with longitudinal data using piecewise analysis that can quantify the rate of change while modelling the relationships. A similar paradigm could be used in dementia research to better characterise the role of auditory functions in AD dementia development in future.

4. Chapter 4: Developing new measures of real-world listening

The SEM results of Chapter 3 suggested an integral model of SIN perception explained by age, PTA, sound grouping and spectrotemporal processing, as well as auditory memory. In this chapter, I present the development of new measures of central sound processing based on the findings of the previous chapters. The first section of this Chapter shows the validation of a word-in-noise task: the British Iowa Test of Consonant Perception. This is a test I developed for both clinical and research use targeting British-English speakers, providing an easy-to-use tool, available as an open-source standalone application to the scientific community for examining word-in-noise perception. Following the word-in-noise test is a new dynamic type of auditory figure-ground test. I designed this new paradigm based on the findings of Section 3.2, which suggested the importance of auditory working memory and varying the frequency components of the figure. The new dynamic figure-ground incorporated the pitch trajectory of natural speech to increase its resemblance to actual speech as well as a pattern discrimination task. They were both devised as tests for SIN analysis that could benefit research as well as clinical practice.

4.1 British version of the Iowa test of consonant perception

4.1.1 Introduction

As discussed in Section 2.3.1, listening to speech in a noisy environment can be challenging. This is especially true for listening to speech from an unfamiliar dialect or accent in a noisy environment. The problem does not only affect non-native speakers: adult speakers of Southern Standard British English, for example, have been found to show lower processing speed when listening to Glaswegian English, especially in adverse listening conditions (Adank et al., 2009). Children tend to struggle more than adults when confronted with accented speech (Bent et al., 2021). Aside from word recognition accuracy, other aspects of speech processing are also affected. More listening effort is needed when people listen to a less familiar accent (Van Engen & Peelle, 2014), and people of different age groups might have different processing strategies. Older adults rely more on cognitive resources compared to young people when processing accented speech (Ingvalson et al., 2017). This highlights a problem with the current implementation of hearing assessments both in research and clinics:

practitioners are often limited by the materials available to them and these materials might not be suitable for the population they test.

This is the case in the UK audiology practice. A large number of commonly used speech tests used in the UK are not available in British English or validated with the British population. A recent survey of British Audiologists and ENTs on current clinical practice for the evaluation of auditory processing disorder (Browne et al., 2024) reported that the QuickSIN was the most commonly used screening tool for adults and the SCAN-3C (Dawes & Bishop, 2007) for children, with both tests recorded in American English. For both children and adults, the most commonly used test was the Listening in Spatialised Noise Sentences Test (LiSN-S, Cameron & Dillon, 2007), available only in American and Australian English. This leads to a concern about overidentification. Researchers found that the American-accent hearing tests used to assess British English speakers could easily misidentify hearing problems by using a uniform standard (Dawes, 2011; Dawes & Bishop, 2007). From the patient's perspective, developing and validating a well-designed SiN test is also in line with patient-identified research priorities in the UK: individuals and families diagnosed with auditory processing disorder report the need for diagnostic tests as one of 3 top priorities (Agrawal et al., 2021); UK patients with mild to moderate hearing loss place the need for "realistic tests" of everyday hearing and potential use of SiN tests for hearing aid rehabilitation within the 15 top research questions that need to be answered (*James Lind Alliance*, 2024).

Making hearing tests available in the appropriate accent is clearly beneficial. From a research perspective, having matched versions of the same tests across accents can bring unique research opportunities. Research on the effect of accent and SiN can be investigated simultaneously. Such studies have been carried out but with individually recorded target stimuli and often a generic babble noise across accents that does not provide effective masking. Having matched tests across accents also presents an opportunity for larger public health-oriented work, taking advantage of the "natural experiments" to assess the efficacy of various remediation approaches. For example, different criteria for medical interventions, such as cochlear implantation, are used in the US and UK. Candidacy for a cochlear implant in the UK, based on NICE guidelines, requires hearing loss ≥ 80 dB HL at two or more commonly measured frequencies (between 500-4000 Hz). Contrastingly, common guidelines in the USA permit implantation when open-set sentence recognition in the best-aided condition is <60%, regardless of the degree of hearing loss. Without equivalent materials across the two dialects, a direct comparison between the two cohorts is difficult. Therefore, a SiN test that allows for better-controlled comparisons between the US and UK populations could be important as it can potentially guide both clinical practices and research.

What type of SiN test would be a valuable addition to the current array of tests? As was reviewed in Chapter 2, the current zeitgeist in the field is to use the most ecologically rich form of speech-in-babble tasks such as HINT or the AzBio. However, the skills employed to perform these tasks are not always auditory or even perceptual. Sentence repetition (even in quiet) is a complex *cognitive* skill requiring lexical access, word recognition, sentence processing, and language production along with embedded skills like working memory (Klem et al., 2015). Supporting this, sentence repetition *in quiet* is often seen as one of the best predictors of Developmental Language Disorder (Wang et al., 2022). Some of these skills may also be affected in people who have hearing loss. For example, language may decline with age even in normal-hearing individuals (Colby & McMurray, 2023; Payne et al., 2014; Waters & Caplan, 2001), or might be disrupted in children developing language with a hearing loss (Tomblin et al., 2015; Dunn et al., 2014). Consequently, single-word tasks—if the words are well balanced from across the phonological space, may serve a valuable role in controlling some of this non-perceptual variability and contributing to the research and clinical resources. In addition, materials of speech in different accents are widely available for sentences (WILDCAT Corpus (Van Engen et al., 2010)) but not for words.

Similarly, open-set responding also poses speech production demands that may be challenging for some populations. In contrast, a closed-set task – in which the response options are carefully chosen to reflect specific phonological dimensions of interest may be able to overcome this, maintaining a reasonable degree of difficulty while allowing the assessor to target particular dimensions of interest more precisely.

The Iowa Test of Consonant Perception (ITCP) was recently developed to address the concerns of imposing non-perceptual elements on speech testing (Geller et al., 2021). It is a single-word, closed-set task that has a good balance of phonetic contrasts (expressed in the response options for each word) which covers the entire phonetic range of the English language. The original test showed very good test-retest reliability, as well as validity based on comparisons with the CNC word recognition test (Lehiste & Peterson, 1959) and the AzBio sentence recognition test (Spahr et al., 2012).

This study sought to develop a British version of the same test using British English speakers with the mainstream Standard Southern British accent. This is the modern equivalent of 'Received Pronunciation', which is widely used in education and the media. The development of ITCP-B aimed at benefiting both clinical practice and research.

To this end, I created a British version of the ITCP (British-ITCP or ITCP-B) for UK English speakers and validated it under laboratory conditions. The ITCP-B leverages the careful work of Geller et al (2021) in identifying an optimal and representative set of items and their response options, and simply replaces the audio with appropriate British accented versions of each stimulus. I evaluated performance accuracy, the test-retest reliability and the cross-talker validity to assess the reliability of the test itself. I also assessed the correlation between the pure-tone audiogram (PTA) and ITCP-B, and the correlation between ITCP-B and a sentence-in-babble (SiB) measure for the convergent validity (Holmes & Griffiths, 2019).

The ITCP-B is free and openly available to the community in the form of a testing APP and scripts that can be easily modified (https://osf.io/53jsg/files/osfstorage). It establishes a phonetically balanced measure of word-in-noise perception that, along with the freely available US ITCP stimuli, will allow direct comparisons between UK and US cohorts using a similar measure, and could facilitate combined studies in the two regions.

4.1.2 Methods

Participants

Forty-six English native speakers born and educated in the UK were recruited for the experiment (30 females, 16 males). Participants were excluded if they had a history of auditory disorders, speech or language disorders, developmental or neurological disorders or were taking psychotropic drugs. Participants were included if they were over 18 years old, and no upper limit was imposed. This is to obtain a representative sample. The PTA averaged across 0.25~8kHz (in the left and right ears) of the sample was 13.92 dB HL, and the standard deviation (SD) was 8.42 dB HL. The average age was 48.65 (SD = 12.18). Out of the 46 participants, more demographic information was collected on 33 participants on their employment status and levels of education. Approximately 39% of participants had full-time employment, 12% had part-time employment, 33% were retired, 6% were still at university, 3% were full-time parents, and 6% were unemployed. In terms of education, 36% had a postgraduate degree, 45% with an undergraduate degree, 9% with A levels, and 9% with GCSEs.

Materials and Design

Recordings were made by two native English speakers (one male and one female) with the Standard Southern British accent. There are many accents in the UK and the received pronunciation was chosen because it is experienced by the majority of the UK population that is exposed to radio and television, even if it is not characteristic of their region.

The word list of the original ITCP test was recorded for each speaker (120 word sets per speaker). These are consonant-vowel-consonant words such as “ball-fall-shawl-wall”. Recordings were made in a sound-proof booth using a large-diaphragm condenser microphone (Rode NT1-A) with a pop filter placed in front. These recordings were made in Audacity (version 3.1.3), with a sampling rate of 44.1 kHz and 16-bit resolution. For both talkers, words were spoken as clearly as possible, at least twice with the carrier “he said [word]” and twice without. This phrase was included to help ensure uniform prosody and rate. Offline, all words were imported into Audacity, the “Clip Fix” function was applied with a 95% threshold for clipping and amplitude reduction overall by 5 dB (to allow for restored peaks). Noise reduction was then applied to the entire recording based on the noise profile for a silent period (with 12 dB reduction, sensitivity set to 6.00 and frequency smoothing set to 3). Each word exemplar was marked for cropping at the zero crossing, exported as a .wav file and then scaled to the same RMS level in Praat (version 6.2.14 (Boersma, 2001)) before being re-exported as a final “cleaned” .wav file. The mean duration of the words used was 0.51 s (±0.086 s)

The noise was extracted from an 8-talker babble soundtrack with 4 male and 4 female voices that lasted for 15 s in total. Importantly, this babble contained British voices. Segments of the babble noise were taken randomly as a masker for the target word, which was always played 1 second before the target sound and stopped at the offset of the target words. The babble noise was mixed with the target sound with a -2 dB signal-to-noise ratio (SNR).

Procedure

The validation testing of ITCP-B was based on two sessions (Session A, and Session B). The order of the two sessions was random, subject to participant availability. The two sessions were typically separated by 10 weeks (median duration = 80 days, range = 5~356 days). In both sessions, researchers carried out all three tests in the following order for all participants across sessions: audiometry, ITCP-B and SiB tests. The two sessions were identical except for the SiB test: Session A tested the longer version of the SiB test and Session B had the same SiB test but shortened by half (this turned out to be unreliable and I did not used it in data analysis). Auditory stimuli were presented using headphones (Sennheiser HD 380 Pro) connected to an external sound card (RME FireFace UC). All computer tasks were programmed in MATLAB (R2021a, Mathworks, Natick, MA, United States).

The ITCP-B task consisted of 120 trials in total (shortened by half compared to the original ITCP task), with three blocks separated by short self-paced breaks. The whole test typically took 15 minutes to finish. Each trial was up to 2 seconds long with a one-second inter-trial interval. Half of the target words were spoken by the female speaker, while the other half was spoken by the male speaker. The order of the words was randomised between participants, but the same words were always spoken by the same speakers. The outcome measure used here for the ITCP-B was the proportion of words correctly identified.

The sentence-in-babble (SiB) test was similar to that used by Holmes & Griffiths (Holmes & Griffiths, 2019). Target sentences were taken from the English version of the Oldenburg sentences and were recorded by a male speaker with Southern British English. Target sentences were structured as name-verb-number-adjective-noun; an example is “Alan brought four small desks”. The background noise was a 16-talker babble that had an onset 500ms before the target sentence. Participants were asked to repeat all five words from the target sentences: they were presented with a 5*10 matrix on the screen and were asked to select each of the five words from a list of 10 options. The test used a one-down one-up adaptive procedure, with starting SNR at 0 dB and a step size at 2 dB for the first 3 reversals and 0.5 dB afterwards. The testing consisted of two interleaved runs, where each run had a different set of target sentences and terminated after 10 reversals. The median SNR of the last 6 reversals was taken for each run and both were averaged to compute participants' thresholds.

Data Analysis

I conducted data analysis using SPSS Statistics 29.0.1.0 and visualised the results in MATLAB R2021a. The results for both sessions were normally distributed, justifying the use of parametric tests. As the overall test design has been established with the previous validation study (Geller et al., 2021), the current study focused on test-retest reliability.

First, as the two sessions were not perfectly counterbalanced, I checked if there were learning effects or other outside influences that could lead to different performances in the two sessions. I compared the accuracy for each test between the two sessions with paired-sample t-tests.

Test-retest reliability was measured the same way as the ITCP validation (Geller et al., 2021), with the intraclass correlation coefficient (ICC), using a two-way random effects model (absolute agreement). ICC considers both correlation like Pearson correlation and bias when assessing reproducibility (Liu et al., 2016). The absolute agreement measures are used to determine the level of agreement of raters, in this instance the scores of two ITCP-B testing sessions (Koo & Li, 2016).

The relationship between ITCP-B and other speech and hearing measures was measured using Pearson correlations. Two pairs of correlations were assessed: PTA and ITCP-B (two sessions) and the ITCP-B and SiB (convergent validity check, for Session A only as the shorter SiB was not as reliable). A further cross-talker validity test was conducted by comparing the responses to either the male or female speakers. A paired-sample t-test was used to assess if people responded differently to the two voices; the ICC further tests if the test can elicit reliable performance across talkers.

4.1.3 Results

The mean performance accuracy and standard deviations were extremely similar between the two sessions of ITCP-B: Mean (Session A) = 0.68 (SD = 0.08), Mean (Session B) = 0.67 (SD = 0.09). There was no significant difference in the mean performance between sessions: M_{diff} = 0.005 (SD = 0.043), t (45) = 0.866, p = 0. 391. The mean SNR for SiB was -1.07 (SD = 1.44) for Session A.

Figure 4.1 shows the correlation between PTA (averaged across 0.25 kHz to 8 kHz) and ITCP-B. Both sessions had large and significant negative correlations with a similar effect size: r (Session A) = -0.62 ($p<0.001$), r (Session B) = -0.56 ($p<0.001$). Note that the negative correlation is predicted since PTA is scaled such that a lower PTA indicates better hearing, while the ITCP-B is scaled such that higher scores indicate better performance.

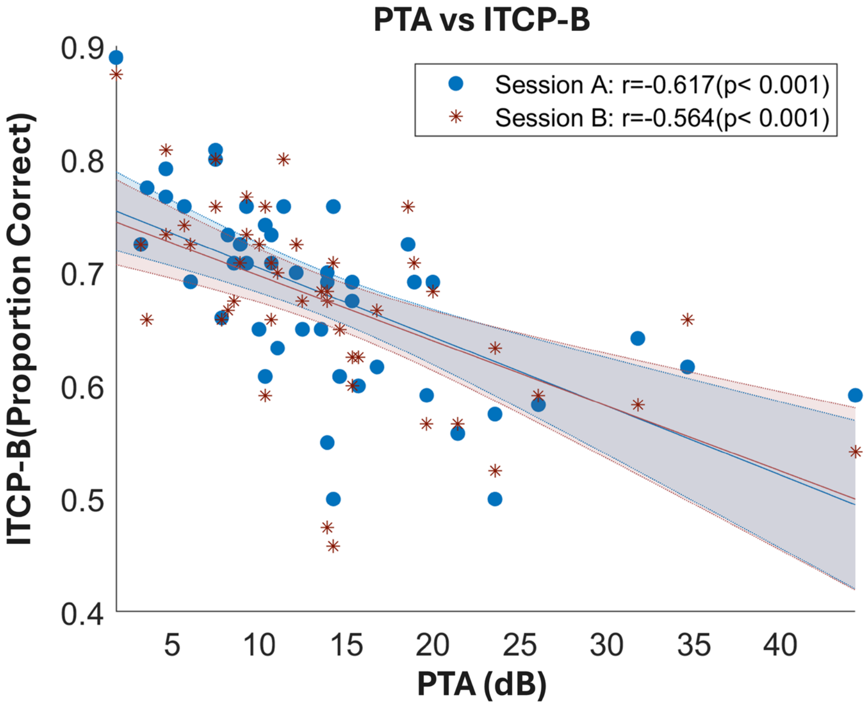

Figure 4.1 Scatterplot of PTA and ITCP-B of the two sessions. The correlation (Pearson) for Session A is in blue and for Session B is in red (the lines of best fit and error areas of the two sessions are in their respective colour as well). PTA results are from Session A. The x-axis plots the PTA results in dB SPL, and the y-axis plots ITCP-B results measured in the proportion of correct answers overall.

Test-Retest Reliability

We next examined the test-retest reliability of ITCP-B by calculating the ICC between the two sessions. The scatterplot (Figure 4.2) displays the close relationship between performance in the two sessions. This is further evidenced by the ICC results (Table 4.1) that showed excellent reliability of $R_{ICC} = 0.93$, which exceeds that of the original ITCP test-retest reliability of $R_{ICC} = 0.80$.

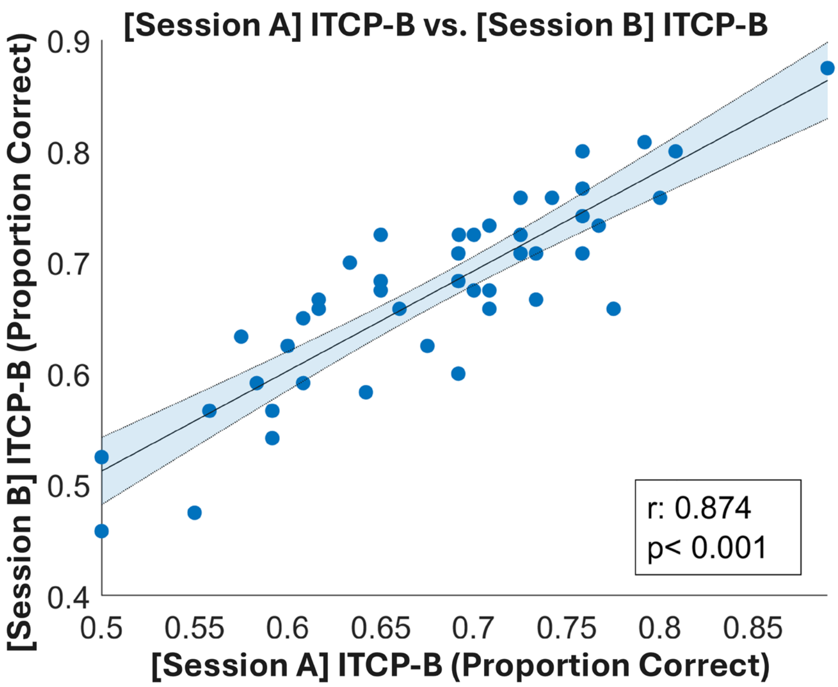

Figure 4.2 The scatterplot shows the association of the performance on ITCP-B in the two sessions. The x-axis represents the scores obtained from Session A and the y-axis represents the scores from Session B. Pearson's r and a p-value for a bivariate correlation are shown on the plot as well. The line of best fit is plotted in black with the error area shaded in blue.

	ICC	CI Lower	CI Upper	P
ITCP-B	0.93	0.88	0.96	$p < 0.001$
ITCP	0.80	0.70	0.86	$p < 0.001$

Table 4.1 A summary of the ICC results from this study (ITCP-B) and the previous validation study (ITCP; Geller et al., 2020) for comparison. ICC is the intraclass correlation coefficient. CI is the confidence interval, and P is the significance level.

Cross-Talker Validity

The cross-talker validity test showed that responses in the two sessions to either the female or the male voice did not differ significantly (M (Female Talker) = 0.68, SD (Female Talker) = 0.07; M (Male Talker) = 0.67, SD (Male Talker) = 0.08; $t(45) = 1.82$, $p = 0.075$). ICC showed a good reliability score as well: $R_{ICC} = 0.79$, $p < 0.001$.

Convergent Validity

The correlation between ITCP-B and SiB was -0.76 ($p < 0.001$), see Figure 4.3 for details. As with the PTA, SiB is scaled as a threshold, so the negative correlation is predicted.

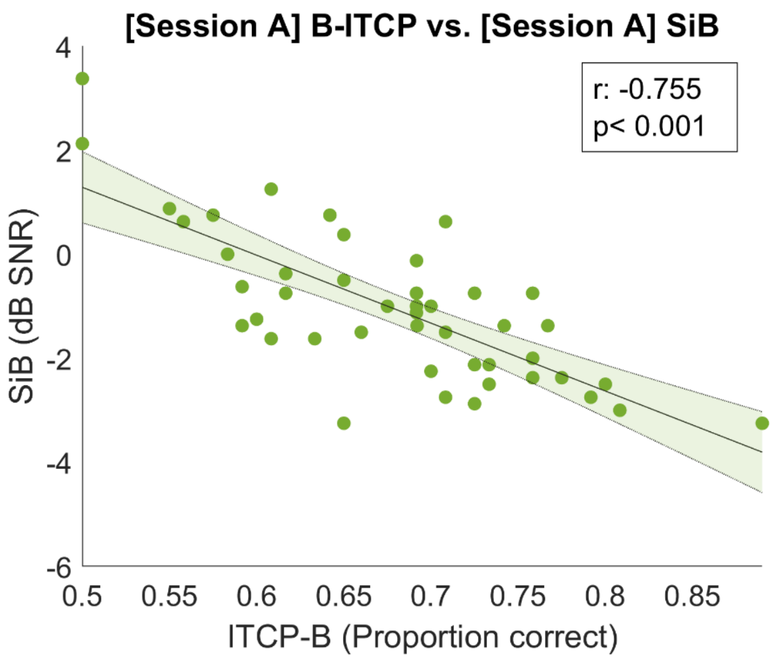

Figure 4.3 Scatterplot for bivariate correlations between ITCP-B and SiB. ITCP-B results are in proportion correct (x-axis) and SiB in dB SNR (y-axis). The line of best fit is plotted in black with the error area shaded in green.

4.1.4 Discussion

The performance data of ITCP-B had a Gaussian distribution and achieved a reasonable level of accuracy (around 68% compared to the 73% reported in Geller et al, 2021). Thus, the ITCP-B meets the minimal criteria for a useful measure. While one of the goals of this study is to establish a test that can elicit comparable results from the UK and the US, the performance accuracy of the current study cannot be directly compared with the ITCP results as the subject cohort and test parameters used here were not tightly matched with the ITCP study (which was validated online and tested all words with four speakers). To develop an equivalent test across the UK-US, further studies are needed which better align the detailed design of the study and the subject populations.

Further, the comparison of the mean accuracy between the two sessions showed no significant difference in the performance of the two sessions. This means that the measure is reliable and stable over time. This is in part due to the unique design features of ITCP in which each of the four items that comprise a response set are used as the target (and they can be used multiple times across talkers). Consequently, subjects cannot learn which item is the correct response for a given set – they must process the stimulus.

We also demonstrated that PTA could predict ITCP-B performance in both testing sessions, which is consistent with our hypothesis and the literature discussed previously (Moore et al., 2020; Besser et al., 2015).

Both the bivariate correlation and the ICC outcome demonstrated excellent test-retest reliability (R_{ICC} = 0.93). This means that the ITCP-B test can obtain a representative and stable assessment of SiN ability over time, which allows for both cross-sectional and longitudinal studies. Again, the ICC score is consistent with the previous results from Iowa (R_{ICC}= 0.80), but higher test-retest reliability was obtained in this study. One potential explanation for the higher ICC score in this study is that the validation for ITCP-B took place in laboratory conditions, but the ITCP validation test was carried out online where audio presentation, background noise and distraction cannot be as well controlled. A comparison of online and lab testing carried out by Bridges and colleagues found that online testing for both visual and auditory modalities tended to generate lower precision and more variability in performance (Bridges et al., 2020). The researchers argued that such a discrepancy in results between the two modalities would not invalidate online auditory research, but it did mean that validating online results was necessary. As the current analysis relies heavily on performance stability, it is expected that a more controlled environment will lead to a higher ICC score. However, the online ITCP still achieved a very good ICC score (R_{ICC} = 0.80), suggesting that the test can be reliably used online as well as in the lab.

The cross-talker validity assessments were carried out to ensure that each talker was representative of the whole. This was important as to obtain a shorter test, half of the stimuli were presented in each voice this contrasts with the original ITCP where the full list of words was heard in both male and female voices). The shortened version is good for time-limited testing in the clinics but raised concern over potentially less balanced results. However, the non-significant t-test showed that the shortened version can provide a reliable assessment of people's SiN ability. Another possibility is that the reduced trial set size in the current study may have been beneficial due to less within-task fatigue.

A further assessment of the validity of ITCP-B against the SiN measure found that ITCP-B correlated strongly with the Oldenburg sentence-in-noise measure. The negative relationship suggested that lower SiN thresholds (better SiN performance) correlated with a higher percentage of correct performance on the ITCP-B task. The strong correlation here suggested that ITCP-B can provide an assessment that is consistent with a well-established sentence measure. This finding is consistent with the ITCP study (Geller et al., 2021), which has established a strong association between ITCP and other standardised SiN tests based on sentences. This consistency in the correlation of ITCP-(B) with other SiN measures in the two validation studies (US and UK) suggests that first, the results are less likely to be due to other non-specific effects such as motivation and arousal. Second, the ITCP-B can give very similar clinical assessment results to patients' real-world listening ability despite that sentence-level SiN measures are thought to be more ecological. The fact that this closed-set word-level SiN test is shorter and engages a 'purer' auditory speech segregation process also adds to the benefit of using the test when sentence-level tests pose a problem.

As highlighted earlier, the development of a comparable speech-in-noise test in the UK and USA would allow for comparisons between two countries with very different criteria for interventions. The ITCP-B and the ITCP potentially represent two tests that can serve this purpose. However, the current experiment only assessed the reliability of the test. To establish age-scaled normative scores, further testing is needed on a wider population, including a wider range of age and hearing sensitivity.

A limitation of the study is that the sample size used is relatively small. Despite having a strong prior, the current sample size is only half of what was used in the original ITCP study (Geller et al., 2021). Further validation studies are needed with a larger sample of normal-hearing adults of all ages to establish normative scores for different age groups.

In conclusion, this study shows that the ITCP-B test has excellent reliability, convergent validity, and cross-talker validity. The shortened version as used in this study provides a good solution for a quick clinical SiN assessment. The full version can be used for research across the UK and US for a more comprehensive test. Both versions are freely available on our OSF page, and researchers can tailor the test based on their preferences.

(This section has been published in 2024: https://doi.org/10.1121/10.0034738.)

4.2 Predicting speech-in-noise ability with static and dynamic auditory figure-ground analysis using structural equation modelling

4.2.1 Introduction

Tracking a target sound in a complex auditory scene is one of the core tasks that the auditory system performs and forms an important part of hearing ability. Complaints about understanding speech in noisy environments are frequently encountered in audiology clinics but are difficult to assess because the pure-tone audiogram does not fully reflect this ability (Merten et al., 2022; Besser et al., 2015; George et al., 2007). Sentences-in-noise and word-in-noise tests have been developed to simulate real-life SIN situations and have been more and more used to assess real-life listening. However, responses to these tests are inevitably influenced by other factors, such as levels of education, accent, and language experience as much as central sound processing (Section 2.3.1). This means the current SIN test resources are difficult to generalise to a wide population, which has motivated this work to develop non-speech measures of SIN based on the figure-ground paradigm. The prototype auditory figure-ground task called the stochastic figure-ground test (SFG) or fixed-frequency auditory figure-ground (AFG-Fixed) was developed by Teki et al. (2013). Modelling suggests that sound segregation is achieved based on the temporal coherence of the figure (Teki et al., 2013). Brain studies implicate a network including the high-level auditory cortex in humans (Teki et al., 2016; Teki & Griffiths, 2016) and in a primate model (Schneider et al., 2018). I also demonstrated in Chapter 3 that the AFG gap discrimination task could predict SIN performance independent of age, PTA, and auditory memory but a combination of these measures could explain 47% of the variance in SIN.

From the findings of Chapter 3, I found that a way to improve AFG was to use a dynamic frequency contour for the auditory figure. What would be the best frequency contour to use? Real-life SIN perception recruits natural frequency changes of speech to better segregate sounds in noise. AFG with changing frequency patterns has been investigated with roving figures following the formants of spoken stimuli (Holmes & Griffiths, 2019). While figures generated from the first three formants of speech did not significantly predict SIN, a stimulus based on the first formant that changed over time coherently did correlate with SIN with a small effect ($r = 0.28$). However, first-formant figure-ground was not a significant predictor of SIN in a multivariate linear regression model including PTA and the classic figure-ground. This suggests that incorporating a dynamic frequency contour into the figure-ground could potentially predict speech perception in noise, but the speech formants might not be the best frequency information to use. Another important frequency contour in speech is the fundamental frequency (F0) that determines pitch perception, which is an important basis for sound segregation (Cheveigné, 2010; Oxenham, 2008a). A more primary role of pitch in sound grouping is suggested by work showing that newborn babies track changes in pitch but not formants as reliably as adults (Arenillas-Alcón et al., 2021). In this study, I aim to assess a type of AFG task in which the frequency components vary over time following the pitch contour of natural speech. This makes the stimulus more speech-like, whilst retaining the overall advantage of the AFG task as a 'pure' measure of grouping relevant to real-life listening without linguistic confounds.

Natural voiced speech contains multiple harmonics related to the fundamental frequency and is associated with pitch. Harmonicity aids hearing in noise (McPherson et al., 2022). Pitch contributes to SIN processing, especially for people with higher language or hearing competence (Llanos et al., 2021; J. Shen & Souza, 2018; Huang et al., 2017). In this study, I generated figures related to the harmonic structure of speech, in contrast to the non-harmonic figures used in the previous studies (Chapter 3, Holmes & Griffiths, 2019). I extracted the fundamental frequency from naturally spoken sentences and developed a new type of dynamic auditory figure-ground stimulus using harmonic complexes. I call this the dynamic figure-ground stimulus (AFG-Dynamic). The harmonic features make the auditory figure-ground more speech-like from an acoustic perspective, without incorporating high-level linguistic cues. To test if the harmonic structure can aid perception, I conducted a pilot study with fixed-frequency harmonic figure-ground and the finding suggested improved figure-ground segregation for the harmonic figure-ground task compared to the nonharmonic one (Appendix I).

Additionally, I created harmonic complexes in different frequency ranges to explore the importance of the frequency range of the figure. Previous studies suggest that high-frequency hearing sensitivity based on the audiogram may be an important determinant of SIN ability (Holmes & Griffiths, 2019; Polspoel et al., 2022; Zadeh et al., 2019) but have not examined complex figures in different frequency ranges. I constrained the frequency range of the AFG-Dynamic stimuli to low-frequency AFG

(AFG-Low) and high-frequency AFG (AFG-High) components to explore how grouping ability in different frequency ranges contributes to SIN perception.

Predictive measures of Speech-in-Noise Perception

The first aim of the study was to investigate if the new dynamic auditory figure-ground tests are predictive of SIN measures. I hypothesise that both versions of AFG-Dynamic tests (AFG-Low and AFG-High) can predict SiN perception and explain an extra variance of SIN independent of the PTA or the prototypical AFG-Fixed. As speech is dynamic in its frequency profile whereas single words have a relatively static frequency pattern, I predict that the fixed-frequency AFG-Fixed can better predict word-level segregation whereas the AFG-Dynamic tests with the changing pitch pattern better predict sentence-level sound segregation.

Modelling the Relationships Among Auditory Figure-Ground Perception, Age, and Speech-in-Noise Perception

The second aim of the study was to describe the relationships among the psychoacoustic measures used in the study and identify the contribution of different factors to SIN perception in a multivariate model using structural equation modelling (SEM). The current study had a complex design investigating different measures of hearing thresholds, auditory figure-ground, and speech-in-noise. This type of design favours the use of SEM compared to regression as it allows having multiple observed variables indicating one latent variable (hypothetical constructs that are not directly measured but can be inferred by their observed variables) and reflects the relative importance of indirect effects, such as the interaction between covariates on outcomes. Three conceptual models based on different outcome variables were therefore constructed with assumptions on the direction of causality according to existing literature. The three outcome variables are: word-in-noise, sentence-in-babble (SiB), and SIN (the two measures combined). As the word- and sentence-level SIN analysis and the self-reported SIN ability tap into different domains of SIN perception, models predicting the three SIN measures separately should provide additional information on the differences between the three domains of SIN analysis when interacting with AFG, PTA and Age. I also investigated the domain-general SIN by combining SiB and WiN.

SSQ was used to assess subjective SIN ability but was not eventually used for analysis as it lacked consistency with the other two SIN tests.

The fixed-frequency AFG test has been shown to predict SIN perception (Chapter 3.2, Holmes & Griffiths, 2019). In this study, I added two additional dynamic AFG measures with high and low frequencies to form an AFG latent variable that predicts SIN perception.

In terms of the exogenous variables, PTA and the participant's age were taken into account. PTA has been shown to predict SIN ability (George et al., 2007; Wong et al., 2008; Besser et al., 2015; Bochner et al., 2015; Holmes & Griffiths, 2019). Age has also been recognised as a key factor impacting both hearing and SIN perception (C. Billings & Madsen, 2018). Deterioration of the auditory periphery – including hair cell and cochlear nerve loss, as well as cochlear synaptopathy (Dias et al., 2024; Liu et al., 2024; Xie et al., 2024) could all lead to decreased real-life listening ability, and these peripheral deteriorations are all tied to ageing (Chadha et al., 2021). Researchers have found a relationship between age and both AFG and SIN performance (Holmes & Griffiths, 2019). Altered auditory peripheral function could result in lowered frequency and temporal resolution, which would inevitably impact the central sound segregation ability measured by the AFG tests. However, the relationships between AFG and SIN perception were retained after accounting for age and PTA, which indicates that figure-ground measures may also index PTA- and age-independent deficits in SIN perception. Thus, I hypothesised that PTA and Age would predict both SIN and AFG with Age impacting PTA, and that AFG can independently predict SIN after accounting for Age and PTA.

4.2.2 Methods

Participants

I recruited a total of 170 participants, of whom 11 were excluded due to data quality as per criteria described later. The final sample size used for analysis was 159. The sample had a wide range of age (mean = 45.24, SD = 18.51, range = 18 – 79) as well as hearing thresholds measured as decibels of hearing level (mean = 13.51 dB HL, SD = 10.05 dB HL, see Figure 4.4 for more detailed audiogram results), with 105 female participants. All participants were neurotypical native English speakers with no history of auditory disorders, no history of speech and language disorders, and who were not currently taking any psychotropic drugs. Informed consent was obtained from participants before the experiments. The study was approved by the Newcastle University Ethics Committee (46225/2023).

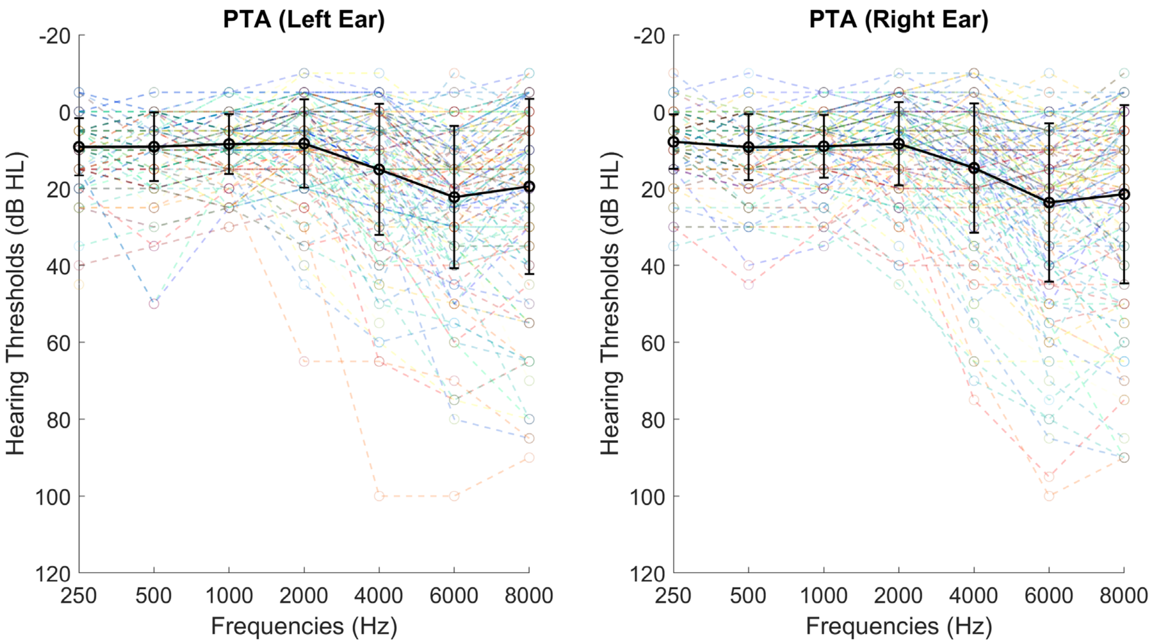

Figure 4.4 The distribution of hearing sensitivity of 250 – 8000 Hz for the left and the right ear separately of all participants. The x-axis shows the frequencies measured, and the y-axis shows hearing thresholds measured in decibels. The coloured lines with circles plot individual audiogram results, and the thicker black line with circles and error bars show the averaged group PTA. The error bars display the standard deviation.

Stimuli and Tasks

Fixed-Frequency Auditory Figure-Ground Gap Discrimination Task

The parameters of the AFG-Fixed gap discrimination task were kept the same as used in Holmes & Griffiths (2019). The AFG-Fixed stimulus consisted of an auditory figure with temporally coherent pure-tone elements (each 50 ms duration) repeating over time. Each figure was 42 chords long with 3 figure components per chord (i.e. coherence level of 3). The figure was superimposed on an auditory ground, which is a tone cloud made of pure-tone elements (also 50 ms duration each) of randomised (or stochastic) frequencies between 180 – 7246 Hz in a logarithmic scale. In each trial, two figure-ground stimuli were presented to the participants, sequentially with an inter-stimulus interval of 400 ms. A gap (6-chords long) was present in either figure. Although, importantly, the ground tones continued through the gap, so participants needed to have segregated the figure from the ground to perform this task. The participants were instructed to choose which of the two figure-ground stimuli contained a gap in the figure. The test used a 1-up 1-down adaptive procedure, starting at signal-to-noise ratio (SNR) of 6 dB and varied systematically across trials. The step size started at 2 dB and went down to 0.5 dB after 3 reversals. Two interleaved runs were presented to each participant with different exemplars, with both runs terminating after 10 reversals. The median of the last 6 reversals for both runs were taken and averaged as a measure of performance. Higher SNR scores indicate worse performance.

Dynamic Auditory Figure-Ground Pattern Discrimination Task

In contrast to the prototype AFG-Fixed which has a fixed-frequency pattern over time, the novel dynamic AFG contains pitch information akin to speech. I extracted the pitch contours from the English Oldenburg sentences read by a male British speaker (Holmes & Griffiths, 2019), using Praat version 6.2.09 with a time step of 0.75/75 Hz (100 pitch values per second), and had a frequency range of 74.94 – 295.44 Hz (M=131.59, SD=15.61). The low-frequency noise (below 10 Hz) and artificial high frequencies (above 300 Hz) introduced by the Praat periodicity analysis were removed to obtain pitch trajectories (see Figure 4.5(a) for an example). There are gaps in natural pitch tracks as shown in Figure 4.5(a). To avoid the participants using these gaps, the natural speech gaps and stops were first removed from the pitch contour. An example of the conjoined signal is shown in Figure 4.5(b). As demonstrated in the plot, the new signal has a general downward trend, and some sharp transitions caused by the removal of the gaps and linear interpolation. To remove the drift from individual signals, I demeaned the signal and applied detrending to the demeaned signal. Low-pass filtering (minimum-order filter with a stopband attenuation of 60 dB) with a 2000 Hz cutoff frequency was then carried out to remove the artificial spikes. The trend and the mean were then added back to the filtered signal to keep the final signal as similar to the original pitch trajectory as possible. An example of the final pitch signal is plotted in red in Figure 4.5(b).

After processing the pitch signals, the resultant frequency profiles were grouped into 50-ms long segments by computing the average to form the figure elements. The F0 contour was multiplied by 2, 3, and 4 to construct the harmonic structure and used as the remaining elements of each chord (see Figure 4.5(c)) for the AFG-Low. The tones were gated with a 10 ms raised-cosine ramp to smooth the onset and offset of the sounds. The high-frequency figure (AFG-High) retained the pitch trajectories of the low-frequency version, but the components were the fundamental frequencies multiplied by 5, 10, 20, and 30. The top frequency of each figure was checked so as not to exceed the masking frequencies. Like the AFG-Fixed stimuli, the auditory ground was composed of randomised pure-tone segments on a logarithmic scale. Although, while ground tones for the AFG-High stimuli used the same range of frequencies as the AFG-Fixed (180 Hz–7246 Hz, scaled logarithmically), AFG-Low stimuli used ground tones with a lower frequency range (90 – 3623 Hz, scaled logarithmically; in other words, half of the upper and lower frequency values from the AFG-High stimuli) to achieve a better masking effect. See Figure 4.5(c) (d) for an illustrated example of the two types of stimuli. The duration of both AFG-Dynamic stimuli varied from 15 to 29 chords (due to differences in sentences' duration) randomised over the trials. Within each trial, the two stimuli were matched in length.

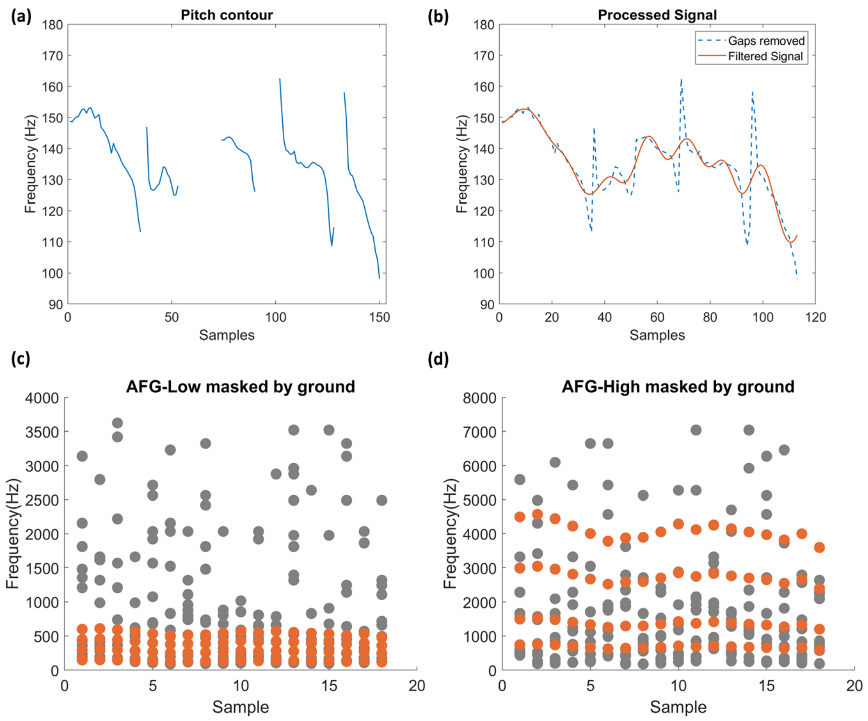

Figure 4.5 The figure shows the extraction of the pitch contour (Figure 4.5(a) (b)) and the AFG stimuli with the pitch contour embedded (Figure 4.5(c) (d)). Figures 4.5(a) and (b) show that the pitch information extracted from the sentence "Alan brought four small desks". The x axis plots the time in seconds and the y axis shows the frequencies in Hz. Figure 4.5(a) shows the raw pitch contour plotted against time. Figure 4.5(b) shows the pitch trajectory after being processed. The blue line is the pitch contour with the gaps removed. The red line shows the final processed signal. The dotted plots illustrate examples of the two different types of AFG-Dynamic stimuli. Figure 4.5(c) shows the lower-frequency dynamic AFG. Figure 4.5(d) on the right side is the high-frequency dynamic AFG. The x-axis shows the time in milliseconds and the y-axis shows the frequency in Hz. Figure elements are depicted in orange while ground elements are depicted in grey.

The two AFG-Dynamic tests (AFG-Low and AFG-High) used the same task design and were counterbalanced across participants. Within each test, both the figure and the ground stimuli were presented per trial, either with the same or a different figure pattern. In the case of a trial with different figure patterns, the durations of the figures were matched but the frequency elements were based on different pitch trajectories. The ground elements were tailored to different figures. The tests used a two-alternative forced-choice task, which required the participants to hold the sounds in memory and decide whether or not the second figure had the same pattern as the first one. The inter-stimulus (within each trial) interval was 0.2 seconds. A two-down one-up staircase procedure was used with a total of 22 reversals. The initial SNR was 12 dB with a step size of 2 dB, which then changed to 0.5 dB after 7 reversals. The trial orders were kept the same across participants. The final score was calculated by taking the median of the dB SNR of the last 6 reversals and a higher SNR would indicate poorer performance. The same design was used for both the low-frequency and the high-frequency versions of the AFG-Dynamic test.

Speech-in-noise measures

Three metrics that reflected the SIN ability were used as the outcome measures, including a word-in-noise test (WiN) (Guo, et al., 2024), a sentence-in-babble test (SiN) (Holmes & Griffiths, 2019), and a subjective self-report measure (The Speech, Spatial and Qualities of Hearing Scale, 'SSQ') (Gatehouse & Noble, 2004).

Word-in-noise test

The WiN test was the ITCP-B described in Section 4.1. Briefly, the target speech sounds were monosyllabic CVC/CVCC words. The babble noise was an 8-talker babble, presented at a -2 dB signal-to-noise ratio (SNR). The onset of the auditory target was 1.0 s before the babble onset. The babble segment of each trial was randomly selected from a 15-second babble stimulus. The length of the words varied from 0.304 – 0.757 s (mean: 0.508 s, SD: 0.086 s). Participants were asked to choose the word they heard out of a list of 4 words displayed on the screen. The proportion of correct responses across trials was taken as the outcome measure for the WiN test. This is the only test that was scored differently as it was not based on an SNR threshold, and a higher score for the WiN test indicates better performance.

Sentence-in-babble test

The SiB test has been described in Section 4.1. The target sentences were English Oldenburg sentences masked by 16-talker babble. The target sentences appeared 500 ms after babble onset and ended 500 ms before babble offset. Participants were shown a 5*10 matrix on the screen, where each word in the sentence had 10 options. The test used a one-down one-up adaptive paradigm with the starting SNR at 0 dB. The total number of reversals was 10 and the step size began at 2 dB and decreased to 0.5 dB after 3 reversals. The task had two interleaved runs. The target sentences were different in each run. The final score was calculated by averaging the dB SNRs of the last 6 reversals across the 2 runs. A lower score on this test indicates better performance.

Speech, Spatial and Qualities of Hearing Scale

The subjective self-report SIN ability was assessed using the Speech, Spatial and Qualities of Hearing Scale-speech hearing (SSQ) (Gatehouse & Noble, 2004). Two of the questions were removed from the shortened speech-hearing questionnaire due to their ambiguity. See Appendix II for the full list of questions used in this questionnaire. Each item has a score from 0 to 10 with the higher score indicating more difficulty in hearing.

Procedure

I carried out an audiometry test first, followed by the 5 computer tasks, which were presented in a fixed order for all participants, except that the order of the AFG-High and the AFG-Low tests were counterbalanced across participants. The tasks were presented in the following order: (1) SiB, (2) AFG-Dynamic test (AFG -High or AFG-Low, determined by counterbalancing across participants), (3) SSQ, (4) WiN, (5) second version of the AFG-Dynamic test (AFG-High or AFG-Low, whichever they had not already completed), (6) AFG-Fixed. Participants were asked to sit in front of a computer monitor (Dell Inc.) used to present the tasks. The auditory stimuli were presented through headphones (Sennheiser HD 380 Pro) linked to a sound card (RME FireFace UC).

Data Analysis

Test of Correlation for AFG-Dynamic

The outcome measures of SiB and the AFG tests were the medians of the last 6 reversals. The performance was considered stable if the performance differences of the last 6 reversals were smaller than ± 5 dB. Participants who did not show stable performance were excluded from the final analysis. Bivariate correlations and hierarchical regressions were carried out to explore the relationship between AFG-Dynamic and SIN tests. Tests of normality (Kolmogorov-Smirnov) showed that AFG-High and WiN were not normally distributed, so Spearman's rho was used to examine the hypotheses regarding the relationships between the three speech measures with AFG-Dynamic (low and high version), AFG-Fixed, PTA and age. Holm-Bonferroni correction was applied to correct for multiple comparisons based on 7*7 pairs of comparison. As linear regression is a more tolerant measure for non-normality, stepwise regression was conducted to check if there were predictive relationships between SIN and AFG as well as specifying the variance explained by individual predictors. These tests were performed using SPSS 29 and visualised with MATLAB R2021a.

Modelling the Inter-Relations of Predictors of SIN

To account for the inter-relationships of the indicator variables, I conducted structural equation modelling (SEM) using the lavaan package (version 0.6-15) in R (version 4.2.1). Maximum likelihood estimation was used with nonnormality correction based on the Satorra-Bentler scaled test statistic. Robust measures were reported in this study (Brosseau-Liard et al., 2012; Brosseau-Liard & Savalei, 2014).

Initial conceptual models (Models 1 and 2) are illustrated in Figure 4.6. Models 1&2 were devised to explore word-level and sentence-level SIN analysis separately. Model 3 illustrates a combined model of all three SIN measures. In all three models, the SIN measures were predicted by AFG indicated by the AFG-Fixed, and two AFG-Dynamic measures. PTA and age also predicted SIN and AFG.

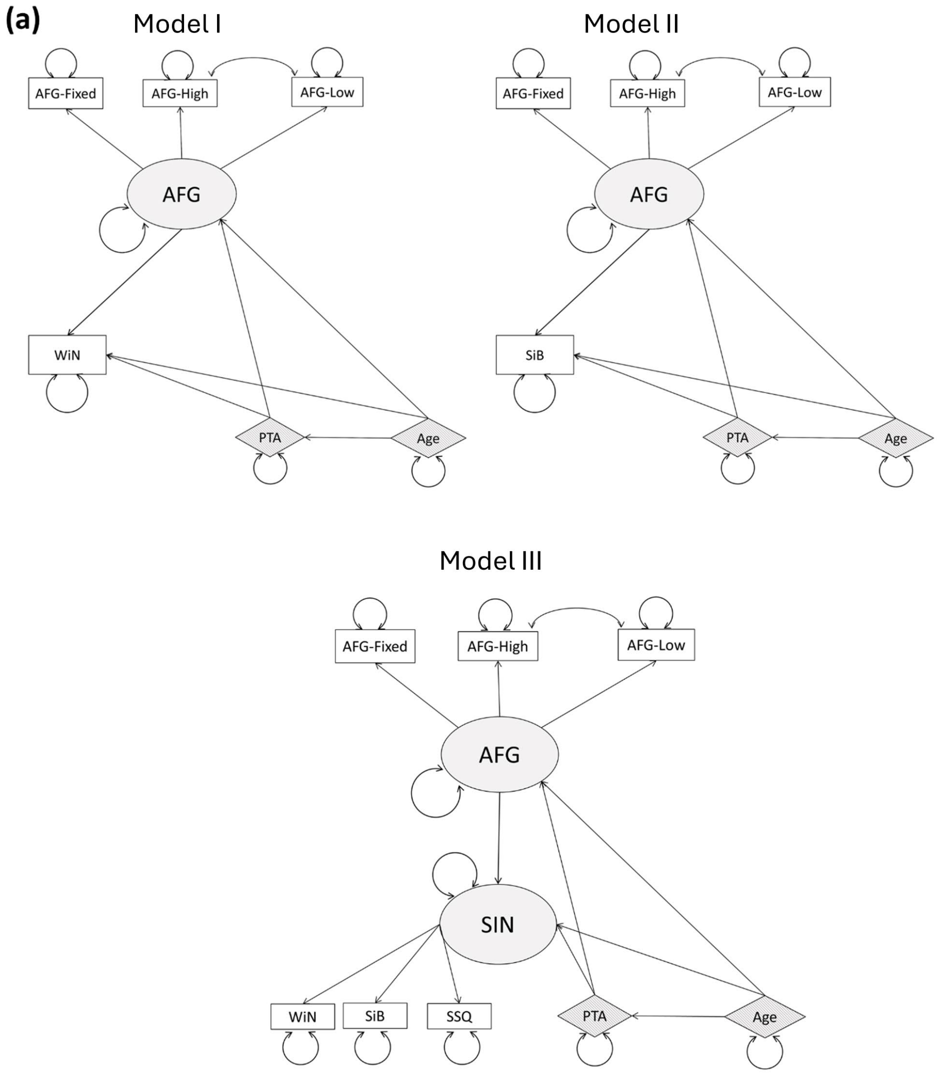

Figure 4.6 shows the conceptual models of WiN (Model I), SiB perception (Model II), and SIN with combined word and sentence perception (Model III). The shaded ovals represent latent variables, the rectangles represent observed variables, and the diamonds with striped shading are exogenous variables. The arrowed circle of each variable represents the error (the size of the circle is not proportional to the radius). The indicators have arrows pointing to them from the latent variables. Exogenous variables point the arrows to the latent variables to suggest a causal effect on the latent variables.

To decide the latent variable structure, I used a confirmatory factor analysis (CFA) to examine the measurement quality with a subset of the data (101 participants) before conducting the final analysis. Figure 4.7 demonstrates the CFA models of the two latent constructs in the three models: SIN and AFG. While there are no rules of thumb defining the acceptable thresholds of a factor loading, SSQ as a measure of functional hearing should predict a large variance of SIN tests similar to the other two SIN indicators. The SSQ however had a visibly weak connection to SIN and thus was removed from further analysis. All three AFG indicators seemed to be acceptable to be entered into the final model. The results of this analysis (Figure 4.7) were used to guide the selection of scaling variables (Bollen et al., 2022). Scaling variables are used to assign scales to latent variables, which is essential when identifying a model. The method used in lavaan is the Fixed Marker (FM) scaling that fixes the loading of the chosen scaling variable to 1 (Lavaan.Org - Model Syntax 2, n.d.). The choice of the scaling indicators can determine the means and variances of the latent variables thus impacting the magnitude of the unstandardised regression path estimates (Klopp & Klößner, 2021) but it is less likely to affect the model fits based on the maximum likelihood estimation (Bollen et al., 2022). The standardised estimates are reported in this study. The path coefficients (abbreviated as β) can be interpreted as: one SD of variable A increase leads to a β SD increase of variable B while all other relevant connections are held constant. The residual or measurement error of the indicators represents variance unexplained by the measure "due to random measurement error, or score unreliability" (Kline, 2015).

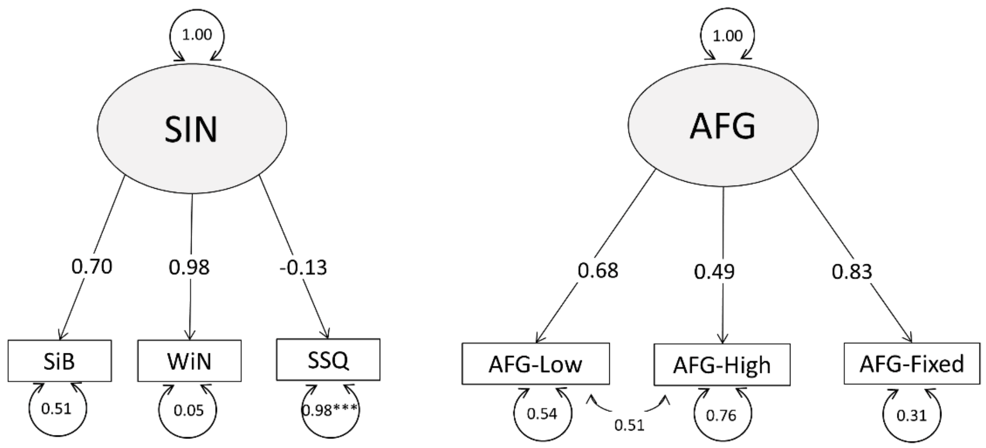

Figure 4.7 CFA with path estimates of SIN and AFG. Shaded ovals represent the latent variables, rectangular boxes are the indicators, and the circles associated with each variable are the residuals. Latent variables are connected to their indicators through arrows pointing to the indicators. The error for SIN and AFG is 1 as they are not subject to any causal influences in this limited model.

The WiN measure was chosen as the scaling variable based on its high path coefficient connecting to the SIN latent variable. WiN was the only test measured by percentage, which resulted in a difference in the scale of the outcome compared to the other tests. This was re-scaled via z-scoring (removing the mean and dividing the results by the standard deviation (SD) of the original scores of WiN). Importantly, contrary to the measures assessed with SNR, a higher score of WiN indicated better performance. This means one SD increase from the mean in WiN would lead to a β SD decrease in SIN. However, since WiN was used as the scaling variable, the SIN latent variable took the scale of WiN, and SiB instead showed a negative path coefficient. The different interpretations of SNR- and percent correct-based scoring would further influence other factors connecting to SIN. To avoid confusion and simplify results interpretation, the WiN results were multiplied by -1 so a higher score would indicate worse performance.

AFG-High, AFG-Low, and AFG-Fixed are the indicators of the latent variable AFG. Similarly, the AFG-Fixed was chosen as the scaling variable due to its close connection with the AFG latent variable. AFG-High and AFG-Low were made with similar parameters except for the frequency range and should tap into very similar mechanisms, hence they covary. The three SEM models further consisted of age and PTA as exogenous variables, which were both configured to predict SIN and AFG.

The model quality was assessed with a number of fit indices as detailed in Section 3.2. The criteria table can also be found below (Table 4.2). Finally, bootstrap analysis was performed by randomly extracting 95% of samples (n = 100 times) to provide a distribution of the estimated RMSEA (Figure 4.9). Confidence intervals (CIs) of the path estimates for all three models were calculated based on the bootstrapped estimates (CI = mean ± margin of error) (Appendix III). The data and SEM analysis scripts are freely available online.

Fit Index	
χ^2 (p)	≥ 0.05
RMSEA	< 0.100
CFI	> 0.90
TLI	> 0.90
SRMR	≤0.08

Table 4.2 Criteria for acceptable model fit.

4.2.3 Results

The descriptive statistics are reported in Table 4.3.

	Mean	Standard Deviation
SiB	-0.880	2.114
WiN	0.673	0.107
AFG-Low	8.991	10.897
AFG-High	7.252	10.447
AFG-Fixed	-14.542	8.200

Table 4.3 The mean and standard deviation of the participant's performance on the five computer tasks.

Relationships between SIN measures and AFG-Dynamic

Both the sentence-level and the word-level SIN tests showed moderate to strong correlations with the dynamic AFG measures (Figure 4.10). After correction, all p-values remained highly significant. SSQ, however, did not show any significant correlation with other speech measures ($p > 0.34$) and was removed from further analysis. The r values and corrected p values of corrections are summarised in Table 4.4.

	WiN	PTA	Age	AFG-High	AFG-Low	AFG-Fixed
SiN	-0.56***	0.57***	0.50***	0.42***	0.47***	0.57***
WiN		-0.67***	-0.73***	-0.39***	-0.47***	-0.61***
PTA			0.72***	0.24**	0.36***	0.59***
Age				0.28**	0.35***	0.55***

Table 4.4 Summary of r values. This table summarises the r values of the correlation test. The p values are reported as asterisks: one asterisk represents $p < 0.05$, two asterisks represent $p < 0.01$, and three asterisks represent $p < 0.001$.

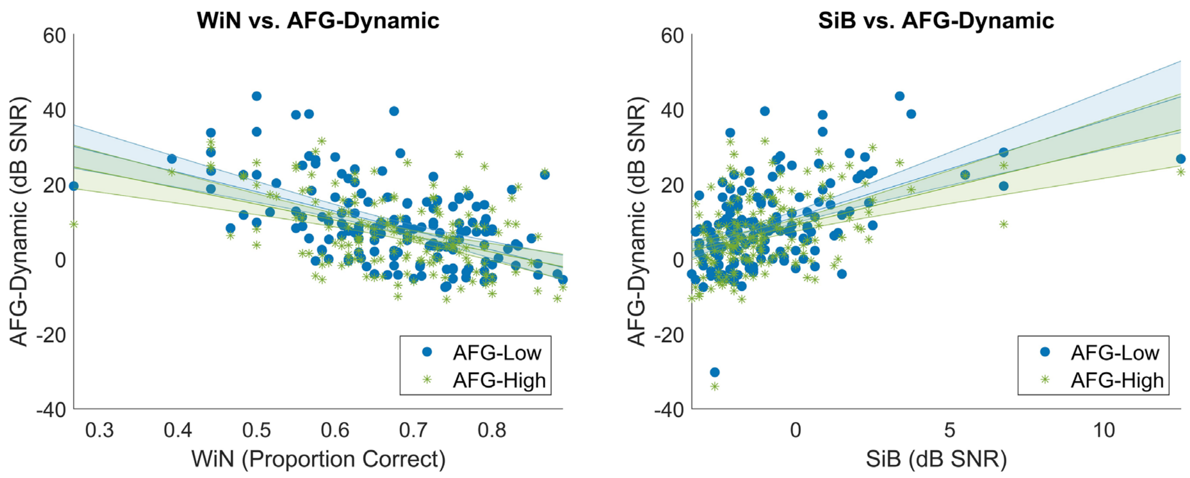

Figure 4.8 Scatterplots of AFG-Dynamic and SIN measures. The lines of best fit are plotted as straight lines in the figure with shaded error bars. The x-axis for the left plot shows the WiN results as proportion correct (number of correct answers divided by the overall number of trials) and the x-axis for the right one shows SiB thresholds measured in dB SNR. The y-axes are the two AFG tasks measured in dB SNR.

The hierarchical regression predicting SiB performance gave three significant predictors, revealing that PTA, AFG-Low, and AFG-Fixed performance significantly predicted SiB performance ($F(3, 155) = 39.879$, $p < .001$). The model accounted for 43.56 % of the variance in SiB. Table 4.5 specifies the variance explained by the significant predictors. For the WiN model, four significant predictors were significant: age, PTA, AFG-Low, and AFG-Fixed ($F(4, 154) = 62.560$, $p < .001$). The model accounted for 61.90% of the variance in WiN. Table 4.5 specifies the variance explained by each predictor. For SiB, PTA was the best predictor explaining about 31% of the model with the AFG-Low adding 9.9% to the model. Whereas, for WiN, age seemed to be the strongest predictor. Of the significant predictors, AFG-Fixed added the least variance to both SiB and WiN, which was about 1%~2% after accounting for the other variables.

SiB	Standardised Coefficients Beta	Adj R^2	p	WiN	Standardised Coefficients Beta	Adj R^2	p
PTA	0.368	0.314	< 0.001	Age	-0.409	0.498	< 0.001
+ AFG-Low	0.253	0.413	< 0.001	+AFG-Low	-0.217	0.580	< 0.001
+ AFG-Fixed	0.197	0.436	0.015	+ PTA	-0.218	0.609	0.003
+AFG-High	0.126	-	0.125	+ AFG-Fixed	-0.134	0.619	0.049
+Age	0.074	-	0.400	+AFG-High	-0.121	-	0.075

Table 4.5 Summary of the regression results. This table displays the adjusted R^2 values and p values of models including an increasing number of predictors that add significant variance to the models predicting either SiB or WiN.

Structural Equation Model of SIN, AFG, Hearing, and Age

The fit indices for the three models are shown in Table 4.6, and path coefficients are plotted in Figure 4.10. The confidence interval of the path estimate of the three models was summarised in Appendix III. All fit indices for Model I and Model II were within our criteria. The path coefficients in Model I were all significant. Model 2 had mostly significant paths with a nonsignificant one of age to SiB. Model III followed the conceptual model structure shown in Figure 4.6 but had the path connecting SSQ to SIN removed as it was not significant. This model met most of the set criteria for an excellent model fit except for the RMSEA. RMSEA incorporates model complexity and models with smaller degrees of freedom tend to obtain a poorer RMSEA (Kenny et al., 2015). This pattern of results is similar to that obtained for Models 1 and 2, which also had excellent fit based on most of the indicators but poorer than expected RMSEA. However, as the combined results of other indicators all showed that the model fits the data very well, I deem that this model is acceptable. All three models were accepted based on the fit criteria. The bootstrapped RMSEA of the three models overlapped over 18% so there was no significant difference between their model fit (Figure 4.9).

Fit Index	Model I	Model II	Model III
χ^2 (p)	9.547 (p=0.067)	7.910 (p=0.122)	20.617 (p=0.009)
RMSEA	0.079	0.065	0.092
CFI	0.990	0.992	0.978
TLI	0.969	0.975	0.949
SRMR	0.028	0.029	0.036
Adj R^2	0.617	0.435	0.889

Table 4.6 Fit indices for Models I, II and III. Adjusted R^2 is also reported in the last row per model.

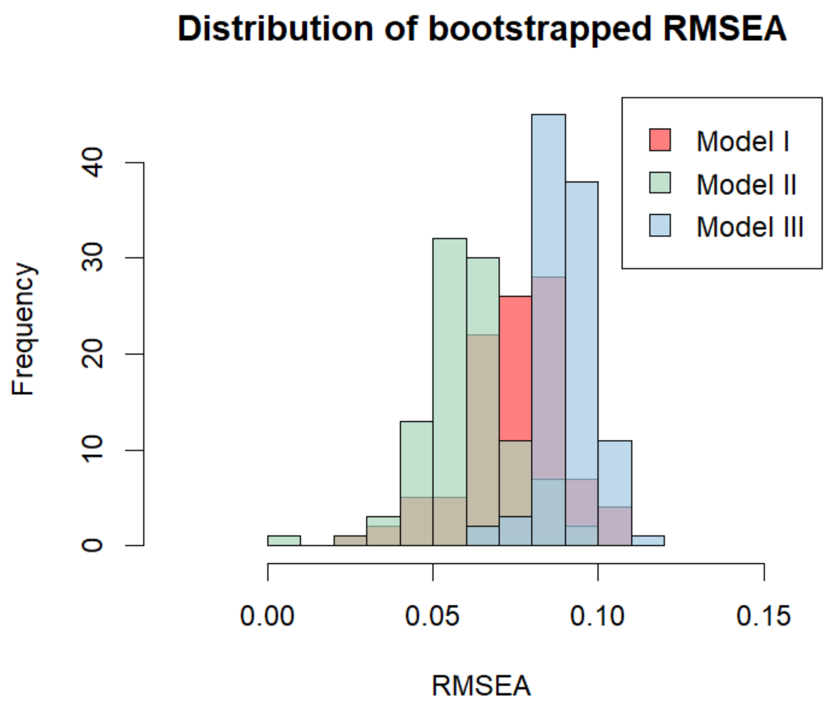

Figure 4.9 Distribution of 100 bootstrapped RMSEA. The x-axis shows the RMSEA values, and the y-axis shows the frequency of the distribution. The three models are represented in different colours as the figure legend specifies.

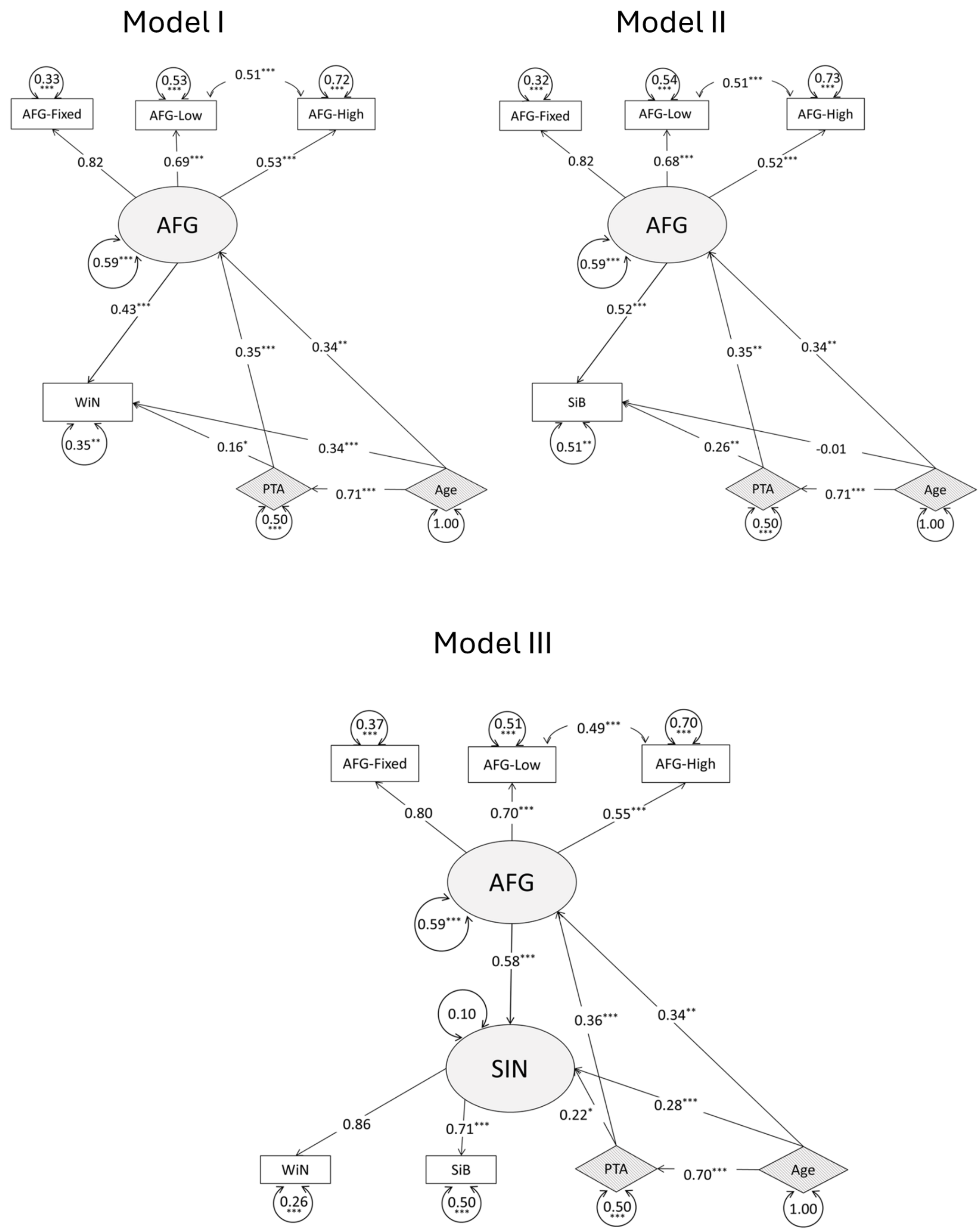

Figure 4.10 SEM models with path estimates. Model I and Model II are presented with either the WiN measure or the SiB measures as the dependent variable, Model III has WiN and SiB combined as the dependent variable. All indicator variables are plotted in rectangles. The oval shape represents the latent variable (AFG) in both models, the exogenous variable is plotted in a diamond shape, and the observed variable not under a latent construct is plotted in a rectangle with rounded edges. The latent variable measured by indicators has arrows pointing towards the indicators. Otherwise, the arrows point from the variable that causes a change in another one. The path coefficients are marked by numbers and error terms are marked by both numbers and a circle around the number. The significance level is marked by asterisks. Three asterisks represent $p < 0.001$, two represent $p < 0.01$, one represents $p < 0.05$. Note that while AFG-Fixed in all three models and WiN in Model III are not marked with asterisks, it is not because they failed to predict the latent variables but because the scaling variables are not estimated in the SEM.

Models I & II reported similar adjusted R^2 as the regression results. As expected, in both models, all three AFG indicators showed significant contributions to the AFG indicator, and the AFG-High and AFG-Low shared significant covariance. AFG-Fixed contributed to AFG with the largest path coefficient ($|\beta| = 0.82$) followed by the two dynamic AFG measures. The latent AFG variable predicted WiN and SiB significantly, with the largest variance compared to PTA and age in both models. PTA had a significant but smaller contribution to each SIN measure. Age only has a significant direct impact on WiN and not on SiB.

Model III explained 86% of the SIN variance (combined word and sentence measure). Similar to Models I and II, AFG explained the largest variance of the latent SIN variable ($\beta = 0.56$) in Model III, compared to age and PTA. Both SiB and WiN showed significant contributions to the latent SIN variable, but WiN had a numerically greater contribution ($|\beta| = 0.86$ for WiN compared to $|\beta| = 0.71$ for SiB). Age was the second largest predictor of SIN, after AFG, and PTA had a smaller (but nevertheless significant) direct impact on SIN. However, both PTA and age had a significant indirect impact on SIN through AFG.

4.2.4 Discussion

Predicting SIN Perception with Dynamic AFG in the Linear Regression Models

This study showed a moderate to strong correlation between all AFG measures and SIN, both on the word and the sentence level. The correlation between AFG-Fixed and SiB reported previously ($r = 0.32$) (Holmes & Griffiths, 2019) was replicated and showed a larger effect ($r = 0.57$). The low-frequency AFG came out as a significant predictor of WiN and SiB, explaining the largest variance in both models after accounting for demographic factors (age or PTA). It is unexpected that even for the WiN model the AFG-Low explained more variance than the static AFG. The dynamic AFG was designed to carry the fundamental frequency patterns and so should better predict sentence-level sound segregation than word-level. AFG-Fixed, on the other hand, had no frequency change over time, which was considered more similar to WiN perception. Based on the regression results, however, it seems that adding the speech pitch pattern to the AFG stimuli only improved its predictive power of SIN in general, not specific to sentence-level perception. This general improvement could be the reason the AFG-Fixed did not explain a higher portion of the variance of SIN as well. Considering that AFG-Low combined both the mechanism of segregating the static figure from the ground employing the figure's temporal coherence feature, and speech-like frequency pattern to aid SIN perception, it is reasonable to see higher variance obtained by AFG-Low in a linear regression model.

One possible explanation for the relationship between AFG-Low and both word and sentence-level SIN is its harmonic structure. AFG-Fixed differed from AFG-Dynamic in two major ways: it is both static and inharmonic. Some of the AFG-Fixed stimuli might contain near-harmonic figures by chance, but most of the stimuli were inharmonic, which elicited weaker pitch perception (Micheyl et al., 2012). Pitch plays an important role in SIN perception (Meha-Bettison et al., 2018; J. Shen & Souza, 2018; Binns & Culling, 2007; Carroll & Zeng, 2007), the mechanism of which was reviewed by Oxenham (2008). This includes not only its strong association with the accent contour of a whole sentence but also other linguistic features such as phonemes in words. The pitch information embedded in the AFG-Low can help with differentiating the envelope fluctuations of the target sound from the background sound, which is key for speech intelligibility. Thus, the stronger pitch strength could be the reason that AFG-Low, while sharing the same basic principles with the static AFG, predicts SiB or WiN better.

The high-frequency dynamic AFG had a numerically weaker correlation as was hypothesised and did not explain additional variance in WiN or SiB after accounting for other tasks. This could be because AFG-High shared a high covariance with AFG-Low due to the similar parameters used for these two tests. The AFG-Low more closely resembles the speech stimuli used in this study with its frequency range being closely configured to the pitch range of speech formants, which might be the reason that AFG-Low outperformed AFG-High in predicting SIN. The design of the two figure-ground conditions differs in their relative frequencies of the target to the ground, making it difficult to compare. The high-frequency condition had a higher ratio of overlaps between the figure and the ground, while the lower frequency condition had a lower ratio of overlaps. As the ground elements were organised logarithmically, the AFG-Low condition was masked with more concentrated ground elements, which makes it better masked compared to AFG-High. This better masking mimics the real-life SIN more as speech segregation relies primarily on the fundamental frequency, not the overtones

(Oxenham, 2008a). This could also explain why AFG-Low predicted SIN better than AFG-High.

The SSQ measure did not correlate with either of the speech measures, which was not a unique finding (Oberfeld & Klöckner-Nowotny, 2016; Ertürk et al., 2023). This could be because the shorter SSQ version does not have enough sensitivity to capture SIN perception as only a few questions were related to speech comprehension in human speech noise. However, as reviewed in Section 2.3.2, past literature has also reported a discrepancy in auditory functions between subjective and objective measures, which is consistent with the current finding.

Predicting SIN Perception in a Multivariate Model

The linear regression models displayed the core contribution of the new dynamic AFG measure as well as the static measure. However, the stepwise procedure did not account for the interaction between variables. The SEM model provided a more comprehensive picture of the experiment that went beyond ranking the important predictors of SIN measures.

Firstly, Models I and II showed that all three AFG predictors have an impact on the SIN performance. This means that when accounting for the interaction and covariance shared between the indicators, all of the AFG predictors should be considered a necessary part of the auditory figure-ground analysis. Interestingly, while the linear measures showed a tighter relationship between AFG-Low and WiN/SiB, AFG-Fixed in the SEM models contributes the most to the AFG latent variable. This suggests that as the 'prototype' AFG, the static AFG that assesses people's ability to pick up the temporally coherent figure from the tone cloud, is still the core of the AFG analysis process. Combined with the regression results, it shows that the dynamic pitch pattern does add an important aspect to AFG and should be used in combination with AFG-Fixed. Based on their individual predictive value of the regression results, in a linear model, when using both measures is not possible, AFG-Low should be a better test to measure SIN ability compared to AFG-Fixed.

The combined AFG measures explained the largest variance (43%, 52%) of both speech measures in Models I & II, compared to age and PTA. This also differs from the linear regression results, where PTA or age was found to be the greatest predictor. This difference suggests that AFG tasks have a greater predictive power of

SIN in combination, whereas each AFG task separately assesses slightly different abilities that are weaker individually than the influence of the demographic factors. The lower path coefficient of PTA compared to AFG indicates that the ability to process speech (either single-word utterances or sentences) in a noisy environment directly relies more on segregating auditory streams and tracking the pattern of the target sounds over time than simply picking up acoustic signals as measured by PTA. However, PTA also had a mediation effect on WiN/SiB through a large path coefficient to AFG. This means that in addition to a relatively small direct impact on SIN perception, deteriorated peripheral hearing could alter functional hearing by modifying central sound processing, which is consistent with our hypothesis.

A mediation effect was also evident with the age-driven impact on SIN perception. Age led to a 71% SD change in PTA in this study, meaning the PTA variance was largely dominated by age-related hearing loss. Age also decreased central sound processing measured by AFG here by 34%, consistent with previous results (Holmes & Griffiths, 2019). However, while Age showed a significant correlation with SiB, it did not modify SiB performance directly in the SEM model, which is consistent with the regression results. WiN is a harder task for people who are older or have higher hearing thresholds. This is because less in the way of compensatory mechanisms can be employed for hearing a short word compared to a sentence that contains a legitimate syntactic structure. While normal ageing can result in deteriorated hearing sensitivity and the perception of other acoustic properties (fine structure or harmonicity), language perception skills are generally preserved (Burke & Mackay, 1997).

The interaction among predictors in Model III did not change much after combining the WiN and SiB into one latent variable. WiN and SiB had a similar level of contribution to the SIN latent factor and the small residual term of SIN suggests that WiN and SiB together provide a holistic measurement of SIN, with a small effect of unmeasured sources of unique variance on the latent variable. It is important to highlight, however, that combining the measures into a latent variable could hide the different effects of other predictors such as age, like in Models I and II.

Limitations and Future Direction

The sample size of the current study might have caused the fit to be suboptimal. There is no golden rule in terms of determining an appropriate sample size for SEM.

Researchers have suggested a variety of standards based on the number of observations (N) per statistical estimates (q) ranging from 20:1 to 5:1 depending on the complexity of the model (Bentler & Chou, 1987; Kline, 2015) or an absolute sample size of 250 if using the Satorra-Bentler scaled method (Hu & Bentler, 1999). The current study has around 8:1 N:q, which is sufficient to find a good solution to meet the convergence criteria, but not optimal. Further studies are needed to validate the model with a larger sample size.

This study also focussed mainly on individuals without symptomatic hearing impairment. The new dynamic measures will need to be tested on different populations such as hearing-impaired or patients with cochlear implants, to see if the results can be replicated with people who struggle with SIN perception. Indeed, recent data suggest that AFG-fixed does predict SIN performance in CI users (Choi et al., 2023), so it is plausible that AFG-dynamic in CI users may explain even further variance. This then can potentially be used for clinical practice to assess patients' dynamic sound segregation. Future research can also incorporate other aspects of SIN perception (e.g., subcortical sound analysis, language ability) and cognitive measures (general intelligence, auditory memory, working memory) to test if the effect of AFG on SIN holds when accounting for these other factors.

Finally, the pattern discrimination task design of the dynamic figure-ground was inherently more challenging than gap detection. While this would ensure figure-tracking and improve its predictive power of SIN perception, it would also involve a higher working memory load, making the task less perceptual. This should be taken into account when choosing which figure-ground task to use. The performance of AFG-Fixed shown in this study might be impacted by fatigue, although this effect should be relatively small. The AFG-Dynamic and AFG-Fixed were always run after the SIN tests: one AFG-Dynamic was run after 25 minutes of SiB testing, and AFG-Fixed after 20 minutes of WiN and AFG-Dynamic testing. This design was to ensure optimal performance on the two speech tasks, but future studies should consider counterbalancing the task orders to minimise fatigue.

In conclusion, these data show that an adequate model of SIN perception needs to account for age, peripheral auditory function, and measures of grouping that I have previously demonstrated to have a brain basis. I introduced new measures of central grouping in this work that incorporate harmonicity and a pitch trajectory taken from natural speech. These measures have improved the prediction of speech in noise in the multivariate model.

(This section has been published in 2025. https://doi.org/10.1098/rspb.2024.2503)

4.2.5 Appendix I: harmonic static figure-ground

Before testing the AFG-Dynamic paradigm, I first examined if harmonicity could be used as a strong grouping cue for the AFG perception. This pilot work was carried out to compare the prototype AFG with AFG with harmonically related components. This section presents the details of its testing methods and results which informed the development of AFG-Dynamic.

The AFG stimuli with harmonically related components (or AFG-Harmonic) were compared with the random-frequency AFG-Fixed. The results showed that AFG-Harmonic can elicit higher detection sensitivity than the prototype AFG-Fixed.

Method

Participants

I tested 12 people aged 20 to 70 of both sexes. Audiometric thresholds were measured for each participant before the main experiment and only people with healthy hearing were included in the study (six frequencies averaged lower than 20dB HL in either ear). In addition to healthy hearing, subjects had no history of auditory disorders (e.g., auditory processing disorders, misophonia, and tinnitus), mental health disorders or traumatic brain injury, and were not taking psychotropic drugs or medication currently. Detailed demographic information and audiograms are shown in Figure 4.13.

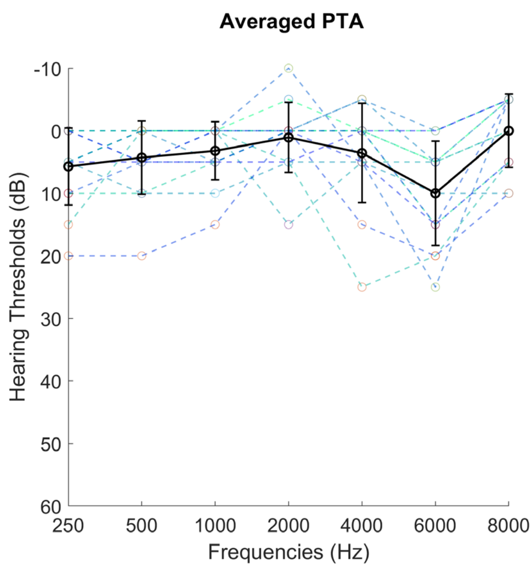

Figure 4.11 Audiometric thresholds at 250-8000 Hz. The thick black line plots the group average with standard deviation bars.

Stimuli

The experimental stimuli were based on the fixed-frequency AFG stimuli developed by Teki et al. and Holmes & Griffiths (Holmes & Griffiths, 2019; Teki et al., 2011). The specific parameters used were the same as the AFG-Fixed condition stated in the previous section. Both AFG-Harmonic and AFG-Fixed (Figure 4.14) were made of auditory figures of 6 chords and a coherence level of 4. The ground was made of randomised spectral elements which overlap in frequency-time space. AFG-Harmonic stimuli were made of frequencies that were positive integer multiples of the fundamental frequency, which took a pseudorandom frequency from a logarithmic scale from 179 Hz to 7246 Hz. To avoid lower frequency bias for the harmonics, the fundamental frequencies were discarded if the fourth harmonics were lower than 800 Hz. AFG-Fixed were constrained not to take absolute harmonic chords to avoid accidental harmonicity. For the gap-detection task, the figures lasted 42 chords, but the ones with a gap contained a 6-chord long silence; the background noise lasted 70 chords. Adaptive procedures were used to detect individual thresholds of 50% for the task-to-mask ratio (TMR). TMR started from 6 dB and increased/decreased by 2 dB each step. The sound level for the stimuli ranged from 72 dB - 72 dB.

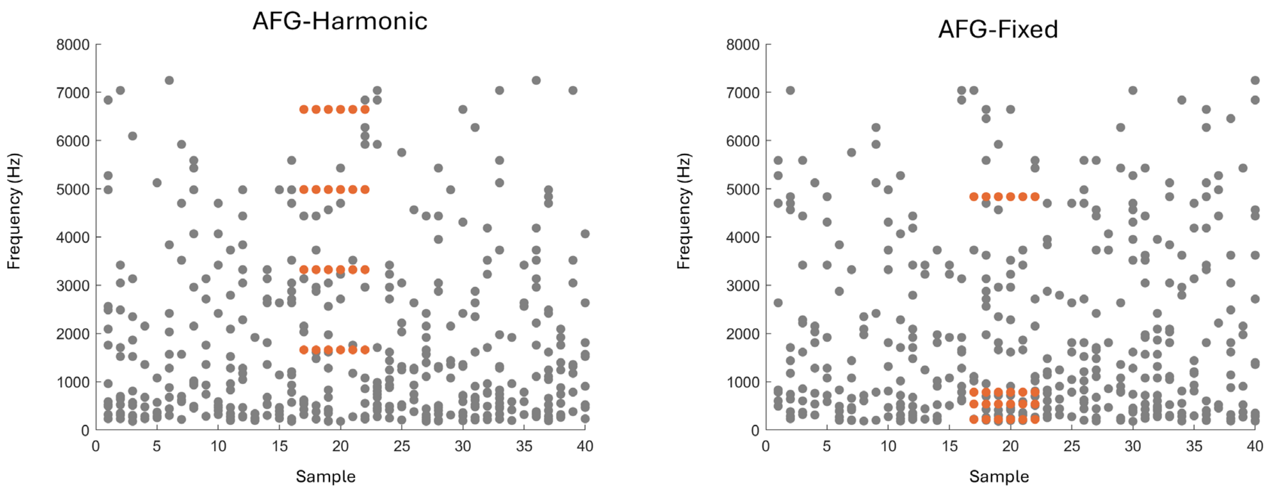

Figure 4.12 AFG stimuli. The figure on the left plots the AFG-Harmonic and the figure on the right is the nonharmonic AFG-Fixed. The x-axes represent the pure-tone elements or samples (50 ms per sample), and the y-axes represent the frequencies in Hz.

Procedures

The experiment was carried out in a soundproof booth. Experimental stimuli were presented using headphones (Sennheiser HD 380 Pro) connected to an external sound card (RME FireFace UC). Participants were asked to sit in front of the LCD display (Dell Inc.) in the booth and respond to the stimuli presented.

Figure-detection tasks with AFG-Harmonic and AFG-Fixed were tested respectively. A trial of the figure-detection task had two sounds, with one containing a figure and one without. A short familiarisation session was given before the main test to acquaint the participants with the AFG sounds. Following the familiarisation session, the main experiment was carried out consisting of a practice trial per condition. Participants were asked to perform a two-alternative forced-choice for both conditions, where they were instructed to choose the sound containing a figure. The order of the condition presentation was randomised.

Data Analysis

The percent correct rate (correct response divided by total trial number) and sensitivity index (d') were used to measure the behavioural results for the figure-detection task. I used a paired-sample t-test to compare the d' and thresholds of AFG-Harmonic and AFG-Fixed.

Results and discussion

The figure-detection task showed statistically significant differences between AFG-Harmonic and AFG-Fixed stimuli for both percent correct and d' (Figure 4.15). Participants performed with a higher rate of correct responses ($Mean_{harmonic}$ = 87.00%, $Mean_{fixed}$ = 78.40%), and higher accuracy ($Mean_{harmonic}$ = 3.659, $Mean_{fixed}$ = 2.789) on harmonic figures compared to non-harmonic stimuli (percent correct: t (9) = 2.713, p = 0.023; d': (t (9) = 2.372, p = 0.042). As predicted, harmonicity was a strong cue for grouping, and configuring the figure components improved the performance in terms of figure-detection accuracy and detection sensitivity.

4.2.6 Appendix II: the SSQ questionnaire

Appendix The SSQ questionnaire

SSQ: Speech

Participant ID:…………………………… Date:……………………………

1. You are talking with one other person and there is a TV on in the same room. Without turning the TV down, can you follow what the person you're talking to says?	Not at all Perfectly 0 1 2 3 4 5 6 7 8 9 10
2. You are talking with one other person in a quiet, carpeted lounge-room. Can you follow what the other person says?	Not at all — Perfectly 0 1 2 3 4 5 6 7 8 9 10
3. You are in a group of about five people, sitting round a table. It is an otherwise quiet place. You can see everyone else in the group. Can you follow the conversation?	Not at all — Perfectly 0 1 2 3 4 5 6 7 8 9 10
4. You are in a group of about five people in a busy restaurant. You can see everyone else in the group. Can you follow the conversation?	Not at all — Perfectly 0 1 2 3 4 5 6 7 8 9 10
5. You are talking with one other person. There is continuous background noise, such as a fan or running water. Can you follow what the person says?	Not at all — Perfectly 0 1 2 3 4 5 6 7 8 9 10
6. You are in a group of about five people in a busy restaurant. You *cannot* see everyone else in the group. Can you follow the conversation?	Not at all — Perfectly 0 1 2 3 4 5 6 7 8 9 10

7. You are talking to someone in a place where there are a lot of echoes, such as a church or railway terminus building. Can you follow what the other person says?	Not at all Perfectly 0 1 2 3 4 5 6 7 8 9 10
8. You are listening to someone talking to you, while at the same time trying to follow the news on TV. Can you follow what both people are saying?	Not at all Perfectly 0 1 2 3 4 5 6 7 8 9 10
9. You are in conversation with one person in a room where there are many other people talking. Can you follow what the person you are talking to is saying?	Not at all Perfectly 0 1 2 3 4 5 6 7 8 9 10
10. You are with a group and the conversation switches from one person to another. Can you easily follow the conversation without missing the start of what each new speaker is saying?	Not at all Perfectly 0 1 2 3 4 5 6 7 8 9 10
11. Can you easily have a conversation on the telephone?	Not at all Perfectly 0 1 2 3 4 5 6 7 8 9 10
12. You are listening to someone on the telephone and someone next to you starts talking. Can you follow what's being said by both speakers?	Not at all Perfectly 0 1 2 3 4 5 6 7 8 9 10

4.2.7 Appendix III: confidence intervals

	Model 1	Model 2		Model 3
AFG =~ AFG-Fixed	[0.815,0.820]	[0.822,0.827]	SIN =~ WiN	[0.860,0.864]
AFG =~ AFG-Low	[0.683,0.688]	[0.679,0.683]	SIN =~ SiN	[0.708,0.712]
AFG =~ AFG-High	[0.528,0.5340]	[0.520,0.525]	AFG =~ AFG-Low	[0.699,0.704]
AFG ~ PTA	[0.348,0.356]	[0.348,0.355]	AFG =~ AFG-High	[0.550,0.556]
AFG ~ Age	[0.338,0.346]	[0.338,0.346]	AFG-Fixed	[0.794,0.798]
PTA ~ Age	[0.703,0.706]	[0.703,0.706]	AFG ~ PTA	[0.348,0.356]
AFG-Low ~~ AFG-High	[0.505,0.511]	[0.512,0.518]	AFG ~ Age	[0.337,0.345]
WiN/SiN ~ AFG	[0.428,0.435]	[0.515,0.523]	PTA ~ Age	[0.703,0.706]
WiN/SiN ~ PTA	[0.156,0.162]	[0.253,0.261]	AFG-Low ~~ AFG-High	[0.484,0.491]
WiN/SiN ~ Age	[0.335,0.341]	[-0.007, -0.001]	SIN ~ AFG	[0.577,0.585]
AFG-Fixed	[0.328,0.336]	[0.316,0.324]	SIN ~ PTA	[0.220,0.227]
AFG-Low	[0.526,0.533]	[0.534,0.539]	SIN ~ Age	[0.279,0.286]
AFG-High	[0.715,0.721]	[0.724,0.729]	WiN	[0.254,0.261]
PTA	[0.501,0.505]	[0.501,0.505]	SiN	[0.493,0.499]

WiN/SiN	[0.342,0.347]	[0.506,0.513]	AFG-Low	[0.504,0.511]
AFG	[0.586,0.592]	[0.587,0.593]	AFG-High	[0.691,0.697]
Age	[1,1]	[1,1]	AFG-Fixed	[0.362,0.370]
			PTA	[0.501,0.505]
			SIN	[0.092,0.100]
			AFG	[0.587,0.593]
			Age	[1,1]

5. Chapter 5: Neural correlates of auditory figure-ground

In this chapter, I move away from psychophysical studies and aim to explore the brain responses of AFG. As reviewed previously, neuroimaging studies have discovered that the fixed-frequency AFG engages high-level mechanisms, some of which are not within traditional auditory areas, including the superior temporal sulcus (STS) bilaterally, the intraparietal sulcus (IPS) and the planum temporale (PT), indicating that auditory grouping does not only involve processes in the early auditory cortices (Teki et al., 2011). Source analysis with EEG also found that object-related negativity (ORN) elicited by SFG was generated in the superior temporal gyrus (STG), IPS, the cingulate gyrus, as well as some frontal regions (Tóth et al., 2016).

While the previous neuroimaging studies have detailed the brain locations responding to the fixed-frequency AFG, they focused on tools with high spatial resolution but low temporal resolution. EEG studies are therefore needed to capture the fast-changing temporal signature of SIN and figure-ground segregations. The neural responses to the new dynamic figure-ground are yet to be researched as well. In this Chapter, I first present an evoked-potential study on the classic AFG with a fixed-frequency pattern, testing the patterns of the elicited EEG amplitude and latency response to both AFG and SIN and propose a testing protocol for clinical use. The second section presents an EEG neural entrainment study exploring the neural tracking of dynamic AFG by mapping the auditory stimuli to the EEG responses using a linear transformation—the temporal response function. This provides an insight into the brain responses to the pitch changes in continuous AFG and SIN sounds as well as the possible generators of the entrainment activities.

5.1 EEG responses to static auditory figure-ground analysis

A previous EEG study on SFG found objective-related negativity and P400 response for figure-ground segregation, which have been associated with segregating two concurrent streams (Tóth et al., 2016). In this study, I further the investigation by testing the neural correlates of distracted SFG vs. SIN listening as well as attended listening using an event-related design. The main aim is to test if the prototype SFG can be used to elicit robust EEG responses compared to SIN. A clinical angle was taken for this study, in which we propose an EEG component as an indicator for sound segregation ability. While behavioural tasks are generally preferred in clinics due to their low cost. However, they require high compliance from the patients, which can be a challenge. Children at a young age, for example, might find it difficult to understand the task and might not give consistent responses. EEG recording allows clinicians to collect brain responses based on a passive listening paradigm, which limits the inaccuracies of human responses. To develop a clinical tool, stable single-subject responses are needed. These individual EEG data were assessed and compared to the group responses. Furthermore, the administration of elaborate testing protocols or expensive neuroimaging techniques is impractical in clinical settings. In order to develop a test for central auditory grouping with simple active tasks and robust and accessible brain recordings in audiology clinics, I assessed the effectiveness of using a single EEG electrode montage referenced to the mastoids similar to that used for brainstem auditory evoked potential (ABR), while carrying out two psychophysical tasks: auditory figure-ground detection and word-in-noise detection. The study demonstrated a vertex response with a delay of greater than 100 ms that can be recorded both in the presence and absence of a relevant task. The results suggested that SFG could provide useful clinical measures of real-world listening ability in patients without having to perform a behavioural task. I also examined ERP responses to a SIN test, from the vertex, which were similar to the SFG evoked responses, but less robust, and not present without an active auditory task. Overall, I propose that EEG responses to auditory figure-ground stimuli could provide a stable measure of real-life listening ability, which could potentially serve as a complementary test to SIN tests.

5.1.1 Materials and methods

Participants

A total of 18 participants (4 male) aged 18 to 53 (mean ± SD: 25.47±10.57) were recruited for the study. Audiometric thresholds were measured and recorded in decibels hearing level (dB HL) for each participant before the main experiment (Figure 5.1). Only people with clinically normal hearing thresholds were included in the study (seven frequencies averaged lower than 20dB HL in either ear). Participants had no history of auditory disorders (e.g., auditory processing disorders, misophonia, or tinnitus), neurological disorders or traumatic brain injuries, and were not taking psychotropic drugs or medication. Experimental procedures were approved by the research ethics committee of Newcastle University and written informed consent was obtained from all participants.

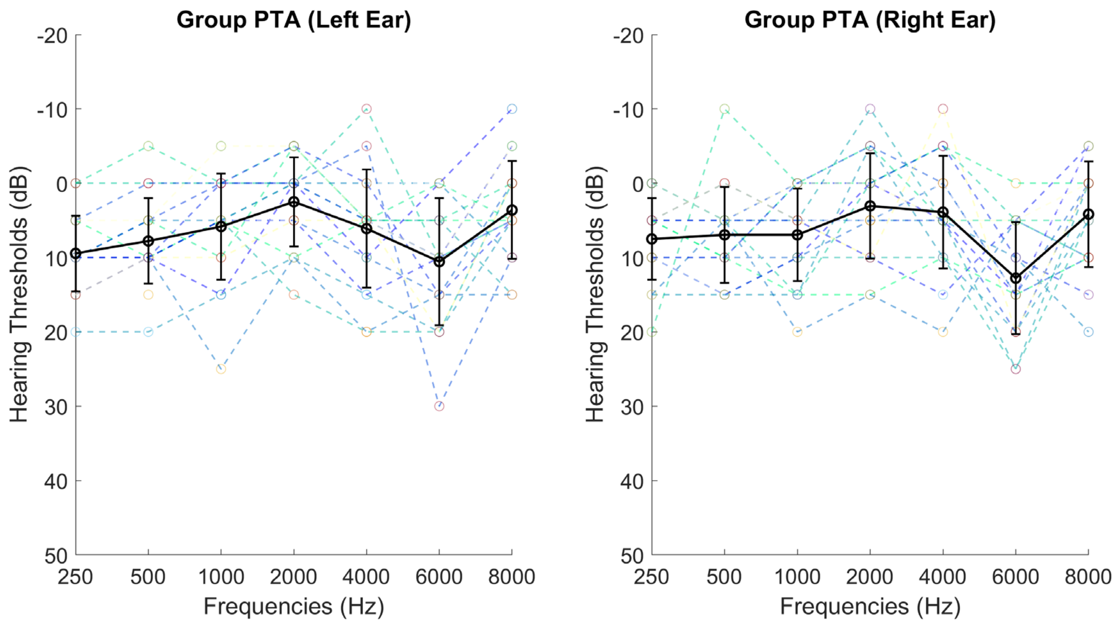

Figure 5.1 Pure-tone audiograms of the participants. The thick black line plots the group average with standard deviation bars.

Stimuli

The auditory stimuli were based on the SIN test used by Holmes & Griffiths (2019) and the SFG stimuli developed by Teki et al. (2011) but were slightly different from what I presented in Chapter 4. Each stimulus comprised a sequence of random chords with 15 pure tone components per chord and a 50 ms duration with 0 ms inter-chord interval. Each stimulus contained two segments; the first segment lasted for 500 ms and was ground-only, while the second segment, also 500 ms long, was divided into two conditions: condition one presented a 10-chords figure (length = 500 ms, coherence=6, 50% of the trials), condition two contained no figure (coherence=0, 50% of the trials). Coherence of 6 has been shown to elicit high detection sensitivity previously so the figure used here is considered highly coherent (Teki et al., 2013). The speech-in-noise stimuli consisted of English names spoken in a British accent and 16-talker babble noise. Similar to the SFG stimulus design, SIN also contained two segments, with the first being only babble noise lasting for about 500 ms and the second with either 50% trials of babble noise or 50% trials of speech (SNR= -3 dB) amidst babble noise. Auditory stimulus onset for both SFG and SIN is defined as 0 ms, and auditory target onset as 500 ms. A distractor visual task was adopted from the Random Dot Kinematograms (RDK) test (Fleming et al., 2018), where white dots were presented on a grey background with a fixation spot at the centre of the screen. The size of the dots was 0.12 degrees (deg) diameter, and they moved at a speed of 5 deg/sec with a density of 30 dots/deg^2. The first segment of RDK was 500ms of random movement. Again, the second segment was divided into two conditions: the first condition had motion coherence of 0.5, creating coherent motion to either the left or right. The coherent condition accounts for 80% of the trials, and the rest of the trials belonged to the random-movement condition, which had motion coherence of 0.

Procedure

The experiment was carried out in a sound-proof booth. Stimuli were presented using headphones (Sennheiser HD 380 Pro) connected to an external sound card (RME FireFace UC). Participants were asked to sit in front of the LCD display (Dell Inc.) in the booth with their eyes about 1 metre away from the screen.

The experiment contained two blocks, first the distractor block and then the active block to reduce participants' learning of the generic properties and structure of the stimuli before doing the active task. During the distractor task, participants were instructed to fixate on the screen and press a key if there was no coherent motion of dots in the RDK task while ignoring the SFG or SIN stimuli during the distractor block. Participants were also shown the visual distractors in the active block, but they were asked to ignore the moving dots and fixate on the fixation point at the centre of the screen and respond when there was no figure or no speech present for the SFG or SIN tasks. The SFG and SIN trials were randomly interleaved, and the inter-trial interval was 1.3 s (1.1-1.5 s, 100 ms steps, uniform distribution). The trial length was 2.3s in total, and there were 200 SFG trials and 200 SIN trials in each block, making 800 trials in total.

Data Acquisition and Analysis

The behavioural response was analysed with a measure of detection sensitivity: d prime (d'). The d' was calculated as the difference between the standardised hit rate and false alarm rate ($d' = z(H) - z(F)$). The extreme values were adjusted by replacing

0 with 0.5/trial number, and 1 with (trial number−0.5)/trial number (Macmillan & Kaplan, 1985). Separate d’ were calculated for SFG and SIN stimuli and for active and distractor tasks. Correlation was performed to check the relationship between PTA and the behavioural as well as neurophysiological measures.

EEG data were acquired using a 128-channel BioSemi system. MATLAB R2021a with EEGLAB version 2019 was used to preprocess the EEG data. Data analysis was carried out with multiple channels as well as with just one channel that can be carried out in clinics (the vertex, A1). For the multiple-channel analysis, the original sampling rate of 2048 Hz was reduced by a factor of 8 to 256 Hz in order to increase the processing speed. The continuous EEG data were filtered from 0.1 - 30 Hz using a highpass Infinite Impulse Response Butterworth filter and then a lowpass band-pass Butterworth filter. The Artifact subspace reconstruction tool was used to detect noisy channels: channels poorly correlated ($r<0.6$) with their random sample consensus reconstruction were rejected and interpolated (8.58 ± 3.67). If over 10% of channels were rejected, the participant was removed from further analysis. This resulted in the rejection of one participant. The data were re-referenced to the common average and epoched from -200 to 1000 ms with a baseline set at 400-500 ms, which is 100 ms before the target stimulus onset. Independent component analysis (ICA) was conducted, and components constituting eye artefacts were rejected via visual inspection. Trial rejection was performed based on probability (>5 SD) and kurtosis (>8). To reduce data loss due to the high montage during trial rejection, temporarily noisy channels were identified and interpolated on a trial-by-trial basis before trial rejection: if a channel exceeded a voltage of 100 mV in a given trial, this channel would be interpolated on that trial only; if more than 3 channels were identified on a given trial, this trial would be rejected from analysis. Event-related potentials (ERPs) were computed across all good trials and across the vertex (A1) and selected neighbouring electrodes (A1, B1, C1, D1, D15, A2, equivalent to a cluster around Cz in a 64-channel system). To calculate the difference at the sensor level in the time domain between the two conditions, Monte Carlo permutation testing was used at the 0-500 ms time window post-target onset (corresponding to the figure/speech stimulus) with 1000 iterations and at 0.025 false alarm rate. Cluster correction (threshold at $p < 0.05$) was also performed to avoid the multiple comparisons problem across time points and channels.

Scalp maps were plotted with cluster-based permutation tests across all electrodes at two time windows (100 - 300 ms and 300 - 500 ms).

For clinical use, after down-sampling and filtering, three channels (A1, D32, B10) were selected for the single-channel analysis. D32 and B10 were used to re-reference the data as substitutes for the mastoids. They are located at a similar position as P9 and P10 in a 64-channel system just behind the ears. Similar to the multi-channel analysis, a probability of 5 and a kurtosis of 8 were used to clean up trials with artefacts. The preprocessed data were then epoched from -200 to 1000 ms with a baseline set at 400-500 ms (henceforth, latencies are defined relative to the auditory target onset), time-locked to the sound onset and ERPs were computed across all good trials at the vertex (channel A1, equivalent to Cz). The amplitude at the vertex over both defined time windows (100 – 300 ms and 300 – 500 ms) was averaged during the active and distractor tasks for the SFG and SIN conditions separately. The amplitude difference between figure and ground, and speech and noise were calculated per participant. A two-way repeated measures Analysis of Variance (ANOVA) was also performed to examine the two within-subject factors, 'Stimulus Type' (SIN vs. SFG) and 'Condition' (active vs. distractor) and their interaction.

5.1.2 Results

The behavioural results showed an average d' of around 2~3 for the two auditory tasks and one visual distractor task (see Table 5.1). Based on the mean statistics, the SFG task elicited a similar detection sensitivity to the SIN task ($t\ (11) = 0.733$, $p=0.473$, Cohen's $d=0.168$). Pure-tone audiograms did not correlate with d' or the EEG amplitudes ($ps>0.50$).

Subject	Active	Active	Distractor	Distractor
1	2.65	3.53	3.71	4.24
2	4.65	4.38	Inf	Inf
3	2.05	2.18	2.74	3.96
4	2.51	2.75	Inf	Inf
5	1.99	2.32	4.20	Inf
6	1.58	1.86	4.12	3.65
7	2.26	2.64	3.17	3.69
8	2.84	2.88	1.00	2.11
9	2.08	2.24	3.12	Inf
10	2.82	3.34	2.32	3.70
11	2.87	3.73	3.60	Inf
12	3.20	3.80	3.17	4.15
13	2.88	3.28	3.09	4.15
14	1.05	1.15	1.69	2.05
15	3.45	2.83	1.43	2.54
16	2.35	2.56	1.56	2.21
17	2.83	2.28	1.51	3.50
18	1.95	2.64	2.24	2.43
Total	2.56	2.82	2.67 (0.98)	3.24 (0.82)

Table 5.1 Detection sensitivity (d') for SFG, SIN and distractor visual tasks. The final row shows the means and standard deviations in brackets.

Multi-channel ERP Topographic Analysis

When inspecting across all channels, central channels showed significantly stronger responses. The scalp maps of figure and ground, speech and noise, and the differences at 100-300 ms and 300-500 ms averaged over time are shown in Figure 5.2. For SFG, the negativity was mostly driven by fronto-central channels, whereas for SIN, the distribution is relatively widespread, and more posterior compared to SFG. A similar topographic distribution of SFG was observed for both conditions at both time windows, but the distractor condition only showed significant differences between the figure and ground at the later time window. The SIN task, however, showed no significant differences between the speech and noise stimuli across channels.

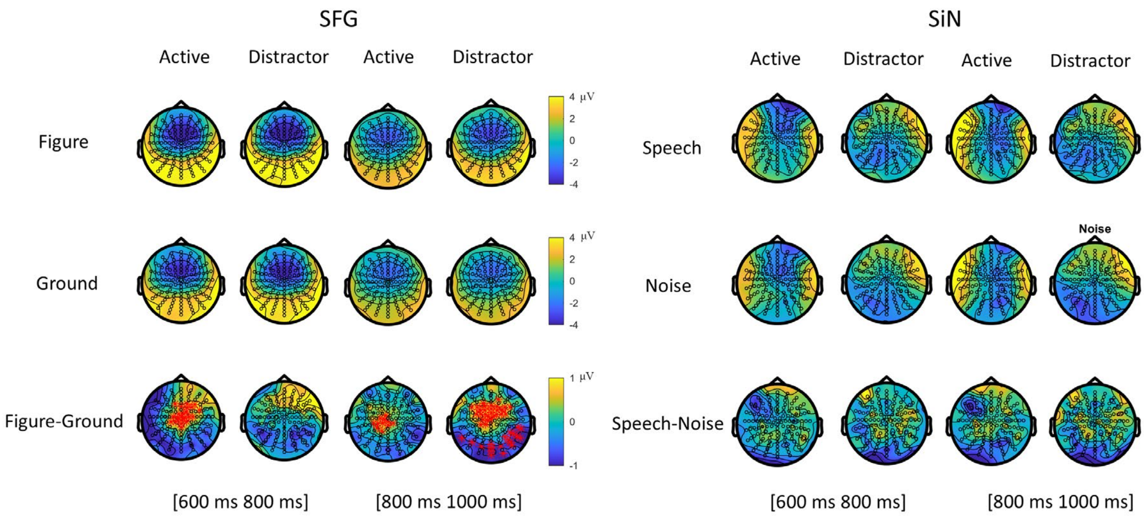

Figure 5.2 Topographic maps of SFG and SIN of the active and distractor condition at 100 - 300 ms and 300 - 500 ms. The bottom panel shows amplitude differences between figure and ground, and speech and noise (calculated as figure minus ground and speech minus noise). Channels that generated significant voltage differences are highlighted in red ($p < 0.05$, cluster-corrected).

Single-Channel Time-Locked Analysis

The ERP group averages for the active and distractor SFG and SIN are illustrated in (Figure 5.3). Through visual inspection, all task conditions showed robust N1 responses to the auditory stimuli. A clear separation elicited by the auditory target from the background was demonstrated post-target onset (i.e., 500 ms) for both SFG and SIN tasks. The auditory targets (figure and speech) elicited greater negativity than the background (ground and noise) alone. Figure tracking started to show significantly enhanced negativity compared to the ground upon the onset of the auditory targets in both active and distractor conditions (approximately 139 ms), peaked around 300 ms after figure onset, and reached statistical significance ($p<0.05$, cluster-corrected) for about 266 ms for both conditions. Such effect was only significant in the figure-ground paradigm, whilst the speech-in-noise paradigm merely elicited a comparable trend. Speech did display significantly less negative amplitude in the active condition at 445 ms post-target onset, which continued to the end of the analysis window ($p<0.05$, cluster-corrected), in the active condition only. This was in the opposite direction to other differences seen, and I interpret this as a rebound overshoot following the initial figure or speech-related negative potential. A similar trend was seen in the active SFG condition.

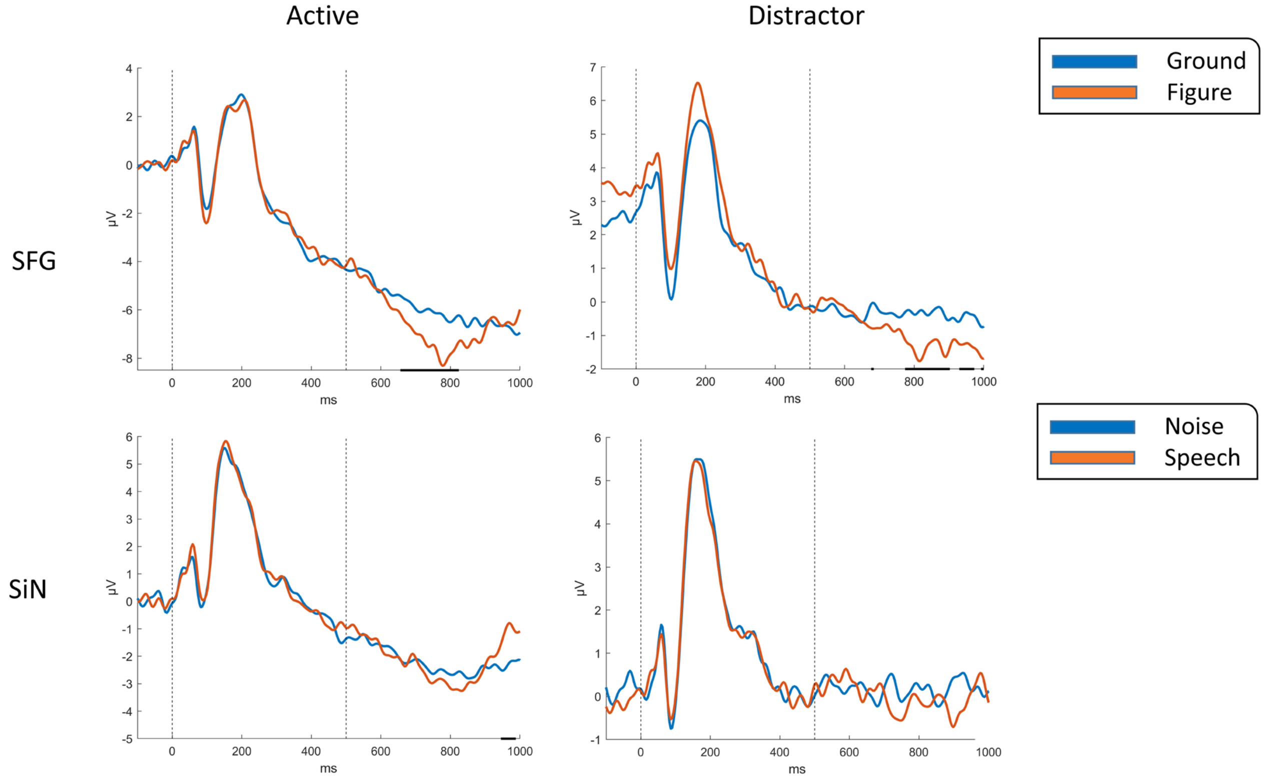

Figure 5.3 Group ERP waveforms at A1 on the active and distractor stochastic figure-ground test and the speech-in-noise test. Dotted lines signal auditory onset (0 ms) and target onset (500 ms). Significance ($p<0.05$) based on non-parametric permutation cluster analysis is highlighted in black above the x-axis.

Individual ERP Analysis

To evaluate the potential for clinical use, where group analysis is not possible, individual data were also examined (Figure 5.4), by taking the average difference between either figure and ground or speech and noise, over the time period 100 to 300 ms post-target onset. On average, participants showed increased negativity when the target sound was present (figure or speech) (mean ± SD; active SFG: -1.09 ± 1.09; distractor SFG: -0.38 ± 1.09; active SIN: -0.27 ± 1.12; distractor SIN: -0.20 ± 0.10). This difference was robustly found across a majority of participants during the active SFG, as can be seen at the top of Figure 5.4, while SIN failed to elicit amplitude differences in over a third of participants. The separation of figure/ground and speech/noise was prominent for most participants. 15 out of 18 participants showed negative values for the amplitude differences of figure and ground in the active condition, 3 weakly showed the opposite pattern, and 3 participants showed very little effect of figure versus ground. The active condition showed a distinctive advantage over the distractor condition regarding the consistency of the activation pattern (15/18 active vs. 10/18 distractor had a negative figure-ground value), but separation was nevertheless evident for most participants (14/18) in the distractor condition. The SIN paradigm showed a similar distribution, but around half of the individual data showed the opposite pattern compared to the group analysis in both conditions. The overall individual data and example waveforms from two selected participants are illustrated in Figure 5.4.

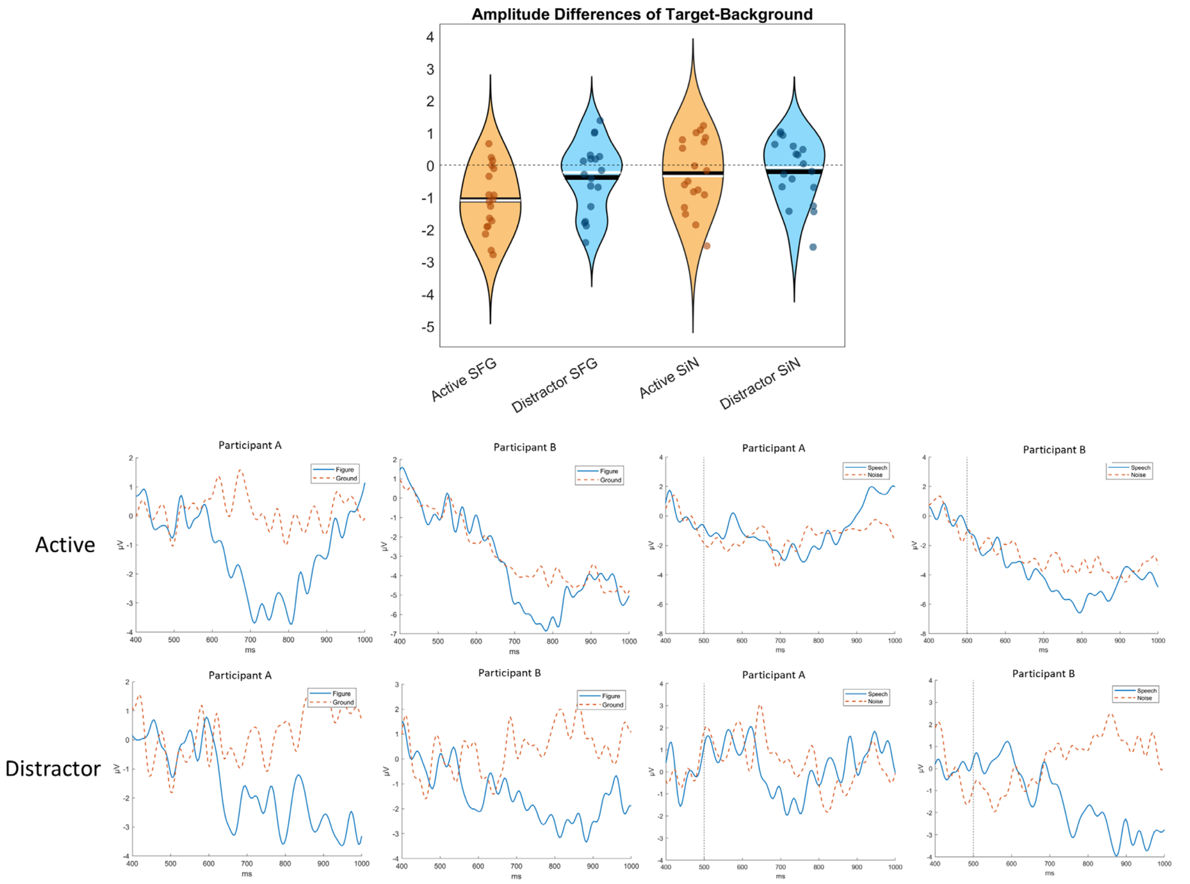

Figure 5.4 Individual data of all 18 participants. Figure 5.4(a) shows the distribution of the voltage differences of SFG (figure-ground) and SIN (speech-noise) over the period of 100 to 300 ms in 18 participants. The mean and the median are highlighted in black and white, respectively. The bottom two rows are example waveforms of two typical participants.

The ANOVA test revealed a significant main effect of ‘Stimulus Type’ ($F(1, 17) = 4.76$, $p=0.04$, $\eta_p^2= 0.22$), which was due to a lower main amplitude difference for SFG than SIN (Table 5.2). The main effect of ‘Condition’ was also significant ($F(1,17) = 9.25$, $p=0.007$, $\eta_p^2=0.35$). The interaction between ‘Stimulus Type’ and ‘Condition’ was not significant ($F(1,17) =1.23$, $p=0.28$, $\eta_p^2=0.07$).

	SIN (M/SD)	SFG(M/SD)
Active	-0.27 (1.12)	-1.09 (1.01)
Distractor	-0.20 (0.10)	-0.38 (1.09)

Table 5.2 Descriptive statistics of the EEG data. These are expressed as speech minus noise and figure minus ground from left to right in active and distractor conditions (top-bottom).

5.1.3 Discussion

The behavioural data demonstrated reliable task performance for all participants in both tasks, with a generally high d' score. This shows that healthy-hearing people could easily detect the auditory target in these tests. When comparing the two active tasks, SFG did not show a significantly higher detection sensitivity (d') than SIN, indicating a comparable SNR level. The visual d's showed higher performance compared to the auditory tasks, which means that the visual distractor paradigm was robust in engaging participants' attention. The audiogram did not show a significant correlation with the outcome measures. This is likely due to the relatively small sample size and the small range of hearing ability from the normal hearing participants.

ERP Responses to Auditory Grouping

The hearing tests demonstrated robust EEG responses of figure and speech with a latency of around less than 200 ms in both active and distractor conditions. The figure evoked greater negativity over the vertex than when it was absent, which was also seen for the speech albeit with a weaker effect. The rapid figure-ground segregation, as well as the slow drift of the SFG responses, were also found in the MEG study (Teki et al., 2011), where the researchers observed short latencies for SFG. These responses are also consistent with the ORN reported by Tóth et al. (2016) in their EEG study. ORN is considered to reflect neural activity that occurs while actively segregating concurrent sounds (Alain et al., 2002). The behavioural data have shown that the visual distractor in this experiment reliably engaged attentional resources, and the brain responses to SIN also exhibited a clear suppression of speech tracking under the distractor paradigm. Conversely, the persistence of figure detection responses under the SFG distractor condition indicates that spectrotemporal grouping could be a pre-attentive process. Similar results were also found in a previous EEG study

(O'Sullivan et al., 2015), where active and passive auditory figure-ground separation demonstrated a similar pattern of neural activation. The SIN test also yielded a pattern of activation that was less consistent on individual analysis than for SFG. The SFG paradigm therefore could potentially provide a more robust neurophysiological measure for central grouping than the SIN test.

The topographic maps of SFG showed distinctive central negativity that is consistent with previous EEG work (Tóth et al., 2016) which localised the brain sources of the spectrotemporal grouping to the superior temporal gyrus and the inferior parietal sulcus, also in line with neuroimaging studies on SFG (Holmes et al., 2021; Teki et al., 2011). Furthermore, a cluster of central channels was revealed to be the major source of activation that powered the figure grouping, which supports the use of a single channel at the vertex for analysis. As the single channel analysis demonstrated very similar waveforms with minor differences in the statistically significant time points, and the recording setup, as well as data analysis procedures, are relatively simple, it is potentially a more optimal measure that could be adapted for clinical use.

The individual data showed that visible figure segregation could be seen in most participants, and a majority of the participants showed a consistent activation pattern with the group-level ERP analysis. This means that the SFG paradigm could be used with EEG recording as a measure for auditory central grouping, and the results could be quantified by extracting a single metric (the average difference between 100-300 ms) from the EEG data and compared to 0. In contrast, the SIN paradigm in the current study did not exhibit reliable neural responses at either the group or individual levels. The ANOVA test showed that SFG also elicited significantly higher negativity compared to SIN suggesting that SFG is a more robust tool for neural responses to auditory grouping.

In conclusion, this study provides proof of principle for a neural measure of figure ground processing suitable for single-subject recordings that might be applied to clinical settings. It could reliably elicit individual behavioural and EEG responses that can easily be obtained in clinical settings with a single channel at the vertex. The visual distractor condition also showed group-level responses, indicating that SFG responses in EEG do not require any specific attention. Further studies are still required to produce a standardised clinical test, and additional steps still required also include studies in older populations, patients with hearing impairment, and performing correlations between SFG behavioural and EEG responses and clinical measures of speech in noise difficulty.

(This section has been published in 2022: https://doi.org/10.1016/j.heares.2022.108524)

5.2 Neural entrainment to pitch changes of auditory targets in noise

5.2.1 Introduction

Segregating and tracking a target sound in complex acoustic environments is an important skill that the auditory system performs to facilitate daily activities. When segregating speech from a noisy environment, humans rely on auditory and cognitive mechanisms to process target speech that stands out due to its acoustic features, even before any language processing. These features include frequency and temporal cues, source location and timbre. Segregation is continuous in the natural environment and engages auditory cognitive mechanisms including perception, working memory and attention (Akeroyd, 2008; Shinn-Cunningham & Best, 2008). In this work I seek to elucidate neural correlates of segregation using stimuli with similar complexity to speech but in the absence of high-level linguistic information. This allows a comparison between pre-linguistic mechanisms for segregation and speech-in-noise (SIN) perception. The work has the potential to suggest a language-independent measure to explain SIN deficits that are not accounted for by peripheral deafness.

The current study depends on further the development of the prototype auditory figure-ground (AFG) task that assesses auditory segregation relevant to SIN perception (Teki, et al., 2011; Teki et al., 2013). Modelling work suggests a figure-tracking mechanism based on the detection of temporal coherence between the component frequencies (Teki et al 2016), which was also evidenced by an electrophysiology work (O'Sullivan et al., 2015) that demonstrated neural tracking of the coherence level of the auditory figure. Brain imaging studies support cortical mechanisms beyond the primary cortex that overlap with those for SIN (Guo et al., 2022; Holmes et al., 2021; Holmes & Griffiths, 2019; O'Sullivan et al., 2015; Schneider et al., 2018; Teki et al., 2016a, 2016b).

While fixed-frequency figure-ground was shown to measure the fundamental sound grouping aspect of SIN processing, natural speech has richer information embedded. One of the most perceptually salient features of natural speech is the pitch, the value of which is determined by the fundamental frequency. Pitch perception plays a crucial role in sound segregation (Dinçer D'Alessandro et al., 2024; Oxenham, 2008b), and training in pitch discrimination improves SIN performance (Gohari et al., 2023; Moossavi et al., 2021). However, pitch contour is highly correlated with other aspects of speech prosody (rhythm and stress contour), making it difficult to isolate the effect of pitch processing in a natural auditory scene containing speech. To address this issue, I have developed a dynamic auditory figure-ground paradigm that simulates the pitch changes in SIN based on a stimulus with isolated pitch changes and no linguistic confounds (Guo et al., 2024). The aim is to measure the behavioural performance and brain substrate for a 'pure' type of pitch tracking in noise as an important precursor to SIN perception.

The dynamic AFG stimulus engages brain mechanisms for tracking a pitch contour derived from speech. Our previous behavioural work showed that dynamic AFG based on the trajectory of fundamental frequency (F0) in human speech (AFG-F0) predicted a large variance of the SIN performance at both the word and sentence level in a multivariate model incorporating hearing sensitivity, age, and both the static and dynamic figure-ground tasks (Section 4.2). However, I do not yet know if the brain parses the AFG information the same way as SIN and if it can reliably track F0 in the AFG stimulus as in natural speech. In this study, I investigate the neural entrainment to both SIN and AFG-dynamic by analysing the EEG temporal response function (TRF) of the frequency profiles embedded in the stimuli. TRF captures more precise characterization of sensory responses to naturalistic speech stimuli than simple correlations between neural and speech signals (Crosse et al., 2016). To further dissect if the entrainment can only be evoked by natural speech pitch patterns or any speech-like frequency contours, I also included a condition with AFG following the 1/F trajectory (AFG-1/F). In addition to EEG sensor-level analysis, the source locations of the TRF peak responses are investigated in the current work to study if the neural generators of the dynamic AFG are similar to that of SIN compared to the previous neuroimaging studies of static AFG and SIN (Holmes et al., 2021; Teki et al., 2016).

This paradigm has potential clinical applications. Currently, available behavioural SIN tests, such as QuickSIN, SCAN-3C, or LiSN-S, rely heavily on verbal responses (Browne et al., 2024; Cameron & Dillon, 2007). These tests therefore exclude people who are not able to give reliable responses (e.g., due to language production deficits or developmental disorders and cognitive impairments). EEG recordings of SIN responses (Panela et al., 2024; Guo et al., 2022; Muncke et al., 2022) allow a measure of brain activity that is not dependent on response. Here, I seek EEG responses to a more fundamental level of auditory processing before linguistic analysis. The work has the potential to isolate 'intermediate' mechanisms for SIN relevant to speakers of any language with any degree of proficiency, between the level of cochlear processing (measured with the audiogram or otoacoustic emissions) and actual speech in noise (measured with speech stimuli).

5.2.2 Methods

Participants

I collected thirty-four participants and excluded one due to poor EEG recording quality. The full inclusion criteria were as follows: participants should be native English speakers with no history of auditory, language, psychological, developmental or neurological disorders, and who were not currently taking any psychotropic drugs. People with mild hearing loss were included as long as they were able to perform all the tests. The final analysis was carried out on 32 participants (13 women) aged from 22 to 67 (mean = 40.19, standard deviation (SD) = 13.68). The pure-tone audiogram (PTA) results are shown in Figure 5.5. This study followed the Helsinki ethical standards and was approved by the Newcastle University Ethics Committee (46225/2023).

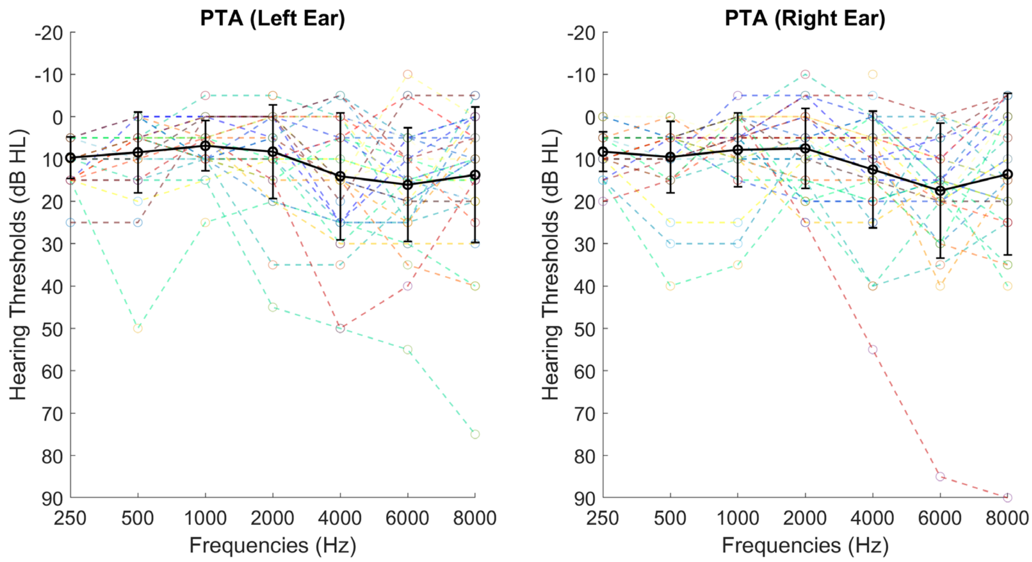

Figure 5.5 Average pure-tone audiogram results of 250 - 8000 Hz of all participants in dashed lines of multiple colours. The thick black line plots the group average with standard deviation bars.

Stimuli and Experimental Design

The AFG-F0 stimuli were adapted from Section 4.2. The auditory target follows the trajectory of the fundamental frequency of speech sentences (Figure 5.6(a)) taken from the SIN task from Holmes & Griffiths (2019) with a frequency range of 74.94 - 295.44 Hz (M=131.59, SD=15.61) using Praat (Boersma, 2001). Any gaps in the frequency contours were removed. The signals were then detrended and lowpass filtered at 3 kHz to remove the sharp transitions that would otherwise be a strong cue for perception (Figure 5.6(a)). The frequency elements were 50 ms each and they were concatenated to form a continuous trajectory. The fundamental trajectory was multiplied by 2, 3, and 4 to form a harmonic structure (see Figure 5.6(b)). These figures were masked by a tone cloud of 10 elements per time point with pseudo-randomly generated frequency elements in the ground from around 90 Hz to around 3623 Hz following a logarithmic scale. The ground stimuli were constrained to have no overlapping frequency elements with the figure. The figure and the ground have the same onset and offset and were played at the same sound intensity level across participants (target-to-masker ratio, or TMR at 0 dB). This ensured that the segregation of the figure from the ground relied strictly on the temporal coherence of the figure as defined by Teki et al. (2016). Each segment of the figure-ground was then concatenated sequentially to make a longer continuous sound.

The figure and ground elements of AFG-1/F stimuli were generated in the same way as the AFG-F0 condition but following artificial pitch trajectories. Briefly, the contour of the 1/F conditions was generated in the frequency domain using a 1/F power spectrum and random phase spectrum. Inverse Fourier transform was performed to obtain the 1/f noise trajectory in the time domain. The frequency series was then normalised and scaled to 74 - 295 Hz to be close to the human speech range. Figure 5.6(c) (d) illustrates an example of the 1/f contour and the AFG-1/F stimulus.

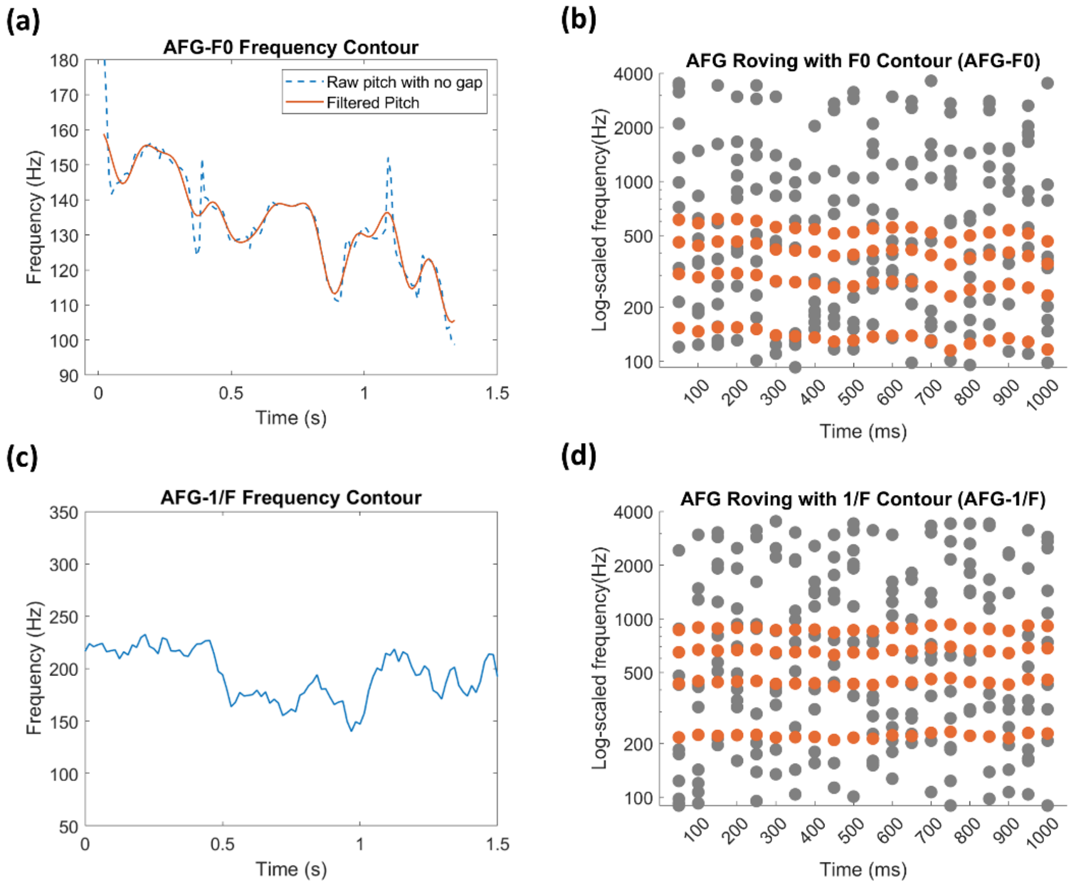

Figure 5.6 The frequency contours of AFG-F0 (Figure 5.6(a)) and 1/F (Figure 5.6(c)) and the figure-ground dotted plots (Figure 5.6(c)(d)). The x-axes of Figure 5.6(a)(c) are time durations in seconds. The y-axes are the frequencies in Hz. Figure 5.6(a) shows the raw pitch (the same as in actual speech) in dotted lines and the filtered pitch contour in red line. Figure 5.6(b)(d) shows examples of AFG-F0 and AFG-1/F respectively. The red dots plot the figure elements, and the grey dots plot the ground elements. The x-axes in Figure 5.6(b)(d) are time durations in milliseconds. The y-axes represent frequencies in Hz.

The trial structure is illustrated in Figure 5.7. The AFG tests consisted of two identical runs presented sequentially with 2 blocks in each trial separated by a self-paced break. The participants were also given a self-paced break between the two trials. Within each block, there were gaps randomly placed in the continuous figure, whilst the ground stimuli continued uninterrupted. These gaps lasted for 600 ms and were randomly placed throughout the testing with each trial containing 30 gaps. Participants were asked to press a button when they could detect a gap. They would need to be able to segregate the figure from the ground continuously during the experiment in order to perform the task, as there was no gap in the ground. This active task was designed to keep the participants' attention level high throughout the recording to maximise the EEG responses.

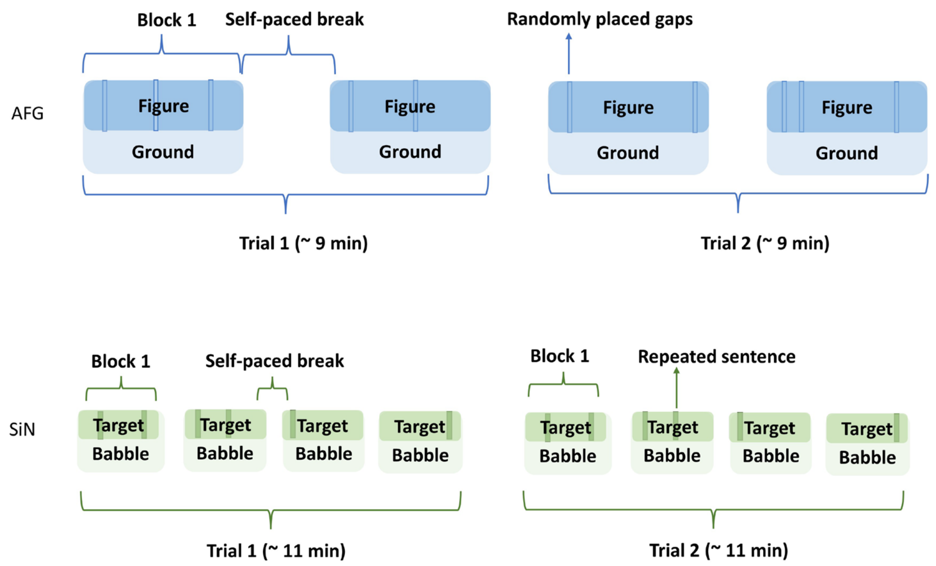

Figure 5.7 Schematics of experiment design. The top plots show the trial structure of the two AFG conditions. The darker blue rectangles are the figures, and the lighter blue are the grounds. The bottom plots show the trial structures for the SIN condition with the darker green as the target sentences and the lighter green the babble noises.

The SIN stimuli were English versions of the Oldenburg (See Chapter 4 for details). The target sentences had five words, which were masked by a 16-talker babble. The signal-to-noise ratio used here was 0 dB. The sentences had the same onset and offset as the background babble to make SIN segments, which were joined together similarly to the AFG stimuli to form a continuous sound. This was done to simulate a naturalistic conversation flow without giving too many semantic and pragmatic cues to the participants.

The SIN condition contained two identical runs, each containing 4 blocks separated by 3 self-paced breaks. There was also an active task for the SIN condition, in which 30 repeated sentences were randomly placed as the response trials. Participants were asked to press the button when they could detect a repeated sentence.

All stimuli were generated off-line with MATLAB R2021a and presented with Psychtoolbox version 3.0.19 through headphones (Sennheiser HD 380 Pro) linked to a sound card (RME FireFace UC).

Procedure

After giving informed consent, participants were taken to a sound-proof booth for an audiometric test. They were then prepared for the EEG recording session. I briefly explained the tasks and specifically asked the participants to pay attention to the target sound throughout the recording. The task instructions were shown on an LCD display. A fixation cross was displayed at the centre of the screen during all three tasks, and participants were told to fixate their gaze on the cross. Feedback was provided whenever participants pressed a button (both for false alarm and correct detection). The EEG session had three tasks following the same order of presentation across participants: the AFG-F0 gap detection task, the SIN repetition detection task, and the AFG-1/F gap detection task. Each task took around 18-25 minutes depending on the duration of the self-paced breaks. Before each experiment, participants were given some example sounds to familiarise themselves with the test stimuli as well as some practice trials that were different from the main experiment. The practice was repeated if the participants failed to do the task until they showed good performance. Participants were also given longer breaks between tasks with refreshments to minimise fatigue.

Data Analysis

Psychophysics

I used D prime (d’) to quantify the performance of the behavioural tasks, which was calculated as: d’ = z(hit) – z (false alarm). The extreme values (0% hit rate or false alarm rate) were replaced by 0.5/n or rates of 100% with (n−0.5)/n where n is the number of signal or noise trials (Macmillan & Kaplan, 1985).

Extracting and Processing the Pitch Information

The input pitch information was processed differently in different conditions. For the two AFG conditions, the frequency information was retained in full, including the gaps. The number of gaps was relatively low and should not have introduced significant distortion to the results. The 30 gaps were filled by the frequency value before the gap onset. I then took the absolute values of the first derivatives of the F0 or 1/F contours to quantify the absolute pitch change of the auditory stimuli (not taking the absolute value would make the assumption that neural responses to pitch decreases were equal and opposite to pitch increases). These were then resampled and aligned to the EEG data.

For the SIN condition, the extraction of F0 contours followed the same method used to create the AFG-F0 figure. As the speech condition also contains the stress contour (specifically the amplitude envelope), which can confound the EEG responses to pitch, it was regressed out from the raw pitch. I took the residual of the linear regression of F0 on the absolute value of the stress contour extracted through the Hilbert transform. This was performed on the target speech only. Finally, the absolute values of the first derivatives of the processed pitch values of all three conditions were taken as the final stimuli aligning to the EEG signals.

EEG Preprocessing

EEG data acquisition was carried out with a 64-channel BioSemi ActiveTwo system. Data analysis was conducted using MATLAB R2021a with EEGLAB version 2019. The continuous EEG data were first referenced to the mastoids (P9 and P10). They were then highpass filtered at 0.1 Hz with a 3rd order Butterworth filter and lowpass filtered at 30 Hz with the same filter. Following the filtering, intervals where participants took long breaks were removed from the data. The remaining data were downsampled to 100 Hz. The Artifact subspace reconstruction tool was used to detect noisy channels: channels poorly correlated ($r<0.6$) with their random sample consensus reconstruction were rejected and interpolated. Independent component analysis (ICA) was applied to remove artifacts such as eye blinks, muscle activity, and heart rate. Up to 16 components were excluded based on visual inspection and classification using the EEGLAB IClabel extension (Pion-Tonachini et al., 2019). After ICA component rejection, the EEG data were re-referenced to a common average reference, following which the data were epoched into 4 epochs for the AFG conditions and 8 epochs for the SIN condition corresponding to their respective experimental blocks. The Cz channel was chosen for analysis based on a previous study, where the researchers found that vertex-to-mastoid analysis could show reliable responses to figure-ground segregation (Guo et al., 2022).

Finally, EEG data were filtered to delta (1-4 Hz) and theta (4-8 Hz) frequency bands using a Butterworth bandpass filter (3rd order). These two frequency bands have been shown to be relevant for the tracking of low (prosody) and higher-level (syntactic) features of speech (Mai & Wang, 2023; Etard & Reichenbach, 2019; Behroozmand et al., 2015; Giraud & Poeppel, 2012).

Computing Temporal Response Forward Model

The temporal response function (TRF, Kegler et al., 2022; Lalor et al., 2006) was used to analyse the relationship between pitch and the EEG responses using the mTRF-Toolbox (Crosse et al., 2021, 2016) and custom scripts developed based on Kegler et al. (2022). A TRF performs a linear transformation between one or more stimulus features and the corresponding EEG responses. This relationship can be mathematically represented as:

$$\mathrm{r(t,n)} = \sum_{\tau=-\tau_{\min}}^{\tau=\tau_{\max}} \mathrm{w}(\tau,\mathrm{n})\mathrm{s}(\mathrm{t}-\tau) + \varepsilon(\mathrm{t},\mathrm{n})$$

where the r(t,n) is the instantaneous EEG response to the stimulus at time t and channel n. Here, s(t) is the absolute of the first derivative of the fundamental frequency |F0’|, and $\varepsilon(\mathrm{t})$ is the residual. The relationship between the response and stimulus is described at a certain range of time lags τ by the TRF weight $\mathrm{w}(\tau)$. The TRF weight, w(τ), can then be estimated by minimising the error between the recorded EEG responses and the predicted responses as below:

$$\mathrm{w} = (\mathrm{S^T S} + \lambda \mathrm{I})^{-1}\mathrm{S^T R}$$

where S is the lagged time series of the stimulus. The λ is the ridge parameter which is defined as $\lambda_\mathrm{n}\, \mathrm{e_m}$, where λ_n is a normalized regularization parameter and e_m is the mean eigenvalue of the covariance matrix (Kegler et al., 2022). I used a fixed normalized regularization parameter of λ_n = 0.1 for all participants. The time lag used to compute the relationship was -200 ms to 500 ms. This range was chosen based on a literature review (results shown in Table 5.5), where the peak latencies reached over 400 ms. The forward model was computed for all participants. The TRF weights were averaged across participants for the group analysis.

Statistical analysis

Group-level statistical significance of the results was assessed with non-parametric permutation testing (1,000 permutations). For each permutation, the stimulus time series was time-shifted by a random value with respect to the EEG data, to abolish any meaningful time relationship between stimulus and response data, whilst preserving all other data features. Any surplus stimulus data beyond one end of the EEG data (start or end) was moved to the other end. These misaligned stimuli were used to compute the TRF forward model for the permutation. Individual TRF weights of channel Cz were averaged across participants. The channel was chosen based on the previous section on single-channel analysis using AFG and SIN stimuli (Section 5.1). Null distributions for TRF weights were created for individual datasets by taking the maximum absolute value across the TRF time series in each permutation (with the 50^{th}-largest value of 1,000 permutations constituting the threshold for detecting significance at an alpha level of 0.05). The peaks of the TRF waveforms of the two AFG conditions as well as their performance were used to correlate with SIN performance using a bivariate Pearson correlation method.

After obtaining a model estimate for all datasets, the quality of the models was assessed by computing the EEG reconstruction accuracy, which is Pearson's correlation between the predicted EEG output of the model and the real EEG data (Crosse et al., 2021). The reconstruction accuracy reflects how well the TRF models capture the encoding of the stimulus. This was compared to the permuted distribution (obtained the same way as described for TRF null distribution) with a pairwise t-test. A two-way (2x2) Analysis of Variance (ANOVA) was also performed on the reconstruction accuracy across two factors: stimulus type (SIN vs. AFG-F0 vs. AFG-1/F) and frequency bands (delta vs. theta). The correlation between the TRF peak amplitude of the AFG condition and SIN d’ was checked with Pearson correlation.

TRF Source Localisation

Previous neuroimaging studies (Holmes et al., 2021; Teki et al., 2016; Teki, et al., 2011) demonstrated high-level brain activities outside the primary auditory cortex at the superior temporal sulcus and intraparietal sulcus for effects of duration and coherence of the figure. In the current study, tracking the frequency patterns during sound segregation could involve potentially different processing mechanisms distinct from pure figure detection in noise. Source localisation was used to explore if the locations driving the surface TRF activities were consistent with previous findings and if they were comparable to speech processing in noise. This analysis was carried out using standardised low-resolution brain electromagnetic tomography using the MNI-152 template (sLORETA, version 20081104) (Pascual-Marqui, 2002). The sLORETA provides a solution (5 mm spatial resolution of 6239 voxels) to the inverse problem at the cortical and hippocampal regions. The significant TRF peaks averaged across participants were transformed into MNI space and tested against the null distribution at the lags of the first and second peaks (Table 5.3). A one-sample t-test was used to compute p-values at each voxel, and the results were corrected with Bonferroni correction at the 0.05 alpha level.

5.2.3 Results

Performance on the active tasks

The d’ results are displayed in Figure 5.8. All three conditions achieved a good level of detection sensitivity (AFG-F0: mean (M) = 2.099, standard deviation (SD) = 0.979; AFG-1/F: M = 2.135, SD = 0.941; SIN: M = 1.971, SD = 0.441). No significant mean difference was found between conditions.

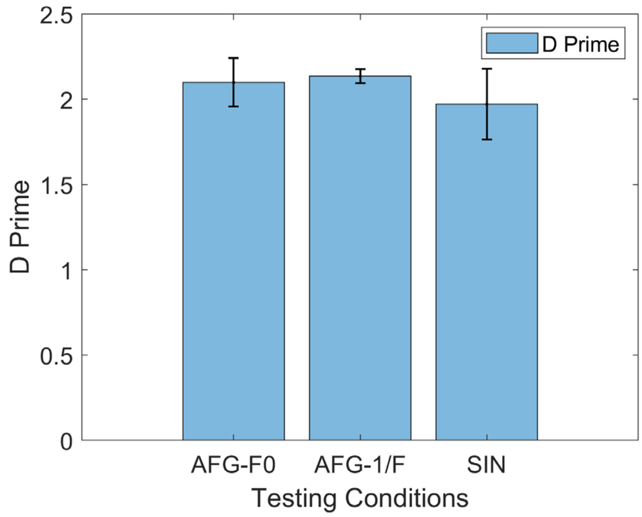

Figure 5.8 Participants' performance in the experiments. The x-axis of the bar plot shows the three conditions as labelled, and the y-axis shows the d' values. The black error bars show the standard error of the mean.

Neural Responses at the Fundamental Frequency with a Single Channel

I examined neural entrainment to the contour of the fundamental frequency in the three experimental conditions by looking into the TRF weights obtained from the forward model on a group level. All group-averaged peak latencies are summarised in Table 5.3. First, the responses for the SIN condition are shown in Figure 5.9(a). The delta band analysis showed a significant early response from 20-110 ms which peaked at 70 ms. This was followed by a later positive response from 180-350 ms that peaked at 260 ms at Cz. The theta-band responses showed a narrower early response range from 80-110 ms that peaked at 90 ms. A significant late response from 160-190 ms that peaked at 190 ms was also observed. The scale of the TRF response was larger for the delta band than the theta band. The topographies of both frequency bands showed either negative or positive activities maximal at the frontal-central electrodes.

TRF \ condition	Delta			Theta		
	SIN	AFG-F0	AFG-1/F	SIN	AFG-F0	AFG-1/F
Peak Latency Early	70	80	60	90	110	90
Peak Latency Late	260	280	220	180	210	170
Peak Amplitude Early	-0.101	0.117	0.011	-0.046	0.016	0.015
Peak Amplitude Late	0.116	-0.081	-0.005	0.047	-0.015	-0.012

Table 5.3 TRF peak time points chosen for the source localisation. These are the group-averaged latencies for the three conditions.

The TRF responses for the AFG-F0 condition are shown in Figure 5.9(b). The delta range for AFG-F0 showed comparable magnitude with the SIN delta condition but the theta range was much smaller (the mean absolute amplitude of AFG-F0 in theta was more than three times smaller than that of SIN). The delta-band response showed a significant positive wave before 150 ms that peaked at 80 ms. The second peak was observed at 280 ms (range 210 ms – 360 ms). The early peak was also found with the theta condition with a range from 70 ms to 130 ms peaking at 110 ms, but the second peak happened earlier compared to the delta band, which was at 210 ms (ranged 180 ms – 240 ms). A third transient peak at 320 ms was also visible for the theta condition (range 290 ms – 340 ms).

The AFG-1/F condition (Figure 5.9(c)) showed a similar pattern to the AFG-F0 condition but with a much smaller magnitude in the Delta band. In addition to the first positive response before 110 ms peaking at 60 ms, and the second 220 ms peak (range: 20 ms – 24 ms), there was later significant negativity at 420 – 540 ms in the delta band that peaked at 480 ms. Theta band showed significant 90 ms (range: 60 ms – 100 ms) and 170 ms peaks (range: 140 ms – 190 ms) similar to AFG-F0 and a transient third peak at 250 ms.

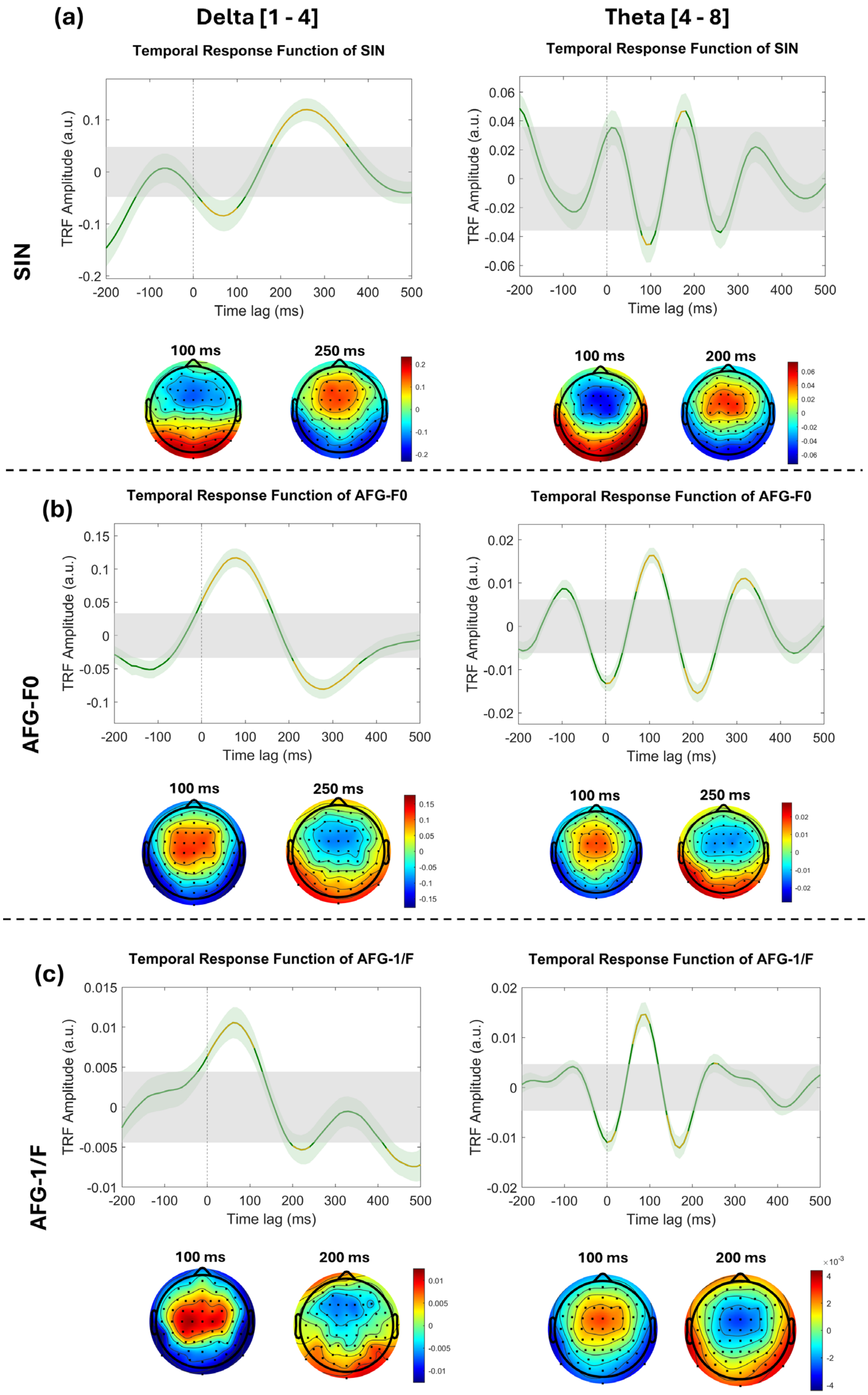

Figure 5.9 TRF responses and topographies for SIN (Figure 5.9(a)) AFG-F0 (Figure 5.9(b)) and AFG-1/F (Figure 5.9(b)) at two frequency bands. The x-axes of the TRF plots show the time lag in milliseconds and the y-axes show the TRF weights in arbitrary units. The TRF waveforms are plotted in a dark green curve with a light green shadow as the standard error. The grey rectangular shadow marks the area of null distribution at 0.05 alpha level, and the yellow curves highlight the significant peaks.

The TRF peak values of the early and late waves were extracted from the AFG conditions. I ran the Pearson correlation between the SIN d' and AFG peak values of the Delta frequency (all data were normally distributed) and found a significant correlation between the early peak of AFG-F0 with SIN ($r = -0.38$, $p = 0.037$) but not the late peak (See Figure 5.10 for more details). The correlation between the TRF peaks of AFG-1/F and SIN was not significant ($p > .206$).

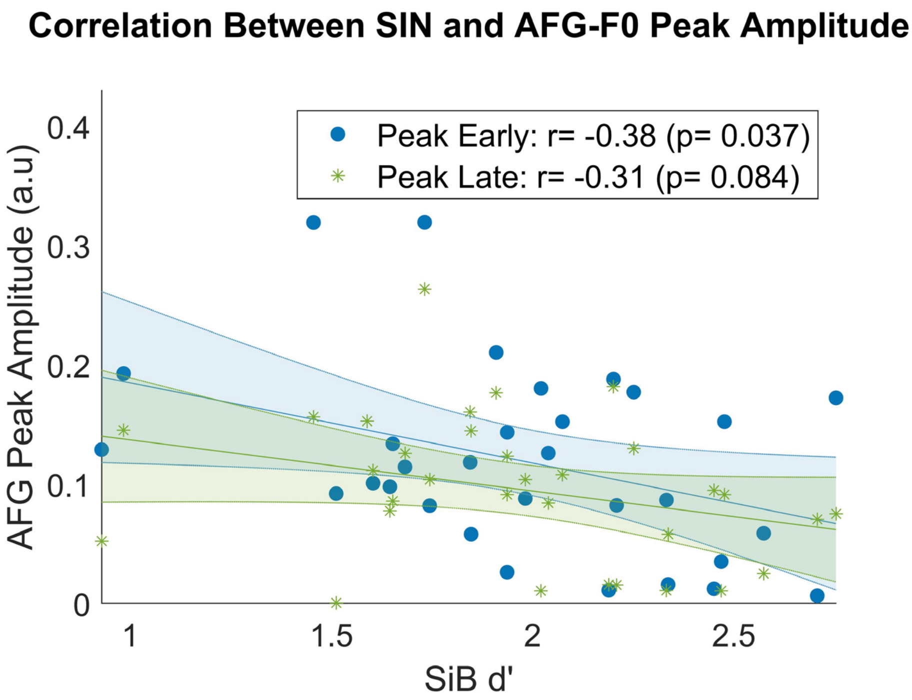

Figure 5.10 Scatterplot of the relationship between SIN d' and the absolute TRF peak amplitudes of AFG-F0 at the early and late peaks (80 ms and 280 ms). The x-axis shows the d' of SIN, and the y-axis shows the peak values of TRF waveforms of the AFG-F0 condition. The shaded area plots the 95% confidence bounds. The legend shows the Pearson correlation coefficients (r) and the p-values.

The reconstruction accuracies of the TRF forward models are shown in Figure 5.11. All accuracies were significant compared to the null distribution (Table 5.4). The ANOVA results indicated a non-significant main effect of stimulus type ($F(2, 62) = 1.90$, $p = 0.158$, effect size: $\eta p^2 = .06$). The main effect of frequency bands was significant ($F(1, 31) = 139.94$, $p < .001$, $\eta p^2 = .82$) due to the lower accuracy of the theta band ($R_{delta} = 0.04$, $R_{theta} = 0.03$). The interaction between stimulus type and frequency bands was significant ($F(2, 62) = 17.23$, $p < .001$, $\eta p^2 = .36$). The interaction was followed up by paired samples t-tests based on the descriptive data (Table 5.4) which showed lower predictive accuracy of the theta band in the SIN condition compared to AFG-F0 (t (31) = -5.72, p < .001), and AFG-F1 (t (31) = -4.83, p < .001), whereas there was no significant difference between the two AFG conditions.

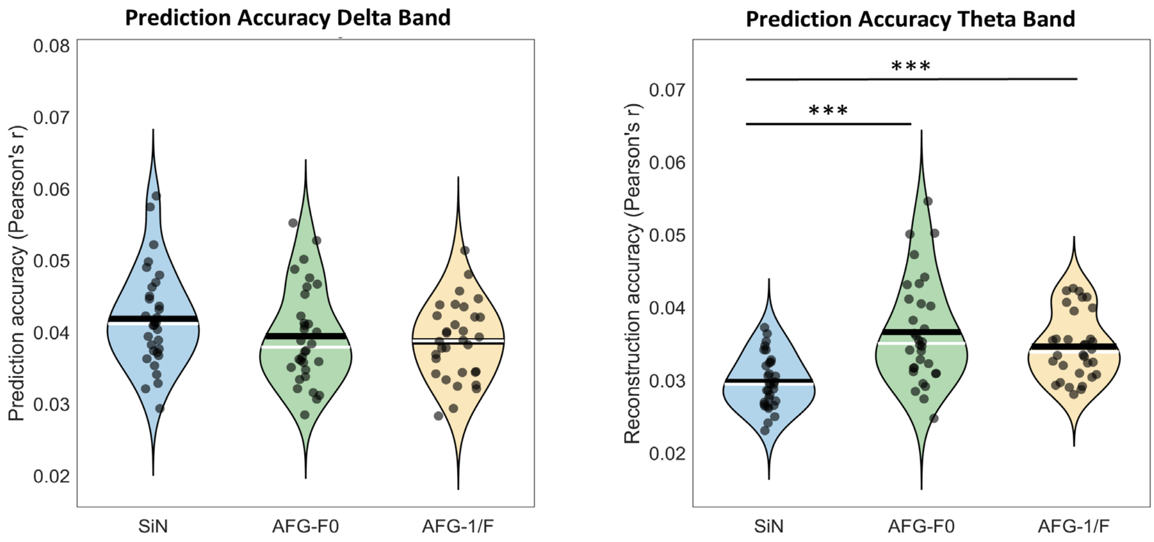

Figure 5.11 Prediction accuracies of TRF models in two frequency bands. The black dots show the reconstruction accuracy of individual participants. The median and the mean are plotted in white and black lines respectively. The asterisks with an underlying line illustrate the significant mean difference between conditions. Three asterisks '***' suggest an alpha level of p< .001.

In order to rule out the possibility of the lower SIN reconstruction accuracy being statistically introduced by our more stringent method of extracting the 'pure' pitch by regressing out the stress contour, I ran the analysis again with the stress contour left in the signal and found that, while the accuracy did improve overall, the SIN condition was still significantly lower than the AFG conditions in the theta-band.

Conditions\ Frequency	Delta			Theta		
	M	SD	t (p)	M	SD	t (p)
SIN	0.042	0.007	17.44 (p< .001)	0.030	0.004	2.02 (p= .023)
AFG-F0	0.040	0.007	13.35 (p< .001)	0.037	0.007	8.92 (p< .001)
AFG-1/F	0.039	0.005	13.85 (p< .001)	0.035	0.005	5.70 (p< .001)

Table 5. 4 Prediction accuracies of TRF models in two frequency bands. The t-test was against the null distribution from the permutation test.

Source Locations

As the reconstruction accuracy showed that theta-band tracking is significantly less accurate than the delta-band, EEG source analysis was only conducted on the delta condition (Figure 5.12). The sLORETA source analysis provided clear localised activities in the SIN and AFG-F0 condition but not the AFG-1/F. The peak time points are summarised in Table 5.3. The SIN and AFG-F0 TRF early delta peaks localised to the superior temporal gyrus (STG), middle temporal gyrus (MTG), and inferior temporal gyrus (ITG) (Figure 5.12). Bilateral source locations were seen in the MTL (hippocampus and parahippocampal region) and insula as well. Outside the temporal lobe, activities in the prefrontal lobe, inferior frontal gyrus (IFG), medial frontal lobe, precentral gyrus and postcentral gyrus, superior parietal lobe (SPL), precuneus, and cuneus were found for both the AFG-1/F and SIN conditions. The SIN peaks showed a more lateralised pattern of activities compared to the AFG-F0 condition. The AFG-F0 first peak showed bilateral tracking but it became left-lateralised at the inferior temporal gyrus, parahippocampus, and the precentral gyrus.

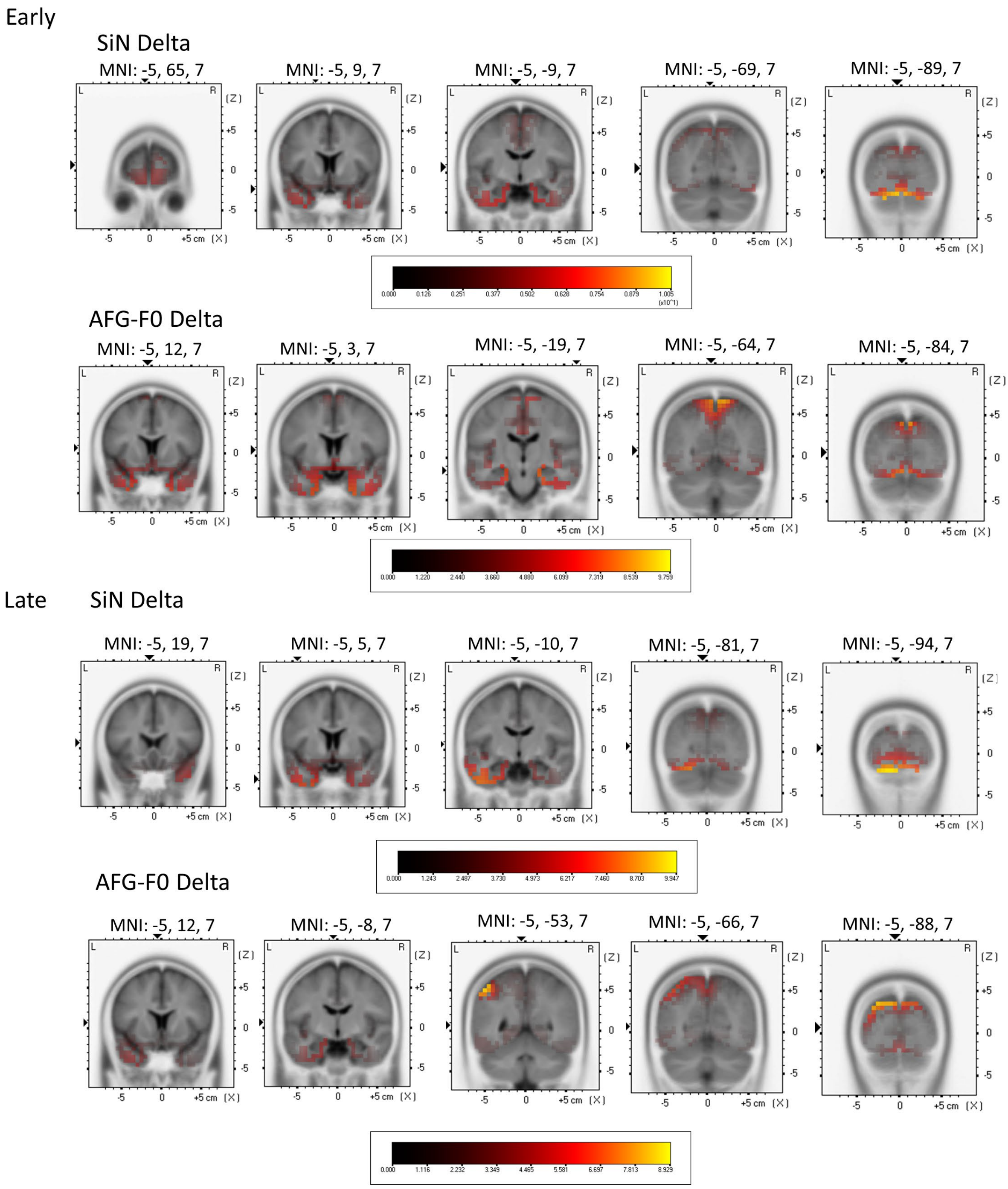

Figure 5.12 EEG source-level neural activities of Delta band TRF peaks to the fundamental frequencies in SiN condition and AFG-F0 condition.

5.2.4 Discussion

The behavioural results showed that participants achieved excellent performance for all three tasks. As the design of the task required constant tracking of the target sound, the high performance indicated that the participants were maintaining their attention on the target sound. The TRF analysis showed that the brain can reliably track F0 or F0-like frequency contour changes in natural or figure-ground stimuli. However, different TRF morphologies and reconstruction accuracies were found both in terms of the type of stimulus and the specific frequency ranges that characterise neural oscillations in EEG.

Similarity in cortical tracking and source locations of synthetic AFG and natural SIN

The similarity between SIN and AFG processing has been demonstrated in both behavioural and neuroimaging domains previously with the stochastic figure-ground stimulus (Holmes et al., 2021b; Holmes & Griffiths, 2019; Schneider et al., 2018; Teki et al., 2011; Teki et al., 2016). O'Sullivan et al. (2015) further explored the neural tracking to the temporal coherence level of a random-frequency dynamic figure-ground stimulus and speculated that the pattern of TRF responses to AFG could be similar to that of SIN. Our results support this assertion, as detailed below.

Firstly, all testing conditions showed significant reconstruction accuracy based on the TRF forward model. This means that the brain can successfully entrain to the frequency changes in either type of stimuli regardless of linguistic content, levels of predictability, or frequency range of neural oscillation (Delta or Theta). In terms of the TRF waveforms, the SIN and AFG-F0 conditions also demonstrated similar temporal encoding, with both conditions showing a peak at ~100 ms and a second peak with inverted polarity at ~250 ms in the delta band, and the same pattern at ~100ms and ~200ms in the theta band. These peak latencies were similar to what was found in previous studies that looked at TRF responses to SIN (Aljarboa et al., 2023; Bachmann et al., 2021; Ding & Simon, 2012a). The similarity in the TRF time signature means that the brain likely is responding to the two stimuli on the same timescale, although the type of responses is not necessarily the same. The detailed TRF morphology will be discussed in the next section.

Further investigation into the source of the significant peaks showed that both SIN and AFG-F0 had generators in the temporal neocortex, parietal cortex, and MTL, which were consistent with previous neuroimaging data (Holmes et al., 2021; Teki et al., 2016). In particular, parietal activities were found in both SIN and AFG conditions.

Based on visual inspection, the superior parietal lobe activities seemed to be stronger and more widespread for AFG than SIN stimuli for the early peak, which was supported by studies comparing SIN to AFG or other non-speech signals in noise (Holmes et al., 2021b; Kulasingham et al., 2021). The intraparietal sulcus (IPS), active during AFG and SIN tracking, has been implicated in stream segregation across multiple sensory domains due to its role in top-down attentional modulation (Calvert, 2001; Cusack, 2005). The engagement of MTL shown here was also found in SIN before as well as another EEG study investigating the source of figure-ground segregation (Tóth et al., 2016). Studies have found that MTL, particularly the hippocampus, is involved not only in auditory working memory but also in extracting complex auditory patterns (See Billig et al., 2022 for a review on the role of the hippocampus in auditory cognition).

Polarity differences in TRF waveform of SIN and AFG

Unlike the wealth of literature on auditory-evoked potentials (AEP), TRF research is relatively new, and the interpretation of the TRF forward model focuses mainly on comparing the absolute amplitudes or prediction accuracies between conditions, but the polarities are rarely discussed. However, the time lags and relative fluctuations in TRF waveforms can provide important information as well. The current study can offer insight into the interpretation of TRF morphologies. Firstly, I demonstrated that the neural tracking of SIN showed response patterns similar to the N1, P2/M200 responses in auditory-evoked potential (hereafter referred to as $N1_{TRF}$ and $P2_{TRF}$ to distinguish from the AEP components), replicating previous findings (Aljarboa et al., 2023; Bachmann et al., 2021; Ding & Simon, 2012a). The AFG condition, on the other hand, showed the opposite polarities at the same lags, which was also consistent with the previous findings on AFG stimuli (O'Sullivan et al., 2015). The opposite polarities of the two conditions were reported by Horton et al. when they compared neural tracking of attended and unattended speech envelopes (Horton et al., 2013). They hypothesised that the inverted polarity seen in the unattended condition could reflect a suppression mechanism during auditory scene analysis, in which the attention network was phase-locked to the inverse of the envelope of the noise. However, studies of a similar design did not find this pattern (O'Sullivan et al., 2015; Power et al., 2012), although researchers did observe lower TRF amplitude and a degree of shifts in the latencies for the unattended stream. It is important to note that in this study, both the types of stimuli (speech vs. pure tone sequence, babble noise vs. tone cloud) used for the two conditions and the tasks (gap detection, repetition detection) were very different, which could result in this polarity inversion. It is therefore uncertain whether the polarity inversion observed here was related to attentional manipulation or performance, or whether it reflects the same neural process, but time-shifted.

Literature on speech tracking in noise suggests that time shifting of TRF peaks is a more likely explanation. I reviewed the recent literature on the neural tracking of continuous speech stimuli using TRF analysis or cross-correlation and found that a wide range of peak latencies have been observed with very similar stimuli and filtering functions (see Table 5.5).

Article	First Peak(ms)	Second Peak(ms)	Following peaks	Stimulus type	Task	Filter	Method
(Panela et al., 2024)	Negative 90-130	Positive ~200		Speech in babble noise	Answering comprehension questions.	10 Hz lowpass	EEG
(Aljarboa et al., 2023)	Negative 50-100	Positive 100-150		Single-talker speech	Answering comprehension questions.	1-30 Hz bandpass	EEG
Same as above	Xcor negative 10-100	Xcor positive 180-150					
(Brodbeck & Simon, 2022)	Normalised peak at 50-100			Single-talker pitch strength and value	Not specified	20-Hz lowpass	MEG
Same as above	Normalised peak at 100~150			Two speech streams pitch strength and value			
(Kegler et al., 2022)	Positive 11			Speech (high-frequency envelope modulation of pitch)	Answering comprehension questions.	50-280 Hz bandpass	EEG
(Muncke et al., 2022)	Negative 100	Positive 200		Speech in noise (Intelligibility)	Passive listening while watching a movie.	1-10 Hz bandpass	EEG
(Bachmann et al., 2021)	Positive 77.07–139.57			Single-talker Relative pitch	Answering comprehension questions.	1-9 Hz bandpass	EEG

(Etard & Reichenbach, 2019)	Xcor Positive 80 TRF Positive 80			Speech in babble noise (envelope of noise)	Answering comprehension questions.	Delta (1-4 Hz)	EEG
Same as above	Positive 90	Negative 390		Target speech Clarity			
Same as above	Positive ~100	Negative 100-230		Target speech comprehension			
(Teoh et al., 2019)	Positive ~160			Single-talker speech (relative Pitch)	Attended listening	Delta (0.2-4 Hz), theta does not encode	EEG
Same as above	Positive ~160 (Delta) Positive ~30(Theta)	Negative 110	Positive 190 Negative 260	Single-talker speech (harmonic resolvability)			
(Broderick et al., 2019)	Positive 100	Negative 400		Competing speech (attended)	Attended listening with a fixation cross	1-8 Hz bandpass	EEG
Same as above	Positive 200	Negative 550-600					
(Ding & Simon, 2013)	Positive 0-80	Negative 80-180		Speech in spectrally matched stationary noise	Attended listening with eyes closed.	1-9 Hz bandpass	MEG
(Horton et al., 2013)	Xcor Positive 0-100	Negative ~200	Positive 250-400	Attended speech	Dichotic listening: chose the direction of the sound source	1-50 Hz bandpass	EEG
Same as above	Xcor no peak	Positive 200		Unattended speech			
(Power et al., 2012)	Positive 50-150	Negative 150-25		Two competing speech streams	Answering comprehension questions.	2-3 Hz bandpass	EEG
(Ding & Simon, 2012a)	Positive ~50	Negative ~100		Two competing streams	Answering comprehension questions.	1-8 Hz bandpass	MEG
(Ding & Simon, 2012b)	Negative 100-200			Two competing streams	Dichotic listening, attended listening with eye closed	1-8 Hz bandpass	MEG

Table 5.5 Latencies and polarities of TRF responses summarised in recent literature. This is not an exclusive list. The peaks reported are mostly TRF peaks, but those marked with Xcor are cross-correlation (Xcor) peaks.

As summarised in Table 5.5, the initial response to tracking a target speech can manifest as a positive TRF peak from 0ms to 160ms, or a negative peak from 0ms-200ms, followed by a peak of the opposite polarity of 100m-390ms. A few studies reported further fluctuations from 190 ms to 400 ms as well. This wide latency range of responses indicates that the definition of a TRF N1 P2 or M50, M100, or M200 based on AEP could be misleading. Unlike the relatively reliable N1 response in evoked potential, TRF literature does not necessarily show a 100 ms negative deflection. What the literature shows is that TRF morphology can vary in the number of peaks and peak latencies when examining very similar stimulus features. The variation could be due to unknown task-specific effects, filtering functions, or attention.

Delta and Theta Bands Encode Different Levels of Acoustic Information

The ANOVA test on the model reconstruction accuracy found that neural tracking of frequency patterns on the theta band had significantly lower accuracy compared to the delta band. Furthermore, the post-hoc t-test showed that the lower SIN accuracy compared to the two AFG conditions was what drove the interaction. The magnitude of the theta responses, however, exhibited the opposite pattern: speech tracking had a higher amplitude compared to AFG.

Cortical speech-tracking has been performed mainly on speech envelope instead of F0, as studies have found relatively low reconstruction accuracy for pitch encoding compared to acoustic envelope encoding, and even non-significant models for pitch in the theta band (Bachmann et al., 2021; Teoh et al., 2019). The current results, however, suggest that low-accuracy pitch-encoding could be a speech-specific effect, as the AFG forward models maintained their prediction accuracies. One possible explanation for the relatively unreliable theta-band neural tracking for SIN is that the theta frequency might encode spectrotemporal information better than complex speech information. Previous literature has broadly related delta tracking to processing high-level features of speech, e.g. semantics and selective attention, whereas the theta band was linked to low-level acoustic processing such as the rhythmic structure of speech (Ding & Simon, 2014; Zion Golumbic et al., 2012; Etard & Reichenbach, 2019; Peelle, 2013). The SIN condition used here encompasses high-level linguistic contents that can have an impact on the EEG responses whereas the AFG conditions only tap into sound segregation based on speech or speech-like pitch contours and harmonicity, which might be preferentially processed by the theta band with greater synchronisation between the neural signals and pitch information. On the other hand, the lack of linguistic information in AFG led to a smaller TRF amplitude. This could be attributed to the effect of listening effort. Enhanced AEP N1 response has been observed for more effortful speech perception (Obleser & Kotz, 2011; Ghani et al., 2020).

Relationship between EEG *responses and behaviour, and potential clinical application*

Finally, I found a significant negative correlation between the early peak of AFG-F0 and SIN performance but not the late peak. Traditionally, the correlation between behavioural SIN and AFG performance has been demonstrated with large samples (n>100) (Guo, et al., 2024; Holmes & Griffiths, 2019). I was therefore not expecting a significant correlation here between the performance of the two tasks themselves. However, a small to moderate negative association was found between AFG-F0 and SIN d’, which could mean that the EEG neural tracking might be more sensitive than behavioural measures in showing this association. Higher amplitude for TRF weights can relate to a variety of auditory cognitive processes. A common finding in SIN perception is that attended streams tend to elicit stronger TRF responses, and the attentional modulation has been found to be strongest at ∼100–250 ms (Horton et al., 2013; Ding & Simon, 2012a; Zion Golumbic et al., 2012). Higher demands for cognitive resources are posed for participants with lower SIN ability as they would need to recruit more attentional or working memory resources to compensate for their impaired fundamental sound grouping ability. A similar effect was found in speech processing in reverberant in a recent study, in which the researchers combined pupillometry recording as well as EEG and found that listening effort and the strength of cortical tracking in the delta band increased with increasing difficulty in SIN perception (Ershaid et al., 2024). Enhanced AEP N1 response has been observed for more effortful speech perception (Obleser & Kotz, 2011; Ghani et al., 2020). While the current design cannot specify if the negative correlation shown was due to listening effort or a general cognitive effect, future studies could incorporate measures of listening effort or attention to test the hypothesis. If the significant correlation can be replicated, the TRF signature of AFG can be potentially used to measure natural listening. The simple setup and efficient recording make it feasible for its usage in clinics. This method has the advantage of posing intrinsically low demands on the patient's ability to do complicated language tasks unlike most of the SIN tests and can dissociate the contribution of auditory processing and linguistic processing, which is influenced by language competence, education, accent, and other social factors.

No correlation was found between the peaks of AFG-1/F and SIN d' despite the same level of reconstruction accuracy derived by the two AFG conditions. The difference could be driven by the distinct levels of stimulus-predictability. Natural sentence trajectories have a level of periodicity with regular recurrence of pitch patterns over time, but the 1/F pattern was mathematically generated and was not configured to have a recurring similar pattern. The lower predictability led to sustained tracking on the target sound to facilitate figure-ground segregation for the AFG-1/F compared to other conditions as evidenced by the significant activities around later latencies (around 500 ms) in delta. I also found pre-zero activities in the SIN and AFG-F0 conditions, whereas there were no significant pre-zero TRF peaks within the 200 ms window before zero for the AFG-1/F condition. The pre-zero activities for the natural speech and AFG-F0 conditions were likely generated by correcting predictions of upcoming pitch contour changes, which were not present for the AFG-1/F condition. This suggests that while an artificial pitch contour can generate the same level of model prediction accuracy, the underlying process might still differ from the processing of natural speech contours.

To conclude, I have successfully demonstrated strong neural tracking of complex frequency patterns including natural pitch contour or speech-like contour in both SIN stimuli and AFG. The pattern of the pitch tracking of AFG-F0 and SIN stimuli showed a high level of similarity in the encoding accuracy, TRF latencies, and source locations of the TRF peaks in the Delta condition. In the theta band, however, the AFG obtained higher model accuracy than speech models with lower magnitude possibly due to lower demand of listening effort. The peak amplitude of the AFG-F0 condition also correlated with SIN performance, suggesting potential clinical use.

A major limitation of this study is that the 'performance' measure of SIN processing was based on a simple task of detecting repetition with a small number of trials. To obtain a more reliable relationship, a proper assessment of SIN performance is needed using speech-based tests and a larger number of trials. Future experiments should be conducted to validate the correlation between the AFG-F0 amplitude and SIN performance.

6. Chapter 6: Conclusion and general discussion

This work summarised the mechanisms of the auditory system involved in supporting speech perception in noise and reviewed the commonly used hearing tests in clinics that can predict SIN ability. The first two chapters identified outstanding questions in the field and led to the two main objectives of this thesis: exploring the inter-relations of the auditory cognitive predictors of SIN, and developing new measures of SIN perception that can both better assess real-life listening and facilitate research into the link between listening and cognition. Driven by the two objectives, experiments were carried out to explore the links between the auditory cognitive predictors of SIN perception using multivariate analysis. New verbal and nonverbal listening tests were developed to better assess different aspects of SIN processing, and EEG responses to sound segregation and target-tracking were investigated to reveal the underlying neural mechanisms of SIN analysis.

6.1 Predictors of speech-in-noise perception

This work identified key predictors of SIN perception and developed a dynamic AFG paradigm that can explain an independent variance of SIN. The comprehensive review presented in Chapter 2 summarised most of the commonly used measures that indicate real-life listening ability. First of all, speech-based tests were often regarded as the best tool to assess real-life listening ability. The review on verbal tests discussed a discrepancy between subjective ratings and objective scores when quantifying a person's real-life listening ability. It was suggested that objective ratings should be used to measure performance, important for assessing one's listening ability and subjective measures should be used to inform a patient's personal experience of hearing aids use or rehabilitation.

The audiogram was shown to be the most used tool by far to describe hearing ability in clinics and research. I conducted a meta-analysis on the relationship between the pure-tone audiometry including the standard and extended-high-frequency audiograms and SIN performance as assessed by various speech-based tests. The results revealed a moderate correlation ($r = 0.450$) between standard-frequency PTA, and a weaker correlation ($r = 0.384$) for the extended-high-frequency PTA (Table 2.2). Age was shown to modulate the relationship but only on extended-high-frequency PTA.

However, the reliability of this result could be influenced by the selective reporting of certain published studies; non-significant results were often omitted from publications and could not be analysed. Other potential factors that can impact the strength of this relationship are hearing sensitivity and sample size. While some evidence suggested that the type of SIN materials could influence the strength of the PTA-SIN relationship (Wilson et al., 2007), this was not found when analysing a larger number of studies.

For temporal processing, tests of temporal acuity were found to be a reliable measure of SIN performance, but temporal ordering could not predict SIN performance. The effect of other domains of temporal processing on SIN has not been well researched. Measures of auditory stream segregation, especially the auditory figure-ground paradigm demonstrated the potential to be a reliable measure of central sound segregation, which is crucial for SIN processing. Researchers demonstrated a moderate effect size of $r = 0.32$ in one study exploring the association between figure-ground and SIN but the result needed further validation (Holmes & Griffiths, 2019). Measures of working memory have been well reviewed and generally showed a small to moderate effect of relationship with SIN especially in processing speed, inhibitory control, and working memory (Dryden et al., 2017). More auditory-specific short-term memory tests for frequency and amplitude precision showed an effect size of around $r = 0.49$, but a significant correlation was not consistently found (Lad et al., 2024, 2020a). In terms of physiological measures, despite the effect of pupil response and facial expressions on revealing listening effort, electrical recordings of brainstem and cortical responses are currently the only reliable tools that can be used to examine SIN performance. Cortical measures such as ASSR and N1 are strong predictors of SIN performance ($r > 0.6$) (Manju et al., 2014). Detection of auditory changes only showed a significant association with SIN performance when elicited by ACC but not MMN.

Informed by the review, I conducted a behavioural study to explore the inter-relationships among some of the most relevant auditory cognitive predictors of SIN perception using the measures that showed a close association with verbal SIN scores. This incorporated measures of auditory streaming/grouping, auditory short-term memory, temporal acuity, as well as phonological working memory, fluid intelligence, musical sophistication, and a test for reading ability or crystallised intelligence. The results demonstrated moderate to strong correlations ($r = 0.3$-0.7) between SIN performance and all the predictors included. Linear regression models revealed that age was the most important predictor of both word-level and sentence-level SIN perception and PTA was not a significant predictor after accounting for age and other central auditory measures. In addition to age, for word-level perception, the auditory short-term memory test for frequency precision explained a large variance of word-in-noise perception (0.046), followed by reading ability, figure-ground gap discrimination, verbal working memory, and figure-ground detection. For sentence-level perception, more important factors were reading ability, verbal working memory, and figure-ground gap discrimination. This revealed critical differences between sentence and word processing when masked by noise: while both require fundamental sound grouping, processing single words needs more precision and short-term memory for frequency information, whereas sentence processing recruits more higher-level cognitive mechanisms, including reading ability and working memory - functions that are less affected by age-related cognitive decline.

Further analysis with age-split data suggested that people of different age groups tackled the SIN tasks differently. Young people consistently showed significant correlations of precisions for frequency & AM rates and gap detection thresholds with SIN tests. In contrast, for older people, the AM precision and gap detection scores did not correlate with sentence-in-noise processing, and only a weak correlation was found with frequency precision. This could suggest different computation strategies for younger and older people. Younger people may rely on acoustic cues when processing speech stimuli with or without context. However, older people with deteriorated perceptual systems might only rely on acoustic cues when no other cues are available (e.g. single-syllable stimuli). When processing sentence-level stimuli, older people might employ more cognitive resources to compensate for the loss of frequency and temporal acuity.

If hearing sensitivity did not predict SIN measures in the linear regression models, how did it influence listening? The structural equation model answered this question (Section 3.2). Similarly to what was found in the linear models, PTA did not predict SIN directly. However, PTA modified both the short auditory processing latent construct (AFG and gap detection) and long auditory processing (auditory-specific memory and verbal working memory), and they both significantly modified SIN performance.

As SIN perception relied significantly on both short-term and long-term central sound processing, a new paradigm was developed that incorporates the auditory figure-ground stimulus and a pattern discrimination task of the auditory memory paradigm. The dynamic auditory figure-ground stimulus further incorporated the fundamental frequencies of speech to better simulate natural speech and avoid the power differences between the figure and the ground in the prototype AFG as the repetitive frequencies can produce more coherent energy at specific frequencies compared to randomly varying or non-repetitive signals. The new dynamic figure-ground was shown to predict both word and sentence perception in noise better than the prototype AFG after accounting for PTA and age ($R^2_{change\ (SiB)} = 0.099$, $R^2_{change\ (WIN)} = 0.082$, Section 4.2). This improvement could be partly attributed to the dynamic pitch contours that are speech-like, and partly attributed to the pattern discrimination paradigm which ensured continuous figure-tracking more than a gap-detection task could and added working memory load that was shown to predict SIN (Section 3.2). This is further evidenced by comparing the structural equation models of Chapters 3 and 4. The SEM models of Section 4.2 with the prototype AFG, two dynamic AFG measures, PTA and age together explained 62%-86% of the variance in SIN perception. The model presented in Section 3.2 incorporating the AUM measures was only able to explain 47% variance in SIN. While a direct comparison of the model fits or adjusted R-squared values between the two SEM models is not justified due to the sample differences and structural differences, a comparison of the outcomes can be explored. The model in Section 3.2 showed that age was the most important predictor of SIN, followed by the long-term central processing latent variable (auditory short-term memory and working memory) and short-term central processing (fundamental grouping and temporal acuity). However, the models in Section 4.2 showed that when combining the dynamic figure-ground with the static figure-ground, the auditory figure-ground alone explained a higher variance than age and PTA. While the samples of the two studies are different, the demographic features are similar in terms of age and hearing sensitivity and there was around 30% overlap between the two samples (same participants who took part in both studies). It is therefore plausible to think that using the dynamic figure-ground with the pattern discrimination paradigm can assess SIN performance better than using the static figure-ground or AUM measures alone, as it can be seen as a combined method of the two tests. Future studies should assess the dynamic figure-ground and AUM in the same model to test this hypothesis. It is also important to note that as the new pattern-discrimination task is very different from the original gap-detection task, the choices of which figure-ground tasks to use should be based on whether the aim is to test “pure” sound grouping or sound grouping over time with pattern analysis.

To further understand the mechanisms of AFG and SIN processing in the brain, two EEG experiments were carried out. An event-related potential design elicited figure-ground segregation response at around 139 ms, peaking at 300 ms post figure-onset. The response to the target figure elicited a significant negative peak compared to the ground under both attended and distracted conditions. For SIN processing, however, no segregation was observed under the distracted condition, suggesting that sound segregation at the cortical level might not be as sensitive for speech stimuli compared to the simpler auditory figure-ground stimuli. In this study, the amplitude or latency of the figure-ground segregation failed to show any association with SIN performance. This could be due to the relatively small sample size (n = 18). It could also reflect the weak association between fixed-frequency figure-detection task and speech recognition in noise as shown in the behavioural studies.

The second EEG study explored two types of dynamic stimuli that combined instantaneous sound grouping and continuous tracking of the pitch contours: dynamic figure-ground stimuli and sentence-in-noise stimuli. The results demonstrated that the pitch changes in both AFG-dynamic and SIN stimuli can be entrained reliably in the delta and theta bands. The early-peak amplitude of tracking F0 in AFG correlated with SIN performance significantly. This suggests that AFG pitch tracking can be a potential biomarker for SIN perception. However, the size of the coefficient was small, and the results need to be validated with proper sentence or word-in-noise tests as the behavioural performance in the study was a simple detection task of repeated sentences that might not reflect real-life listening well.

6.2 Exploring the relationship between listening and cognition

The hypothesis linking listening and cognitive decline stemmed from research associating hearing loss and dementia, which showed that the relationship between SIN and cognition decline was stronger than peripheral hearing to cognitive decline (Hoff et al., 2023; Mamo & Helfer, 2021). A hypothesis on the neural mechanisms explaining the link between hearing and dementia also suggested that, during effortful listening, the heightened activity of the MTL increases AD pathology due to SIN difficulty. A recent study on mice found that induced deafness led to an increase of amyloid-β plaques in the hippocampus and temporal cortices, and it was also associated with decreased hippocampal synaptic density as well as cognitive decline (Pan et al., 2024). This suggests that the interaction between hearing and dementia or cognitive decline might rest somewhere higher than the auditory periphery, likely at the hippocampus or the cortex.

Based on the data from Chapter 3, an exploratory analysis was conducted testing the hypothesis that hearing or listening (central sound processing and SIN) can modify cognitive performance. The structural equation model showed that central sound processing measured by figure-ground gap discrimination, auditory short-term memory for amplitude precision, and gap detection, predicted general cognition with the strongest effect (path coefficient = 0.82), which was even higher than age (path coefficient = 0.44). Neither SIN measures nor PTA had a significant path leading to general cognition, but PTA affected cognition indirectly through central sound processing. If causal relationships between hearing loss and cognitive decline could be established through experimental manipulation, this data could show that the core mechanisms driving this relationship rest at the central auditory system involved in non-verbal SIN perception.

While I cannot conclude causal relationships based on this analysis, it provided guidance for future research. Central sound processing should be considered a key aspect of the research investigating hearing loss causing cognitive decline or dementia. One tool that I suggested in this work is the dynamic figure-ground paradigm. The EEG source analysis detailed in Section 5.2 showed that the significant neural tracking of the AFG frequency changes could be generated by the medial temporal lobe, temporal cortex, and parietal cortex, all of which were found for SIN processing as well. The MTL was proposed as a processing hub relevant to both hearing loss and dementia. As EEG source localisations have poor spatial resolution, the most immediate step is to examine this paradigm with intracranial or fMRI recordings that allow a more detailed examination of the hippocampus. If similar generators can be found with these methods, dynamic figure-ground could be used as an important tool for investigating the relationship between central sound segregation and cognition. Patient studies could be carried out comparing people with cognitive impairment and AD dementia with healthy controls of similar hearing sensitivity to see if they can perform AFG pattern discrimination or track the auditory figure in a similar way. Longitudinal studies recording cognitive performance over time in both healthy ageing people and those with mild cognitive decline could also be useful to reveal if AFG pattern discrimination can indicate the rate of cognitive decline and thus can be used as a helpful tool to identify people with a high risk of developing dementia.

6.3 Future directions

Considering the close predictive relationship of AFG and SIN perception, further studies could be conducted to develop AFG into a clinical diagnostic tool. AFG could be used as a complementary hearing assessment. It can test patients' peripheral and central sound processing, independent of language ability, and provide useful information on their hearing profile for audiologists to determine the effect of an intervention. As suggested by the EEG findings in Chapter 5, EEG evoked potentials of figure-ground segregation or neural entrainment to figure tracking were robust under different conditions, and the analysis could be based on a very simple vertex-to-mastoid configuration. This suggests the potential of developing them for clinical usage, which needs simple setups that allow time-limited testing. The TRF peaks of dynamic figure-ground, especially, demonstrated a significant correlation with SIN performance, which means that the EEG responses to dynamic figure-ground can be used to assess not only the fundamental sound grouping ability but also SIN ability directly. Future work should be carried out to validate the test-retest reliability with a larger sample and shorter recording length. It is particularly important to examine people with hearing disorders in addition to establishing a normative response pattern. Future studies can inspect if the test is robust with CI users or hearing-aids users, for example, to assess their real-life listening ability.

In addition to validating the AFG tests for clinical diagnosis of SIN listening difficulty, training strategies based on the AFG paradigm could be developed. It is conceivable that improvement in AFG ability may predict SIN improvement, and a behavioural training scheme could be devised based on the paradigm proposed in Section 4.2. Researchers could identify people with SIN complaints and provide training to improve their ability to perform simple sound segregation based on the figure-ground gap-detection task and improve their figure-tracking and pattern analysis ability based on the dynamic figure-ground pattern-discrimination task. They may potentially improve SIN perception or reduce the listening effort during SIN perception. The training programme can be delivered through an online platform or a desktop application, with attention-checking tasks built into the programme to ensure engagement. This would be particularly beneficial for individuals who are unable to train on SIN tests due to the complexity of the stimuli, such as people with receptive aphasia or speech and language disorders. The tasks are also relatively simple to perform at home so the delivery of the training programme would not rely on a specialist.

It would be interesting to explore other forms of training strategies as well, such as neurofeedback. Neurofeedback is a form of biofeedback therapy that requires the participants to control their brain functions based on the feedback signals. Current use of this technique has focused on employing a generic form of training to treat a variety of diseases, such as control over certain brainwave frequencies to treat insomnia, epilepsy, anxiety, learning difficulty, and so forth, with inconclusive effects (Marzbani et al., 2016). In light of the question about its efficacy, the training method should be tailored to the training goal specifically to achieve a better effect. For instance, Section 5.2 demonstrated the feasibility of using EEG to record neural entrainment to figure-tracking with relatively high reliability. A similar paradigm can be used for training SIN ability by providing immediate neurofeedback on, for example, the prediction accuracy of the TRF waveform, or the SNR of attended auditory figure or speech compared to the unattended stream. This type of training could improve control of selective attention and short-term memory, thus having a global benefit in speech perception and cognition.

References

Adank, P., Evans, B. G., Stuart-Smith, J., & Scott, S. K. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology. Human Perception and Performance*, *35*(2), 520–529. https://doi.org/10.1037/a0013552

Agrawal, D., Dritsakis, G., Mahon, M., Mountjoy, A., & Bamiou, D. E. (2021). Experiences of Patients With Auditory Processing Disorder in Getting Support in Health, Education, and Work Settings: Findings From an Online Survey. *Frontiers in Neurology*, *12*, 607907. https://doi.org/10.3389/fneur.2021.607907

Aiken, S. J., & Picton, T. W. (2008). Envelope and spectral frequency-following responses to vowel sounds. *Hearing Research*, *245*(1), 35–47. https://doi.org/10.1016/j.heares.2008.08.004

Akeroyd, M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? A survey of twenty experimental studies with normal and hearing-impaired adults. *International Journal of Audiology*, *47 Suppl 2*, S53-71. https://doi.org/10.1080/14992020802301142

Alagramam, K. N., & Weisz, C. J. C. (2023). Auditory System, Central. In *Reference Module in Neuroscience and Biobehavioral Psychology*. Elsevier. https://doi.org/10.1016/B978-0-323-95702-1.00054-3

Alain, C., Du, Y., Bernstein, L. J., Barten, T., & Banai, K. (2018). Listening under difficult conditions: An activation likelihood estimation meta-analysis. *Human Brain Mapping*, *39*(7), 2695–2709. https://doi.org/10.1002/hbm.24031

Alain, C., Schuler, B. M., & McDonald, K. L. (2002). Neural activity associated with distinguishing concurrent auditory objects. *Journal of the Acoustical Society of America*, *111*(2), 990–995. Scopus. https://doi.org/10.1121/1.1434942

Alemi, R., & Lehmann, A. (2019). Middle Latency Responses to Optimized Chirps in Adult Cochlear Implant Users. *Journal of the American Academy of Audiology*, *30*(5), 396–405. https://doi.org/10.3766/jaaa.18014

Alexander, W. H., & Brown, J. W. (2018). Frontal cortex function as derived from hierarchical predictive coding. *Scientific Reports*, *8*(1), 3843. https://doi.org/10.1038/s41598-018-21407-9

Aljarboa, G. S., Bell, S. L., & Simpson, D. M. (2023). Detecting cortical responses to continuous running speech using EEG data from only one channel. *International Journal of Audiology*, *62*(3), 199–208. https://doi.org/10.1080/14992027.2022.2035832

Ananthanarayana, R. M., Buss, E., & Monson, B. B. (2024). Band importance for speech-in-speech recognition in the presence of extended high-frequency cues. *The Journal of the Acoustical Society of America*, *156*(2), 1202–1213. https://doi.org/10.1121/10.0028269

Anderson, K., & Smaldino, J. (1999). Listening Inventories for Education: A classroom measurement tool. *The Hearing Journal*, *52*(10), 74.

Anderson, L. A., & Linden, J. F. (2011). Physiological differences between histologically defined subdivisions in the mouse auditory thalamus. *Hearing Research*, *274*(1), 48–60. https://doi.org/10.1016/j.heares.2010.12.016

Anderson, S., & Kraus, N. (2010). Objective neural indices of speech-in-noise perception. *Trends in Amplification*, *14*(2), 73–83. https://doi.org/10.1177/1084713810380227

Anderson, S., Skoe, E., Chandrasekaran, B., Zecker, S., & Kraus, N. (2010). Brainstem correlates of speech-in-noise perception in children. *Hearing Research*, *270*(1–2), 151–157. https://doi.org/10.1016/j.heares.2010.08.001

Anderson, S., White-Schwoch, T., Parbery-Clark, A., & Kraus, N. (2013). A dynamic auditory-cognitive system supports speech-in-noise perception in older adults. *Hearing Research*, *300*, 18–32. https://doi.org/10.1016/j.heares.2013.03.006

Andersson, K. E., Neher, T., & Christensen, J. H. (2023). Ecological momentary assessments of real-world speech listening are associated with heart rate and acoustic condition. *Frontiers in Audiology and Otology*, *1*. https://doi.org/10.3389/fauot.2023.1275210

Antons, J.-N., Schleicher, R., Arndt, S., Moller, S., Porbadnigk, A. K., & Curio, G. (2012). Analyzing Speech Quality Perception Using Electroencephalography. *IEEE Journal of Selected Topics in Signal Processing*, *6*(6), 721–731. IEEE Journal of Selected Topics in Signal Processing. https://doi.org/10.1109/JSTSP.2012.2191936

Arenillas-Alcón, S., Costa-Faidella, J., Ribas-Prats, T., Gómez-Roig, M. D., & Escera, C. (2021). Neural encoding of voice pitch and formant structure at birth as revealed by frequency-following responses. *Scientific Reports*, *11*(1), 6660. https://doi.org/10.1038/s41598-021-85799-x

Bachmann, F. L., MacDonald, E. N., & Hjortkjær, J. (2021). Neural Measures of Pitch Processing in EEG Responses to Running Speech. *Frontiers in Neuroscience*, *15*, 738408. https://doi.org/10.3389/fnins.2021.738408

Bartlett, E. L. (2013). The organization and physiology of the auditory thalamus and its role in processing acoustic features important for speech perception. *Brain and Language*, *126*(1), 29–48. https://doi.org/10.1016/j.bandl.2013.03.003

Behroozmand, R., Ibrahim, N., Korzyukov, O., Robin, D. A., & Larson, C. R. (2015). Functional role of delta and theta band oscillations for auditory feedback processing during vocal pitch motor control. *Frontiers in Neuroscience*, *9*. https://doi.org/10.3389/fnins.2015.00109

Bent, T., Holt, R. F., Van Engen, K. J., Jamsek, I. A., Arzbecker, L. J., Liang, L., & Brown, E. (2021). How pronunciation distance impacts word recognition in children and adults. *The Journal of the Acoustical Society of America*, *150*(6), 4103–4117. https://doi.org/10.1121/10.0008930

Bentler, P. M., & Bonett, D. (1980). Significance Tests and Goodness-of-Fit in Analysis of Covariance Structures. *Psychological Bulletin*, *88*, 588–606. https://doi.org/10.1037/0033-2909.88.3.588

Bentler, P. M., & Chou, C.-P. (1987). *Practical Issues in Structural Modeling*. https://journals.sagepub.com/doi/10.1177/0049124187016001004

Berger, J. I., Gander, P. E., Kim, S., Schwalje, A. T., Woo, J., Na, Y.-M., Holmes, A., Hong, J. M., Dunn, C. C., Hansen, M. R., Gantz, B. J., McMurray, B., Griffiths, T. D., & Choi, I. (2023). Neural Correlates of Individual Differences in Speech-in-Noise Performance in a Large Cohort of Cochlear Implant Users. *Ear and Hearing*, *44*(5), 1107–1120. https://doi.org/10.1097/aud.0000000000001357

Bernard, M. W., Koohi, N., & Bamiou, D.-E. (2024). Auditory processing disorder: An online survey of hearing healthcare professionals' knowledge and practices. *International Journal of Audiology*, *0*(0), 1–10. https://doi.org/10.1080/14992027.2024.2321155

Besser, J., Festen, J. M., Goverts, S. T., Kramer, S. E., & Pichora-Fuller, M. K. (2015). Speech-in-speech listening on the LiSN-S test by older adults with good audiograms depends on cognition and hearing acuity at high

frequencies. *Ear and Hearing*, *36*(1), 24–41. https://doi.org/10.1097/AUD.0000000000000096
Bhatt, I. S., & Wang, J. (2019). Evaluation of dichotic listening performance in normal-hearing, noise-exposed young females. *Hearing Research*, *380*, 10–21. https://doi.org/10.1016/j.heares.2019.05.008
Bidelman, G., & Momtaz, S. (2021). Subcortical rather than cortical sources of the frequency-following response (FFR) relate to speech-in-noise perception in normal-hearing listeners. *Neuroscience Letters*, *746*. https://doi.org/10.1016/j.neulet.2021.135664
Billig, A. J., Lad, M., Sedley, W., & Griffiths, T. D. (2022). The hearing hippocampus. *Progress in Neurobiology*, *218*, 102326. https://doi.org/10.1016/j.pneurobio.2022.102326
Billings, C. J., McMillan, G. P., Penman, T. M., & Gille, S. M. (2013). Predicting Perception in Noise Using Cortical Auditory Evoked Potentials. *JARO: Journal of the Association for Research in Otolaryngology*, *14*(6), 891–903. https://doi.org/10.1007/s10162-013-0415-y
Billings, C. J., Tremblay, K. L., Stecker, G. C., & Tolin, W. M. (2009). Human evoked cortical activity to signal-to-noise ratio and absolute signal level. *Hearing Research*, *254*(1), 15–24. https://doi.org/10.1016/j.heares.2009.04.002
Billings, C., & Madsen, B. (2018). A perspective on brain-behavior relationships and effects of age and hearing using speech-in-noise stimuli. *Hearing Research*, *369*. https://doi.org/10.1016/j.heares.2018.03.024
BinKhamis, G., Elia Forte, A., Reichenbach, T., O'Driscoll, M., & Kluk, K. (2019). Speech Auditory Brainstem Responses in Adult Hearing Aid Users: Effects of Aiding and Background Noise, and Prediction of Behavioral Measures. *Trends in Hearing*, *23*, 2331216519848297. https://doi.org/10.1177/2331216519848297
Binns, C., & Culling, J. F. (2007). The role of fundamental frequency contours in the perception of speech against interfering speech. *The Journal of the Acoustical Society of America*, *122*(3), 1765–1776. https://doi.org/10.1121/1.2751394
Bishop, C. W., & Miller, L. M. (2009). A Multisensory Cortical Network for Understanding Speech in Noise. *Journal of Cognitive Neuroscience*, *21*(9), 1790–1805. https://doi.org/10.1162/jocn.2009.21118
Bishop, D. V. M., & Hardiman, M. J. (2010). Measurement of mismatch negativity in individuals: A study using single-trial analysis. *Psychophysiology*, *47*(4), 697. https://doi.org/10.1111/j.1469-8986.2009.00970.x
Blankenship, C. M., Jareen, M.-D., & Fawen, Z. (2022). Within- and across-frequency temporal processing and speech perception in cochlear implant users. *PloS One*, *17*(10). https://doi.org/10.1371/journal.pone.0275772
Blankenship, C., Zhang, F., & Keith, R. (2016). Behavioral Measures of Temporal Processing and Speech Perception in Cochlear Implant Users. *Journal of the American Academy of Audiology*, *27*(9), 701–713. https://doi.org/10.3766/jaaa.15026
Blazer, D. G., & Tucci, D. L. (2019). Hearing loss and psychiatric disorders: A review. *Psychological Medicine*, *49*(6), 891–897. https://doi.org/10.1017/S0033291718003409
Bochner, J. H., Garrison, W. M., & Doherty, K. A. (2015). The NTID speech recognition test: NSRT(®). *International Journal of Audiology*, *54*(7), 490–498. https://doi.org/10.3109/14992027.2014.991976

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International 5:9/10, 341-345.*
Bollen, K. A., & Pearl, J. (2013). Eight Myths About Causality and Structural Equation Models. In S. L. Morgan (Ed.), *Handbook of Causal Analysis for Social Research* (pp. 301–328). Springer Netherlands. https://doi.org/10.1007/978-94-007-6094-3_15
Bonetti, L., Fernández-Rubio, G., Carlomagno, F., Dietz, M., Pantazis, D., Vuust, P., & Kringelbach, M. L. (2024). Spatiotemporal brain hierarchies of auditory memory recognition and predictive coding. *Nature Communications*, *15*(1), 4313. https://doi.org/10.1038/s41467-024-48302-4
Borch Petersen, E., Lunner, T., Vestergaard, M. D., & Sundewall Thorén, E. (2016). Danish reading span data from 283 hearing-aid users, including a sub-group analysis of their relationship to speech-in-noise performance. *International Journal of Audiology*, *55*(4), 254–261. https://doi.org/10.3109/14992027.2015.1125533
Bramhall, N. (2021). Use of the auditory brainstem response for assessment of cochlear synaptopathy in humansa). *The Journal of the Acoustical Society of America*, *150*(6), 4440–4451. https://doi.org/10.1121/10.0007484
Bramhall, N., Beach, E. F., Epp, B., Le Prell, C. G., Lopez-Poveda, E. A., Plack, C. J., Schaette, R., Verhulst, S., & Canlon, B. (2019). The search for noise-induced cochlear synaptopathy in humans: Mission impossible? *Hearing Research*, *377*, 88–103. https://doi.org/10.1016/j.heares.2019.02.016
Bregman, A. S. (1994). *Auditory Scene Analysis: The Perceptual Organization of Sound*. MIT Press.
Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology*, *89*(2), 244–249. https://doi.org/10.1037/h0031163
Brenowitz, W. D., Filshtein, T. J., Yaffe, K., Walter, S., Ackley, S. F., Hoffmann, T. J., Jorgenson, E., Whitmer, R. A., & Glymour, M. M. (2020). Association of genetic risk for Alzheimer disease and hearing impairment. *Neurology*, *95*(16), e2225–e2234. https://doi.org/10.1212/WNL.0000000000010709
Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ*, *8*, e9414. https://doi.org/10.7717/peerj.9414
Bright, T., & Pallawela, D. (2016). Validated Smartphone-Based Apps for Ear and Hearing Assessments: A Review. *JMIR Rehabilitation and Assistive Technologies*, *3*(2), e13. https://doi.org/10.2196/rehab.6074
British Society of Audiology. (2018). Recommended Procedure Pure Tone Audiometry. *British Society of Audiology*. https://www.thebsa.org.uk/wp-content/uploads/2024/01/Recommended-Procedure-Pure-Tone-Audiometry-2018.pdf
Brodbeck, C., & Simon, J. Z. (2022). Cortical tracking of voice pitch in the presence of multiple speakers depends on selective attention. *Frontiers in Neuroscience*, *16*. https://doi.org/10.3389/fnins.2022.828546
Broderick, M. P., Anderson, A. J., & Lalor, E. C. (2019). Semantic Context Enhances the Early Auditory Encoding of Natural Speech. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *39*(38), 7564–7575. https://doi.org/10.1523/JNEUROSCI.0584-19.2019

Brosseau-Liard, P. E., & Savalei, V. (2014). Adjusting Incremental Fit Indices for Nonnormality. *Multivariate Behavioral Research*, *49*(5), 460–470. https://doi.org/10.1080/00273171.2014.933697

Brosseau-Liard, P. E., Savalei, V., & Li, L. (2012). An Investigation of the Sample Performance of Two Nonnormality Corrections for RMSEA. *Multivariate Behavioral Research*, *47*(6), 904–930. https://doi.org/10.1080/00273171.2012.715252

Browne, M., Koohi, N., & Bamiou, D. E. (2024). Auditory processing disorder: An online survey of hearing healthcare professionals' knowledge and practices. *Int J Audiology, In Press*. https://doi.org/10.1080/14992027.2024.2321155

Browne, M. W., & Cudeck, R. (1992). Alternative Ways of Assessing Model Fit. *Sociological Methods & Research*, *21*(2), 230–258. https://doi.org/10.1177/0049124192021002005

Brownsett, S., & Wise, R. (2010). The contribution of the parietal lobes to speaking and writing. *Cerebral Cortex (New York, N.Y. : 1991)*, *20*(3). https://doi.org/10.1093/cercor/bhp120

Brückmann, M., Pagliarin, K. C., & Garcia, M. V. (2021). Mismatch negativity in older adults and its relationship with the cognitive and behavioral aspects of central auditory processing. *Clinics*, *76*, e1830. https://doi.org/10.6061/clinics/2021/e1830

Burke, D. M., & Mackay, D. G. (1997). Memory, language, and ageing. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *352*(1363), 1845–1856. https://doi.org/10.1098/rstb.1997.0170

Kohrman, D., Wan, G., Cassinotti, L., & Corfas, G. (2020). Hidden Hearing Loss: A Disorder with Multiple Etiologies and Mechanisms. *Cold Spring Harbor Perspectives in Medicine*, *10*(1), a035493. https://doi.org/10.1101/cshperspect.a035493

Calvert, G. A. (2001). Crossmodal Processing in the Human Brain: Insights from Functional Neuroimaging Studies. *Cerebral Cortex*, *11*(12), 1110–1123. https://doi.org/10.1093/cercor/11.12.1110

Cameron, S., & Dillon, H. (2007). Development of the Listening in Spatialized Noise-Sentences Test (LISN-S). *Ear and Hearing*, *28*(2), 196–211. https://doi.org/10.1097/AUD.0b013e318031267f

Carroll, J., & Zeng, F.-G. (2007). Fundamental frequency discrimination and speech perception in noise in cochlear implant simulations. *Hearing Research*, *231*(1–2), 42–53. https://doi.org/10.1016/j.heares.2007.05.004

Casale, J., Kandle, P. F., Murray, I. V., & Murr, N. I. (2024). Physiology, Cochlear Function. In *StatPearls*. StatPearls Publishing. http://www.ncbi.nlm.nih.gov/books/NBK531483/

Ceccato, J.-C., Duran, M.-J., Swanepoel, D. W., Smits, C., De Sousa, K. C., Gledhill, L., Venail, F., & Puel, J.-L. (2021). French Version of the Antiphasic Digits-in-Noise Test for Smartphone Hearing Screening. *Frontiers in Public Health*, *9*, 725080. https://doi.org/10.3389/fpubh.2021.725080

Cesur, S., & Derinsu, U. (2020). Temporal Processing and Speech Perception Performance in Postlingual Adult Users of Cochlear Implants. *Journal of the American Academy of Audiology*, *31*, 129–136. https://doi.org/10.3766/jaaa.19002

Chadha, S., Kamenov, K., & Cieza, A. (2021). The world report on hearing, 2021. *Bulletin of the World Health Organization*, *99*(4). https://doi.org/10.2471/BLT.21.285643

Chaitidis, N., Theocharis, P., Festas, C., & Aritzi, I. (2020). Association of rheumatoid arthritis with hearing loss: A systematic review and meta-analysis. *Rheumatology International*, *40*(11), 1771–1779. https://doi.org/10.1007/s00296-020-04609-1
Chakrabarty, D., & Elhilali, M. (2019). A Gestalt inference model for auditory scene segregation. *PLoS Computational Biology*, *15*(1), e1006711. https://doi.org/10.1371/journal.pcbi.1006711
Chandrasekaran, B., Hornickel, J., Skoe, E., Nicol, T., & Kraus, N. (2009). Context-dependent encoding in the human auditory brainstem relates to hearing speech in noise: Implications for developmental dyslexia. *Neuron*, *64*(3), 311. https://doi.org/10.1016/j.neuron.2009.10.006
Chen, Z. (2005). Stochastic correlative firing for figure-ground segregation. *Biological Cybernetics*, *92*(3), 192–198. https://doi.org/10.1007/s00422-005-0544-4
Cherry, E. C. (1953). Some Experiments on the Recognition of Speech, with One and with Two Ears. *The Journal of the Acoustical Society of America*, *25*(5), 975–979. https://doi.org/10.1121/1.1907229
Cheveigné, A. de. (2010). Pitch perception. In C. J. Plack (Ed.), *Oxford Handbook of Auditory Science: Hearing* (p. 0). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199233557.013.0004
Chierchia, G., Fuhrmann, D., Knoll, L. J., Pi-Sunyer, B. P., Sakhardande, A. L., & Blakemore, S.-J. (2019). The matrix reasoning item bank (MaRs-IB): Novel, open-access abstract reasoning items for adolescents and adults. *Royal Society Open Science*, *6*(10), 190232. https://doi.org/10.1098/rsos.190232
Chintanpalli, A., Jennings, S. G., Heinz, M. G., & Strickland, E. A. (2012). Modeling the Anti-masking Effects of the Olivocochlear Reflex in Auditory Nerve Responses to Tones in Sustained Noise. *Journal of the Association for Research in Otolaryngology*, *13*(2), 219–235. https://doi.org/10.1007/s10162-011-0310-3
Choi, I., Gander, P. E., Berger, J. I., Woo, J., Choy, M. H., Hong, J., Colby, S., McMurray, B., & Griffiths, T. D. (2023). Spectral Grouping of Electrically Encoded Sound Predicts Speech-in-Noise Performance in Cochlear Implantees. *Journal of the Association for Research in Otolaryngology: JARO*, *24*(6), 607–617. https://doi.org/10.1007/s10162-023-00918-x
Choi, J. E., Moon, I. J., Baek, S.-Y., Kim, S. W., & Cho, Y.-S. (2019). Discrepancies between self-reported hearing difficulty and hearing loss diagnosed by audiometry: Prevalence and associated factors in a national survey. *BMJ Open*, *9*(4), e022440. https://doi.org/10.1136/bmjopen-2018-022440
Christensen, H., Mackinnon, A. J., Korten, A., & Jorm, A. F. (2001). The 'common cause hypothesis' of cognitive aging: Evidence for not only a common factor but also specific associations of age with vision and grip strength in a cross-sectional analysis. *Psychology and Aging*, *16*(4), 588–599. https://doi.org/10.1037//0882-7974.16.4.588
Christensen, J. H., Saunders, G. H., Porsbo, M., & Pontoppidan, N. H. (2021). The everyday acoustic environment and its association with human heart rate: Evidence from real-world data logging with hearing aids and wearables. *Royal Society Open Science*, *8*(2), 201345. https://doi.org/10.1098/rsos.201345
Coene, M., Krijger, S., Meeuws, M., De Ceulaer, G., & Govaerts, P. J. (2016). Linguistic Factors Influencing Speech Audiometric Assessment. *BioMed Research International*, *2016*, e7249848. https://doi.org/10.1155/2016/7249848

Coffey, E. B. J., Herholz, S. C., Chepesiuk, A. M. P., Baillet, S., & Zatorre, R. J. (2016). Cortical contributions to the auditory frequency-following response revealed by MEG. *Nature Communications*, *7*(1), Article 1. https://doi.org/10.1038/ncomms11070

Coffey, E. B. J., Nicol, T., White-Schwoch, T., Chandrasekaran, B., Krizman, J., Skoe, E., Zatorre, R. J., & Kraus, N. (2019). Evolving perspectives on the sources of the frequency-following response. *Nature Communications*, *10*(1), 5036. https://doi.org/10.1038/s41467-019-13003-w

Çolak, H., Aydemir, B. E., Sakarya, M. D., Çakmak, E., Alniaçik, A., & Türkyilmaz, M. D. (2024). Subcortical Auditory Processing and Speech Perception in Noise Among Individuals With and Without Extended High-Frequency Hearing Loss. *Journal of Speech, Language, and Hearing Research: JSLHR*, *67*(1), 221–231. https://doi.org/10.1044/2023_JSLHR-23-00023

Colby, S. E., & McMurray, B. (2023). Efficiency of spoken word recognition slows across the adult lifespan. *Cognition*, *240*, 105588. https://doi.org/10.1016/j.cognition.2023.105588

Corona, A. P., Ferrite, S., Bright, T., & Polack, S. (2020). Validity of hearing screening using hearTest smartphone-based audiometry: Performance evaluation of different response modes. *International Journal of Audiology*, *59*(9), 666–673. https://doi.org/10.1080/14992027.2020.1731767

Coslett, H. B., & Schwartz, M. F. (2018). Chapter 18—The parietal lobe and language. In G. Vallar & H. B. Coslett (Eds.), *Handbook of Clinical Neurology* (Vol. 151, pp. 365–375). Elsevier. https://doi.org/10.1016/B978-0-444-63622-5.00018-8

Cox, R. M., & Alexander, G. C. (1995). The abbreviated profile of hearing aid benefit. *Ear and Hearing*, *16*(2), 176–186. https://doi.org/10.1097/00003446-199504000-00005

Crosse, M. J., Di Liberto, G. M., Bednar, A., & Lalor, E. C. (2016). The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli. *Frontiers in Human Neuroscience*, *10*. https://www.frontiersin.org/articles/10.3389/fnhum.2016.00604

Crosse, M. J., Zuk, N. J., Di Liberto, G. M., Nidiffer, A. R., Molholm, S., & Lalor, E. C. (2021). Linear Modeling of Neurophysiological Responses to Speech and Other Continuous Stimuli: Methodological Considerations for Applied Research. *Frontiers in Neuroscience*, *15*. https://doi.org/10.3389/fnins.2021.705621

Cui, M. E., & Herrmann, B. (2023). Eye Movements Decrease during Effortful Speech Listening. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *43*(32), 5856–5869. https://doi.org/10.1523/JNEUROSCI.0240-23.2023

Cusack, R. (2005). The intraparietal sulcus and perceptual organization. *Journal of Cognitive Neuroscience*, *17*(4), 641–651. https://doi.org/10.1162/0898929053467541

Cvijanović, N., Kechichian, P., Janse, K., & Kohlrausch, A. (2017). Effects of noise on arousal in a speech communication setting. *Speech Communication*, *88*, 127–136. https://doi.org/10.1016/j.specom.2017.02.001

Bruns, A. (2021). *Middle Ear Function: Overview, What is Sound?, External Ear*. https://emedicine.medscape.com/article/874456-overview

Davies, K., & Sugano, Y. (2020). *The Auditory Pathway—Structures of the Ear—Auditory Transduction—TeachMeAnatomy*. https://teachmeanatomy.info/neuroanatomy/pathways/auditory-pathway/
Dawes, P. (2011). The SCAN-A in testing for auditory processing disorder in a sample of British adults. *International Journal of Audiology*, *50*(2), 107–111. https://doi.org/10.3109/14992027.2010.527861
Dawes, P., & Bishop, D. V. M. (2007). The SCAN-C in testing for auditory processing disorder in a sample of British children. *International Journal of Audiology*, *46*(12), 780–786. https://doi.org/10.1080/14992020701545906
de Boer, J., Thornton, A. R. D., & Krumbholz, K. (2012). What is the role of the medial olivocochlear system in speech-in-noise processing? *Journal of Neurophysiology*, *107*(5), 1301–1312. https://doi.org/10.1152/jn.00222.2011
De Sousa, K. C., Swanepoel, D. W., Moore, D. R., Myburgh, H. C., & Smits, C. (2020). Improving Sensitivity of the Digits-In-Noise Test Using Antiphasic Stimuli. *Ear and Hearing*, *41*(2), 442–450. https://doi.org/10.1097/AUD.0000000000000775
Decruy, L., Vanthornhout, J., & Francart, T. (2020). Hearing impairment is associated with enhanced neural tracking of the speech envelope. *Hearing Research*, *393*, 107961. https://doi.org/10.1016/j.heares.2020.107961
Delecrode, C. R., Cardoso, A. C. V., Frizzo, A. C. F., & Guida, H. L. (2014). Pitch pattern sequence and duration pattern tests in Brazil: Literature review. *Revista CEFAC*, *16*(1), 283–293. https://doi.org/10.1590/1982-021620143912
DeMetropolis, S. M., Schoepflin, J., & Raphael, L. J. (2021). Relationship of gap detection and speech recognition in noise in younger versus older listeners. *Proceedings of Meetings on Acoustics*, *23*(1), 060009. https://doi.org/10.1121/2.0001449
Deroche, M. L. D., Nguyen, D. L., & Gracco, V. L. (2017). Modulation of Speech Motor Learning with Transcranial Direct Current Stimulation of the Inferior Parietal Lobe. *Frontiers in Integrative Neuroscience*, *11*. https://doi.org/10.3389/fnint.2017.00035
Dias, J., McClaskey, C., Alvey, A., Lawson, A., Matthews, L., Dubno, J. R., & Harris, K. (2024). Effects of age and noise exposure history on auditory nerve response amplitudes: A systematic review, study, and meta-analysis. *Hearing Research*, *447*. https://doi.org/10.1016/j.heares.2024.109010
Dias, J. W., McClaskey, C. M., & Harris, K. C. (2021). Early auditory cortical processing predicts auditory speech in noise identification and lipreading. *Neuropsychologia*, *161*, 108012. https://doi.org/10.1016/j.neuropsychologia.2021.108012
Diedesch, A. C., Bock, S. J. A., & Gallun, F. J. (2021). Clinical Importance of Binaural Information: Extending Auditory Assessment in Clinical Populations Using a Portable Testing Platform. *American Journal of Audiology*, *30*(3), 655–668. https://doi.org/10.1044/2021_AJA-20-00168
Dinçer D'Alessandro, H., Nicastri, M., Portanova, G., Giallini, I., Russo, F. Y., Magliulo, G., Greco, A., & Mancini, P. (2024). Low-frequency pitch coding: Relationships with speech-in-noise and music perception by pediatric populations with typical hearing and cochlear implants. *European Archives of Oto-Rhino-Laryngology*, *281*(7), 3475–3482. https://doi.org/10.1007/s00405-023-08445-4

Ding, N., & Simon, J. Z. (2012a). Emergence of neural encoding of auditory objects while listening to competing speakers. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(29), 11854–11859. JSTOR.
Ding, N., & Simon, J. Z. (2012b). Neural coding of continuous speech in auditory cortex during monaural and dichotic listening. *Journal of Neurophysiology*, *107*(1), 78–89. https://doi.org/10.1152/jn.00297.2011
Ding, N., & Simon, J. Z. (2013). Adaptive Temporal Encoding Leads to a Background-Insensitive Cortical Representation of Speech. *Journal of Neuroscience*, *33*(13), 5728–5735. https://doi.org/10.1523/JNEUROSCI.5297-12.2013
Ding, N., & Simon, J. Z. (2014). Cortical entrainment to continuous speech: Functional roles and interpretations. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00311
Ding, T., Yan, A., & Liu, K. (2019). What is noise-induced hearing loss? *British Journal of Hospital Medicine (London, England: 2005)*, *80*(9), 525–529. https://doi.org/10.12968/hmed.2019.80.9.525
Drennan, W. R. (2022). Identifying Subclinical Hearing Loss: Extended Audiometry and Word Recognition in Noise. *Audiology & Neuro-Otology*, *27*(3), 217–226. https://doi.org/10.1159/000518962
Dryden, A., Allen, H. A., Henshaw, H., & Heinrich, A. (2017). The Association Between Cognitive Performance and Speech-in-Noise Perception for Adult Listeners: A Systematic Literature Review and Meta-Analysis. *Trends in Hearing*, *21*, 233121651774467. https://doi.org/10.1177/2331216517744675
Du, X., Xu, H., Song, P., Zhai, Y., Ye, H., Bao, X., Huang, Q., Tanigawa, H., Tu, Z., Chen, P., Zhao, X., Rauschecker, J. P., & Yu, X. (2024). Beyond Auditory Relay: Dissecting the Inferior Colliculus's Role in Sensory Prediction, Cognitive Decision-Making, and Reward Prediction. *eLife*, *13*. https://doi.org/10.7554/eLife.101142.1
Dunn, C. C., Walker, E. A., Oleson, J., Kenworthy, M., Van Voorst, T., Tomblin, J. B., Ji, H., Kirk, K. I., McMurray, B., Hanson, M., & Gantz, B. J. (2014). Longitudinal speech perception and language performance in pediatric cochlear implant users: The effect of age at implantation. *Ear and Hearing*, *35*(2), 148–160. https://doi.org/10.1097/AUD.0b013e3182a4a8f0
Dykiert, D., & Deary, I. J. (2013). Retrospective validation of WTAR and NART scores as estimators of prior cognitive ability using the Lothian Birth Cohort 1936. *Psychological Assessment*, *25*(4), 1361–1366. https://doi.org/10.1037/a0033623
Eckert, M. A., Matthews, L. J., & Dubno, J. R. (2017). Self-Assessed Hearing Handicap in Older Adults With Poorer-Than-Predicted Speech Recognition in Noise. *Journal of Speech, Language, and Hearing Research: JSLHR*, *60*(1), 251–262. https://doi.org/10.1044/2016_JSLHR-H-16-0011
Eggerrmont, J. (2015). Temporal modulation transfer functions. In J. J. Eggermont (Ed.), *Auditory Temporal Processing and its Disorders* (p. 0). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198719090.003.0005
Ehret, G. (2009). Tonotopic Organization (Maps). In M. D. Binder, N. Hirokawa, & U. Windhorst (Eds.), *Encyclopedia of Neuroscience* (pp. 4083–4088). Springer. https://doi.org/10.1007/978-3-540-29678-2_6042
Eichenbaum, H. (2017). The role of the hippocampus in navigation is memory. *Journal of Neurophysiology*, *117*(4), 1785–1796. https://doi.org/10.1152/jn.00005.2017

Elhilali, M., & Shamma, S. A. (2008). A cocktail party with a cortical twist: How cortical mechanisms contribute to sound segregation. *The Journal of the Acoustical Society of America*, *124*(6), 3751–3771. https://doi.org/10.1121/1.3001672

Ensink, Camp, & Cremers. (1998). Mitochondrial inherited hearing loss. *Clinical Otolaryngology & Allied Sciences*, *23*(1), 3–8. https://doi.org/10.1046/j.1365-2273.1998.00089.x

Ershaid, H., Lizarazu, M., McLaughlin, D., Cooke, M., Simantiraki, O., Koutsogiannaki, M., & Lallier, M. (2024). Contributions of listening effort and intelligibility to cortical tracking of speech in adverse listening conditions. *Cortex*, *172*, 54–71. https://doi.org/10.1016/j.cortex.2023.11.018

Ertürk, P., Aslan, F., & Türkyılmaz, M. D. (2023). Listening to speech-in-noise with hearing aids: Do the self-reported outcomes reflect the behavioral speech perception task performance? *European Archives of Oto-Rhino-Laryngology*. https://doi.org/10.1007/s00405-023-08193-5

Etard, O., & Reichenbach, T. (2019). Neural Speech Tracking in the Theta and in the Delta Frequency Band Differentially Encode Clarity and Comprehension of Speech in Noise. *The Journal of Neuroscience*, *39*, 1828–18. https://doi.org/10.1523/JNEUROSCI.1828-18.2019

Etymotic Research. (2005). *BKB-SINTM Speech-in-Noise Test Version 1.03. Etymotic Research, Inc., Elk Grove Village, IL*.

Fancello, V., Fancello, G., Palma, S., Monzani, D., Genovese, E., Bianchini, C., & Ciorba, A. (2023). The Role of Primary Mitochondrial Disorders in Hearing Impairment: An Overview. *Medicina*, *59*(3), 608. https://doi.org/10.3390/medicina59030608

Felix, R. A., Gourévitch, B., & Portfors, C. V. (2018). Subcortical pathways: Towards a better understanding of auditory disorders. *Hearing Research*, *362*, 48–60. https://doi.org/10.1016/j.heares.2018.01.008

Feng, Y., Yin, S., Kiefte, M., & Wang, J. (2010). Temporal Resolution in Regions of Normal Hearing and Speech Perception in Noise for Adults with Sloping High-Frequency Hearing Loss. *Ear & Hearing*, *31*(1), 115–125. https://doi.org/10.1097/AUD.0b013e3181bb69be

Fine, I., & Park, J.-M. (2018). Blindness and Human Brain Plasticity. *Annual Review of Vision Science*, *4*, 337–356. https://doi.org/10.1146/annurev-vision-102016-061241

Fishman, Y. I., Reser, D. H., Arezzo, J. C., & Steinschneider, M. (2001). Neural correlates of auditory stream segregation in primary auditory cortex of the awake monkey. *Hearing Research*, *151*(1–2), 167–187. https://doi.org/10.1016/s0378-5955(00)00224-0

Fishman, Y. I., Steinschneider, M., & Micheyl, C. (2014). Neural representation of concurrent harmonic sounds in monkey primary auditory cortex: Implications for models of auditory scene analysis. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *34*(37), 12425–12443. https://doi.org/10.1523/JNEUROSCI.0025-14.2014

Fleming, S. M., van der Putten, E. J., & Daw, N. D. (2018). Neural mediators of changes of mind about perceptual decisions. *Nature Neuroscience*, *21*(4), 617–624. https://doi.org/10.1038/s41593-018-0104-6

Ford, A. H., Hankey, G. J., Yeap, B. B., Golledge, J., Flicker, L., & Almeida, O. P. (2018). Hearing loss and the risk of dementia in later life. *Maturitas*, *112*, 1–11. https://doi.org/10.1016/j.maturitas.2018.03.004

Friedman, N. P., & Robbins, T. W. (2022). The role of prefrontal cortex in cognitive control and executive function. *Neuropsychopharmacology*, *47*(1), 72–89. https://doi.org/10.1038/s41386-021-01132-0
Fuchs, P. A., & Lauer, A. M. (2019). Efferent Inhibition of the Cochlea. *Cold Spring Harbor Perspectives in Medicine*, *9*(5), a033530. https://doi.org/10.1101/cshperspect.a033530
Füllgrabe, C., Moore, B. C. J., & Stone, M. A. (2015). Age-group differences in speech identification despite matched audiometrically normal hearing: Contributions from auditory temporal processing and cognition. *Frontiers in Aging Neuroscience*, *6*. https://doi.org/10.3389/fnagi.2014.00347
Füllgrabe, C., & Rosen, S. (2016). On The (Un)importance of Working Memory in Speech-in-Noise Processing for Listeners with Normal Hearing Thresholds. *Frontiers in Psychology*, *7*, 1268. https://doi.org/10.3389/fpsyg.2016.01268
Furman, A. C., Kujawa, S. G., & Liberman, M. C. (2013). Noise-induced cochlear neuropathy is selective for fibers with low spontaneous rates. *Journal of Neurophysiology*, *110*(3), 577–586. https://doi.org/10.1152/jn.00164.2013
Gafoor, S. A., & Uppunda, A. K. (2023). Speech Perception in Noise and Medial Olivocochlear Reflex: Effects of Age, Speech Stimulus, and Response-Related Variables. *Journal of the Association for Research in Otolaryngology*, *24*(6), 619–631. https://doi.org/10.1007/s10162-023-00919-w
Gafoor, S. A., & Uppunda, A. K. (2024). Role of the medial olivocochlear efferent auditory system in speech perception in noise: A systematic review and meta-analyses. *International Journal of Audiology*, *63*(8), 561–569. https://doi.org/10.1080/14992027.2023.2260951
Gatehouse, S. (2000). The Glasgow Hearing Aid Benefit Profile: What it measures and how to use it. *The Hearing Journal*, *53*(3), 10.
Gatehouse, S., & Noble, W. (2004). The Speech, Spatial and Qualities of Hearing Scale (SSQ). *International Journal of Audiology*, *43*(2), 85–99.
Geller, J., Holmes, A., Schwalje, A., Berger, J. I., Gander, P. E., Choi, I., & McMurray, B. (2021). Validation of the Iowa Test of Consonant Perception. *The Journal of the Acoustical Society of America*, *150*(3), 2131–2153. https://doi.org/10.1121/10.0006246
George, E. L. J., Festen, J. M., & Houtgast, T. (2006). Factors affecting masking release for speech in modulated noise for normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *120*(4), 2295–2311. https://doi.org/10.1121/1.2266530
George, E. L. J., Zekveld, A. A., Kramer, S. E., Goverts, S. T., Festen, J. M., & Houtgast, T. (2007). Auditory and nonauditory factors affecting speech reception in noise by older listeners. *The Journal of the Acoustical Society of America*, *121*(4), 2362–2375. https://doi.org/10.1121/1.2642072
Ghani, U., Signal, N., Niazi, I. K., & Taylor, D. (2020). ERP based measures of cognitive workload: A review. *Neuroscience & Biobehavioral Reviews*, *118*, 18–26. https://doi.org/10.1016/j.neubiorev.2020.07.020
Gilmore, G. C., Spinks, R. A., & Thomas, C. W. (2006). Age effects in coding tasks: Componential analysis and test of the sensory deficit hypothesis. *Psychology and Aging*, *21*(1), 7–18. https://doi.org/10.1037/0882-7974.21.1.7
Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511–517. https://doi.org/10.1038/nn.3063

Gnanateja, G. N., Rupp, K., Llanos, F., Remick, M., Pernia, M., Sadagopan, S., Teichert, T., Abel, T. J., & Chandrasekaran, B. (2021). Frequency-Following Responses to Speech Sounds Are Highly Conserved across Species and Contain Cortical Contributions. *eNeuro*, *8*(6), ENEURO.0451. https://doi.org/10.1523/ENEURO.0451-21.2021

Gohari, N., Dastgerdi, Z. H., Rouhbakhsh, N., Afshar, S., & Mobini, R. (2023). Training Programs for Improving Speech Perception in Noise: A Review. *Journal of Audiology & Otology*, *27*(1), 1–9. https://doi.org/10.7874/jao.2022.00283

Gordon, -Salant Sandra, & Fitzgibbons, P. J. (1997). Selected Cognitive Factors and Speech Recognition Performance Among Young and Elderly Listeners. *Journal of Speech, Language, and Hearing Research*, *40*(2), 423–431. https://doi.org/10.1044/jslhr.4002.423

GrÉGoire, J., & Van Der Linden, M. (1997). Effect of age on forward and backward digit spans. *Aging, Neuropsychology, and Cognition*, *4*(2), 140–149. https://doi.org/10.1080/13825589708256642

Griffiths, T. D., Lad, M., Kumar, S., Holmes, E., McMurray, B., Maguire, E. A., Billig, A. J., & Sedley, W. (2020). How Can Hearing Loss Cause Dementia? *Neuron*, *108*(3), 401–412. https://doi.org/10.1016/j.neuron.2020.08.003

Guo, X., Benzaquén, E., Holmes, E., Berger, J. I., Brühl, I., Sedley, W., Rushton, S., & Griffiths, T. (2024). *Predicting speech-in-noise ability with static and dynamic auditory figure-ground analysis using structural equation modelling* (p. 2024.09.08.611859). bioRxiv. https://doi.org/10.1101/2024.09.08.611859

Guo, X., Benzaquén, E., Holmes, E., Choi, I., McMurray, B., Bamiou, D.-E., Berger, J. I., & Griffiths, T. D. (2024). *British Version of the Iowa Test of Consonant Perception* (p. 2024.09.04.611204). bioRxiv. https://doi.org/10.1101/2024.09.04.611204

Guo, X., Dheerendra, P., Benzaquén, E., Sedley, W., & Griffiths, T. D. (2022). *EEG Responses to Auditory Figure-Ground Perception* (p. 2022.03.03.482346). bioRxiv. https://doi.org/10.1101/2022.03.03.482346

Gutschalk, A., & Dykstra, A. R. (2014). Functional imaging of auditory scene analysis. *Hearing Research*, *307*, 98–110. https://doi.org/10.1016/j.heares.2013.08.003

Gutschalk, A., Micheyl, C., Melcher, J. R., Rupp, A., Scherg, M., & Oxenham, A. J. (2005). Neuromagnetic Correlates of Streaming in Human Auditory Cortex. *The Journal of Neuroscience*, *25*(22), 5382–5388. https://doi.org/10.1523/JNEUROSCI.0347-05.2005

Hain, T. C. (2007). Chapter 12 - Cranial Nerve VIII: Vestibulocochlear System. In C. G. Goetz (Ed.), *Textbook of Clinical Neurology (Third Edition)* (pp. 199–215). W.B. Saunders. https://doi.org/10.1016/B978-141603618-0.10012-8

Hayduk, L. A., & Littvay, L. (2012). Should researchers use single indicators, best indicators, or multiple indicators in structural equation models? *BMC Medical Research Methodology*, *12*(1), 159. https://doi.org/10.1186/1471-2288-12-159

Heeke, P., Vermiglio, A. J., Bulla, E., Velappan, K., & Fang, X. (2018). The Relationship between Random Gap Detection and Hearing in Noise Test Performances. *Journal of the American Academy of Audiology*, *29*(10), 948–954. https://doi.org/10.3766/jaaa.18005

Heinrich, A., Alain, C., & Schneider, B. A. (2004). Within- and between-channel gap detection in the human auditory cortex. *Neuroreport*, *15*(13), 2051–2056. https://doi.org/10.1097/00001756-200409150-00011

Hennessy, S., Mack, W. J., & Habibi, A. (2022). Speech-in-noise perception in musicians and non-musicians: A multi-level meta-analysis. *Hearing Research*, *416*, 108442. https://doi.org/10.1016/j.heares.2022.108442
Hoben, R., Easow, G., Pevzner, S., & Parker, M. A. (2017). Outer Hair Cell and Auditory Nerve Function in Speech Recognition in Quiet and in Background Noise. *Frontiers in Neuroscience*, *0*. https://doi.org/10.3389/fnins.2017.00157
Hoff, M., Skoog, J., Bodin, T. H., Tengstrand, T., Rosenhall, U., Skoog, I., & Sadeghi, A. (2023). Hearing Loss and Cognitive Function in Early Old Age: Comparing Subjective and Objective Hearing Measures. *Gerontology*, *69*(6), 694–705. https://doi.org/10.1159/000527930
Holmes, E., & Griffiths, T. D. (2019). 'Normal' hearing thresholds and fundamental auditory grouping processes predict difficulties with speech-in-noise perception. *Scientific Reports*, *9*(1), Article 1. https://doi.org/10.1038/s41598-019-53353-5
Holmes, E., To, G., & Johnsrude, I. S. (2021). How Long Does It Take for a Voice to Become Familiar? Speech Intelligibility and Voice Recognition Are Differentially Sensitive to Voice Training. *Psychological Science*, *32*(6), 903–915. https://doi.org/10.1177/0956797621991137
Holmes, E., Zeidman, P., Friston, K. J., & Griffiths, T. D. (2021). Difficulties with Speech-in-Noise Perception Related to Fundamental Grouping Processes in Auditory Cortex. *Cerebral Cortex (New York, N.Y.: 1991)*, *31*(3), 1582–1596. https://doi.org/10.1093/cercor/bhaa311
Hong, R. S., & Turner, C. W. (2006). Pure-tone auditory stream segregation and speech perception in noise in cochlear implant recipients. *The Journal of the Acoustical Society of America*, *120*(1), 360–374. https://doi.org/10.1121/1.2204450
Hoover, E., Pasquesi, L., & Souza, P. (2015). Comparison of clinical and traditional gap detection tests. *Journal of the American Academy of Audiology*, *26*(6), 540–546. https://doi.org/10.3766/jaaa.14088
Hornickel, J., Skoe, E., Nicol, T., Zecker, S., & Kraus, N. (2009). Subcortical differentiation of stop consonants relates to reading and speech-in-noise perception. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(31), 13022. https://doi.org/10.1073/pnas.0901123106
Horton, C., D'Zmura, M., & Srinivasan, R. (2013). Suppression of competing speech through entrainment of cortical oscillations. *Journal of Neurophysiology*, *109*(12), 3082–3093. https://doi.org/10.1152/jn.01026.2012
Howarth, A., & Shone, G. R. (2006). Ageing and the auditory system. *Postgraduate Medical Journal*, *82*(965), 166–171. https://doi.org/10.1136/pgmj.2005.039388
Hribar, M., Suput, D., Carvalho, A. A., Battelino, S., & Vovk, A. (2014). Structural alterations of brain grey and white matter in early deaf adults. *Hearing Research*, *318*, 1–10. https://doi.org/10.1016/j.heares.2014.09.008
Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118
Huang, Y. T., Newman, R. S., Catalano, A., & Goupell, M. J. (2017). Using prosody to infer discourse prominence in cochlear-implant users and normal-hearing listeners. *Cognition*, *166*, 184–200. https://doi.org/10.1016/j.cognition.2017.05.029

Iliadou, V. V., Weihing, J., Chermak, G. D., & Bamiou, D. E. (2018). Otoacoustic emission suppression in children diagnosed with central auditory processing disorder and speech in noise perception deficits. *International Journal of Pediatric Otorhinolaryngology*, *111*, 39–46. https://doi.org/10.1016/j.ijporl.2018.05.027

Ingvalson, E. M., Dhar, S., Wong, P. C. M., & Liu, H. (2015). Working memory training to improve speech perception in noise across languages. *The Journal of the Acoustical Society of America*, *137*(6), 3477–3486. https://doi.org/10.1121/1.4921601

Ingvalson, E. M., Lansford, K. L., Fedorova, V., & Fernandez, G. (2017). Cognitive factors as predictors of accented speech perception for younger and older adults. *The Journal of the Acoustical Society of America*, *141*(6), 4652–4659. https://doi.org/10.1121/1.4986930

Irace, A. L., Sharma, R. K., Reed, N. S., & Golub, J. S. (2021). Smartphone-Based Applications to Detect Hearing Loss: A Review of Current Technology. *Journal of the American Geriatrics Society*, *69*(2), 307–316. https://doi.org/10.1111/jgs.16985

Irwin, R. J., & McAuley, S. F. (1987). Relations among temporal acuity, hearing loss, and the perception of speech distorted by noise and reverberation. *The Journal of the Acoustical Society of America*, *81*(5), 1557–1565. https://doi.org/10.1121/1.394508

James Lind Alliance. (2024, February 26). https://www.jla.nihr.ac.uk/priority-setting-partnerships/mild-to-moderate-hearing-loss/top-10-priorities/

Jansen, S., Luts, H., Dejonckere, P., van Wieringen, A., & Wouters, J. (2014). Exploring the sensitivity of speech-in-noise tests for noise-induced hearing loss. *International Journal of Audiology*, *53*(3), 199–205. https://doi.org/10.3109/14992027.2013.849361

Jansen, S., Luts, H., Wagener, K. C., Frachet, B., & Wouters, J. (2010). The French digit triplet test: A hearing screening tool for speech intelligibility in noise. *International Journal of Audiology*, *49*(5), 378–387. https://doi.org/10.3109/14992020903431272

Jasmin, R., Alina, S., & Tobias, R. (2024). No Evidence of Musical Training Influencing the Cortical Contribution to the Speech-Frequency-Following Response and Its Modulation through Selective Attention. *eNeuro*, *11*(9). https://doi.org/10.1523/ENEURO.0127-24.2024

Kaandorp, Smits, P, M., St, G., & Jm, F. (2015). Assessing speech recognition abilities with digits in noise in cochlear implant and hearing aid users. *International Journal of Audiology*, *54*(1). https://doi.org/10.3109/14992027.2014.945623

Kegler, M., Weissbart, H., & Reichenbach, T. (2022). The neural response at the fundamental frequency of speech is modulated by word-level acoustic and linguistic information. *Frontiers in Neuroscience*, *16*, 915744. https://doi.org/10.3389/fnins.2022.915744

Keith, R. W. (2000). Random gap detection test. *Auditec, St Louis (MO)*, *13*.

Kelly, D. P. (2009). Chapter 70—HEARING IMPAIRMENT. In W. B. Carey, A. C. Crocker, W. L. Coleman, E. R. Elias, & H. M. Feldman (Eds.), *Developmental-Behavioral Pediatrics (Fourth Edition)* (pp. 687–697). W.B. Saunders. https://doi.org/10.1016/B978-1-4160-3370-7.00070-5

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The Performance of RMSEA in Models With Small Degrees of Freedom. *Sociological Methods & Research*, *44*(3), 486–507. https://doi.org/10.1177/0049124114543236
Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., & Banerjee, S. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America*, *116*(4 Pt 1), 2395–2405. https://doi.org/10.1121/1.1784440
Kim, J.-R. (2015). Acoustic Change Complex: Clinical Implications. *Journal of Audiology & Otology*, *19*(3), 120. https://doi.org/10.7874/jao.2015.19.3.120
Kılıç, S., Sendesen, E., Aslan, F., Erbil, N., Aydın, Ö., & Türkyılmaz, D. (2024). Investigating Sensitivity to Auditory Cognition in Listening Effort Assessments: A Simultaneous EEG and Pupillometry Study. *Brain and Behavior*, *14*(11), e70135. https://doi.org/10.1002/brb3.70135
Klem, M., Melby-Lervåg, M., Hagtvet, B., Lyster, S.-A. H., Gustafsson, J.-E., & Hulme, C. (2015). Sentence repetition is a measure of children's language skills rather than working memory limitations. *Developmental Science*, *18*(1), 146–154. https://doi.org/10.1111/desc.12202
Kline, R. B. (2015). *Principles and Practice of Structural Equation Modeling, Fourth Edition: Fourth Edition* (4th edition). Guilford Press.
Koelewijn, T., Zekveld, A. A., Festen, J. M., & Kramer, S. E. (2012). Pupil Dilation Uncovers Extra Listening Effort in the Presence of a Single-Talker Masker. *Ear and Hearing*, *33*(2), 291. https://doi.org/10.1097/AUD.0b013e3182310019
Koerner, T. K., Zhang, Y., Nelson, P. B., Wang, B., & Zou, H. (2016). Neural indices of phonemic discrimination and sentence-level speech intelligibility in quiet and noise: A mismatch negativity study. *Hearing Research*, *339*, 40–49. https://doi.org/10.1016/j.heares.2016.06.001
Köhler, W. (1967). Gestalt psychology. *Psychologische Forschung*, *31*(1), XVIII–XXX. https://doi.org/10.1007/BF00422382
Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012
Kösem, A., Dai, B., McQueen, J. M., & Hagoort, P. (2023). Neural tracking of speech envelope does not unequivocally reflect intelligibility. *NeuroImage*, *272*, 120040. https://doi.org/10.1016/j.neuroimage.2023.120040
Kozou, H., Kujala, T., Shtyrov, Y., Toppila, E., Starck, J., Alku, P., & Näätänen, R. (2005). The effect of different noise types on the speech and non-speech elicited mismatch negativity. *Hearing Research*, *199*(1), 31–39. https://doi.org/10.1016/j.heares.2004.07.010
Kragel, J. E., Schuele, S., VanHaerents, S., Rosenow, J. M., & Voss, J. L. (2021). Rapid coordination of effective learning by the human hippocampus. *Science Advances*, *7*(25), eabf7144. https://doi.org/10.1126/sciadv.abf7144
Kujawa, S. G., & Liberman, M. C. (2009). Adding insult to injury: Cochlear nerve degeneration after 'temporary' noise-induced hearing loss. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *29*(45), 14077–14085. https://doi.org/10.1523/JNEUROSCI.2845-09.2009
Kulasingham, J. P., Joshi, N. H., Rezaeizadeh, M., & Simon, J. Z. (2021). Cortical Processing of Arithmetic and Simple Sentences in an Auditory Attention Task. *Journal of Neuroscience*, *41*(38), 8023–8039. https://doi.org/10.1523/JNEUROSCI.0269-21.2021

Kumar, S., Gander, P. E., Berger, J. I., Billig, A. J., Nourski, K. V., Oya, H., Kawasaki, H., Howard, M. A., & Griffiths, T. D. (2021). Oscillatory correlates of auditory working memory examined with human electrocorticography. *Neuropsychologia*, *150*, 107691. https://doi.org/10.1016/j.neuropsychologia.2020.107691

Kumar, S., Joseph, S., Gander, P. E., Barascud, N., Halpern, A. R., & Griffiths, T. D. (2016). A Brain System for Auditory Working Memory. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *36*(16), 4492–4505. https://doi.org/10.1523/JNEUROSCI.4341-14.2016

Kwak, C., & Han, W. (2020). Towards Size of Scene in Auditory Scene Analysis: A Systematic Review. *Journal of Audiology & Otology*, *24*(1), 1–9. https://doi.org/10.7874/jao.2019.00248

Lad, M., Billig, A. J., Kumar, S., & Griffiths, T. D. (2022). A specific relationship between musical sophistication and auditory working memory. *Scientific Reports*, *12*(1), 3517. https://doi.org/10.1038/s41598-022-07568-8

Lad, M., Holmes, E., Chu, A., & Griffiths, T. D. (2020a). Speech-in-noise detection is related to auditory working memory precision for frequency. *Scientific Reports*, *10*(1), 13997. https://doi.org/10.1038/s41598-020-70952-9

Lad, M., Holmes, E., Chu, A., & Griffiths, T. D. (2020b). Speech-in-noise detection is related to auditory working memory precision for frequency. *Scientific Reports*, *10*(1), Article 1. https://doi.org/10.1038/s41598-020-70952-9

Lad, M., Taylor, J.-P., & Griffiths, T. D. (2024). The contribution of short-term memory for sound features to speech-in-noise perception and cognition. *Hearing Research*, *451*. https://doi.org/10.1016/j.heares.2024.109081

Lalor, E. C., Pearlmutter, B. A., Reilly, R. B., McDarby, G., & Foxe, J. J. (2006). The VESPA: A method for the rapid estimation of a visual evoked potential. *NeuroImage*, *32*(4), 1549–1561. https://doi.org/10.1016/j.neuroimage.2006.05.054

Lautenschlager, L., Tochetto, T., & Costa, M. J. (2015). Recognition of speech in noise and relations with suppression of otoacoustic emissions and the acoustic reflex. *Brazilian Journal of Otorhinolaryngology*, *77*(1), 115. https://doi.org/10.1590/S1808-86942011000100019

Lazard, D. S., Innes-Brown, H., & Barone, P. (2014). Adaptation of the communicative brain to post-lingual deafness. Evidence from functional imaging. *Hearing Research*, *307*, 136–143. https://doi.org/10.1016/j.heares.2013.08.006

Lehiste, I., & Peterson, G. E. (1959). Linguistic Considerations in the Study of Speech Intelligibility. *The Journal of the Acoustical Society of America*, *31*(3), 280–286. https://doi.org/10.1121/1.1907713

Liberman, M. C. (2017). Noise-induced and age-related hearing loss: New perspectives and potential therapies. *F1000Research*, *6*, 927. https://doi.org/10.12688/f1000research.11310.1

Liberman, M. C., & Kujawa, S. G. (2017). Cochlear synaptopathy in acquired sensorineural hearing loss: Manifestations and mechanisms. *Hearing Research*, *349*, 138–147. https://doi.org/10.1016/j.heares.2017.01.003

Lichtenhan, J. T., Wilson, U. S., Hancock, K. E., & JJ Guinan, J. (2015). Medial Olivocochlear Efferent Reflex Inhibition of Human Cochlear Nerve Responses. *Hearing Research*, *333*, 216. https://doi.org/10.1016/j.heares.2015.09.001

Lin, F. R., Metter, E. J., O'Brien, R. J., Zonderman, A. B., & Ferrucci, L. (2012). *Hearing Loss and Incident Dementia*. 14.

Lin, F. R., Pike, J. R., Albert, M. S., Arnold, M., Burgard, S., Chisolm, T., Couper, D., Deal, J. A., Goman, A. M., Glynn, N. W., Gmelin, T., Gravens-Mueller, L., Hayden, K. M., Huang, A. R., Knopman, D., Mitchell, C. M., Mosley, T., Pankow, J. S., Reed, N. S., … Coresh, J. (2023). Hearing intervention versus health education control to reduce cognitive decline in older adults with hearing loss in the USA (ACHIEVE): A multicentre, randomised controlled trial. *The Lancet*, *402*(10404), 786–797. https://doi.org/10.1016/S0140-6736(23)01406-X

Lin, F. R., Yaffe, K., Xia, J., Xue, Q.-L., Harris, T. B., Purchase-Helzner, E., Satterfield, S., Ayonayon, H. N., Ferrucci, L., Simonsick, E. M., & Health ABC Study Group, for the. (2013). Hearing Loss and Cognitive Decline in Older Adults. *JAMA Internal Medicine*, *173*(4), 293. https://doi.org/10.1001/jamainternmed.2013.1868

Lin, H. W., Furman, A. C., Kujawa, S. G., & Liberman, M. C. (2011). Primary Neural Degeneration in the Guinea Pig Cochlea After Reversible Noise-Induced Threshold Shift. *Journal of the Association for Research in Otolaryngology*, *12*(5), 605–616. https://doi.org/10.1007/s10162-011-0277-0

Lindenberger, U., & Baltes, P. B. (1994). Sensory functioning and intelligence in old age: A strong connection. *Psychology and Aging*, *9*(3), 339–355. https://doi.org/10.1037//0882-7974.9.3.339

Lister, J. J., Maxfield, N. D., & Pitt, G. J. (2007). Cortical evoked response to gaps in noise: Within-channel and across-channel conditions. *Ear and Hearing*, *28*(6), 862–878. https://doi.org/10.1097/AUD.0b013e3181576cba

Lister, J. J., Maxfield, N. D., Pitt, G. J., & Gonzalez, V. B. (2011). Auditory evoked response to gaps in noise: Older adults. *International Journal of Audiology*, *50*(4), 211–225. https://doi.org/10.3109/14992027.2010.526967

Lister, J. J., Roberts, R. A., Shackelford, J., & Rogers, C. L. (2006). An adaptive clinical test of temporal resolution. *American Journal of Audiology*, *15*(2), 133–140. https://doi.org/10.1044/1059-0889(2006/017)

Liu, J., Stohl, J., & Overath, T. (2024). Hidden hearing loss: Fifteen years at a glance. *Hearing Research*, *443*, 108967. https://doi.org/10.1016/j.heares.2024.108967

Liu, J., Tang, W., Chen, G., Lu, Y., Feng, C., & Tu, X. M. (2016). Correlation and agreement: Overview and clarification of competing concepts and measures. *Shanghai Archives of Psychiatry*, *28*(2), 115–120. https://doi.org/10.11919/j.issn.1002-0829.216045

Livingston, G., Huntley, J., Liu, K. Y., Costafreda, S. G., Selbæk, G., Alladi, S., Ames, D., Banerjee, S., Burns, A., Brayne, C., Fox, N. C., Ferri, C. P., Gitlin, L. N., Howard, R., Kales, H. C., Kivimäki, M., Larson, E. B., Nakasujja, N., Rockwood, K., … Mukadam, N. (2024). Dementia prevention, intervention, and care: 2024 report of the Lancet standing Commission. *Lancet (London, England)*, *404*(10452), 572–628. https://doi.org/10.1016/S0140-6736(24)01296-0

Livingston, G., Sommerlad, A., Orgeta, V., Costafreda, S. G., Huntley, J., Ames, D., Ballard, C., Banerjee, S., Burns, A., Cohen-Mansfield, J., Cooper, C., Fox, N., Gitlin, L. N., Howard, R., Kales, H. C., Larson, E. B., Ritchie, K., Rockwood, K., Sampson, E. L., … Mukadam, N. (2017). Dementia prevention, intervention, and care. *The Lancet*, *390*(10113), 2673–2734. https://doi.org/10.1016/S0140-6736(17)31363-6

Llanos, F., German, J. S., Gnanateja, G. N., & Chandrasekaran, B. (2021). The neural processing of pitch accents in continuous speech. *Neuropsychologia*, *158*, 107883. https://doi.org/10.1016/j.neuropsychologia.2021.107883
Loughrey, D. G., Kelly, M. E., Kelley, G. A., Brennan, S., & Lawlor, B. A. (2018). Association of Age-Related Hearing Loss With Cognitive Function, Cognitive Impairment, and Dementia: A Systematic Review and Meta-analysis. *JAMA Otolaryngology-- Head & Neck Surgery*, *144*(2), 115–126. https://doi.org/10.1001/jamaoto.2017.2513
MacCutcheon, D., Füllgrabe, C., Eccles, R., van der Linde, J., Panebianco, C., & Ljung, R. (2020). Investigating the Effect of One Year of Learning to Play a Musical Instrument on Speech-in-Noise Perception and Phonological Short-Term Memory in 5-to-7-Year-Old Children. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02865
Mackersie, C. L., MacPhee, I. X., & Heldt, E. W. (2015). Effects of Hearing Loss on Heart-Rate Variability and Skin Conductance Measured During Sentence Recognition in Noise. *Ear and Hearing*, *36*(1), 145. https://doi.org/10.1097/AUD.0000000000000091
Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin*, *98*(1), 185–199.
Mai, G., & Wang, W. S. -Y. (2023). Distinct roles of delta- and theta-band neural tracking for sharpening and predictive coding of multi-level speech features during spoken language processing. *Human Brain Mapping*, *44*(17), 6149–6172. https://doi.org/10.1002/hbm.26503
Maillard, E., Joyal, M., Murray, M. M., & Tremblay, P. (2023). Are musical activities associated with enhanced speech perception in noise in adults? A systematic review and meta-analysis. *Current Research in Neurobiology*, *4*, 100083. https://doi.org/10.1016/j.crneur.2023.100083
Malmierca, M. S. (2015). Chapter 29—Auditory System. In G. Paxinos (Ed.), *The Rat Nervous System (Fourth Edition)* (pp. 865–946). Academic Press. https://doi.org/10.1016/B978-0-12-374245-2.00029-2
Malmierca, M. S., & Smith, P. H. (2009). Cochlear Nucleus. In M. D. Binder, N. Hirokawa, & U. Windhorst (Eds.), *Encyclopedia of Neuroscience* (pp. 772–778). Springer. https://doi.org/10.1007/978-3-540-29678-2_1097
Mamo, S. K., & Helfer, K. S. (2021). Speech Understanding in Modulated Noise and Speech Maskers as a Function of Cognitive Status in Older Adults. *American Journal of Audiology*, *30*(3), 642–654. https://doi.org/10.1044/2021_AJA-20-00177
Manju, V., Gopika, K. K., & Nambi, P. M. A. (2014). Association of Auditory Steady State Responses with Perception of Temporal Modulations and Speech in Noise. *ISRN Otolaryngology*, *2014*, 374035. https://doi.org/10.1155/2014/374035
Manting, C. L., Gulyas, B., Ullén, F., & Lundqvist, D. (2021). Auditory steady-state responses during and after a stimulus: Cortical sources, and the influence of attention and musicality. *NeuroImage*, *233*, 117962. https://doi.org/10.1016/j.neuroimage.2021.117962
Martin, B., Tremblay, K., & Stapells, D. (2007). *Principles and applications of cortical auditory evoked potentials* (pp. 482–507).
Marzbani, H., Marateb, H. R., & Mansourian, M. (2016). Neurofeedback: A Comprehensive Review on System Design, Methodology and Clinical

Applications. *Basic and Clinical Neuroscience*, *7*(2), 143–158. https://doi.org/10.15412/J.BCN.03070208
Matthews, L. J., Lee, F. S., Mills, J. H., & Schum, D. J. (1990). Audiometric and subjective assessment of hearing handicap. *Archives of Otolaryngology--Head & Neck Surgery*, *116*(11), 1325–1330. https://doi.org/10.1001/archotol.1990.01870110097013
McClaskey, C. M., Dias, J. W., & Harris, K. C. (2019). Sustained envelope periodicity representations are associated with speech-in-noise performance in difficult listening conditions for younger and older adults. *Journal of Neurophysiology*, *122*(4), 1685–1696. https://doi.org/10.1152/jn.00845.2018
McCroskey, R. L., & Keith, R. W. (1997). Auditory Fusion Test – Revised. *Auditec, St Louis (MO)*.
McGuire, K., Firestone, G. M., Zhang, N., & Zhang, F. (2021). The Acoustic Change Complex in Response to Frequency Changes and Its Correlation to Cochlear Implant Speech Outcomes. *Frontiers in Human Neuroscience*, *15*, 757254. https://doi.org/10.3389/fnhum.2021.757254
McKay, C. M. (2021). No Evidence That Music Training Benefits Speech Perception in Hearing-Impaired Listeners: A Systematic Review. *Trends in Hearing*, *25*, 2331216520985678. https://doi.org/10.1177/2331216520985678
McLaughlin, D. J., Baese-Berk, M. M., Bent, T., Borrie, S. A., & Van Engen, K. J. (2018). Coping with adversity: Individual differences in the perception of noisy and accented speech. *Attention, Perception & Psychophysics*, *80*(6), 1559–1570. https://doi.org/10.3758/s13414-018-1537-4
McPherson, M. J., Grace, R. C., & McDermott, J. H. (2022). Harmonicity aids hearing in noise. *Attention, Perception & Psychophysics*, *84*(3), 1016–1042. https://doi.org/10.3758/s13414-021-02376-0
Meha-Bettison, K., Sharma, M., Ibrahim, R. K., & Mandikal Vasuki, P. R. (2018). Enhanced speech perception in noise and cortical auditory evoked potentials in professional musicians. *International Journal of Audiology*, *57*(1), 40–52. https://doi.org/10.1080/14992027.2017.1380850
Mendel, L. L. (2007). Objective and subjective hearing aid assessment outcomes. *American Journal of Audiology*, *16*(2). https://doi.org/10.1044/1059-0889(2007/016)
Mendoza, J. E. (2011). Cochlear Nuclei (Dorsal and Ventral). In J. S. Kreutzer, J. DeLuca, & B. Caplan (Eds.), *Encyclopedia of Clinical Neuropsychology* (pp. 607–608). Springer. https://doi.org/10.1007/978-0-387-79948-3_720
Mepani, A. M., Verhulst, S., Hancock, K. E., Garrett, M., Vasilkov, V., Bennett, K., de Gruttola, V., Liberman, M. C., & Maison, S. F. (2021). Envelope following responses predict speech-in-noise performance in normal-hearing listeners. *Journal of Neurophysiology*, *125*(4), 1213–1222. https://doi.org/10.1152/jn.00620.2020
Merten, N., Boenniger, M. M., Herholz, S. C., & Breteler, M. M. B. (2022). The Associations of Hearing Sensitivity and Different Cognitive Functions with Perception of Speech-in-Noise. *Ear and Hearing*, *43*(3), 984–992. https://doi.org/10.1097/AUD.0000000000001154
Merten, N., Fischer, M. E., Tweed, T. S., Breteler, M. M. B., & Cruickshanks, K. J. (2020). Associations of Hearing Sensitivity, Higher-Order Auditory Processing, and Cognition Over Time in Middle-Aged Adults. *The Journals of Gerontology: Series A*, *75*(3), 545–551. https://doi.org/10.1093/gerona/glz189

Micheyl, C., Ryan, C. M., & Oxenham, A. J. (2012). Further evidence that fundamental-frequency difference limens measure pitch discrimination. *The Journal of the Acoustical Society of America*, *131*(5), 3989–4001. https://doi.org/10.1121/1.3699253

Milne, A. E., Bianco, R., Poole, K. C., Zhao, S., Oxenham, A. J., Billig, A. J., & Chait, M. (2021). An online headphone screening test based on dichotic pitch. *Behavior Research Methods*, *53*(4), 1551–1562. https://doi.org/10.3758/s13428-020-01514-0

Mishra, S. K., & Lutman, M. E. (2014). Top-Down Influences of the Medial Olivocochlear Efferent System in Speech Perception in Noise. *PLOS ONE*, *9*(1), e85756. https://doi.org/10.1371/journal.pone.0085756

Mitchell, B. L., Thorp, J. G., Evans, D. M., Nyholt, D. R., Martin, N. G., & Lupton, M. K. (2020). Exploring the genetic relationship between hearing impairment and Alzheimer's disease. *Alzheimer's & Dementia : Diagnosis, Assessment & Disease Monitoring*, *12*(1), e12108. https://doi.org/10.1002/dad2.12108

Mitchnick, K. A., Ahmad, Z., Mitchnick, S. D., Ryan, J. D., Rosenbaum, R. S., & Freud, E. (2022). Damage to the human dentate gyrus impairs the perceptual discrimination of complex, novel objects. *Neuropsychologia*, *172*, 108238. https://doi.org/10.1016/j.neuropsychologia.2022.108238

Monge, Z. A., & Madden, D. J. (2016). Linking cognitive and visual perceptual decline in healthy aging: The information degradation hypothesis. *Neuroscience & Biobehavioral Reviews*, *69*, 166–173. https://doi.org/10.1016/j.neubiorev.2016.07.031

Moore, D. R., Zobay, O., & Ferguson, M. A. (2020). Minimal and Mild Hearing Loss in Children: Association with Auditory Perception, Cognition, and Communication Problems. *Ear and Hearing*, *41*(4), 720–732. https://doi.org/10.1097/AUD.0000000000000802

Moossavi, A., Mehrkian, S., Gohari, N., Nazari, M. A., Bakhshi, E., & Alain, C. (2021). The effect of harmonic training on speech perception in noise in hearing-impaired children. *International Journal of Pediatric Otorhinolaryngology*, *149*, 110845. https://doi.org/10.1016/j.ijporl.2021.110845

Motlagh Zadeh, L., Silbert, N. H., Swanepoel, D. W., & Moore, D. R. (2021). Improved Sensitivity of Digits-in-Noise Test to High-Frequency Hearing Loss. *Ear and Hearing*, *42*(3), 565–573. https://doi.org/10.1097/AUD.0000000000000956

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PloS One*, *9*(2), e89642. https://doi.org/10.1371/journal.pone.0089642

Muncke, J., Kuruvila, I., & Hoppe, U. (2022). Prediction of Speech Intelligibility by Means of EEG Responses to Sentences in Noise. *Frontiers in Neuroscience*, *16*, 876421. https://doi.org/10.3389/fnins.2022.876421

Murphy, C. F. B., Hashim, E., Dillon, H., & Bamiou, D. E. (2019). British children's performance on the listening in spatialised noise-sentences test (LISN-S). *International Journal of Audiology*, *58*(11), 754–760. https://doi.org/10.1080/14992027.2019.1627592

Musiek, F. E., Shinn, J. B., Jirsa, R., Bamiou, D.-E., Baran, J. A., & Zaida, E. (2005). GIN (Gaps-In-Noise) test performance in subjects with confirmed central auditory nervous system involvement. *Ear and Hearing*, *26*(6), 608–618. https://doi.org/10.1097/01.aud.0000188069.80699.41

Musiek, F. E., Shinn, J., Chermak, G. D., & Bamiou, D.-E. (2017). Perspectives on the Pure-Tone Audiogram. *Journal of the American Academy of Audiology*, *28*(07), 655–671. https://doi.org/10.3766/jaaa.16061

Musiek, F., & Nagle, S. (2018). The Middle Latency Response: A Review of Findings in Various Central Nervous System Lesions. *Journal of the American Academy of Audiology*, *29*(9), 855–867. https://doi.org/10.3766/jaaa.16141

Narne, V. K. (2013). Temporal Processing and Speech Perception in Noise by Listeners with Auditory Neuropathy. *PLOS ONE*, *8*(2), e55995. https://doi.org/10.1371/journal.pone.0055995

Natarajan, N., Batts, S., & Stankovic, K. M. (2023). Noise-Induced Hearing Loss. *Journal of Clinical Medicine*, *12*(6), Article 6. https://doi.org/10.3390/jcm12062347

Newman, C. W., Weinstein, B. E., Jacobson, G. P., & Hug, G. A. (1990). The Hearing Handicap Inventory for Adults: Psychometric adequacy and audiometric correlates. *Ear and Hearing*, *11*(6), 430–433. https://doi.org/10.1097/00003446-199012000-00004

Niebur, E., Hsiao, S. S., & Johnson, K. O. (2002). Synchrony: A neuronal mechanism for attentional selection? *Current Opinion in Neurobiology*, *12*(2), 190–194. https://doi.org/10.1016/s0959-4388(02)00310-0

NIH. (2019, January 28). *Line Up! To Reverse Hearing Loss, New Hair Cells Need to Stand in Formation | Noisy Planet*. https://www.noisyplanet.nidcd.nih.gov/have-you-heard/new-hair-cells-need-to-stand-in-formation-to-reverse-hearing-loss

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *The Journal of the Acoustical Society of America*, *95*(2), 1085–1099. https://doi.org/10.1121/1.408469

Oberfeld, D., & Klöckner-Nowotny, F. (2016). Individual differences in selective attention predict speech identification at a cocktail party. *eLife*, *5*, e16747. https://doi.org/10.7554/eLife.16747

Obleser, J., & Kotz, S. A. (2011). Multiple brain signatures of integration in the comprehension of degraded speech. *NeuroImage*, *55*(2), 713–723. https://doi.org/10.1016/j.neuroimage.2010.12.020

O'Sullivan, J. A., Shamma, S. A., & Lalor, E. C. (2015). Evidence for Neural Computations of Temporal Coherence in an Auditory Scene and Their Enhancement during Active Listening. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *35*(18), 7256–7263. https://doi.org/10.1523/JNEUROSCI.4973-14.2015

Oxenham, A. J. (2008a). Pitch Perception and Auditory Stream Segregation: Implications for Hearing Loss and Cochlear Implants. *Trends in Amplification*, *12*(4), 316–331. https://doi.org/10.1177/1084713808325881

Oxenham, A. J. (2008b). Pitch Perception and Auditory Stream Segregation: Implications for Hearing Loss and Cochlear Implants. *Trends in Amplification*, *12*(4), 316–331. https://doi.org/10.1177/1084713808325881

Oxenham, A. J. (2013). Revisiting place and temporal theories of pitch. *Acoustical Science and Technology / Edited by the Acoustical Society of Japan*, *34*(6), 388–396.

Pan, L., Li, C., Meng, L., Zhang, G., Zou, L., Tian, Y., Chen, S., Sun, Y., Su, D., Zhang, X., Xiong, M., Xiao, T., Xia, D., Hong, Z., & Zhang, Z. (2024). GDF1

ameliorates cognitive impairment induced by hearing loss. *Nature Aging*, *4*(4), 568–583. https://doi.org/10.1038/s43587-024-00592-5
Panela, R. A., Copelli, F., & Herrmann, B. (2024). Reliability and generalizability of neural speech tracking in younger and older adults. *Neurobiology of Aging*, *134*, 165–180. https://doi.org/10.1016/j.neurobiolaging.2023.11.007
Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009). Musician enhancement for speech-in-noise. *Ear and Hearing*, *30*(6), 653–661. https://doi.org/10.1097/AUD.0b013e3181b412e9
Parkkonen, L., Fujiki, N., & Mäkelä, J. P. (2009). Sources of auditory brainstem responses revisited: Contribution by magnetoencephalography. *Human Brain Mapping*, *30*(6), 1772. https://doi.org/10.1002/hbm.20788
Parmar, B. J., Rajasingam, S. L., Bizley, J. K., & Vickers, D. A. (2022). Factors Affecting the Use of Speech Testing in Adult Audiology. *American Journal of Audiology*, *31*(3), 528. https://doi.org/10.1044/2022_AJA-21-00233
Pascual-Marqui, R. D. (2002). Standardized low resolution brain electromagnetic. *Clinical Pharmacology*.
Payne, B. R., Grison, S., Gao, X., Christianson, K., Morrow, D. G., & Stine-Morrow, E. A. L. (2014). Aging and individual differences in binding during sentence understanding: Evidence from temporary and global syntactic attachment ambiguities. *Cognition*, *130*(2), 157–173. https://doi.org/10.1016/j.cognition.2013.10.005
Pedersen, K., & Rosenhall, U. (1991). Correlations Between Self-Assessed Hearing Handicap and Standard Audiometric Tests in Elderly Persons. *Scandinavian Audiology*, *20*(2), 109–116. https://doi.org/10.3109/01050399109070799
Peelle, J. E. (2013). Cortical responses to degraded speech are modulated by linguistic predictions. *The Journal of the Acoustical Society of America*, *133*(5), 3387–3387. https://doi.org/10.1121/1.4805857
Peelle, J. E., & Wingfield, A. (2016). The neural consequences of age-related hearing loss. *Trends in Neurosciences*, *39*(7), 486–497. https://doi.org/10.1016/j.tins.2016.05.001
Peter, M. G., Mårtensson, G., Postma, E. M., Nordin, L. E., Westman, E., Boesveldt, S., & Lundström, J. N. (2020). Morphological changes in secondary, but not primary, sensory cortex in individuals with life-long olfactory sensory deprivation. *NeuroImage*, *218*, 117005. https://doi.org/10.1016/j.neuroimage.2020.117005
Phillips, D. P., & Smith, J. C. (2004). Correlations among within-Channel and between-Channel Auditory Gap-Detection Thresholds in Normal Listeners. *Perception*, *33*(3), 371–378. https://doi.org/10.1068/p5116
Phillips, D. P., Taylor, T. L., Hall, S. E., Carr, M. M., & Mossop, J. E. (1997). Detection of silent intervals between noises activating different perceptual channels: Some properties of ‘central’ auditory gap detection. *The Journal of the Acoustical Society of America*, *101*(6), 3694–3705. https://doi.org/10.1121/1.419376
Pichora-Fuller, M. K., Kramer, S. E., Eckert, M. A., Edwards, B., Hornsby, B. W. Y., Humes, L. E., Lemke, U., Lunner, T., Matthen, M., Mackersie, C. L., Naylor, G., Phillips, N. A., Richter, M., Rudner, M., Sommers, M. S., Tremblay, K. L., & Wingfield, A. (2016). Hearing Impairment and Cognitive Energy: The Framework for Understanding Effortful Listening (FUEL). *Ear and Hearing*, *37 Suppl 1*, 5S-27S. https://doi.org/10.1097/AUD.0000000000000312

Pichora-Fuller, M. K., & MacDonald, E. (2009). Sensory Aging: Hearing. *Encyclopedia of Neuroscience*, 645–650.
Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *NeuroImage*, *198*, 181–197. https://doi.org/10.1016/j.neuroimage.2019.05.026
Polspoel, S., Kramer, S. E., van Dijk, B., & Smits, C. (2022). The Importance of Extended High-Frequency Speech Information in the Recognition of Digits, Words, and Sentences in Quiet and Noise. *Ear and Hearing*, *43*(3), 913–920. https://doi.org/10.1097/AUD.0000000000001142
Potgieter, J.-M., Swanepoel, D. W., Myburgh, H. C., & Smits, C. (2018). The South African English Smartphone Digits-in-Noise Hearing Test: Effect of Age, Hearing Loss, and Speaking Competence. *Ear and Hearing*, *39*(4), 656–663. https://doi.org/10.1097/AUD.0000000000000522
Power, A. J., Foxe, J. J., Forde, E.-J., Reilly, R. B., & Lalor, E. C. (2012). At what time is the cocktail party? A late locus of selective attention to natural speech. *European Journal of Neuroscience*, *35*(9), 1497–1503. https://doi.org/10.1111/j.1460-9568.2012.08060.x
Pronk, M., Lissenberg-Witte, B. I., van der Aa, H. P. A., Comijs, H. C., Smits, C., Lemke, U., Zekveld, A. A., & Kramer, S. E. (2019). Longitudinal Relationships Between Decline in Speech-in-Noise Recognition Ability and Cognitive Functioning: The Longitudinal Aging Study Amsterdam. *Journal of Speech, Language, and Hearing Research: JSLHR*, *62*(4S), 1167–1187. https://doi.org/10.1044/2018_JSLHR-H-ASCC7-18-0120
Purdy, S. C., & Kelly, A. S. (2016). Change in Speech Perception and Auditory Evoked Potentials over Time after Unilateral Cochlear Implantation in Postlingually Deaf Adults. *Seminars in Hearing*, *37*(1), 62. https://doi.org/10.1055/s-0035-1570329
Reynard, P., Lagacé, J., Joly, C.-A., Dodelé, L., Veuillet, E., & Thai-Van, H. (2022). Speech-in-Noise Audiometry in Adults: A Review of the Available Tests for French Speakers. *Audiology and Neurotology*, *27*(3), 185–199. https://doi.org/10.1159/000518968
Roberts, K. L., & Allen, H. A. (2016). Perception and Cognition in the Ageing Brain: A Brief Review of the Short- and Long-Term Links between Perceptual and Cognitive Decline. *Frontiers in Aging Neuroscience*, *8*, 39. https://doi.org/10.3389/fnagi.2016.00039
Rutherford, B. R., Brewster, K., Golub, J. S., Kim, A. H., & Roose, S. P. (2018). Sensation and Psychiatry: Linking Age-Related Hearing Loss to Late-Life Depression and Cognitive Decline. *The American Journal of Psychiatry*, *175*(3), 215–224. https://doi.org/10.1176/appi.ajp.2017.17040423
Rutherford, M. A., von Gersdorff, H., & Goutman, J. D. (2021). Encoding sound in the cochlea: From receptor potential to afferent discharge. *The Journal of Physiology*, *599*(10), 2527–2557. https://doi.org/10.1113/JP279189
Sanchez, V. A., Arnold, M. L., Moore, D. R., Clavier, O., & Abrams, H. B. (2022). Speech-in-noise testing: Innovative applications for pediatric patients, underrepresented populations, fitness for duty, clinical trials, and remote services. *The Journal of the Acoustical Society of America*, *152*(4), 2336–2356. https://doi.org/10.1121/10.0014418
Sanju, H. K., Jain, T., & Kumar, P. (2023). Acoustic Change Complex as a Neurophysiological Tool to Assess Auditory Discrimination Skill: A Review.

International Archives of Otorhinolaryngology, *27*(2), e362–e369. https://doi.org/10.1055/s-0042-1743202
Sarant, J., Harris, D., Busby, P., Maruff, P., Schembri, A., Lemke, U., & Launer, S. (2020). The Effect of Hearing Aid Use on Cognition in Older Adults: Can We Delay Decline or Even Improve Cognitive Function? *Journal of Clinical Medicine*, *9*(1), E254. https://doi.org/10.3390/jcm9010254
Sardone, R., Battista, P., Panza, F., Lozupone, M., Griseta, C., Castellana, F., Capozzo, R., Ruccia, M., Resta, E., Seripa, D., Logroscino, G., & Quaranta, N. (2019). The Age-Related Central Auditory Processing Disorder: Silent Impairment of the Cognitive Ear. *Frontiers in Neuroscience*, *13*, 619. https://doi.org/10.3389/fnins.2019.00619
Schneider, F., Dheerendra, P., Balezeau, F., Ortiz-Rios, M., Kikuchi, Y., Petkov, C. I., Thiele, A., & Griffiths, T. D. (2018). Auditory figure-ground analysis in rostral belt and parabelt of the macaque monkey. *Scientific Reports*, *8*(1), 17948. https://doi.org/10.1038/s41598-018-36903-1
Scinicariello, F., Przybyla, J., Carroll, Y., Eichwald, J., Decker, J., & Breysse, P. N. (2019). Age and sex differences in hearing loss association with depressive symptoms: Analyses of NHANES 2011–2012. *Psychological Medicine*, *49*(6), 962–968. https://doi.org/10.1017/S0033291718001617
Shamma, S. A., Elhilali, M., & Micheyl, C. (2011). Temporal coherence and attention in auditory scene analysis. *Trends in Neurosciences*, *34*(3), 114–123. https://doi.org/10.1016/j.tins.2010.11.002
Shamma, S., Elhilali, M., Ma, L., Micheyl, C., Oxenham, A. J., Pressnitzer, D., Yin, P., & Xu, Y. (2013). Temporal Coherence and the Streaming of Complex Sounds. *Advances in Experimental Medicine and Biology*, *787*, 535–543. https://doi.org/10.1007/978-1-4614-1590-9_59
Shen, J., & Souza, P. E. (2018). On Dynamic Pitch Benefit for Speech Recognition in Speech Masker. *Frontiers in Psychology*, *9*, 1967. https://doi.org/10.3389/fpsyg.2018.01967
Shen, Y. (2014). Gap Detection and Temporal Modulation Transfer Function as Behavioral Estimates of Auditory Temporal Acuity Using Band-Limited Stimuli in Young and Older Adults. *Journal of Speech, Language, and Hearing Research : JSLHR*, *57*(6), 2280. https://doi.org/10.1044/2014_JSLHR-H-13-0276
Shen, Y., & Richards, V. M. (2013). Temporal modulation transfer function for efficient assessment of auditory temporal resolution. *The Journal of the Acoustical Society of America*, *133*(2), 1031. https://doi.org/10.1121/1.4773271
Shi, L., Chang, Y., Li, X., Aiken, S., Liu, L., & Wang, J. (2016). Cochlear Synaptopathy and Noise-Induced Hidden Hearing Loss. *Neural Plasticity*, *2016*, e6143164. https://doi.org/10.1155/2016/6143164
Shi, L., Liu, L., He, T., Guo, X., Yu, Z., Yin, S., & Wang, J. (2013). Ribbon Synapse Plasticity in the Cochleae of Guinea Pigs after Noise-Induced Silent Damage. *PLOS ONE*, *8*(12), e81566. https://doi.org/10.1371/journal.pone.0081566
Shinn, J. B. (2003). Temporal processing: The basics. *The Hearing Journal*, *56*(7), 52. https://doi.org/10.1097/01.HJ.0000292557.52409.67
Shinn-Cunningham, B. G., & Best, V. (2008). Selective Attention in Normal and Impaired Hearing. *Trends in Amplification*, *12*(4), 283–299. https://doi.org/10.1177/1084713808325306

Shokuhifar, G., Javanbakht, M., Vahedi, M., Mehrkian, S., & Aghadoost, A. (2024). The relationship between speech in noise perception and auditory working memory capacity in monolingual and bilingual adults. *International Journal of Audiology*, 1–8. https://doi.org/10.1080/14992027.2024.2328556

Shoushtarian, M., Weder, S., Innes-Brown, H., & McKay, C. M. (2019). Assessing hearing by measuring heartbeat: The effect of sound level. *PLoS ONE*, *14*(2), e0212940. https://doi.org/10.1371/journal.pone.0212940

Sinha, S. K., & Basavaraj, V. (2011). Speech Evoked Auditory Brainstem Responses: A New Tool to Study Brainstem Encoding of Speech Sounds. *Indian Journal of Otolaryngology and Head & Neck Surgery*, *62*(4), 395. https://doi.org/10.1007/s12070-010-0100-y

Slade, K., Plack, C. J., & Nuttall, H. E. (2020). The Effects of Age-Related Hearing Loss on the Brain and Cognitive Function. *Trends in Neurosciences*, *43*(10), 810–821. https://doi.org/10.1016/j.tins.2020.07.005

Smith, S. B., & Cone, B. (2021). Efferent unmasking of speech-in-noise encoding? *International Journal of Audiology*, *60*(9), 677–686. https://doi.org/10.1080/14992027.2020.1862425

Smith, S. B., Krizman, J., Liu, C., White-Schwoch, T., Nicol, T., & Kraus, N. (2019). Investigating peripheral sources of speech-in-noise variability in listeners with normal audiograms. *Hearing Research*, *371*, 66–74. https://doi.org/10.1016/j.heares.2018.11.008

Smits, C., & Houtgast, T. (2005). Results from the Dutch speech-in-noise screening test by telephone. *Ear and Hearing*, *26*(1), 89–95. https://doi.org/10.1097/00003446-200502000-00008

Smits, S, T. G., & Jm, F. (2013). The digits-in-noise test: Assessing auditory speech recognition abilities in noise. *The Journal of the Acoustical Society of America*, *133*(3). https://doi.org/10.1121/1.4789933

Snell, K. B., & Frisina, D. R. (2000). Relationships among age-related differences in gap detection and word recognition. *The Journal of the Acoustical Society of America*, *107*(3), 1615–1626. https://doi.org/10.1121/1.428446

Snell, K. B., Mapes, F. M., Hickman, E. D., & Frisina, D. R. (2002). Word recognition in competing babble and the effects of age, temporal processing, and absolute sensitivity. *The Journal of the Acoustical Society of America*, *112*(2), 720–727. https://doi.org/10.1121/1.1487841

Snyder, J. S., & Alain, C. (2007). Toward a neurophysiological theory of auditory stream segregation. *Psychological Bulletin*, *133*(5), 780–799. https://doi.org/10.1037/0033-2909.133.5.780

Song, Q., Shen, P., Li, X., Shi, L., Liu, L., Wang, J., Yu, Z., Stephen, K., Aiken, S., Yin, S., & Wang, J. (2016). Coding deficits in hidden hearing loss induced by noise: The nature and impacts. *Scientific Reports*, *6*(1), 25200. https://doi.org/10.1038/srep25200

Spahr, A. J., Dorman, M. F., Litvak, L. M., Van Wie, S., Gifford, R. H., Loizou, P. C., Loiselle, L. M., Oakes, T., & Cook, S. (2012). Development and validation of the AzBio sentence lists. *Ear and Hearing*, *33*(1), 112–117. https://doi.org/10.1097/AUD.0b013e31822c2549

Spandita, H. L., & Jain, C. (2024). Comparison of listening efforts in typically developing children and children with central auditory processing disorders. *International Journal of Pediatric Otorhinolaryngology*, *184*, 112058. https://doi.org/10.1016/j.ijporl.2024.112058

Squire, L. R., & Zola-Morgan, S. (1991). The Medial Temporal Lobe Memory System. *Science*, *253*(5026), 1380–1386. https://doi.org/10.1126/science.1896849
Stachenfeld, K. L., Botvinick, M. M., & Gershman, S. J. (2017). The hippocampus as a predictive map. *Nature Neuroscience*, *20*(11), 1643–1653. https://doi.org/10.1038/nn.4650
Stevenson, J. S., Clifton, L., Kuźma, E., & Littlejohns, T. J. (2021). Speech-in-noise hearing impairment is associated with an increased risk of incident dementia in 82,039 UK Biobank participants. *Alzheimer's & Dementia*, *n/a*(n/a). https://doi.org/10.1002/alz.12416
Stevenson, J. S., Clifton, L., Kuźma, E., & Littlejohns, T. J. (2022). Speech-in-noise hearing impairment is associated with an increased risk of incident dementia in 82,039 UK Biobank participants. *Alzheimer's & Dementia*, *18*(3), 445–456. https://doi.org/10.1002/alz.12416
Sun, Y.-S., Fang, W.-H., Kao, T.-W., Yang, H.-F., Peng, T.-C., Wu, L.-W., Chang, Y.-W., Chou, C.-Y., & Chen, W.-L. (2015). Components of Metabolic Syndrome as Risk Factors for Hearing Threshold Shifts. *PloS One*, *10*(8), e0134388. https://doi.org/10.1371/journal.pone.0134388
Teipel, S. J., Grothe, M., Lista, S., Toschi, N., Garaci, F. G., & Hampel, H. (2013). Relevance of Magnetic Resonance Imaging for Early Detection and Diagnosis of Alzheimer Disease. *Medical Clinics of North America*, *97*(3), 399–424. https://doi.org/10.1016/j.mcna.2012.12.013
Teki, S., Barascud, N., Picard, S., Payne, C., Griffiths, T. D., & Chait, M. (2016). Neural Correlates of Auditory Figure-Ground Segregation Based on Temporal Coherence. *Cerebral Cortex (New York, N.Y.: 1991)*, *26*(9), 3669–3680. https://doi.org/10.1093/cercor/bhw173
Teki, S., Chait, M., Kumar, S., Kriegstein, K. von, & Griffiths, T. D. (2011). Brain Bases for Auditory Stimulus-Driven Figure–Ground Segregation. *Journal of Neuroscience*, *31*(1), 164–171. https://doi.org/10.1523/JNEUROSCI.3788-10.2011
Teki, S., Chait, M., Kumar, S., Shamma, S., & Griffiths, T. D. (2013). Segregation of complex acoustic scenes based on temporal coherence. *eLife*, *2*, e00699. https://doi.org/10.7554/eLife.00699
Teki, S., Chait, M., Kumar, S., von Kriegstein, K., & Griffiths, T. D. (2011). Brain bases for auditory stimulus-driven figure-ground segregation. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*(1), 164–171. https://doi.org/10.1523/JNEUROSCI.3788-10.2011
Teki, S., & Griffiths, T. D. (2016). Brain Bases of Working Memory for Time Intervals in Rhythmic Sequences. *Frontiers in Neuroscience*, *10*, 239. https://doi.org/10.3389/fnins.2016.00239
Teoh, E. S., Cappelloni, M. S., & Lalor, E. C. (2019). Prosodic pitch processing is represented in delta-band EEG and is dissociable from the cortical tracking of other acoustic and phonetic features. *European Journal of Neuroscience*, *50*(11), 3831–3842. https://doi.org/10.1111/ejn.14510
Thompson, E. C., Krizman, J., White-Schwoch, T., Nicol, T., Estabrook, R., & Kraus, N. (2019). Neurophysiological, linguistic, and cognitive predictors of children's ability to perceive speech in noise. *Developmental Cognitive Neuroscience*, *39*, 100672. https://doi.org/10.1016/j.dcn.2019.100672
Tomblin, J. B., Harrison, M., Ambrose, S. E., Walker, E. A., Oleson, J. J., & Moeller, M. P. (2015). Language Outcomes in Young Children with Mild to Severe

Hearing Loss. *Ear and Hearing*, *36 Suppl 1*(0 1), 76S-91S. https://doi.org/10.1097/AUD.0000000000000219
Tóth, B., Kocsis, Z., Háden, G. P., Szerafin, Á., Shinn-Cunningham, B. G., & Winkler, I. (2016). EEG signatures accompanying auditory figure-ground segregation. *NeuroImage*, *141*, 108–119. https://doi.org/10.1016/j.neuroimage.2016.07.028
Trau-Margalit, A., Fostick, L., Harel-Arbeli, T., Nissanholtz-Gannot, R., & Taitelbaum-Swead, R. (2023). Speech recognition in noise task among children and young-adults: A pupillometry study. *Frontiers in Psychology*, *14*, 1188485. https://doi.org/10.3389/fpsyg.2023.1188485
Tremblay, K. L., & Kraus, N. (2002). Auditory Training Induces Asymmetrical Changes in Cortical Neural Activity. *Journal of Speech, Language, and Hearing Research*, *45*(3), 564–572. https://doi.org/10.1044/1092-4388(2002/045)
Tremblay, K. L., Pinto, A., Fischer, M. E., Klein, B. E. K., Klein, R., Levy, S., Tweed, T. S., & Cruickshanks, K. J. (2015). Self-Reported Hearing Difficulties Among Adults With Normal Audiograms: The Beaver Dam Offspring Study. *Ear and Hearing*, *36*(6), e290-299. https://doi.org/10.1097/AUD.0000000000000195
Trine, A., & Monson, B. B. (2020). Extended High Frequencies Provide Both Spectral and Temporal Information to Improve Speech-in-Speech Recognition. *Trends in Hearing*, *24*, 2331216520980299. https://doi.org/10.1177/2331216520980299
Uchida, Y., Sugiura, S., Nishita, Y., Saji, N., Sone, M., & Ueda, H. (2019). Age-related hearing loss and cognitive decline—The potential mechanisms linking the two. *Auris Nasus Larynx*, *46*(1), 1–9. https://doi.org/10.1016/j.anl.2018.08.010
Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M., & Bradlow, A. R. (2010). The Wildcat Corpus of native-and foreign-accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and Speech*, *53(4), 510–540*.
Van Engen, K. J., & Peelle, J. E. (2014). Listening effort and accented speech. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00577
Van Hirtum, T., Somers, B., Verschueren, E., Dieudonné, B., & Francart, T. (2023). Delta-band neural envelope tracking predicts speech intelligibility in noise in preschoolers. *Hearing Research*, *434*, 108785. https://doi.org/10.1016/j.heares.2023.108785
Velluti, R. A. (2018). Chapter 4—Auditory Information Processing During Sleep. In R. A. Velluti (Ed.), *The Auditory System in Sleep (Second Edition)* (pp. 91–111). Academic Press. https://doi.org/10.1016/B978-0-12-810476-7.00004-X
Venegas, J., & Clark, E. (2011). Wechsler Test of Adult Reading. In J. S. Kreutzer, J. DeLuca, & B. Caplan (Eds.), *Encyclopedia of Clinical Neuropsychology* (pp. 2693–2694). Springer. https://doi.org/10.1007/978-0-387-79948-3_1500
Venkitakrishnan, S., & Wu, Y.-H. (2023). Facial Expressions as an Index of Listening Difficulty and Emotional Response. *Seminars in Hearing*, *44*, 166–187. https://doi.org/10.1055/s-0043-1766104
Vermiglio, A. J., & Fang, X. (2021). The World Health Organization (WHO) hearing impairment guidelines and a speech recognition in noise (SRN) disorder. *International Journal of Audiology*, 1–8. https://doi.org/10.1080/14992027.2021.1976424

Vermiglio, A. J., Soli, S. D., Freed, D. J., & Fisher, L. M. (2012). The relationship between high-frequency pure-tone hearing loss, hearing in noise test (HINT) thresholds, and the articulation index. *Journal of the American Academy of Audiology*, *23*(10), 779–788. https://doi.org/10.3766/jaaa.23.10.4

Vonck, B. M. D., van Heteren, J. A. A., Lammers, M. J. W., de Jel, D. V. C., Schaake, W. A. A., van Zanten, G. A., Stokroos, R. J., & Versnel, H. (2022). Cortical potentials evoked by tone frequency changes can predict speech perception in noise. *Hearing Research*, *420*, 108508. https://doi.org/10.1016/j.heares.2022.108508

Walton, J. P., & Burkard, R. (2001). 40—Neurophysiological Manifestations of Aging in the Peripheral and Central Auditory Nervous System. In P. R. Hof & C. V. Mobbs (Eds.), *Functional Neurobiology of Aging* (pp. 581–595). Academic Press. https://doi.org/10.1016/B978-012351830-9/50042-1

Wang, D., Zheng, L., Lin, Y., Zhang, Y., & Sheng, L. (2022). Sentence Repetition as a Clinical Marker for Mandarin-Speaking Preschoolers With Developmental Language Disorder. *Journal of Speech, Language, and Hearing Research: JSLHR*, *65*(4), 1543–1560. https://doi.org/10.1044/2021_JSLHR-21-00401

Warzybok, A., Brand, T., Wagener, K. C., & Kollmeier, B. (2015). How much does language proficiency by non-native listeners influence speech audiometric tests in noise? *International Journal of Audiology*, *54*(sup2), 88–99. https://doi.org/10.3109/14992027.2015.1063715

Waters, G. S., & Caplan, D. (2001). Age, working memory, and on-line syntactic processing in sentence comprehension. *Psychology and Aging*, *16*(1), 128–144. https://doi.org/10.1037/0882-7974.16.1.128

Wechsler, D. (1955). *Manual for the Wechsler Adult Intelligence Scale* (pp. vi, 110). Psychological Corp.

Weißgerber, T., Baumann, U., Brand, T., & Neumann, K. (2013). German Oldenburg Sentence Test for Children: A Useful Speech Audiometry Tool for Hearing-Impaired Children at Kindergarten and School Age. *Folia Phoniatrica et Logopaedica*, *64*(5), 227–233. https://doi.org/10.1159/000342414

Whitlock, J. R., Heynen, A. J., Shuler, M. G., & Bear, M. F. (2006). Learning Induces Long-Term Potentiation in the Hippocampus. *Science*, *313*(5790), 1093–1097. https://doi.org/10.1126/science.1128134

Wilson, R. H. (2003). Development of a speech-in-multitalker-babble paradigm to assess word-recognition performance. *Journal of the American Academy of Audiology*, *14*(9), 453–470.

Wilson, R. H. (2011). Clinical experience with the words-in-noise test on 3430 veterans: Comparisons with pure-tone thresholds and word recognition in quiet. *Journal of the American Academy of Audiology*, *22*(7), 405–423. https://doi.org/10.3766/jaaa.22.7.3

Wilson, R. H., McArdle, R. A., & Smith, S. L. (2007). An Evaluation of the BKB-SIN, HINT, QuickSIN, and WIN Materials on Listeners With Normal Hearing and Listeners With Hearing Loss. *Journal of Speech, Language, and Hearing Research*, *50*(4), 844–856.

Winston, T., & Lei, S. (2023). Role of mitochondrial dysfunction and oxidative stress in sensorineural hearing loss. *Hearing Research*, *434*. https://doi.org/10.1016/j.heares.2023.108783

Winter, I. M. (2015). Physiology of Hearing: Basic principles of audiology. In *Hearing*. CRC Press.

Wong, L. L. N., Cheung, C., & Wong, E. C. M. (2008). Comparison of hearing thresholds obtained using pure-tone behavioral audiometry, the Cantonese Hearing in Noise Test (CHINT) and cortical evoked response audiometry. *Acta Oto-Laryngologica*, *128*(6), 654–660. https://doi.org/10.1080/00016480701642189

Wöstmann, M., Erb, J., Kreitewolf, J., & Obleser, J. (2021). Personality captures dissociations of subjective versus objective hearing in noise. *Royal Society Open Science*, *8*(11), 210881. https://doi.org/10.1098/rsos.210881

Xie, D., Luo, J., Chao, X., Li, J., Liu, X., Fan, Z., Wang, H., & Xu, L. (2022). Relationship Between the Ability to Detect Frequency Changes or Temporal Gaps and Speech Perception Performance in Post-lingual Cochlear Implant Users. *Frontiers in Neuroscience*, *16*. https://doi.org/10.3389/fnins.2022.904724

Xie, L., Das, S. R., Wisse, L. E. M., Ittyerah, R., Yushkevich, P. A., Wolk, D. A., & Initiative, for the A. D. N. (2018). Early Tau Burden Correlates with Higher Rate of Atrophy in Transentorhinal Cortex. *Journal of Alzheimer's Disease*, *62*(1), 85–92. https://doi.org/10.3233/JAD-170945

Xie, R., Wang, M., & Zhang, C. (2024). Mechanisms of age-related hearing loss at the auditory nerve central synapses and postsynaptic neurons in the cochlear nucleus. *Hearing Research*, *442*, 108935. https://doi.org/10.1016/j.heares.2023.108935

Yang, Y., Lee, J., & Kim, G. (2020). Integration of locomotion and auditory signals in the mouse inferior colliculus. *eLife*, *9*, e52228. https://doi.org/10.7554/eLife.52228

Yi, H. G., Leonard, M. K., & Chang, E. F. (2019). The Encoding of Speech Sounds in the Superior Temporal Gyrus. *Neuron*, *102*(6), 1096–1110. https://doi.org/10.1016/j.neuron.2019.04.023

Zadeh, L., Silbert, N. H., Sternasty, K., Swanepoel, D. W., Hunter, L. L., & Moore, D. R. (2019). Extended high-frequency hearing enhances speech perception in noise. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(47), 23753–23759. https://doi.org/10.1073/pnas.1903315116

Zekveld, A. A., Kramer, S. E., & Festen, J. M. (2010). Pupil response as an indication of effortful listening: The influence of sentence intelligibility. *Ear and Hearing*, *31*(4), 480–490. https://doi.org/10.1097/AUD.0b013e3181d4f251

Zendel, B. R., West, G. L., Belleville, S., & Peretz, I. (2019). Musical training improves the ability to understand speech-in-noise in older adults. *Neurobiology of Aging*, *81*, 102–115. https://doi.org/10.1016/j.neurobiolaging.2019.05.015

Zheng, J., Shen, W., He, D. Z. Z., Long, K. B., Madison, L. D., & Dallos, P. (2000). Prestin is the motor protein of cochlear outer hair cells. *Nature*, *405*(6783), 149–155. https://doi.org/10.1038/35012009

Zion Golumbic, E. M., Poeppel, D., & Schroeder, C. E. (2012). Temporal context in speech processing and attentional stream selection: A behavioral and neural perspective. *Brain and Language*, *122*(3), 151–161. https://doi.org/10.1016/j.bandl.2011.12.010

List of abbreviations

A1: primary auditory cortex

ABR: auditory brainstem response

AD: Alzheimer's disease

AFG: auditory figure-ground

AFG-Dynamic: dynamic auditory figure-ground

AFG-Fixed: fixed-frequency auditory figure-ground

AFG-F0: auditory figure-ground with fundamental frequency

AFG-1/F: auditory figure-ground with 1/f contour

AFG-High: high-frequency auditory-figure-ground

AFG-Low: low-frequency auditory figure-ground

AM: amplitude modulation

ANF: auditory nerve fibre

ANOVA: Analysis of Variance

ASSR: auditory state-state response

AUM: auditory memory

AUM-Amp: auditory memory for amplitude precision

AUM-Freq: auditory memory for frequency precision

CAEPs: cortical auditory evoked potentials

CAF: confirmatory factor analysis

CAP: central auditory processing

CFI: comparative fit index

CI: cochlear implant

CNS: central nervous system

CPL: central sound processing long

CPS: central sound processing short

dB: decibels

DiN: digit-in-noise

DS: digit span

EEG: electroencephalography

ERP: event-related potential

FFR: frequency-following response

fMRI: functional magnetic resonance imaging

GCog: general cognition

GDT/GAP-Det: gap detection

GIN: Gap in Noise

HINT: Hearing in Noise Test

RMSEA: root-mean-square error of approximation

RGDT: Random Gap Detection Test

R-SPIN: Revised Speech Perception in Noise Test

ICA: independent component analysis

ICC: intraclass correlation coefficient

IFG: inferior frontal gyrus

IHC: inner hair cells

IPS: intraparietal sulcus

ITCP-B: British Iowa Test of Consonant Perception

ITD: interaural timing difference

LiSN-S: Listening in Spatialized Noise – Sentences Test

MEG: magnetoencephalography

MMN: mismatch negativity

MOC: medial olivocochlear

MSI: Goldsmith musical sophistication index

MTL: medial temporal lobe

MLR: middle latency response

mtDNA: mitochondrial DNA

MTX: matrix reasoning

OAES: otoacoustic emissions

OHC: outer hair cells

ORN: object-related negativity

PAS: peripheral auditory system

PFC: prefrontal cortex

PT: planum temporale

PTA: pure-tone audiogram/audiometry

PTA_EHF: extended-high-frequency PTA

PTA_SF: standard-frequency PTA

QuickSIN: Quick Speech in Noise

SD: standard deviation

SEM: structural equation modelling/structural equation model

SFG: stochastic figure-ground

STG: superior temporal gyrus

SiB: sentence-in-babble

SIN: speech-in-noise

SNR: signal-to-noise ratio

SRMR: standardised root mean squared residual

SSQ: Speech, Spatial and Qualities of Hearing Scale

TRF: temporal response function

TLI: Tucker-Lewis Index

TMR: target-to-masker ratio

TMT: Trail-Making Test

TMTF: temporal modulation transfer function

WIN: word-in-noise

WTAR: The Wechsler Test of Adult Reading

List of materials

A. British-ITCP test

The electronic materials can be found in the OSF repository https://osf.io/53jsg/files/osfstorage. The list of words used are as follows.

ball	fall	shawl	wall
fall	ball	shawl	wall
shawl	ball	fall	wall
wall	ball	fall	shawl
ban	man	van	than
man	ban	van	than
van	ban	man	than
than	ban	van	man
lash	bash	dash	gash
bash	lash	dash	gash
dash	lash	bash	gash
gash	lash	bash	dash
patch	thatch	match	batch
thatch	patch	match	batch
match	patch	latch	batch
batch	patch	latch	match
lead	mead	weed	need
mead	lead	weed	need
weed	lead	mead	need
need	lead	mead	weed
beer	gear	dear	tier
gear	beer	dear	tier
dear	beer	gear	tier
tier	beer	gear	dear
yet	vet	get	net
vet	yet	get	net
get	yet	vet	net
net	yet	vet	get
bill	till	gill	dill
till	bill	gill	dill
gill	bill	till	dill
dill	bill	till	gill
bob	cob	sob	gob
cob	bob	sob	gob
sob	bob	cob	gob
gob	bob	cob	sob
boom	doom	womb	room
doom	boom	womb	room
womb	boom	doom	room
room	boom	doom	womb

boon	dune	noon	moon
dune	boon	noon	moon
noon	boon	dune	moon
moon	boon	dune	noon
cop	pop	top	shop
pop	cop	top	shop
top	cop	pop	shop
shop	cop	pop	top
that	vat	cat	sat
vat	that	cat	sat
cat	vat	that	sat
sat	vat	that	vat
caught	taught	fought	thought
taught	caught	fought	thought
fought	caught	taught	thought
thought	caught	taught	fought
tell	cell	shell	yell
cell	tell	shell	yell
shell	tell	cell	yell
yell	tell	cell	shell
chute	coot	suit	toot
coot	chute	suit	toot
suit	chute	coot	toot
toot	chute	coot	suit
took	look	cook	rook
look	took	cook	rook
cook	took	look	rook
rook	took	look	cook
cool	pool	fool	ghoul
pool	cool	fool	ghoul
fool	cool	pool	ghoul
ghoul	cool	pool	fool
watt	lot	rot	yacht
lot	watt	rot	yacht
rot	watt	lot	yacht
yacht	watt	lot	rot
dab	fab	gab	nab
fab	dab	gab	nab
gab	dab	fab	nab
nab	dab	fab	gab
said	dead	red	led
dead	said	red	led
red	said	dead	led
led	said	dead	red
kneel	meal	veal	feel
meal	kneel	veal	feel

veal	kneel	meal	feel
feel	kneel	meal	veal
sin	shin	kin	thin
shin	sin	kin	thin
kin	sin	shin	thin
thin	sin	shin	kin
zip	lip	yip	rip
lip	zip	yip	rip
yip	zip	lip	rip
rip	zip	lip	yip
ken	pen	then	zen
pen	ken	then	zen
then	ken	pen	zen
zen	ken	pen	then
king	ping	thing	zing
ping	king	thing	zing
thing	king	ping	zing
zing	king	ping	thing
sit	zit	lit	mitt
zit	sit	lit	mitt
lit	sit	zit	mitt
mitt	sit	zit	lit
lock	rock	mock	wok
rock	lock	mock	wok
mock	lock	rock	wok
wok	lock	rock	mock
more	pour	tore	shore
pour	more	tore	shore
tore	more	pour	shore
shore	more	pour	tore
pong	tong	thong	song
tong	pong	thong	song
thong	pong	tong	song
song	pong	tong	thong

B. Sentence-in-noise test

The sentences are English Oldenburg sentence; the whole list is as follows.

1	Peter	got	two	large	desks	Peter got two large desks
2	Kathy	sees	three	small	chairs	Kathy sees three small chairs
3	Lucy	brought	four	old	tables	Lucy brought four old tables
4	Alan	gives	seven	dark	toys	Alan gives seven dark toys
5	Rachel	sold	eight	heavy	spoons	Rachel sold eight heavy spoons
6	William	prefers	nine	green	windows	William prefers nine green windows
7	Steven	has	twelve	cheap	sofas	Steven has twelve cheap sofas
8	Thomas	kept	fifteen	pretty	rings	Thomas kept fifteen pretty rings

9	Doris	ordered	nineteen	red	flowers	Doris ordered nineteen red flowers
10	Nina	wants	sixty	white	houses	Nina wants sixty white houses
11	Peter	sees	four	dark	spoons	Peter sees four dark spoons
12	Kathy	brought	seven	heavy	windows	Kathy brought seven heavy windows
13	Lucy	gives	eight	green	sofas	Lucy gives eight green sofas
14	Alan	sold	nine	cheap	rings	Alan sold nine cheap rings
15	Rachel	prefers	twelve	pretty	flowers	Rachel prefers twelve pretty flowers
16	William	has	fifteen	red	houses	William has fifteen red houses
17	Steven	kept	nineteen	white	desks	Steven kept nineteen white desks
18	Thomas	ordered	sixty	large	chairs	Thomas ordered sixty large chairs
19	Doris	wants	two	small	tables	Doris wants two small tables
20	Nina	got	three	old	toys	Nina got three old toys
21	Peter	brought	eight	cheap	flowers	Peter brought eight cheap flowers
22	Kathy	gives	nine	pretty	houses	Kathy gives nine pretty houses
23	Lucy	sold	twelve	red	desks	Lucy sold twelve red desks
24	Alan	prefers	fifteen	white	chairs	Alan prefers fifteen white chairs
25	Rachel	has	nineteen	large	tables	Rachel has nineteen large tables
26	William	kept	sixty	small	toys	William kept sixty small toys
27	Steven	ordered	two	old	spoons	Steven ordered two old spoons
28	Thomas	wants	three	dark	windows	Thomas wants three dark windows
29	Doris	got	four	heavy	sofas	Doris got four heavy sofas
30	Nina	sees	seven	green	rings	Nina sees seven green rings
31	Peter	gives	twelve	white	tables	Peter gives twelve white tables
32	Kathy	sold	fifteen	large	toys	Kathy sold fifteen large toys
33	Lucy	prefers	nineteen	small	spoons	Lucy prefers nineteen small spoons
34	Alan	has	sixty	old	windows	Alan has sixty old windows
35	Rachel	kept	two	dark	sofas	Rachel kept two dark sofas
36	William	ordered	three	heavy	rings	William ordered three heavy rings
37	Steven	wants	four	green	flowers	Steven wants four green flowers
38	Thomas	got	seven	cheap	houses	Thomas got seven cheap houses
39	Doris	sees	eight	pretty	desks	Doris sees eight pretty desks
40	Nina	brought	nine	red	chairs	Nina brought nine red chairs
41	Peter	sold	nineteen	old	sofas	Peter sold nineteen old sofas
42	Kathy	prefers	sixty	dark	rings	Kathy prefers sixty dark rings
43	Lucy	has	two	heavy	flowers	Lucy has two heavy flowers
44	Alan	kept	three	green	houses	Alan kept three green houses
45	Rachel	ordered	four	cheap	desks	Rachel ordered four cheap desks
46	William	wants	seven	pretty	chairs	William wants seven pretty chairs
47	Steven	got	eight	red	tables	Steven got eight red tables
48	Thomas	sees	nine	white	toys	Thomas sees nine white toys
49	Doris	brought	twelve	large	spoons	Doris brought twelve large spoons
50	Nina	gives	fifteen	small	windows	Nina gives fifteen small windows
51	Peter	prefers	two	green	desks	Peter prefers two green desks
52	Kathy	has	three	cheap	chairs	Kathy has three cheap chairs
53	Lucy	kept	four	pretty	tables	Lucy kept four pretty tables
54	Alan	ordered	seven	red	toys	Alan ordered seven red toys

55	Rachel	wants	eight	white	spoons	Rachel wants eight white spoons
56	William	got	nine	large	windows	William got nine large windows
57	Steven	sees	twelve	small	sofas	Steven sees twelve small sofas
58	Thomas	brought	fifteen	old	rings	Thomas brought fifteen old rings
59	Doris	gives	nineteen	dark	flowers	Doris gives nineteen dark flowers
60	Nina	sold	sixty	heavy	houses	Nina sold sixty heavy houses
61	Peter	has	four	red	spoons	Peter has four red spoons
62	Kathy	kept	seven	white	windows	Kathy kept seven white windows
63	Lucy	ordered	eight	large	sofas	Lucy ordered eight large sofas
64	Alan	wants	nine	small	rings	Alan wants nine small rings
65	Rachel	got	twelve	old	flowers	Rachel got twelve old flowers
66	William	sees	fifteen	dark	houses	William sees fifteen dark houses
67	Steven	brought	nineteen	heavy	desks	Steven brought nineteen heavy desks
68	Thomas	gives	sixty	green	chairs	Thomas gives sixty green chairs
69	Doris	sold	two	cheap	tables	Doris sold two cheap tables
70	Nina	prefers	three	pretty	toys	Nina prefers three pretty toys
71	Peter	kept	eight	small	flowers	Peter kept eight small flowers
72	Kathy	ordered	nine	old	houses	Kathy ordered nine old houses
73	Lucy	wants	twelve	dark	desks	Lucy wants twelve dark desks
74	Alan	got	fifteen	heavy	chairs	Alan got fifteen heavy chairs
75	Rachel	sees	nineteen	green	tables	Rachel sees nineteen green tables
76	William	brought	sixty	cheap	toys	William brought sixty cheap toys
77	Steven	gives	two	pretty	spoons	Steven gives two pretty spoons
78	Thomas	sold	three	red	windows	Thomas sold three red windows
79	Doris	prefers	four	white	sofas	Doris prefers four white sofas
80	Nina	has	seven	large	rings	Nina has seven large rings
81	Peter	ordered	twelve	heavy	tables	Peter ordered twelve heavy tables
82	Kathy	wants	fifteen	green	toys	Kathy wants fifteen green toys
83	Lucy	got	nineteen	cheap	spoons	Lucy got nineteen cheap spoons
84	Alan	sees	sixty	pretty	windows	Alan sees sixty pretty windows
85	Rachel	brought	two	red	sofas	Rachel brought two red sofas
86	William	gives	three	white	rings	William gives three white rings
87	Steven	sold	four	large	flowers	Steven sold four large flowers
88	Thomas	prefers	seven	small	houses	Thomas prefers seven small houses
89	Doris	has	eight	old	desks	Doris has eight old desks
90	Nina	kept	nine	dark	chairs	Nina kept nine dark chairs
91	Peter	wants	nineteen	pretty	sofas	Peter wants nineteen pretty sofas
92	Kathy	got	sixty	red	rings	Kathy got sixty red rings
93	Lucy	sees	two	white	flowers	Lucy sees two white flowers
94	Alan	brought	three	large	houses	Alan brought three large houses
95	Rachel	gives	four	small	desks	Rachel gives four small desks
96	William	sold	seven	old	chairs	William sold seven old chairs
97	Steven	prefers	eight	dark	tables	Steven prefers eight dark tables
98	Thomas	has	nine	heavy	toys	Thomas has nine heavy toys
99	Doris	kept	twelve	green	spoons	Doris kept twelve green spoons
100	Nina	ordered	fifteen	cheap	windows	Nina ordered fifteen cheap windows

101	Peter	got	four	dark	spoons	Peter got four dark spoons
102	Kathy	sees	seven	heavy	windows	Kathy sees seven heavy windows
103	Lucy	brought	eight	green	sofas	Lucy brought eight green sofas
104	Alan	gives	nine	cheap	rings	Alan gives nine cheap rings
105	Rachel	sold	twelve	pretty	flowers	Rachel sold twelve pretty flowers
106	William	prefers	fifteen	red	houses	William prefers fifteen red houses
107	Steven	has	nineteen	white	desks	Steven has nineteen white desks
108	Thomas	kept	sixty	large	chairs	Thomas kept sixty large chairs
109	Doris	ordered	two	small	tables	Doris ordered two small tables
110	Nina	wants	three	old	toys	Nina wants three old toys
111	Peter	sees	nine	red	chairs	Peter sees nine red chairs
112	Kathy	brought	twelve	white	tables	Kathy brought twelve white tables
113	Lucy	gives	fifteen	large	toys	Lucy gives fifteen large toys
114	Alan	sold	nineteen	small	spoons	Alan sold nineteen small spoons
115	Rachel	prefers	sixty	old	windows	Rachel prefers sixty old windows
116	William	has	two	dark	sofas	William has two dark sofas
117	Steven	kept	three	heavy	rings	Steven kept three heavy rings
118	Thomas	ordered	four	green	flowers	Thomas ordered four green flowers
119	Doris	wants	seven	cheap	houses	Doris wants seven cheap houses
120	Nina	got	eight	pretty	desks	Nina got eight pretty desks
121	Peter	brought	nineteen	dark	flowers	Peter brought nineteen dark flowers
122	Kathy	gives	sixty	heavy	houses	Kathy gives sixty heavy houses
123	Lucy	sold	two	green	desks	Lucy sold two green desks
124	Alan	prefers	three	cheap	chairs	Alan prefers three cheap chairs
125	Rachel	has	four	pretty	tables	Rachel has four pretty tables
126	William	kept	seven	red	toys	William kept seven red toys
127	Steven	ordered	eight	white	spoons	Steven ordered eight white spoons
128	Thomas	wants	nine	large	windows	Thomas wants nine large windows
129	Doris	got	twelve	small	sofas	Doris got twelve small sofas
130	Nina	sees	fifteen	old	rings	Nina sees fifteen old rings
131	Peter	gives	three	red	windows	Peter gives three red windows
132	Kathy	sold	four	white	sofas	Kathy sold four white sofas
133	Lucy	prefers	seven	large	rings	Lucy prefers seven large rings
134	Alan	has	eight	small	flowers	Alan has eight small flowers
135	Rachel	kept	nine	old	houses	Rachel kept nine old houses
136	William	ordered	twelve	dark	desks	William ordered twelve dark desks
137	Steven	wants	fifteen	heavy	chairs	Steven wants fifteen heavy chairs
138	Thomas	got	nineteen	green	tables	Thomas got nineteen green tables
139	Doris	sees	sixty	cheap	toys	Doris sees sixty cheap toys
140	Nina	brought	two	pretty	spoons	Nina brought two pretty spoons
141	Peter	sold	eight	dark	tables	Peter sold eight dark tables
142	Kathy	prefers	nine	heavy	toys	Kathy prefers nine heavy toys
143	Lucy	has	twelve	green	spoons	Lucy has twelve green spoons
144	Alan	kept	fifteen	cheap	windows	Alan kept fifteen cheap windows
145	Rachel	ordered	nineteen	pretty	sofas	Rachel ordered nineteen pretty sofas
146	William	wants	sixty	red	rings	William wants sixty red rings

147	Steven	got	two	white	flowers	Steven got two white flowers
148	Thomas	sees	three	large	houses	Thomas sees three large houses
149	Doris	brought	four	small	desks	Doris brought four small desks
150	Nina	gives	seven	old	chairs	Nina gives seven old chairs
151	Peter	prefers	fifteen	red	houses	Peter prefers fifteen red houses
152	Kathy	has	nineteen	white	desks	Kathy has nineteen white desks
153	Lucy	kept	sixty	large	chairs	Lucy kept sixty large chairs
154	Alan	ordered	two	small	tables	Alan ordered two small tables
155	Rachel	wants	three	old	toys	Rachel wants three old toys
156	William	got	four	dark	spoons	William got four dark spoons
157	Steven	sees	seven	heavy	windows	Steven sees seven heavy windows
158	Thomas	brought	eight	green	sofas	Thomas brought eight green sofas
159	Doris	gives	nine	cheap	rings	Doris gives nine cheap rings
160	Nina	sold	twelve	pretty	flowers	Nina sold twelve pretty flowers
161	Peter	has	two	dark	sofas	Peter has two dark sofas
162	Kathy	kept	three	heavy	rings	Kathy kept three heavy rings
163	Lucy	ordered	four	green	flowers	Lucy ordered four green flowers
164	Alan	wants	seven	cheap	houses	Alan wants seven cheap houses
165	Rachel	got	eight	pretty	desks	Rachel got eight pretty desks
166	William	sees	nine	red	chairs	William sees nine red chairs
167	Steven	brought	twelve	white	tables	Steven brought twelve white tables
168	Thomas	gives	fifteen	large	toys	Thomas gives fifteen large toys
169	Doris	sold	nineteen	small	spoons	Doris sold nineteen small spoons
170	Nina	prefers	sixty	old	windows	Nina prefers sixty old windows
171	Peter	kept	seven	red	toys	Peter kept seven red toys
172	Kathy	ordered	eight	white	spoons	Kathy ordered eight white spoons
173	Lucy	wants	nine	large	windows	Lucy wants nine large windows
174	Alan	got	twelve	small	sofas	Alan got twelve small sofas
175	Rachel	sees	fifteen	old	rings	Rachel sees fifteen old rings
176	William	brought	nineteen	dark	flowers	William brought nineteen dark flowers
177	Steven	gives	sixty	heavy	houses	Steven gives sixty heavy houses
178	Thomas	sold	two	green	desks	Thomas sold two green desks
179	Doris	prefers	three	cheap	chairs	Doris prefers three cheap chairs
180	Nina	has	four	pretty	tables	Nina has four pretty tables
181	Peter	ordered	twelve	dark	desks	Peter ordered twelve dark desks
182	Kathy	wants	fifteen	heavy	chairs	Kathy wants fifteen heavy chairs
183	Lucy	got	nineteen	green	tables	Lucy got nineteen green tables
184	Alan	sees	sixty	cheap	toys	Alan sees sixty cheap toys
185	Rachel	brought	two	pretty	spoons	Rachel brought two pretty spoons
186	William	gives	three	red	windows	William gives three red windows
187	Steven	sold	four	white	sofas	Steven sold four white sofas
188	Thomas	prefers	seven	large	rings	Thomas prefers seven large rings
189	Doris	has	eight	small	flowers	Doris has eight small flowers
190	Nina	kept	nine	old	houses	Nina kept nine old houses
191	Peter	wants	sixty	red	rings	Peter wants sixty red rings
192	Kathy	got	two	white	flowers	Kathy got two white flowers

193	Lucy	sees	three	large	houses	Lucy sees three large houses
194	Alan	brought	four	small	desks	Alan brought four small desks
195	Rachel	gives	seven	old	chairs	Rachel gives seven old chairs
196	William	sold	eight	dark	tables	William sold eight dark tables
197	Steven	prefers	nine	heavy	toys	Steven prefers nine heavy toys
198	Thomas	has	twelve	green	spoons	Thomas has twelve green spoons
199	Doris	kept	fifteen	cheap	windows	Doris kept fifteen cheap windows
200	Nina	ordered	nineteen	pretty	sofas	Nina ordered nineteen pretty sofas
201	Peter	got	seven	cheap	flowers	Peter got seven cheap flowers
202	Kathy	sees	eight	pretty	houses	Kathy sees eight pretty houses
203	Lucy	brought	nine	red	desks	Lucy brought nine red desks
204	Alan	gives	twelve	white	chairs	Alan gives twelve white chairs
205	Rachel	sold	fifteen	large	tables	Rachel sold fifteen large tables
206	William	prefers	nineteen	small	toys	William prefers nineteen small toys
207	Steven	has	sixty	old	spoons	Steven has sixty old spoons
208	Thomas	kept	two	dark	windows	Thomas kept two dark windows
209	Doris	ordered	three	heavy	sofas	Doris ordered three heavy sofas
210	Nina	wants	four	green	rings	Nina wants four green rings
211	Peter	sees	fifteen	heavy	houses	Peter sees fifteen heavy houses
212	Kathy	brought	nineteen	green	desks	Kathy brought nineteen green desks
213	Lucy	gives	sixty	cheap	chairs	Lucy gives sixty cheap chairs
214	Alan	sold	two	pretty	tables	Alan sold two pretty tables
215	Rachel	prefers	three	red	toys	Rachel prefers three red toys
216	William	has	four	white	spoons	William has four white spoons
217	Steven	kept	seven	large	windows	Steven kept seven large windows
218	Thomas	ordered	eight	small	sofas	Thomas ordered eight small sofas
219	Doris	wants	nine	old	rings	Doris wants nine old rings
220	Nina	got	twelve	dark	flowers	Nina got twelve dark flowers
221	Peter	brought	three	old	desks	Peter brought three old desks
222	Kathy	gives	four	dark	chairs	Kathy gives four dark chairs
223	Lucy	sold	seven	heavy	tables	Lucy sold seven heavy tables
224	Alan	prefers	eight	green	toys	Alan prefers eight green toys
225	Rachel	has	nine	cheap	spoons	Rachel has nine cheap spoons
226	William	kept	twelve	pretty	windows	William kept twelve pretty windows
227	Steven	ordered	fifteen	red	sofas	Steven ordered fifteen red sofas
228	Thomas	wants	nineteen	white	rings	Thomas wants nineteen white rings
229	Doris	got	sixty	large	flowers	Doris got sixty large flowers
230	Nina	sees	two	small	houses	Nina sees two small houses
231	Peter	gives	nine	large	chairs	Peter gives nine large chairs
232	Kathy	sold	twelve	small	tables	Kathy sold twelve small tables
233	Lucy	prefers	fifteen	old	toys	Lucy prefers fifteen old toys
234	Alan	has	nineteen	dark	spoons	Alan has nineteen dark spoons
235	Rachel	kept	sixty	heavy	windows	Rachel kept sixty heavy windows
236	William	ordered	two	green	sofas	William ordered two green sofas
237	Steven	wants	three	cheap	rings	Steven wants three cheap rings
238	Thomas	got	four	pretty	flowers	Thomas got four pretty flowers

239	Doris	sees	seven	red	houses	Doris sees seven red houses
240	Nina	brought	eight	white	desks	Nina brought eight white desks
241	Peter	sold	sixty	red	tables	Peter sold sixty red tables
242	Kathy	prefers	two	white	toys	Kathy prefers two white toys
243	Lucy	has	three	large	spoons	Lucy has three large spoons
244	Alan	kept	four	small	windows	Alan kept four small windows
245	Rachel	ordered	seven	old	sofas	Rachel ordered seven old sofas
246	William	wants	eight	dark	rings	William wants eight dark rings
247	Steven	got	nine	heavy	flowers	Steven got nine heavy flowers
248	Thomas	sees	twelve	green	houses	Thomas sees twelve green houses
249	Doris	brought	fifteen	cheap	desks	Doris brought fifteen cheap desks
250	Nina	gives	nineteen	pretty	chairs	Nina gives nineteen pretty chairs
251	Peter	prefers	seven	cheap	toys	Peter prefers seven cheap toys
252	Kathy	has	eight	pretty	spoons	Kathy has eight pretty spoons
253	Lucy	kept	nine	red	windows	Lucy kept nine red windows
254	Alan	ordered	twelve	white	sofas	Alan ordered twelve white sofas
255	Rachel	wants	fifteen	large	rings	Rachel wants fifteen large rings
256	William	got	nineteen	small	flowers	William got nineteen small flowers
257	Steven	sees	sixty	old	houses	Steven sees sixty old houses
258	Thomas	brought	two	dark	desks	Thomas brought two dark desks
259	Doris	gives	three	heavy	chairs	Doris gives three heavy chairs
260	Nina	sold	four	green	tables	Nina sold four green tables
261	Peter	has	fifteen	heavy	spoons	Peter has fifteen heavy spoons
262	Kathy	kept	nineteen	green	windows	Kathy kept nineteen green windows
263	Lucy	ordered	sixty	cheap	sofas	Lucy ordered sixty cheap sofas
264	Alan	wants	two	pretty	rings	Alan wants two pretty rings
265	Rachel	got	three	red	flowers	Rachel got three red flowers
266	William	sees	four	white	houses	William sees four white houses
267	Steven	brought	seven	large	desks	Steven brought seven large desks
268	Thomas	gives	eight	small	chairs	Thomas gives eight small chairs
269	Doris	sold	nine	old	tables	Doris sold nine old tables
270	Nina	prefers	twelve	dark	toys	Nina prefers twelve dark toys
271	Peter	kept	three	old	windows	Peter kept three old windows
272	Kathy	ordered	four	dark	sofas	Kathy ordered four dark sofas
273	Lucy	wants	seven	heavy	rings	Lucy wants seven heavy rings
274	Alan	got	eight	green	flowers	Alan got eight green flowers
275	Rachel	sees	nine	cheap	houses	Rachel sees nine cheap houses
276	William	brought	twelve	pretty	desks	William brought twelve pretty desks
277	Steven	gives	fifteen	red	chairs	Steven gives fifteen red chairs
278	Thomas	sold	nineteen	white	tables	Thomas sold nineteen white tables
279	Doris	prefers	sixty	large	toys	Doris prefers sixty large toys
280	Nina	has	two	small	spoons	Nina has two small spoons
281	Peter	ordered	nine	large	sofas	Peter ordered nine large sofas
282	Kathy	wants	twelve	small	rings	Kathy wants twelve small rings
283	Lucy	got	fifteen	old	flowers	Lucy got fifteen old flowers
284	Alan	sees	nineteen	dark	houses	Alan sees nineteen dark houses

285	Rachel	brought	sixty	heavy	desks	Rachel brought sixty heavy desks
286	William	gives	two	green	chairs	William gives two green chairs
287	Steven	sold	three	cheap	tables	Steven sold three cheap tables
288	Thomas	prefers	four	pretty	toys	Thomas prefers four pretty toys