Cortical and subcortical speech-evoked responses in young and older adults: Effects of background noise, arousal states, and neural excitability

Guangting Mai

A dissertation submitted in fulfilment
of the requirements for the degree of
Doctor of Philosophy of **UCL**

Department of Experimental Psychology

University College London

2020

Declaration

I, *Guangting Mai*, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

This thesis investigated how the brain processes speech signals in human adults across a wide age-range in the sensory auditory systems using electroencephalography (EEG). Two types of speech-evoked phase-locked responses were focused on: (i) cortical responses (theta-band phase-locked responses) that reflect processing of low-frequency slowly-varying envelopes of speech; (ii) subcortical/peripheral responses (frequency-following responses; FFRs) that reflect encoding of speech periodicity and temporal fine structure information. The aims are to elucidate how these neural activities are affected by different internal (aging, hearing loss, level of arousal and neural excitability) and external (background noise) factors during our daily life through three studies.

Study 1 investigated theta-band phase-locking and FFRs in noisy environments in young and older adults. It investigated how aging and hearing loss affect these activities under quiet and noisy environments, and how these activities are associated with speech-in-noise perception. The results showed that ageing and hearing loss affect speech-evoked phase-locked responses through different mechanisms, and the effects of aging on cortical and subcortical activities take different roles in speech-in-noise perception.

Study 2 investigated how level of arousal, or consciousness, affects phase-locked responses in young and older adults. The results showed that both theta-band phase-locking and FFRs decreases following decreases in the level of arousal. It was further found that neuro-regulatory role of sleep spindles on theta-band phase-locking is distinct between young and older adults, indicating that the mechanisms of neuro-regulation for phase-locked responses in different arousal states are age-dependent.

Study 3 established a causal relationship between the auditory cortical excitability and FFRs using combined transcranial direct current stimulation (tDCS) and EEG. FFRs were measured before and after tDCS was applied over the auditory cortices. The results showed that changes in neural excitability of the right auditory cortex can alter FFR magnitudes along the contralateral pathway. This shows important theoretical and clinical implications that causally link functions of auditory cortex with neural encoding of speech periodicity.

Taken together, findings of this thesis will advance our understanding of how speech signals are processed via neural phase-locking in our everyday life across the lifespan.

Impact Statement

Speech communication is a major human activity during daily life. It is critical to understand how speech is processed to achieve successful speech perception. Using techniques of neurophysiological processing and non-invasive brain stimulation, this thesis addresses questions on how speech is processed by the human brain across various internal and external factors in daily life, such as aging, hearing loss, background noise, and different physiological states like arousal and neural excitability. The thesis thus should have great impacts in the academic field of speech and neuroscience research.

Studies conducted in the thesis have been brought about via dissemination to high-impact publications and presentations in conferences and workshops in the specialized field of auditory perception and neuroscience. Publications include *Journal of the Acoustical Society of America* (Chapter 2) and *NeuroImage* (Chapter 3):

Mai, G., Tuomainen, J., & Howell, P. (2018). Relationship between speech-evoked neural responses and perception of speech in noise in older adults. *The Journal of the Acoustical Society of America*, *143*(3), 1333-1345. https://doi.org/10.1121/1.5024340

Mai, G., Schoof, T., & Howell, P. (2019). Modulation of phase-locked neural responses to speech during different arousal states is age-dependent. *NeuroImage*, *189*, 734-744. https://doi.org/10.1016/j.neuroimage.2019.01.049

Conferences and workshops include *International Conference in Auditory Cortex 2017* in Banff, Canada (Chapter 2; supported by a *Trainee Travel Grant*), *FFR Workshop 2019* in London, UK (Chapter 3) and *Neuroscience 2019* in Chicago, USA (Chapter 4, supported by a *Trainee Professional Development Award*).

Contents

Causal relationship between the right auditory cortex and frequency-following responses: A combined tDCS and EEG study 67

General Discussion 82

List of Figures

List of Tables

List of abbreviations

ANOVA Analysis of variance

BbN Babble noise

BOLD Blood oxygen level dependent

DV Dependent variable

EEG Electroencephalography

F_0 Fundamental frequency

F_0-ENV F_0-rate envelope

F1 The first formant

F2 The second formant

F3 The third formant

FFR Frequency-following response

fMRI Functional magnetic resonance imaging

GABA Gamma-Aminobutyric acid

H1 The first harmonic

H2 The second harmonic

Hz Hertz

IC Inferior colliculus

PLV Phase locking value

MEG Magnetoencephalography

ms millisecond

nREM Non-rapid eye movement

PDC Partial directed coherence

PTA Pure-tone audiometric threshold

SiN Speech-in-noise

Slow-ENV Slowly-varying envelope

SNR Signal-to-noise ratio

SpN Speech-shaped noise

SRT Speech reception threshold

tDCS Transcranial direct current stimulation

TFS Temporal fine structure

TMS Transcranial magnetic stimulation

Acknowledgements

There has been a lot to be thankful for when completing a PhD thesis during a special time in life like this, in the midst of the pandemic.

I would first like to express my thanks to my supervisors. First thanks go to Prof Peter Howell. Pete's initial interests in FFRs guided me to start the projects for this thesis and I greatly appreciate his supervision along the way. I would like to thank Dr Jyrki Tuomainen for his energetic support, his encouragement and intellectual advice. I would not even pass the PhD upgrade without his support. I would like to thank Dr Ilias Tachtsidis for his generosity to be my interdepartmental sponsor for my Graduate Cross-disciplinary Training scholarship. I thank his very fruitful guidance on mastering the fNIRS technique and conducting fNIRS experiments (although this is not included in the thesis, it formed an important part of my PhD).

I appreciate my colleagues for their intellectual and technical support for the EEG and tDCS data collection: Mr Andrew Clark, Dr Naheem Bashir and Dr Eryk Walczak. I thank Dr Tim Schoof for her collaboration in sharing her data and her help on the arousal study. For the fNIRS project not included in this thesis, I owe my big thanks to Dr Tim Green and Prof Stuart Rosen for their knowledgeable and technical guidance. I would also like to thank Dr Matt Davis at Cambridge who has been keen to sponsor me for the exhausting postdoc fellowship applications during my final stage of PhD. His support (as well as generosity and patience during the course) has been keeping my faith in academics in my difficult times.

Thanks also go to participants who took part in the experiments, especially older adults. They are all warm-hearted and generous seniors who show great willingness and enthusiasm to contribute to the aging research.

My biggest thanks should be given to my family. It would have been a long journey with stress and difficulties that I could only face by myself if they were not there. But they are always there. Without them, it is impossible for me to walk this far. Thank you to my parents who are giving me the incredibly tremendous support both emotionally and financially. No matter what decisions I make, wherever I am, they always watch my back. I hope I have been their proud child who leads a fruitful and independent life. Thank you to my parents-in-law, the dearest grandparents of our child who have given him (and all of us) the greatest happiness in his life. I will always owe them for their love and sacrifice in the difficult times. My biggest loving thanks to my adorable wife, Fang. I simply cannot imagine how it would be possible for me to make this journey (and the adventure ahead!) without her company, sacrifice, love and support she is providing every single day. Last but not least, I thank our tiny naughty little boy Eddie, the crystallisation and biggest star of our lives. You have shown us that we still have a bright and brand new life even without a PhD!

Chapter 1

General Introduction

Neural phase-locked activities refer to alignment of brain activities with external input stimuli and play a crucial role in fundamental brain functions (Schroeder and Lakatos 2009). 'Phase' refers to time moments that occur every cycle of periodic brain activities. The time points of each phase indicate times at which the moments of firing and excitability of neural populations are determined (i.e., different phases correspond to different excitability states) (Schroeder and Lakatos 2009). Events in input stimuli which are aligned at the high-excitability phase are amplified and optimally processed (Henry and Obleser, 2012; Zoefel and VanRullen, 2017). The brain responds to the stimuli by synchronizing the high-excitability phases to the informative moments within the stimuli (e.g., energy peaks of sounds or moments of visual events that need to be attended to) (Zoefel and VanRullen, 2017).

The current thesis focuses on speech, which is a stimulus with many periodic attributes that extend over a wide frequency range. They include low-frequency envelopes that reflect slowly-fluctuating energy variations at approximately syllable rates, through higher frequencies that correspond to periodic vibrations of the vocal folds and extend to even higher frequencies that represent temporal fine structures that characterize pitch and formants (Rosen, 1992). The thesis investigates neural phase-locked responses to these attributes of speech and how these responses change over the lifespan and impact on our perception of speech. Further details about phase-locked activity to speech are presented in the following two sections before the specific questions addressed by the thesis are presented.

1.1 Low-frequency neural phase-locked activity for speech perception

Neural activities phase-lock, or align, to specific acoustic properties of speech signals in order to achieve successful speech understanding (Peelle and Davis, 2012). Speech signals are sounds with complex acoustic attributes. The primary attribute within the speech signal is the slowly-fluctuating envelope profile (Slow-ENV) that modulates at a low frequency (normally < 10 Hz) that represents the general profile of speech energy variation over time (Rosen, 1992; see **Figure 1.1**). Essentially, it reflects energy modulations in spoken sentences at rates that correspond approximately to syllable rates (Greenberg et al., 2003; Peelle and Davis, 2012). Psychoacoustic experiments have provided confirmatory evidence that shows that Slow-ENV is the primary cues for speech understanding. For example, noise- or tone-vocoded speech, in which Slow-ENV is preserved in just a few spectral bands with fine structure information

replaced by Gaussian noise or pure tones, can lead to ~90% correct phoneme and word recognition (e.g., Shannon et al., 1995; Arai et al., 1999; Souza and Rosen, 2008).
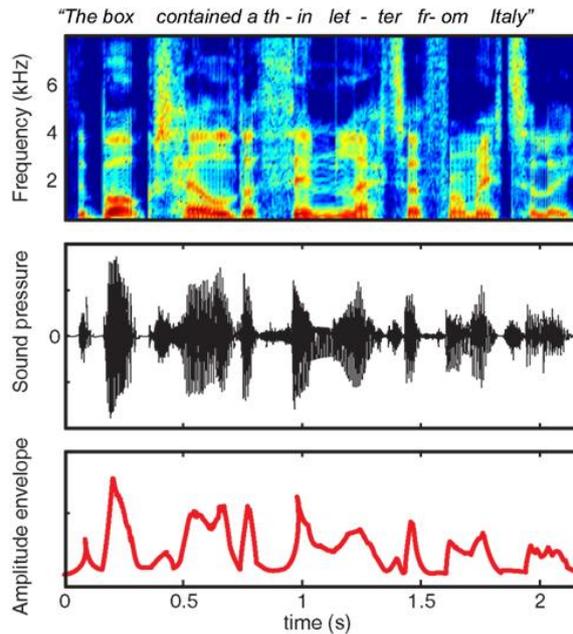
Figure 1.1 Illustration of Slow-ENV of speech signals. *The top and mid panels show the spectrogram and the corresponding waveforms of a spoken sentence. The lower panel shows the Slow-ENV that represents the slow-varying envelope profile (<10 Hz) of the sentence, corresponding approximately to syllables in the speech. The graph is adapted from Peelle and Davis (2012).*

Slow-ENV cues dominate the modulation spectral power of speech in the theta band (4–8 Hz) corresponding to syllable rates of natural speech (Greenberg et al., 2003). Research has accumulated concerning how neural phase-locking to Slow-ENV at the corresponding frequency range contributes to speech perception. These studies used neuroimaging techniques of magnetoencephalography (MEG) or electroencephalography (EEG) that quantify oscillatory brain activities at the millisecond level. They have shown that theta-band phase-locking to Slow-ENV can index intelligibility of speech (Ahissar et al., 2001; Luo and Poeppel, 2007; Gross et al., 2013; Peelle et al., 2013; Doelling et al., 2014; Mai et al., 2016). Ahissar et al. (2001) provided the first evidence associating neural phase-locking to Slow-ENV with speech perception. They used MEG to investigate the relationship between neural phase-locked responses at syllable rates (within the theta-band range) of sentences spoken at faster rates that lead to reduced speech intelligibility. It was found that greater neural phase-locking in the auditory cortex can predict speech intelligibility. Luo and Poeppel (2007) used MEG to record brain responses to noise-vocoded spoken sentences in which speech intelligibility was modulated by changing the spectral resolution (i.e., number of spectral bands) via noise-vocoding. It was found that features of theta-band (4–8 Hz) inter-trial phase-locking at the auditory cortex can be used to reliably classify different sentences and that the classification ability was positively correlated

with speech intelligibility. Peelle et al. (2013) used a similar noise-vocoding method and showed that MEG phase-locking to Slow-ENV at 4–7 Hz was significantly greater when participants listened to 16-band (intelligible) as opposed to when they listened to single-band (unintelligible) noise-vocoded sentences. Specifically, such effects were present in the higher-order linguistic region of left middle temporal gyrus. In the study by Doelling et al. (2014), Slow-ENVs at 2–9 Hz were removed in different spectral bands in noise-vocoded sentences. As a result, MEG phase-locking at the corresponding frequency decreased and was accompanied by reduced speech intelligibility. Gross et al. (2013) (MEG) and Mai et al. (2016) (EEG) also provided evidence which showed that theta-band phase-locking to Slow-ENV is greater in intelligible speech (unprocessed speech) than in unintelligible speech (time-reversed speech).

These neurophysiological (MEG/EEG) studies have thus shown strong associations between low-frequency (especially theta-band) neural phase-locking to Slow-ENV and speech perception. However, a causal relationship has not been established in these studies. To address this issue, several studies have used transcranial alternating current stimulation (tACS), a non-invasive brain stimulation technique that applies alternating currents that perturb the phase relationship between neural activities and external stimuli. Using tACS, studies have provided evidence for the causal relationship between neural phase-locking and speech intelligibility (Zoefel et al., 2018, 2020; Riecke et al., 2018; Wilsch et al., 2018; Keshavarzi et al., 2020; Keshavarzi and Reichenbach, 2020). Zoefel et al. (2018) conducted a combined tACS and functional magnetic resonance imaging (fMRI) study. The authors used tACS to alter the phase relationship between neural oscillations and Slow-ENV of spoken sentences at the syllable rate (~3 Hz) when participants listened to both intelligible and unintelligible noise-vocoded sentences in the MRI scanner. They showed that the tACS phase manipulation on intelligible sentences can modulate haemodynamic responses in the superior temporal gyrus, while such findings were absent for the manipulation on unintelligible sentences. Riecke et al. (2018) and Wilsch et al. (2018) also used tACS to alter the phase relationship between neural oscillations and Slow-ENV of speech (syllable rate of 4 Hz in Riecke et al. (2018) and frequencies < 10 Hz in Wilsch et al. (2018)). They found that, compared to sham stimulation (stimulation that is only applied transiently at the start of an experiment session), such manipulations can significantly modulate intelligibility of sentences heard under noisy environments. Zoefel et al., (2020) found that the tACS phase manipulation upon noise-vocoded speech at the syllable rate can modulate speech intelligibility compared to sham. These studies (Riecke et al., 2018; Wilsch et al., 2018; Zoefel et al., 2020), however, only showed effects of tACS that decrease, rather than increase, speech intelligibility. More recent studies, on the other hand, have shown that tACS manipulation on Slow-ENV at theta-band, but not delta-band, can lead to improved intelligibility of spoken sentences in noise compared to sham (Keshavarzi et al., 2020; Keshavarzi and Reichenbach, 2020).

The current evidence, therefore, showed that neural phase-locking to Slow-ENV, especially at the frequency range of theta, plays an important role for understanding speech. Furthermore, brain stimulation studies have shown that such phase-locking is not merely a

consequence, or product, of the change in speech intelligibility, but can also have causal influences on speech perception.

1.2 Neural phase-locked responses to speech attributes of periodicity and temporal fine-structure information

Slow-ENV has been evidenced to be the primary cues for speech perception (e.g., Shannon et al., 1995). However, while Slow-ENV is sufficient for understanding speech in quiet (Shannon et al., 1995; Arai et al., 1999; Souza and Rosen, 2008), it is not sufficient for speech perception in noisy listening environments. For example, when background noise is present, word recognition of noise-vocoded speech that only preserves Slow-ENV cues decreases significantly compared to unprocessed speech or when higher-frequency attributes are also preserved (e.g., Zeng et al., 2005). This thus indicates that, other attributes of speech signals have additional impact on speech understanding in difficult listening situations.
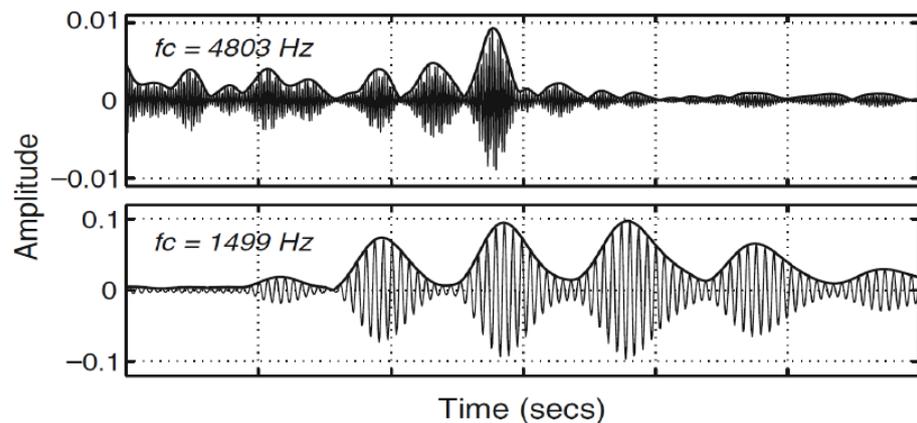
Figure 1.2 Illustration of F0-ENV and TFS of speech signals. *The two panels illustrate the waveforms of the sound 'en' in 'sense'. The thick lines represent the F_0-ENV (envelope at F_0-rate, or periodicity) superimposed on the rapidly-varying TFS (thin lines) at centre frequencies (f_c) of 4803 (top) and 1499 Hz (lower). The graph is adapted from Moore (2008).*

As well as Slow-ENV which is acoustic modulations of speech at the low-frequency range, modulations at high-frequency ranges play further essential roles. These include cues of periodicity (fundamental frequency, or F_0-envelope (F_0-ENV) at F_0 rate of 100–300Hz) and temporal fine structures (TFS; >300 Hz) information (Rosen, 1992; Moore, 2008). **Figure 1.2** shows how F_0-ENV is superimposed on rapidly-varying TFS as carriers in a speech segment. Psychoacoustic experiments have provided a large body of evidence that shows the importance of these acoustic attributes for successful speech comprehension. For example, it has been shown that F_0-ENV is an important cue for segregating the target speech from background

competing sounds (Bregman, 1994; Arehart et al., 1997; Bird & Darwin, 1998; Binns & Culling, 2005, 2007). TFS, which is important for perception of formant structure (Moore, 2014) as well as voicing and pitch (Rosen, 1992; Smith et al., 2002), significantly benefits speech comprehension when background noise is present (Zeng et al., 2005; Stickney et al., 2007; Eaves et al., 2011).

Like neural phase-locking to Slow-ENV, phase-locking to acoustic features of sound at the frequency range of F_0-ENV and TFS can be captured with non-invasive neurophysiological tools of EEG and MEG. Such neural phase-locked activity is called the Frequency-Following Response (FFR) (Coffey et al., 2019). The FFR is elicited using a repeatedly-presented short auditory stimulus (usually with the length of tens to a few hundred milliseconds) such as pure-tone or complex sounds (e.g., a single vowel, syllable or a musical note) (Aiken and Picton, 2008; Skoe and Kraus, 2010). The EEG/MEG signals are then temporally averaged across all sweeps of stimuli to obtain the evoked responses that reflect the phase-locked neural encoding of sounds at the range of F_0-ENV and TFS (Aiken and Picton, 2008). FFR_{ENV} (FFR that represents F_0-ENV and its harmonics) can be obtained by adding responses to sweeps with positive/original and negative (i.e., inversion of the original waveform) polarities to minimize the responses to TFS, while FFR_{TFS} (FFR that represents TFS and formant features) is obtained by subtracting responses to the two polarities to minimize the responses to envelope components (Aiken and Picton, 2008). The FFR resembles the acoustic features of the speech signals by which it is elicited hence reflecting the neural 'fidelity' of these features. **Figure 1.3** gives an example of FFR elicited by a single vowel /i/ (the vowel used to obtain FFR is shown as **Figure 1.3A** and **1.3B**; the resultant FFR_{ENV} and FFR_{TFS} are shown as **Figure 1.3C** and **1.3D**, respectively).

The first study that observed human FFRs dated back to the 1970s and showed that phase-locked responses to pure-tones above 200 Hz can be obtained via scalp-recorded EEG (Moushegian et al., 1973). The first human FFRs elicited by speech stimulus were obtained by Galbraith et al. (1995). It reported that FFRs, which were elicited by word stimuli in young normal-hearing listeners, were perceived as intelligible speech when they were reproduced as auditory stimuli. This showed that speech-evoked FFR can reflect the fidelity of speech encoding in the brain. The speech-evoked FFR has since then been studied over the past 25 years and has been shown to be present across the lifespan from infants to aging adults (see reviews and tutorials: Skoe and Kraus, 2010; Krizman and Kraus, 2019; Coffey et al., 2019). While neural phase-locking to Slow-ENV reflects brain processing of speech at the cortical level, FFRs reflect the neural processing primarily in the brainstem (Chandrasekaran and Kraus, 2010; Bidelman, 2018), although recent research found that FFRs could also have neural sources in the auditory cortex (Coffey et al., 2016).
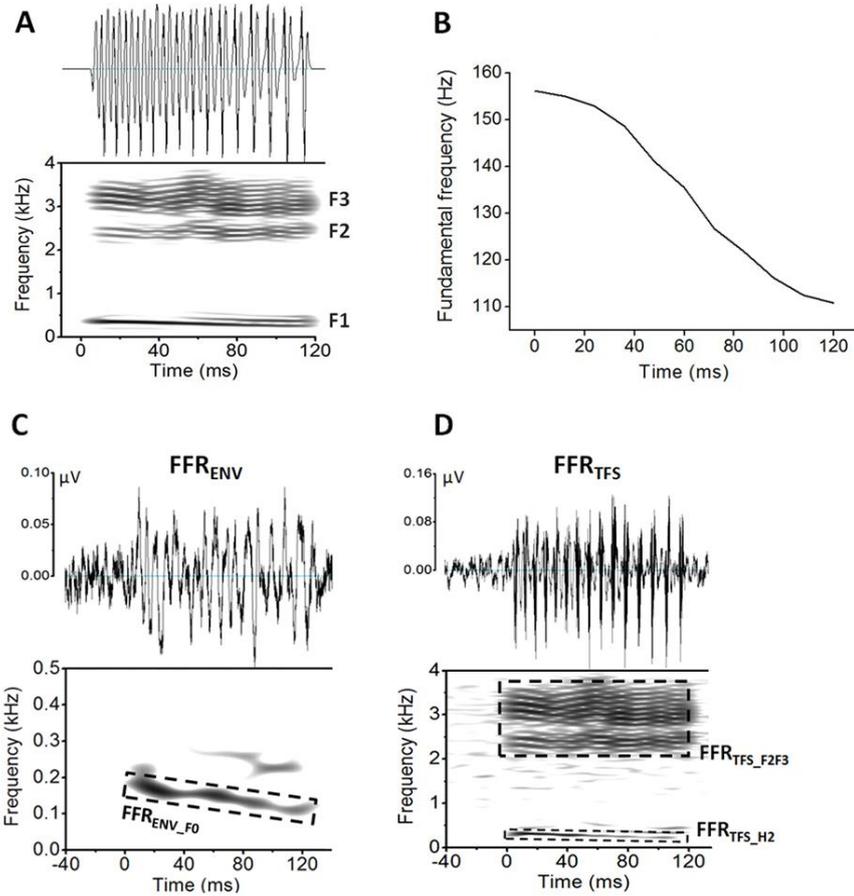
Figure 1.3 Example of FFR that is elicited by a repeatedly presented vowel /i/. A) The waveform of the vowel (top) and the corresponding spectrogram (bottom) which shows three formants F1, F2 and F3 within the range between 0.1 and 4 kHz. **B)** F_0 contour of the vowel showing that it has a falling pitch from ~160 to 110 Hz. **C)** FFR_{ENV} obtained by adding responses to sweeps with positive/original and negative polarities. A falling trend of FFR_{ENV_F0} that resembles the pitch contour can be clearly seen in the spectrogram (indicated by a dashed box). **D)** FFR_{TFS} obtained by subtracting responses to sweeps with the two polarities. The dashed boxes indicate the responses to H2 (FFR_{TFS_H2}, the second harmonic that represents F1) and F2 and F3 (FFR_{TFS_F2F3}). The graphs are taken from Mai et al. (2018).

Echoing the important role of acoustic F_0-ENV and TFS for speech perception in noisy listening environments, it has been found that speech-evoked FFRs may play important roles in speech-in-noise perception (Anderson et al., 2011; Song et al., 2011; Parbery-Clark et al., 2011; Fujihira and Shiraishi, 2015). Specifically, Song et al. (2011) and Parbery-Clark et al. (2011) found that the strength of speech-evoked FFRs at F_0-ENV (FFR_{ENV_F0}) correlated significantly with performances of word recognition in speech in noise in young normal-hearing adults. Such results were replicated in older normal-hearing adults (Anderson et al., 2011). Fujihira and Shiraishi (2015) also tested the relationship between speech-evoked FFR and speech perception in older normal-hearing adults, but in reverberant environments. They found that

magnitude of FFR_{TFS} at around the formant frequency of the vowel stimulus correlated significantly with performances of word recognition of speech with reverberation.

As well as the important association with speech-in-noise perception, FFR is also an important index for various hearing and language functions. First, since it reflects encoding of F_0-ENV, FFR is a neural index for pitch perception. For example, FFR representations of Mandarin tones (neural tracking of lexical tone contours in Mandarin) were stronger in native Mandarin speakers than English speakers (who are non-tonal language speakers), indicating that FFR can reflect the perception of linguistic pitch that is differentiated by linguistic experience (Krishnan et al., 2004, 2005, 2009). Another example is that FFRs can be strengthened by long-term musical experience that is related to better pitch perception (Musacchia et al., 2007; Strait et al., 2009; Bidelman et al., 2011). While the strength of FFRs is modulated by musical experience, musicians who are non-tonal language speakers (e.g., English) have better FFR representations of lexical tone contours than non-musicians (Wong et al., 2007). Furthermore, FFRs can be used to predict various auditory, language and cognitive disorders. For instance, due to the capacity to evaluate neural fidelity of complex acoustic stimuli, abnormal FFRs have been argued to be associated with hearing deficits such as cochlear synaptopathy (e.g., Encina-Llamas et al., 2019) and auditory processing disorders (e.g., Schochat et al., 2017). FFRs are also suggested to be biomarkers for learning disorders and cognitive impairments in children, such as learning difficulties in literacy (Cunningham et al., 2001; Banai et al., 2007; White-Schwoch et al., 2015), dyslexia (Hornickel et al., 2013) and autism (Russo et al., 2008), arguing for functional impairments at the brainstem level along with these problems.

1.3 Introduction of the thesis

This thesis focuses on these critical neural phase-locked responses (low-frequency phase-locked responses and FFRs to high-frequency acoustic properties[1]) and aims to investigate how they interact with various factors during our everyday life and across the lifespan. This thesis focuses on the following factors: (1) background noise, aging and hearing loss (Chapter 2); (2) state of arousal (Chapter 3); and (3) auditory cortical neural excitability (Chapter 4).

1.3.1 Effects of aging and hearing loss on speech-evoked phase-locked responses and their impacts on SiN perception

---

[1] The term 'FFR' conventionally refers to phase-locked responses to acoustic signals oscillate at relatively high frequencies (at or above periodicity, normally >100 Hz). However, low-frequency phase-locked responses are also frequency following responses but occur at the low-frequency range of Slow-ENV (<10 Hz). Throughout this thesis, the term 'FFR' specifically refers to this conventional meaning separately from the 'low-frequency phase-locked responses'.

Older adults often experience increased difficulty with speech-in-noise (SiN) perception compared to young adults (Hume and Dubno, 2010). Hearing loss, which affects 12 million people across the UK (most of whom are older adults according to Action on Hearing Loss), further worsens SiN perception. Daily communications, such as meeting up with friends and family, following conversations in crowded public places, are often difficult for older adults. Factors governing this SiN difficulty due to aging and hearing loss have been studied extensively. For example, aging and hearing loss is related to declines in cognitive functions such as working memory and attention (Lin et al., 2013). Such declines in, e.g., attention ability, can lead to reduced ability to ignore distractive auditory information (Andres et al., 2006) and poor processing of target speech sounds in noisy environments (Tun et al., 2002, 2009). There are also other factors related to SiN perception that older adults have poorer ability such as frequency selectivity (Sommers and Gehr, 1998), sensitivity of temporal fine structure (TFS) information (Hopkins and Moore, 2011; Fullgrabe et al., 2015) and gap detection (Schneider and Hamstra, 1999; Humes et al., 2009; Harris et al., 2010).

Chapter 2 focuses on how aging and hearing loss affects speech-evoked phase-locked responses to Slow-ENV (low-frequency phase-locked responses), F_0-ENV and TFS (FFRs), which have been shown to take important roles in speech perception.

For low-frequency phase-locked responses to Slow-ENV which reflect cortical processing of speech, previous research showed that it is enhanced in older compared to young adults (Presacco et al., 2016a). Such findings are consistent with previous research showing that theta-band phase-locking in response to amplitude-modulated tones increases with age (Tlumak et al., 2015; Goossens et al., 2016). This is also consistent with findings that demonstrated increased auditory-evoked responses in older adults compared to young adults (Alain et al., 2014; Herrmann et al., 2013, 2016). It has been shown that greater low-frequency phase-locked responses to auditory stimuli reflect greater neural firing (Ng et al., 2013) and haemodynamic responses in the auditory cortex (Oya et al., 2018). This indicates that increases in low-frequency phase-locking could reflect hyperexcitability of auditory cortex in older adults (Caspary et al., 2008). It has been argued that such hyperexcitablity may change the inhibitory and excitatory balance between auditory and cognitive processes, i.e., lower-level auditory hyperexcitability may hinder allocation for higher-level cognitive resources that could further impair SiN perception (Presacco et al., 2016a).

For phase-locked responses to F_0-ENV and TFS (FFRs), previous studies have shown that older adults have smaller FFR magnitudes compared to young adults (FFR_{ENV_F0} and FFR_{TFS}, Anderson et al., 2012; FFR_{ENV_F0}, Presacco et al., 2016a). This is argued to be one of the most important reasons that older adults have worsened temporal precision in brainstem processing of speech signals that leads to impaired SiN perception (Anderson et al., 2012). Furthermore, FFRs are associated with SiN perception in older adults (Anderson et al., 2011; Fujihira and Shiraishi, 2015). Specifically, greater FFR_{ENV_F0} magnitude was associated with better SiN perception with speech-shaped noise (Anderson et al., 2011). Greater magnitude of FFR_{TFS} in the resolved harmonics region has also been associated with better SiN perception

when there is reverberation (Fujihira and Shiraishi, 2015). These indicate that age-related changes in FFRs may be related to impaired SiN perception. Hearing loss, on the other hand, was found to result in reduced neural inhibition that leads to greater encoding of F_0-rate envelope modulations in both animals (Kale and Heinz, 2010; Henry et al., 2014; Zhong et al., 2014) and humans (Anderson et al., 2013; Goossens et al., 2019). This exaggerated neural encoding has been argued to act as a distraction from neural processing of other important acoustic features (such as temporal fine structures) (Kale and Heinz, 2010; Henry et al., 2014). These findings together indicate that aging and hearing loss may influence SiN perception by distinct mechanisms.

Despite these reported effects of age and hearing loss on neural phase-locked responses, it is still unclear at present how these changes are associated with impaired SiN perception. Presacco et al. (2016a) argued that increased low-frequency phase-locking to speech Slow-ENV in older adults may reflect a loss of excitation-inhibition balance which may, as a result, impair SiN perception. This is, however, not in line with the results that have been reported showing that greater low-frequency phase-locking to speech is positively related to better speech perception (Ahissar et al., 2001; Luo and Poeppel, 2007; Peelle et al., 2013; Doelling et al., 2014). These findings are consistent with other studies showing that greater magnitudes of cortical auditory evoked potentials (CAEPs) can predict better SiN perception in both young and older adults (Billings et al., 2015). Also, Presacco et al. (2016a) did not find correlations between low-frequency phase-locking and SiN performances. Indeed, the lack of correlation in Presacco et al. (2016a) may be due to different types of background noise used for the neural recording and when SiN performances were measured (single-talker background for neural recording and the background of four-talker babble noise in SiN perception tasks; see Presacco et al. (2016a)).

For speech-evoked FFRs, although strengths of FFRs are associated with SiN perception in older adults (Anderson et al., 2011; Fujihira and Shiraishi, 2015), there is no definitive evidence that has clarified how age effects and/or effects of hearing loss on FFRs are related to impaired SiN perception. For instance, recent studies (Presacco et al. 2016a; Schoof and Rosen, 2016) tested the relationship between FFR magnitudes and SiN perception in young (< 30 years old) and older (> 60 years old) adults. Age-related declines in FFRs were shown in these studies which was consistent with previous reports (Anderson et al., 2012), but neither study found significant correlation between FFR and SiN perception. Furthermore, older participants in these studies (Presacco et al. 2016a; Schoof and Rosen, 2016) all had relatively normal-hearing (thresholds < 30 dB HL at frequencies ≤ 4 kHz), hence it did not provide information about effects of hearing loss. A more recent study by Presacco et al. (2019) recruited three groups of participants of young normal-hearing adults, older normal-hearing adults and older adults with hearing loss to disentangle the effects of aging and hearing loss. Consistent with previous research, this study found that aging can lead to greater low-frequency phase-locked responses to speech Slow-ENV and smaller FFR magnitudes; no effects of hearing loss were observed for either type of the phase-locked responses. Also, similar to

previous studies, this study did not find significant correlations between neural phase-locked responses and SiN performances.

It is thus still unclear nowadays how the effects of aging and hearing loss on speech-evoked phase-locked responses are associated with SiN perception. To address this issue, Chapter 2 (Study 1) of this thesis investigated low-frequency phase-locking and FFRs during SiN perception in both young and older adults. Study 1 examined young and older adults over a wide age range (19–75 years), where older adults had hearing ranging from normal to mild-to-moderate hearing loss. This, to a greater extent compared to previous studies in which all participants were relatively normal-hearing (Anderson et al., 2011, 2012; Presacco et al. 2016a; Schoof and Rosen, 2016), resembled the ecological distributions of hearing in aging populations in the real society (Gopinath et al., 2009; Humes et al., 2010). Statistical analyses disentangled the effects of aging and hearing loss and tested how effects of aging and hearing loss on speech-evoked phase-locked responses are associated with SiN perception. Furthermore, compared to previous studies that used different types of background noise in the neural (i.e., when phase-locked responses were recorded) and behavioural tasks (i.e., SiN perception tasks) (Presacco et al. 2016a, 2019), the present Study 1 was conducted with a better design in which the same types of background noise were used for the two tasks.

1.3.2 Arousal state and its possible effects on speech-evoked phase-locked responses

Arousal, or consciousness, is an important physiological and psychological status in our everyday life across the lifespan (Picchioni et al., 2014). Arousal is associated with a gating mechanism in the thalamus that controls the flow of sensory information from lower-level systems (periphery and brainstem) to the cortex (Steriade et al., 1993; McCormick and Bal, 1994; 1997). Reduced arousal leads to sensory deafferentation in terms of reduced thalamocortical connectivity that affects the brain's perception of input stimuli (Spoormaker et al., 2010, 2011; Picchioni et al., 2014). Studying the effect of arousal is thus a good way to understand how sensory systems (from brainstem to sensory cortex) serve as the 'gate' to regulate the brain's ability to process sensory inputs that shape human perceptions.

In the auditory domain, auditory signals can be processed by the brain in low arousal states (such as during sleep) (Issa and Wang, 2008; Nir et al., 2015). These studies showed that neural processing of sounds in primary auditory cortex of mammals during sleep is comparable to responses during wakefulness. However, in humans, neural responses to auditory stimuli can be reduced during low arousal compared to high arousal states in subcortical (Portas et al., 2000) and cortical regions (Czisch et al., 2002, 2004; Davis et al., 2007; Wilf et al., 2016). Portas et al., (2000) was the one of the first studies to investigate how neural responses to auditory stimuli change according to changes in arousal state in human participants. This study used beep sounds and participants' first name as stimuli and tested fMRI signals in normal-hearing adults during wakefulness and sleep BOLD responses were significantly reduced during sleep compared to wakefulness in the thalamus and parietal, frontal

and cingulate cortical regions. Subsequent studies by Czisch et al. (2002, 2004) used speech (read text) stimuli and showed reduced BOLD responses in the auditory cortex during sleep compared to wakefulness. Davis et al. (2007) used noise stimuli (signal-correlated noise) and meaningful speech with different semantic ambiguity and tested the changes in BOLD responses according to different levels of sedation (wakefulness, light sedation and deep sedation). The effects of semantic ambiguity were present at higher-order temporal and frontal language regions only during wakefulness but not during sedation. The speech-specific effect (responses comparing speech vs. noise) was significantly greater during the light, than the deep, sedation and responses to speech decreased with the sedation level in various temporal, parietal and frontal areas (Davis et al., 2007). A more recent study by Wilf et al., (2016) used speech materials with different linguistic hierarchical contents (comprehensible speech, pseudo-words and scrambled speech). It investigated how BOLD responses change according to arousal states (wakefulness vs. sleep). It also investigated the interaction between arousal and linguistic hierarchy, i.e., whether arousal affected BOLD responses differently across stimuli with different linguistic hierarchical contents. There was significantly reduced responses during sleep than wakefulness in the primary auditory cortex for all types of stimuli and in high-level linguistic regions (Wernicke's and Broca's areas) for phonetically/phonologically valid speech (comprehensible and pseudo-word speech). Significant interactions between the linguistic hierarchy and the arousal state were found in the high-level linguistic regions but not in the primary auditory cortex. These studies (Portas et al., 2000; Czisch et al., 2002, 2004; Davis et al., 2007; Wilf et al., 2016) thus confirmed that brain processing of auditory stimuli, especially speech signals, is significantly affected by the level of arousal in different cortical/subcortical regions that are responsive to speech and language.

Speech-evoked phase-locked responses, which are the focus of the current thesis, also change according to changes in arousal states. Makov et al. (2017) studied relationships between speech-evoked phase-locked responses measured via EEG at different linguistic levels (syllables, words, phrases and sentences) and arousal states (wakefulness vs. sleep). Phase-locked responses at rates corresponding to higher-order linguistic units (words, phrases and sentences) were statistically greater in wakefulness than in sleep, but not at the rates corresponding to those of lower-order units (syllables). Despite this result, decreases in phase-locked responses at syllable rates (similar to phase-locked responses to speech Slow-ENV) were still seen (Makov et al., 2017). Due to the small sample size (~15 participants) in this study, it is not clear whether a statistical effect could be present in a study with an adequate number of participants. Also, currently evidence is lacking concerning whether phase-locked responses to more fine-grained speech acoustic attributes (such as FFRs to F_0–ENV cues) are affected by arousal. Furthermore, it is unclear whether the effect of arousal can be influenced by other important factors such as age which affects speech-evoked phase-locked responses (as discussed in **1.3.1**) and the properties of sleep status.

Chapter 3 (Study 2) examined the effects of arousal on speech-evoked phase-locked responses. It also considered other factors that are important for processing phase-locked

responses at different levels of arousal. These factors included: (1) age; and (2) sleep spindles. Sleep spindles are bursts of oscillatory neural activity at frequencies of 12–16 Hz occurring at Stage 2 sleep, i.e., non-rapid eye movement (nREM) sleep (Warby et al., 2014) that last successively for at least 0.5 seconds (De Gennaro and Ferrara, 2003). Occurrence of spindles can distinguish the stages of wakefulness and nREM/light sleep, i.e., eye-closed wakefulness and, the stage between wakefulness and deep sleep (Warby et al., 2014). Spindle activity, such as spindle density (frequency of occurrence of spindles across time), has been used to indicate the level of arousal and sleep stability (Kim et al., 2012). Spindles are transmitted to the cortex over thalamo-thalamic and thalamo-cortical loops where they modulate neural sensitivity to auditory stimuli (Dang-Vu et al., 2011; Schabus et al., 2012). Furthermore, the properties of spindles are influenced by age, where the magnitude, duration and density of spindles decrease in older compared to young adults (Martin et al., 2013; Mander et al., 2017). Therefore, sleep spindles should modulate auditory activity at cortical and subcortical levels and the neuromodulation could differ across ages.

Speech-evoked phase-locked responses originate from both auditory cortical (low-frequency phase-locking to Slow-ENV; Peelle and Davis, 2012) and subcortical (FFRs to F_0-ENV; Chandrasekaran and Kraus, 2010; Bidelman, 2018) regions. Auditory activities in these regions are affected by arousal (Portas et al., 2000; Czisch et al., 2002, 2004; Davis et al., 2007; Wilf et al., 2016). Study 2 examined the links between arousal, sleep spindle density and the speech-evoked phase-locked responses in adults across a wide age range (19–75 years old). It thus aimed to elucidate the effects of arousal states and how sleep spindle properties can regulate early-stage speech processing in the brain across the lifespan.

1.3.3 Effect of cortical neural excitability on speech-evoked FFRs

One of the most important topics in FFR research is the neural origins of FFR and the relationship between FFR and the cortical activity (Coffey et al., 2019). It has been argued that the main neural sources of FFRs are in the inferior colliculus (IC) at the brainstem (Chandrasekaran and Kraus, 2010; Bidelman, 2015, 2018). This argument is long evidenced by the fact that the short latency of FFR captured by electrophysiological recordings such as EEG (usually between 5 to 10 ms) is consistent with the first spike latency in IC (Langner and Schreiner, 1988). Earlier studies also provided evidence that neural deactivations in IC can lead to eradication of FFR. For example, an animal study showed that cryogenic cooling can lead to disappearance of FFRs that can be previously observed in both IC and scalp before the cooling (Smith et al., 1975). In humans, it was found that FFR disappeared after focal lesions of IC (Sohmer and Pratt, 1977). Recent efforts using source localisation for FFR recorded via scalp EEG in humans showed that the main source of FFR is IC (Bidelman, 2015, 2018).

Some other recent studies, on the other hand, showed additional sources of FFR at the cortical level in humans (Coffey et al., 2016, 2017a). Coffey et al. (2016, 2017a) used MEG to localize FFRs. They showed that, besides IC, FFR also has sources in the right primary auditory cortex that are associated with musical experience, pitch discrimination ability and speech-in-

noise perception. FFR strength was further shown to be associated with right-lateralized auditory cortical activity (Coffey et al., 2017b). This study (Coffey et al., 2017b) combined EEG that recorded FFRs and fMRI that measured cortical activations. It was shown that FFR strength was correlated with BOLD responses in the right auditory cortex which were replicated with two different acoustic stimuli (a speech syllable and a musical note). Although this study did not provide evidence that FFR has additional origins in the cortex, it emphasized the close relationship between FFR and hemodynamic responses in the auditory cortex that has not been illustrated by previous research. These results are thus consistent with the relative specialization of right auditory cortex for pitch and tonal processing (Zatorre and Berlin, 2001; Patterson et al., 2002; Hyde et al., 2008; Albouy et al., 2013; Cha et al., 2016) that is reflected by the strength of FFRs (Musacchia et al., 2007; Strait et al., 2009; Bidelman et al., 2011).

The previous findings have therefore demonstrated the potential cortical contributions to FFRs, by either localising the neural origins in the auditory cortex (Coffey et al., 2016, 2017a) or using multimodal imaging that assessed the correlation between the FFR and neural activations in the auditory cortex (Coffey et al., 2017b). Despite these results, a further important question is whether such contributions are causal, which has not been clarified. Chapter 4 (Study 3) aimed to establish whether there is a causal relationship between auditory cortex and speech-evoked FFRs. It applied a non-invasive brain stimulation, transcranial direct current stimulation (tDCS), to change the neural excitability in the auditory cortex and tested for the after-effects of tDCS on the strengths of speech-evoked FFRs. tDCS is a brain stimulation technique that changes the cortical excitability (Jacobson et al., 2012). tDCS can cause depolarization (via anodal) and hyperpolarization (via cathodal) of neurons by applying direct currents over the scalp. This leads, respectively, to neural excitation and inhibition in proximal parts of the cortex (Nitsche and Paulus, 2001). Previous studies found that applying tDCS over the auditory cortex can significantly change performances of pitch discrimination (Mathys et al., 2010; Matsushita et al., 2015). Mathys et al. (2010) found that cathodal tDCS over both the left and right auditory cortices can impair pitch discrimination ability compared to sham stimulation, with the effects being significantly stronger in the right hemisphere than in the left. Matsushita et al. (2015) studied how tDCS can affect the learning process of pitch discrimination. They showed that anodal stimulation on the right auditory cortex adversely affected the learning effect compared to sham. These results thus support the causal role of the right auditory cortex for pitch perception. However, such causality has not been established for neurophysiological signatures like FFRs. Hence, this was investigated in Study 3 of this thesis. Indeed, using the approach of brain stimulation cannot confirm the neural sources of FFRs in the cortex. However, the causal contributions at the cortical level could be established to advance our understanding of how FFRs are associated with the auditory cortical processing and cortical lateralization of pitch and speech perception.

In sum, this thesis addresses the neural phase-locked responses to different levels of speech attributes (Slow-ENV, F_0-ENV and TFS) and different stages in the auditory systems (subcortical and cortical levels) that are important for speech perception. Crucially, this thesis will illustrate how these responses are influenced by important factors during our everyday life

across the lifespan and how these influences may relate to speech perception. These factors include internal factors such as aging, hearing loss and changes in physiological status like state of arousal and neural excitability as well as an external factor of background noise. The thesis should thus provide us with a better understanding of how speech is perceived in our daily lives.

Chapter 2

Relationship between speech-evoked phase-locked neural responses and speech-in-noise perception in young and older adults

2.1 Introduction

Older adults often suffer from understanding speech in noisy listening environments even in those who have normal hearing (Hume and Dubno, 2010). The present study (Study 1 of this thesis) focused on the recent claim that impaired SiN perception in older adults is due to degraded temporal neural encoding of speech sounds (Anderson et al., 2011, 2012; Presacco et al., 2016a).

Speech-evoked phase-locked responses, including theta-band phase-locking and FFRs typically represent the temporal neural encoding of speech sounds. Theta-band phase-locking reflects cortical tracking and/or evoked responses to amplitude variations of speech Slow-ENV and is associated with speech perception (Luo and Poeppel, 2007; Howard and Poeppel, 2010; Peelle et al., 2013). Changes in theta-band phase-locking to Slow-ENV have also been shown to have causal relationship with speech intelligibility (Zoefel et al., 2018, 2020; Riecke et al., 2018; Wilsch et al., 2018; Keshavarzi et al., 2020; Keshavarzi and Reichenbach, 2020). Theta-band phase-locking has been found to increase with age (Tlumak et al., 2015; Goossens et al., 2016) that could reflect the hyperexcitability in the auditory cortex (Caspary et al., 2008). The hyperexcitability may alter the balance between inhibitory and excitatory neural processes in older adults that changes network connectivity and over-represents speech envelopes relative to other speech features (Presacco et al., 2016a). Subcortically, speech-evoked FFRs that originate primarily from the auditory brainstem (Chandrasekaran and Kraus, 2010; Bidelman 2018) require precise temporal processing of F_0 (FFR_{ENV_F0}) and TFS for higher harmonics in speech (FFR_{TFS}) (Aiken and Picton, 2008; Skoe and Kraus, 2010). FFR magnitudes decrease with age (Anderson et al., 2012; Presacco et al., 2016a) and greater FFR magnitudes are associated with higher SiN accuracies in older adults (Anderson et al., 2011; Fujihira and Shiraishi, 2015).

Besides the cortical and subcortical responses to speech, functional connectivity between phase-locked cortical and subcortical activities may also play an important role in SiN perception. Previous research showed that greater afferent connectivity between auditory cortex (cortical evoked potentials) and brainstem (FFRs) in responses to speech is associated with

better SiN perception (Bidelman et al., 2018). Furthermore, such functional connectivity was shown to be poorer for older than for young adults (Bidelman et al., 2019). This indicates that the effects of aging on cortico-subcortical connectivity may make impact on SiN perception.

At present, however, there has been no definitive evidence to show whether changes in speech-evoked phase-locked responses according to aging are related to poor SiN perception. Previous research used both young and older adults failed to find relations between speech-evoked responses and SiN performances (Presacco et al., 2016a) partly because different types of background noise were used for neural recording and SiN perception tasks. Also, previous studies (e.g., Anderson et al., 2011; Schoof and Rosen, 2016; Presacco et al., 2016a) were conducted with older adults with relatively normal audiometric hearing. Only using normal-hearing adults in these studies thus had not represented the wide range of hearing losses typically observed in aging populations (Gopinath et al., 2009; Humes et al., 2010).

The present study addressed whether age effects on subcortical/cortical phase-locked encoding of speech were associated with impaired SiN perception. Behavioural and neural assessments were conducted in healthy adults across a wide age-range (19–75 years). Older adults in the present study had audiometric thresholds at frequencies between 2 and 4 kHz indicative of normal hearing to mild/moderate hearing loss. Therefore individual variability associated with peripheral hearing losses that occur during normal aging was present in the sample (Gopinath et al., 2009; Humes et al., 2010). For the behavioural assessment, participants completed SiN perception tasks under two types of background noise: steady-state speech-shaped noise (SpN) and 16-talker babble noise (BbN). For the neural assessments, participants listened to a repeated syllable under the same types of noise as in the behavioural assessment, whilst speech-evoked phase-locked activity was recorded at both cortical (theta-band PLV) and subcortical (FFRs) levels using scalp-electroencephalography (EEG). SiN perception and the neural signatures were compared across the two age groups and multiple linear regressions were conducted to investigate whether the age-related neural signatures were associated statistically with SiN perception.

Based on past evidence, it was predicted that older, relative to young, adults would have: (1) smaller subcortical (FFRs) magnitudes (Anderson et al., 2012; Presacco et al., 2016a); (2) greater cortical (theta-band PLV) phase-locked responses to speech (Presacco et al. 2016a); and (3) decreased SiN perception (Hume and Dubno, 2010). The predictions for testing the hypotheses that age effects on neural measures relate to behavioural performance are that decreased SiN perception with age should be statistically associated with: (1) reduced FFR magnitudes, (2) greater theta-band PLV, and (3) reduced cortico-subcortical connectivity. At the same time, this study explores which neural (cortical and/or subcortical) signatures optimally model SiN perception which is an issue that is not clear to date.

2.2 Methods

The present study followed the same procedure and used parts of the older adults' data from Mai et al. (2018)[2].

2.2.1 Participants
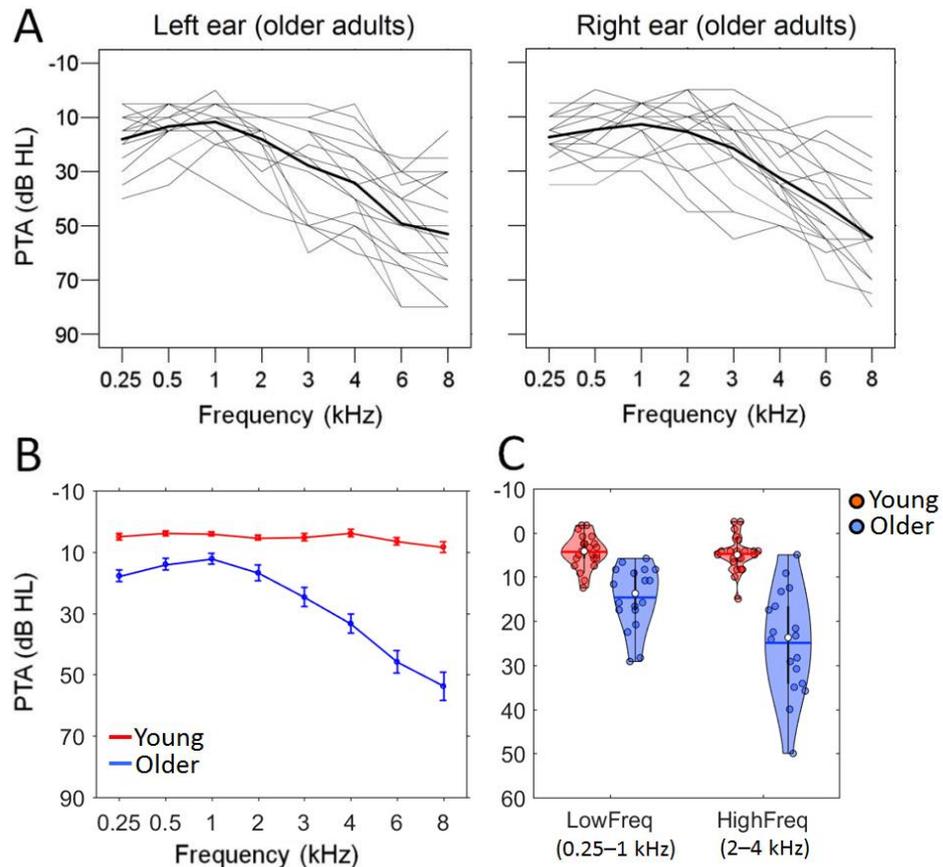
Figure 2.1 Audiograms of participants. (A) *Individual pure-tone audiograms (PTA) for the older group in both ears for the range from 0.25 to 8 kHz. The bold lines represent the grand-averages across participants[3].* ***(B)*** *PTAs for the young (red) and older (blue) groups averaged across ears. Data for 6 and 8 kHz (dashed lines) were not used in the subsequent statistics because spectral distribution of the speech stimuli used in the present study only extended to 4 kHz. Error bars indicate the standard errors of the means.* ***(C)*** *The violin plots[4] for PTAs at the low (0.25–1 kHz) and high frequency (2–4 kHz) ranges.*

---

[2] The present study used the data of non-hearing-aid older adults in Group 2 of Mai et al. (2018), where participants listened to the same acoustic stimuli as in the present study whilst neural recordings were made. Data from hearing aid users in Mai et al. (2018) were not included here, since PTAs could not be measured precisely in these participants and hearing aids may introduce additional effects.

[3] Three older participants had PTAs that were higher than the measurable limit of the audiometer (85 dB) at 8 kHz (one in the left ear and two in the right ear) and thresholds for them were set at 85 dB when calculating the grand-averages.

[4] Violin plots throughout this thesis indicate the distribution of datasets. In each plot, the white circle represents the median points; the vertical (in black) and horizontal (coloured) lines represents the 1.5 times the interquartile range and the mean value of the dataset, respectively.

Participants comprised 23 young (19-42 years; Mean ± SD = 26.3 ± 5.5 years; 15 males) and 18 older adults (53-75 years; Mean ± SD = 67.0 ± 5.6 years; 7 males). All were native UK English speakers with no reports of neurological diseases, language-related or psychiatric problems. **Figure 2.1** shows the pure-tone audiometric (PTAs) data for frequencies 0.25–8 kHz measured using an MA41 Audiometer (MAICO Diagnostics, Germany). All young participants had normal hearing (PTA ≤ 25 dB HL). In older participants, inter-individual variability was high particularly at frequencies of 2 kHz and above (see individual curves in **Figure 2.1A**). The older adults showed significantly higher low-frequency PTAs (PTA_{Low}; averaged across 0.25–1 kHz) and high-frequency PTAs (PTA_{High}; averaged across 2–4 kHz)[5] compared to the young group (both $p < 10^{-6}$). A two-way mixed-design ANOVA was conducted for PTA with factors of Frequency (PTA_{High} vs. PTA_{Low}) and Age Group (young vs. older). A significant [Frequency × Age Group] interaction occurred ($F(1, 39) = 12.579$, $p = 0.001$), indicating that older adults had significantly greater declines in hearing at the high compared to the low frequencies, reflecting the typical characteristic of age-related sensorineural hearing loss at high frequencies. Also, the boxplot (**Figure 2.1C**) indicated that high-frequency hearing in the older group ranged from normal (≤ 25 dB HL) to mild-to-moderate (25–50 dB HL) hearing loss, comparable with the distribution pattern reported in other older samples (Gopinath et al., 2009; Humes et al., 2010). Since PTAs differed across frequencies and age groups, PTA_{Low} and PTA_{High} were used as well as PTA averaged across the wider frequency range (0.25–4 kHz; PTA_{Wide}) as separate covariates and predictors during statistical analyses (see *2.4*).

2.2.2 Behavioural experiment

SiN perception tasks involved participants listening to BKB sentences (Bench et al., 1979) under two types of background noise: steady-state speech-shaped (SpN) and 16-talker babble (BbN) noise. All sentences were pre-recorded utterances spoken by a male British English speaker whose absolute range in F_0 spanned from ~80 to ~200 Hz. Each sentence included three key (content) words, e.g., "The clown has a funny face" with key words "clown", "funny" and "face". BbN was a mixture of 16 different utterances spoken by 16 male British English speakers with similar voice quality to the target speaker. SpN was formed by randomizing the phases of the long-term spectrum of BbN and transforming the spectrum back to the time domain. As a result, SpN has the same long-term power spectrum as BbN and stable time-domain properties (Rosen et al., 2013).

Participants were seated comfortably in a sound-treated booth facing a Fostex 6301B loudspeaker (Canford Group Ltd.) at zero-degree azimuth. Distance between the loudspeaker and participants' ears was constant at 1 meter. After eight trials of practice, participants listened to two different sets of 30 sentences (for the backgrounds of SpN and BbN, respectively) at an intensity at this distance of 70 dB SPL. Participants repeated as many words as they could from each sentence. Sentences were presented via Matlab 2010a (Mathwork, USA) and SNR varied

---

[5] Since the spectral distribution in the speech stimulus used in both behavioural and neural assessments extended to ≤ 4 kHz, PTAs at 6 and 8 kHz were not included in the statistical analyses in the present study.

adaptively to track for the speech reception threshold (SRT, Plomp and Mimpen, 1979) at which 50% of words were correct. For each background type, the first sentence was played at a relatively high SNR (8 and 10 dB for SpN and BbN, respectively). SNR was decreased by 4 dB for subsequent sentences until < 50% words correct (i.e., < 2 words) were reported. SNR was then increased/decreased by 2 dB when word correctness was less/more than 50% in each of the following sentences. SRT was calculated by linear interpolation using the two SNRs which had > 50% and < 50% correct across the minimal step distance (i.e., 2 dB).

2.2.3 EEG experiment

Acoustic stimuli

Participants listened to a repeatedly-presented, 120-ms-long /i/ syllable produced by a male speaker (**Figure 2.2A**). The F_0 contour of the syllable fell from ~ 160 to ~110 Hz (**Figure 2.2B**). The F_0 contour covered a similar frequency range and direction of change as those in the F_0s of the target speaker in the BKB sentences used in the SiN perception tasks (BKB sentences are narratives that generally have a falling F_0 contour). The three formants in the syllable were at ~ 280 Hz (F1), ~ 2400 Hz (F2) and ~ 3100 Hz (F3). The amplitude envelope profile was stable except that 5-ms-long rising and falling cosine windows were applied at the onset and offset to avoid transients.
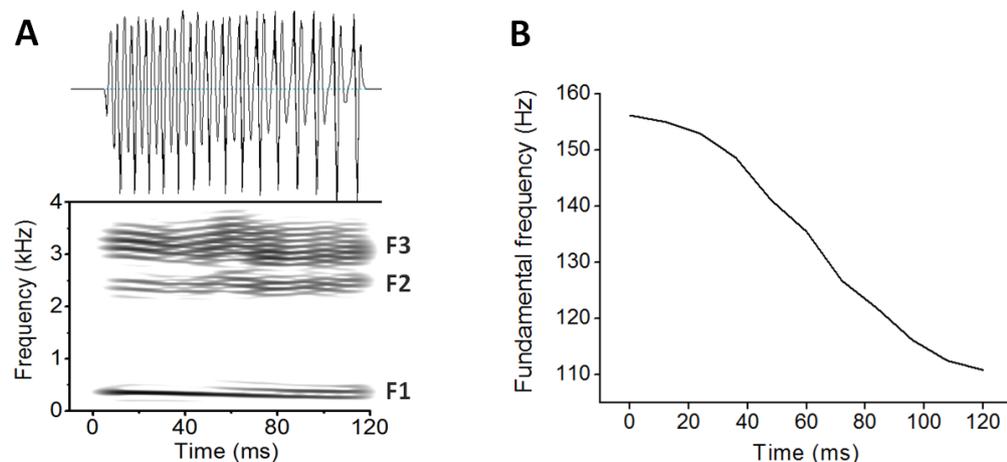
Figure 2.2 The syllable /i/ used during EEG recording. (A) The temporal waveform (top) and spectrogram (bottom) of the syllable. F1, F2 and F3 frequencies are around 280, 2400 and 3100 Hz, respectively. **(B)** The falling F_0 contour ranging from around 160 to 110 Hz obtained by autocorrelation. The waveform, spectrogram and F_0 were generated via PRAAT (Boersma and Weenink, 2013).

The syllable was presented repeatedly at both original (positive) and inverted (negative) polarities in random order with inter-stimulus intervals (ISIs) that varied randomly between 60 and 120 ms (syllable repetition rate was approximately 5 syllables per second). The stimuli were

presented in quiet, SpN and 16-talker BbN backgrounds (the last two were the same backgrounds that were used in the SiN perception tasks). The SNRs were set at -1 dB, which led to neural responses that correlated significantly with SiN perception in older adults (Mai et al., 2018). There were 6400 sweeps under each background type (3200 sweeps for each polarity). Recordings at each background type were split into 16 segments of equal duration giving 48 segments in total with 400 sweeps per segment. The segments were played in succession in an intermixed order.

EEG data acquisition

Scalp-EEGs were recorded on an ActiveTwo system (Biosemi, The Netherlands) at a sampling rate of 16384 Hz. Three active electrodes were placed at Cz (vertex), C3 and C4 according to the 10/20 configuration. Cz was used to obtain FFRs (Skoe and Kraus, 2010). Cortical responses were measured on C3 and C4 that reflects activity in the auditory cortex (Carpenter and Shahin, 2013; Noguchi et al., 2015) and allows reliable cortical phase-locked activity that is significantly associated with SiN perception to be recorded (Mai et al., 2018). Bilateral earlobes were used as the reference. Ground electrodes were CMS/DRL. Electrode impedance was kept below 35 mV. The experiment was conducted in an electromagnetic-shielded and sound-treated booth. The stimuli were played via a Rogers LS3/5A loudspeaker (Falcon Acoustics, UK) at zero-degree horizontal azimuth relative to participants' heads when they were reclined (the chair was adjustable). The stimulus level (measured across time including ISIs) at the distance between the loudspeaker and participants' ears (constant at 1 meter) was calibrated at 74.5 dB before background noise was added. The stimulus level was at 79.5 dB after either SpN or BbN was added.

Participants were instructed to relax, close their eyes and keep still in order to avoid movement artefacts. They did not have to make any response to the stimuli (passive listening) and they were not stopped from falling asleep. A webcam monitored the participants throughout the test and no significant changes in head or body position were observed. Participants were not stopped from falling asleep because another purpose of the current experiment was to study the effects of arousal on speech-evoked neural processing across ages (Mai et al., 2019). This investigation was separate from the present paper. As Mai et al. (2019) found that arousal significantly affected the phase-locked responses, only EEG data from periods with high arousal were used here (see *2.4.3* for details).

2.2.4 Signal processing for EEG data

The signal processing procedure used Matlab 2014a (Mathwork, USA).

Frequency following responses (FFRs)

EEGs at Cz were re-referenced to the average of bilateral earlobes and bandpass filtered between 70 and 4000 Hz using a zero-phase 2nd-order Butterworth filter. Baseline was adjusted using the pre-stimulus period of 50 ms. Sweeps exceeding ± 25 µV were rejected to

exclude movement artefacts. FFRs with positive (FFR$_{pos}$) and negative (FFR$_{neg}$) polarities were obtained by averaging across sweeps with their respective polarities. FFRs that represent envelope modulations (FFR$_{ENV}$) and TFS (FFR$_{TFS}$) respectively were obtained by addition and subtraction of FFR$_{pos}$ and FFR$_{neg}$ that were then divided by 2 (Aiken and Picton, 2008). **Figure 2.3** shows an example of FFRs obtained in the present study (FFRs of a single participant recorded in BbN).

Three FFR magnitudes were measured: (1) FFR$_{ENV\_F0}$ that represents neural encoding of envelope modulations at F$_0$, quantified as the magnitude along the F$_0$ trajectory using FFR$_{ENV}$ (**Figure 2.3A**); (2) FFR$_{TFS\_H2}$ that represents neural encoding of TFS at the resolved harmonics region (2nd harmonics H2 at 220–330 Hz in the neighbourhood of F1), quantified as the magnitudes along the H2 trajectory using FFR$_{TFS}$ (**Figure 2.3B**); and (3) FFR$_{TFS\_F2F3}$ that represents neural encoding of TFS in the unresolved harmonics region (frequency range around F2 and F3), quantified as the magnitudes along the F2 and F3 trajectories using FFR$_{TFS}$ (**Figure 2.3B**). For (3), it is noteworthy that neural phase-locking ability at such high frequencies of F2/F3 is weak (Verschooten et al., 2019). We cannot rule out the possible contributions of electrical artefacts (generated by the stimulation of acoustic waveforms, see Skoe and Kraus, 2010) to FFR$_{TFS\_F2F3}$[6]. The procedures for spectral magnitude calculations followed Mai et al. (2018).

First, F$_0$-ENV (F$_0$ based on the acoustic envelope), H2, F2 and F3 trajectories of the /i/ syllable were calculated. To obtain the F$_0$-ENV trajectory, a set of 40-ms sliding windows (1-ms per step) was applied to the syllable's Hilbert envelope. Each 40-ms segment was Hanning-windowed, zero-padded to 1 second (to achieve 1 Hz frequency resolution) and Fourier-transformed. The frequency with the highest Fourier magnitude between 110 and 160 Hz (the F$_0$ range) was chosen as the F$_0$ value at each step. H2, F2 and F3 trajectories were obtained in the same way, except that: 1) sliding windows were applied to the syllable rather than the Hilbert envelope; 2) H2 values were selected within the H2 range (220 ~ 320 Hz); 3) instead of choosing values based on the Fourier spectrum after zero-padding, F2 and F3 values were chosen based on the spectral profile via cepstral smoothing (Proakis and Manolakis, 2007) in the F2 and F3 ranges respectively (2200 ~ 2600 Hz and 2800 ~ 3500 Hz). Second, to calculate the FFR$_{ENV\_F0}$ magnitude, the same set of 40-ms sliding windows was applied to FFR$_{ENV}$ with Hanning-windowing, zero-padding and Fourier-transforms. The mean log-magnitude was measured across a 20 Hz bandwidth centered at the frequency of the F$_0$-ENV trajectory at that step. The magnitudes were then averaged across all steps along the F$_0$-ENV trajectory. FFR$_{TFS\_H2}$ and FFR$_{TFS\_F2F3}$ magnitudes were obtained in the same way, except that: 1) the procedure was applied on FFR$_{TFS}$ along the H2 (for FFR$_{TFS\_H2}$) and the F2 and F3 (for FFR$_{TFS\_F2F3}$) trajectories; 2) instead of obtaining magnitudes based on the Fourier spectrum

---

[6] Although efforts were made to try to minimize such contributions by using insert earphones with plastic tubes to transduce sounds to the ears, artefacts could also be generated through hardware circuitry between the computer that presented the acoustic stimuli and the EEG recording system. This is a caveat that was not tested in the present study, which needs to be kept in mind.

after zero-padding, FFR_{TFS_F2F3} magnitude at each step was the summed magnitude of the spectral profile (via cepstral smoothing) across a 150 Hz and 300 Hz bandwidth respectively centred at F2 and F3 of the syllable at that step.
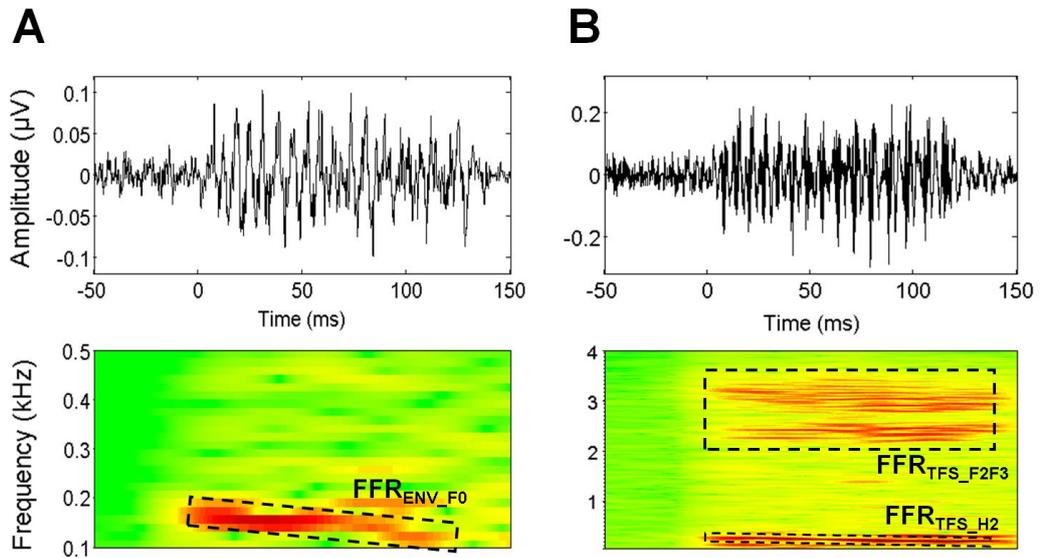
Figure 2.3 FFR waveforms (top) and spectrograms (bottom) of a single participant recorded in BbN. (A) *FFR_{ENV} (bandpass filtered at 70–2000 Hz); **(B)** FFR_{TFS} (bandpass filtered at 70–4000 Hz). The waveforms were based on sweeps with normalized numbers (1450–1550) during the high arousal state (see 2.4.4 and 2.4.5). FFR_{ENV_F0} (at F_0 range between 160 and 110 Hz), FFR_{TFS_H2} (at H2 range between 220 and 320 Hz) and FFR_{TFS_F2F3} (at F2-F3 range between 2000 and 4000 Hz) are indicated by the boxes surrounded by dashes and their labels. '0' corresponds to the syllable onset.*

In addition, neural transmission from the cochlea to the auditory brainstem for FFR_{ENV} takes between 5 and 10 ms (Chandrasekaran and Kraus, 2010; Skoe and Kraus, 2010), while FFR_{TFS} occurs at earlier stages in the auditory periphery (Aiken and Picton, 2008). Hence, the maximum magnitude for time lags in the range 8 to 13 ms and 3 to 8 ms (at 1-ms steps; including an additional 3 ms of air transmission from the loudspeaker to the cochlea) were used as the final FFR_{ENV_F0} and FFR_{TFS} (FFR_{TFS_H2} and FFR_{TFS_F2F3}) magnitudes, respectively.

As well as FFR magnitudes, inter-trial phase-locking values (PLV) at the F_0 (FFR_{PLV_F0}) were also calculated. This was because FFR_{PLV_F0} reflects pure phase-locking that excludes the influence of single-trial spectral magnitudes and has a better signal-to-noise ratio than does FFR magnitudes (Zhu et al., 2013). FFR_{PLV_F0} was calculated in a similar way to FFR_{ENV_F0} along the F_0-ENV trajectory, except that, after zero-padding in each step (without Hanning-windowing), PLV was calculated (Morillon et al., 2012) instead of spectral magnitudes:

$$PLV = \frac{1}{n}\left|\sum_{i=1}^{n} e^{j\phi_i}\right|$$

where n denotes the total number of sweeps, ϕ_i denotes the Fourier phase value at the frequency of the F_0-ENV trajectory for the ith sweep at that step, and j is $\sqrt{-1}$. As PLV is restricted to values between 0 and 1, FFR_{PLV_F0} at each step was quantified by logit-transforming PLV to $[-\infty, +\infty]$, making it appropriate for linear regression analysis (Waschke et al., 2017):

$$FFR_{PLV\_F0} = ln\frac{PLV}{1-PLV}$$

FFR_{PLV_F0} values were then averaged across all steps along the F_0-ENV trajectory. The final FFR_{PLV_F0} value was taken as the maximal value for the time lags between 8 and 13 ms as in measurement of the FFR_{ENV_F0} magnitude.

Cortical responses

Cortical responses were measured as theta-band (4–6 Hz, to correspond to the stimulus repetition rate of 5 syllables per second) phase-locking values (theta-band PLV) at C3 and C4. EEGs were decimated to 1024 Hz, re-referenced to the average of the bilateral earlobes and bandpass filtered (4–6 Hz) using a 2nd-order zero-phase Butterworth filter. Sweeps exceeding ± 15 μV on either electrode were rejected (Mai et al., 2018). Lower rejection threshold was used than with FFRs (± 25 μV) because the theta-band signal normally does not have excessively high amplitude since it occupies a relatively narrow frequency range (4–6 Hz). More than 80% of the sweeps were retained in all participants after artefact rejection. Theta-band PLV time series (*PLV(t)*) were calculated and then logit-transformed:

$$PLV(t) = \frac{1}{n}\left|\sum_{i=1}^{n} e^{j\phi_i(t)}\right|$$

Logit-theta-band PLV $(t) = ln\frac{PLV(t)}{1-PLV(t)}$

where n denotes the total number of sweeps, $\phi_i(t)$ denotes the Hilbert phase series of the filtered EEG of the ith sweep time-locked to the syllable onset and j is $\sqrt{-1}$. Hilbert phase was used as it reflects phase-locking to stimuli even when EEG amplitude variation occurs due to relaxation and eye closure (Thatcher, 2012). Logit-theta-band PLV(t) values were then averaged across the stimulus period (120 ms). Neural transmission from cochlea to auditory cortex takes 10 to 30 ms in primates (Lakatos et al., 2007). Hence Logit-theta-band PLV was taken as the maximum value for time lags between 13 and 33 ms (at 1-ms steps) with the 3 ms for air transmission included. Finally, the Logit-theta-band PLV(t) was averaged across the two electrodes.

Cortico-subcortical connectivity

Cortico-subcortical connectivity was conducted using Partial Directed Coherence (PDC) (Baccala and Sameshima, 2001; Schelter et al., 2005) which is a Granger Causality based method that can quantify directed connectivity between subcortical and cortical signals. Subcortical and cortical signals were obtained by bandpass filtering EEGs at 100–180 Hz (covering the F_0 range) at Cz and 1–40 Hz at C3 and C4, respectively, using a 2nd-order zero-phase Butterworth filter. The Hilbert envelope of the subcortical signals were then obtained and further filtered at 1–40 Hz (same frequency range as that of the cortical signals) so that PDC can be applied. The filtered signals were segmented every 3 seconds and segments in which either subcortical or cortical signals exceeded ± 20 µV were rejected. PDCs were calculated based on the framework of vector autoregression (VAR) model for each segment:

$$X(t) = \sum_{r=1}^{p} a(r)X(t-r) + \varepsilon(t)$$

where $X(t)$ denotes the vector $(X_{Subcort}(t), X_{Cort}(t))^{\mathsf{T}}$ at time point t, where $X_{Subcort}(t)$ and $X_{Cort}(t)$ are the zero-mean serials of the subcortical (at Cz) and cortical signals (at C3/C4); $a(r)$ is the coefficient matrix of the VAR model at delayed time step r; $\varepsilon(t)$ is the error vector; p is the order of the VAR model. Here, the order p was set at the length of 210 ms that corresponds to the average cycle of the syllable occurrence. $a(r)$ were estimated via multivariate least squares (MLS). The estimated $a(r)$ was then transformed to the frequency domain:

$$A(\omega) = I - \sum_{r=1}^{p} a(r)\, e^{-j\omega r}$$

which denotes the difference between the identity matrix I and the Fourier transform of the coefficient series of $a(r)$. PDC was then calculated as:

$$PDC_{j \to i}(\omega) = \frac{|A_{ij}(\omega)|}{\sqrt{\sum_k |A_{kj}(\omega)|}}$$

where $j \to i$ refers to the directed flows from signal j (subcortical signal at Cz or cortical signal at C3/C4) to signal i (cortical signal at C3/C4 or subcortical signal at Cz). The PDC values at each 3-second segment were taken as those at the frequency (i.e., ω) of 4.7619 Hz (corresponding 210 ms of order p) and averaged across C3 and C4. PDCs (i.e., $PDC_{Subcort \to Cort}$ and $PDC_{Cort \to Subcort}$, reflecting flows from subcortical to cortical and cortical to subcortical signals, respectively) were finally obtained by averaging values across all segments for different noise types.

Classification of arousal states

Participants were not required to remain awake during the EEG recording because, as mentioned, a separate purpose of the present experiment was to investigate the effect of

arousal on speech-evoked responses (Mai et al., 2019). Mai et al. (2019) reported that both subcortical and cortical responses showed significant suppression in low arousal/nREM states compared to high arousal states. This accords with earlier functional imaging work that showed that neural responses to speech in subcortical (Portas et al., 2000) and cortical (Czisch et al., 2004; Wilf et al., 2016) auditory regions reduce during sleep compared to wakefulness. Also, significant correlations between behavioural measures and EEG parameters (FFRs and theta-band PLV) were only found in the high arousal state (Mai et al., 2018). Hence, only EEG data from periods with high arousal were used to avoid any influences that arousal has on the neural-behavioural relationship.

Sleep spindles were used to determine arousal state (Martin et al., 2013). Sections of the Cz EEG recordings were categorized into three types of epochs (all epochs were 21-second long) based on the occurrence of sleep spindles: (1) epochs in high arousal states (wakefulness or nREM Stage 1); (2) epochs in low arousal states (nREM Stage 2); and (3) epochs in transition between (1) and (2). After the experiment, participants gave a subjective ranking concerning how much they had slept. There was a significant correlation between the sleep ranking and the percentage of epochs classified as 'low arousal' ($p = 0.002$), which validated the spindle-based method. Further methodological details about the classification will be described in Chapter 3 (which specifically studied the effect of arousal on speech-evoked phase-locked responses).

Normalization of sweep numbers

Robust FFRs require around 1500 artefact-free sweeps (c.f., Dajani et al., 2005; Wong et al., 2007). Hence, participants' data for a particular background type were not included in subsequent analyses if there were < 1450 artefact-free sweeps in high arousal epochs for that background type. This resulted in 17%–30% of participants being rejected from further analyses (depending on the types of analyses conducted; see *3.2* and *3.3* for details). Moreover, as magnitudes of phase-locked activity are sensitive to the number of sweeps (Aviyente et al., 2011), problems can arise during statistical analyses if the number of sweeps differs significantly across participants. Therefore, the number of sweeps was normalized to around 1500 for both FFRs and theta-band PLV for each participant in each background type. Normalization was achieved by selecting high arousal epochs at random that included 1450 to 1550 sweeps and EEG signatures (magnitudes of FFR_{ENV_F0}, FFR_{TFS_H2} and FFR_{TFS_F2F3}, Logit-theta-band PLV and PDCs) were obtained from the selected epochs. The random selection procedure was repeated 100 times, giving 100 estimates for each EEG signature. Averages over the 100 estimates were used in the final statistical analyses. Therefore, this process ensured EEG signatures were based on around 1500 sweeps regardless of artefact rejection rates or different number of epochs of the three arousal states across participants (Mai et al., 2018).

Confirming FFR robustness

FFR magnitudes are small and their robustness was tested by statistically comparing the FFR magnitudes with the EEG noise floors using pairwise t-tests. The noise floors were quantified as the EEG magnitudes at the corresponding frequency range (110–160 Hz (F_0) for FFR_{ENV_F0}; 220–320 Hz (H2) for FFR_{TFS_H2}; 150-Hz bandwidth centred at 2400 Hz (F2) and 300-Hz bandwidth centred at 3100 Hz (F3) for FFR_{TFS_F2F3}) at the 50-ms FFR pre-stimulus period (Mai et al., 2018). The quantification procedure was similar to that used in calculating FFR magnitudes, in which a set of 40-ms sliding windows was applied on FFR_{ENV} (for FFR_{ENV_F0}) or FFR_{TFS} (for FFR_{TFS_H2} and FFR_{TFS_F2F3}) which used 1-ms steps over the pre-stimulus period. Magnitudes of noise floors were measured as the spectral magnitudes (summed magnitude of the cepstrally-smoothed profile for FFR_{TFS_F2F3}) across the corresponding frequency ranges averaged across all steps. These were all conducted along with the calculations of FFR magnitudes during the processes for normalization of sweep numbers.

2.2.5 Statistical analyses

Statistical analyses were conducted using SPSS 13.0 (SPSS Inc., USA).

Linear mixed-effect regressions

Linear mixed-effect regressions were conducted with the behavioural (SRT in the SiN perception tasks) and EEG signatures (FFR_{ENV_F0}, FFR_{PLV_F0}, FFR_{TFS_H2}, FFR_{TFS_F2F3} and Logit-theta-band PLV) as the dependent variables, Noise Type (SpN and BbN for SRT; Quiet, SpN and BbN for EEG) and Age Group (young vs. older) as the fixed-effect factors, and Participant as the random-effect factor. The type of covariance matrix that was chosen was the one that generated the smallest Bayesian Information Criterion (BIC) value (Wang et al., 2007). Post-hoc t-tests were conducted if a significant [Noise Type × Age Group] interaction occurred. These analyses were conducted for two reasons: (1) main effects of Age Group were tested to look into the effects of age on SiN perception and neural responses to speech; (2) since older adults experience more difficulties in SiN perception under BbN compared to other types of noise (Helfer and Freyman, 2008; Schoof and Rosen, 2014), testing the [Noise Type × Age Group] interaction should help reveal what neural mechanisms underlie this.

Additional linear mixed-effect regressions were conducted that included PTAs (PTA_{Low}, PTA_{High} and PTA_{Wide}, averaged across 0.25–1 kHz, 2–4 kHz and 0.25–4 kHz, respectively; all mean-centred) as covariates for the EEG signatures with the same fixed- and random-effect factors as in the previous analyses. These tested for age effects after the variability in peripheral hearing loss was controlled for. PTA_{Low} and PTA_{High} reflect hearing loss at low and high frequencies, respectively, while PTA_{Wide} reflects the combined effect of both. The three PTA variables were used as covariates in separate analyses to avoid the risk of collinearity. A concern about these additional mixed-effect regressions is that PTAs and Age are correlated (effects of PTA and Age can overlap), hence including PTAs as covariates may carry risks of

partially diminishing the Age effects despite controlling for hearing loss. For this reason, the two types of linear mixed-effect regressions (with and without PTAs as covariates) were conducted separately. Alpha values for testing significance were not adjusted.

Multiple linear regressions

Multiple linear regressions were then conducted to test for any behavioural-neural relationship using data from both the young and older groups. EEG signatures (FFR$_{ENV\_F0}$, FFR$_{PLV\_F0}$, FFR$_{TFS\_H2}$, and FFR$_{TFS\_F2F3}$, Logit-theta-band PLV and PDCs) were used as predictors of SiN perception (SRTs; dependent variables) for the corresponding noise types. Specifically, EEG signatures obtained in SpN were used to predict SRT in SpN, while EEG signatures obtained in BbN were used to predict SRT in BbN. This avoided problems that could arise due to behavioural and neural recordings being made under different types of noise (Presacco et al., 2016a). Additionally, since FFRs in quiet have been suggested to be associated with SiN perception (Anderson et al., 2011), we also used EEG signatures in quiet to predict SRTs. Age was not included as a predictor, since the regressions investigated the contributions of age-related factors (which were identified by significant main effects of Age in the ANOVAs), rather than age itself, to SiN perception.

The Best-Subset Regression approach was used that selected predictors of EEG signatures that generated the lowest BIC value. This approach provided the optimal model with best goodness of fit and least chance of overfitting (Burnham and Anderson, 2003). PTAs (PTA$_{Low}$, PTA$_{High}$ and PTA$_{Wide}$) were also included as predictors to generate the Best-Subsets which take into account the effects of peripheral hearing loss. To avoid any spurious regression results caused by multicollinearity, subsets with variance inflation factors (VIFs) > 1.5 were excluded (c.f., Stine, 1995). It is noteworthy that Best-Subset Regression is exploratory in nature and is often used when there is lack of a priori theory for a given topic. Here, it is not clear which neural parameter(s) can optimally model SiN perception, and such an approach was used to answer this question and aid identification of the neural substrates that underlie SiN perception.

After regression analyses using data from both young and older adults, the analyses were further conducted separately for the young and the older group, to evaluate whether they employ different neural mechanisms for SiN perception.

2.3 Results

2.3.1 Behavioural results

Linear mixed-effect regression was conducted for SRT with Noise Type (SpN vs. BbN) and Age Group (young vs. older) as the fixed-effect factors and Participant as the random-effect

factor. The SRTs are plotted as violin plots in **Figure 2.4** for SpN and BbN of the 23 young and 18 older participants and the statistics are summarized in **Table 2.1**. The significant main effect of Noise Type ($F_{(1, 39)}$ = 382.850, $p < 10^{-21}$; $SRT_{SpN} < SRT_{BbN}$) is consistent with previous finding that speech is better recognized in SpN than in BbN (Rosen et al., 2013). The significant main effect of Age Group ($F_{(1, 39)}$ = 5.527, p = 0.024; $SRT_{Young} < SRT_{Older}$) showed that young adults had better performance than older adults. A significant [Noise Type × Age Group] interaction occurred ($F_{(1, 39)}$ = 10.010, p = 0.003) and post-hoc t-tests showed that young adults had significantly better performance than older adults in BbN ($t_{(26.667)}$ = -3.399, p = 0.002, Cohen's d = 1.132; equal variances not assumed), but not in SpN ($t_{(39)}$ = -0.135, p = 0.893, Cohen's d = 0.043; equal variances assumed).
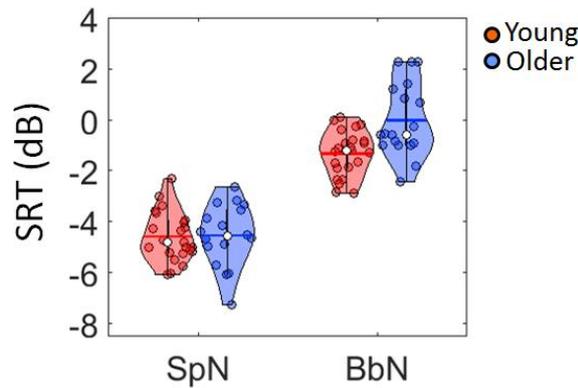
Figure 2.4. Violin plots for SRT as a function of Noise Type (SpN vs. BbN) and Age Group (young vs. older). *Low SRTs represent good SiN perception.*

Table 2.1. Statistical result of linear mixed-effect regression for SRT with Noise Type (SpN vs. BbN) and Age Group (young vs. older) as the fixed-effect factors and Participant as the random-effect factor. DV, df, F, and p refer to the dependent variable, degrees of freedom, F values, p values, respectively. Significant p values (< 0.05) are indicated in bold. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

DV	Fixed-effect factors	*df1*	*df2*	*F*	*p*
SRT	Noise Type	1	39	382.850	**$< 10^{-21}$***
	Age Group	1	39	5.527	**0.024***
	Noise Type × Age Group	1	39	10.010	**0.003****

2.3.2 Neural results

Robustness of FFR was first confirmed by using pairwise t-tests that assessed whether FFR magnitudes were statistically greater than their corresponding EEG noise floors (see *Confirming FFR robustness* in *2.2.4*). It was shown that spectral magnitudes of FFRs were all

significantly greater than noise floors (FFR$_{ENV\_F0\_Quiet}$: $p < 10^{-8}$; FFR$_{ENV\_F0\_SpN}$ and FFR$_{ENV\_F0\_BbN}$: $p < 0.001$; FFR$_{TFS\_H2\_Quiet}$: $p < 10^{-11}$; FFR$_{TFS\_H2\_SpN}$ and FFR$_{TFS\_H2\_BbN}$: $p < 10^{-10}$; FFR$_{TFS\_F2F3\_Quiet}$, FFR$_{TFS\_F2F3\_SpN}$ and FFR$_{TFS\_F2F3\_BbN}$: $p < 10^{-6}$). Furthermore, the response signal-to-noise ratio (SNR, i.e., difference in magnitudes between FFRs and the corresponding noise floors) did not differ between age groups (SNR for FFR$_{ENV\_F0\_Quiet}$: $p > 0.07$; SNRs for other FFR signatures: $p > 0.2$), indicating that SNR would not be a good index for measuring age differences. Additional simulations were also conducted in the present study showing that FFR magnitudes can more reliably quantify the FFR fidelity compared to response SNRs (see *Appendix 2*).

Linear mixed-effect regressions were then conducted for EEG signatures (FFR$_{ENV\_F0}$, FFR$_{PLV\_F0}$, FFR$_{TFS\_H2}$, FFR$_{TFS\_F2F3}$, Logit-theta-band PLV and PDCs) including Noise Type (Quiet, SpN and BbN) and Age Group (young vs. older) as fixed-effect factors and Participant as the random-effect factor. Participants' data were not included if artefact-free sweeps in the high arousal periods (wakefulness and nREM stage 1) were < 1450 in Quiet, SpN or BbN to ensure good EEG signal quality (see *2.4.5*). 29 (15) participants were retained (i.e., 12 (8) were rejected; rejection rate 29% where the numbers in brackets represent the numbers of young adults). Time series of the EEG signatures are given in **Figure 2.5** and the boxplots are shown in **Figure 2.6** (FFRs and Logit-theta-band PLV) and **Figure 2.7** (PDCs).

Statistics on the linear mixed-effect regressions are summarized in **Table 2.2**. For FFRs, significant main effects of Noise Type and [Noise Type × Age Group] interactions were found for both FFR$_{ENV\_F0}$ and FFR$_{PLV\_F0}$. Post-hoc comparisons (via pairwise t-tests) following the main effects showed that FFR$_{ENV\_F0}$ and FFR$_{PLV\_F0}$ were significantly greater in Quiet than in noise (FFR$_{ENV\_F0\_Quiet}$ > FFR$_{ENV\_F0\_SpN}$, $p < 10^{-7}$, Cohen's $d = 1.297$; FFR$_{ENV\_F0\_Quiet}$ > FFR$_{ENV\_F0\_BbN}$, $p < 10^{-8}$, Cohen's $d = 1.651$; FFR$_{PLV\_F0\_Quiet}$ > FFR$_{PLV\_F0\_SpN}$, $p < 10^{-6}$, Cohen's $d = 1.201$; FFR$_{PLV\_F0\_Quiet}$ > FFR$_{PLV\_F0\_BbN}$, $p < 10^{-7}$, Cohen's $d = 1.346$), but they did not differ between SpN and BbN (both $p > 0.5$). No main effects or interactions were found for FFR$_{TFS\_H2}$ or FFR$_{TFS\_F2F3}$.

For Logit-theta-band PLV, there were significant main effects of Noise Type and Age Group, but no significant [Noise Type × Age Group] interaction. Post-hoc comparisons found that Logit-theta-band PLV was greater in Quiet than in noise (Logit-theta-band PLV$_{Quiet}$ > Logit-theta-band PLV$_{SpN}$, $p = 0.002$, Cohen's $d = 0.654$; Logit-theta-band PLV$_{Quiet}$ > Logit-theta-band PLV$_{BbN}$, $p < 10^{-6}$, Cohen's $d = 1.334$), and greater in SpN than in BbN ($p < 10^{-4}$, Cohen's $d = 0.961$); Logit-theta-band PLV was greater for older than young adults ($p < 0.001$, Cohen's $d = 1.504$).

For PDCs, there were significant main effects of Noise Type for both PDC$_{Subcort \to Cort}$ and PDC$_{Cort \to Subcort}$, but no significant effects of Age Group or [Noise Type × Age Group] interactions. Post-hoc comparisons found that PDC$_{Subcort \to Cort}$ was greater in Quiet than in noise (PDC$_{Subcort \to Cort\_Quiet}$ > PDC$_{Subcort \to Cort\_SpN}$, $p = 0.040$, Cohen's $d = 0.579$; PDC$_{Subcort \to Cort\_Quiet}$ > PDC$_{Subcort \to Cort\_BbN}$, $p = 0.007$, Cohen's $d = 0.787$); on the other hand, PDC$_{Cort \to Subcort}$ was smaller in Quiet than in BbN ($p = 0.013$, Cohen's $d = 0.726$), and smaller in SpN than in BbN ($p = 0.032$, Cohen's $d = 0.587$).

Linear mixed-effect regressions that included PTAs (PTA_{Low}, PTA_{High} and PTA_{Wide}) as covariates were conducted next. For FFR_{ENV_F0}, an additional significant main effect of Age Group was found ($F_{(1, 28.214)} = 8.038$, $p = 0.008$, $FFR_{ENV\_F0\_Young} > FFR_{ENV\_F0\_Older}$) when PTA_{Low} was the covariate. The main effect of PTA_{Low} was also significant ($F_{(1, 28.214)} = 11.765$, $p = 0.002$), where higher PTA_{Low} correlated with greater FFR_{ENV_F0} magnitude (see **Table 2.3**). No significant main effects of Age or PTA were found when PTA_{High} or PTA_{Wide} were used as covariates (all $p > 0.2$) (see **Tables A2-A3** in *Appendix 1*). For Logit-theta-band PLV, the significant main effect of Age Group was maintained ($F_{(1, 32.467)} = 4.793$, $p = 0.036$ and $F_{(1, 32.812)} = 6.520$, $p = 0.016$) when PTA_{Low} and PTA_{High} were used as the covariates, respectively, however, this became non-significant ($F_{(1, 32.368)} = 3.753$, $p = 0.061$) when PTA_{Wide} was used as the covariate. No significant main effects of PTA_{Low}, PTA_{High}, or PTA_{Wide} occurred (all $p > 0.2$) (see **Tables A1-A3** in *Appendix 1*). For FFR_{PLV_F0}, FFR_{TFS_H2}, FFR_{TFS_F2F3} and PDCs, no significant main effects of Age or PTAs occurred (all $p > 0.1$) (see **Tables A1–A3** in *Appendix 1*). The overall results with PTAs as covariates showed: (1) FFR_{ENV_F0} declined with age when PTA_{Low} was controlled for, where higher PTA_{Low} was related to greater FFR_{ENV_F0}; (2) the age effect for Logit-theta-band PLV was maintained when PTAs were controlled for (though dropped below significance when PTA_{Wide} was used as the covariate) and the increased Logit-theta-band PLV cannot be statistically explained by increased PTA.
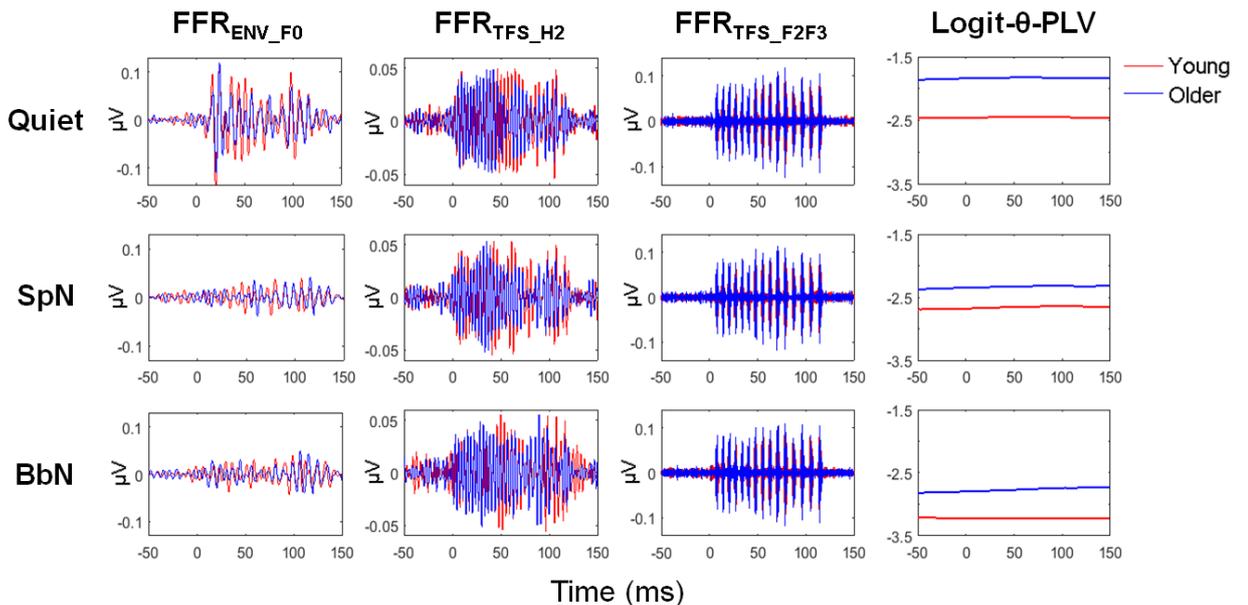
Figure 2.5. Time series of FFR_{ENV_F0}, FFR_{TFS_H2}, FFR_{TFS_F2F3} and Logit-theta-band PLV in Quiet (upper row), SpN (mid row) and BbN (lower row). *The series were based on sweeps with normalized numbers (1,450–1,550) during the high arousal periods (see 2.4.4 and 2.4.5) averaged across young (red) and older (blue) adults. Series of FFR_{ENV_F0} are shown as FFR_{ENV} bandpass filtered at 90–180 Hz (corresponding to the F_0 range). Series of FFR_{TFS_H2} and FFR_{TFS_F2F3} are shown as FFR_{TFS} bandpass filtered at 200–340 Hz (corresponding to the H2 range), and at 2000–4000 Hz (corresponding to the F2 and F3 range), respectively. '0' represents the syllable onset.*
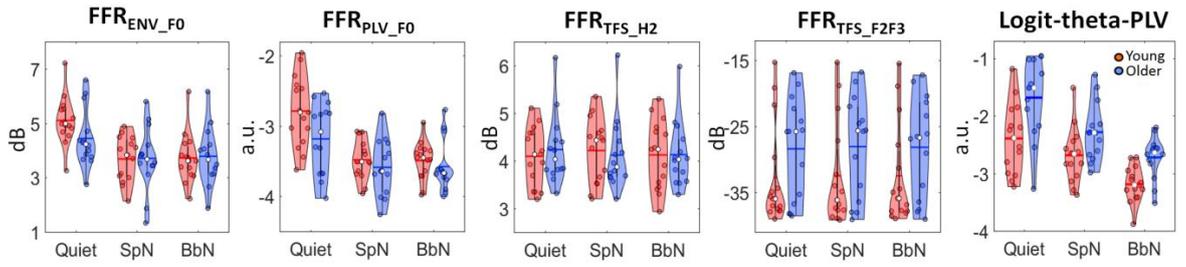
Figure 2.6. Violin plots for the five EEG signatures as a function of Noise Type (Quiet, SpN and BbN) and Age Group (young vs. older). *FFR_{ENV_F0} and FFR_{TFS_H2} were measured as log-power (dB); FFR_{TFS_F2F3} was measured as power of cepstral spectrum (dB) at F2 and F3 range; FFR_{PLV_F0} and Logit-theta-band PLV were measured as logit-transformed phase-locking values (a.u., arbitrary units). Young and older adults are plotted in red and blue respectively.*
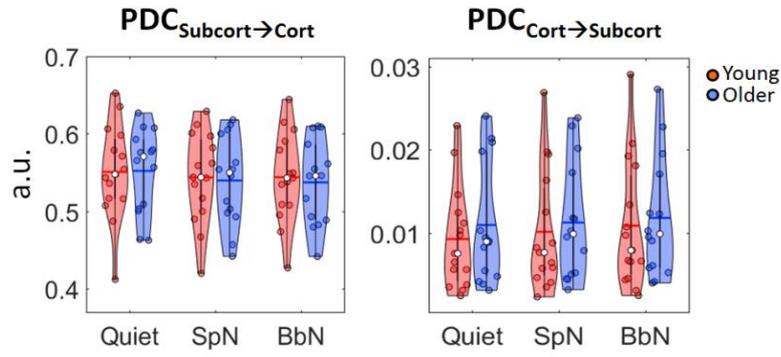
Figure 2.7. Violin plots for the connectivity between subcortical and cortical activities using Partial Directed Coeherence (PDC) as a function of Noise Type and Age Group. *Left: $PDC_{Subcort \rightarrow Cort}$ (directed flows from subcortical to cortical signals); right: $PDC_{Cort \rightarrow Subcort}$ (directed flows from cortical to subcortical signals).*

Table 2.2. Statistical results for linear mixed-effect regressions for the EEG signatures with Noise Type (Quiet, SpN and BbN) and Age Group (young vs. older) as the fixed-effect factors and Participant as the random-effect factor. DVs, df, F, *p* refer to the dependent variable, degrees of freedom, F values, and *p* values (uncorrected), respectively. Significant *p* values (< 0.05) are indicated in bold. *p < 0.05; **p < 0.01; ***p < 0.001.

DVs	Fixed-effect factors	*df1*	*df2*	*F*	*p*
FFR_{ENV_F0}	Noise Type	2	54	49.536	**< 10^{-12}***
	Age Group	1	27	0.326	0.573

	Noise Type × Age Group	2	54	5.608	**0.006**
FFR_{PLV_F0}	Noise Type	2	29.044	25.587	**$< 10^{-6}$***
	Age Group	1	31.424	2.680	0.112
	Noise Type × Age Group	2	29.044	1.998	0.154
FFR_{TFS_H2}	Noise Type	2	54	0.324	0.725
	Age Group	1	27	0.005	0.946
	Noise Type × Age Group	2	54	2.061	0.137
FFR_{TFS_F2F3}	Noise Type	2	54	1.592	0.213
	Age Group	1	27	2.479	0.127
	Noise Type × Age Group	2	54	0.051	0.950
Logit-theta-band PLV	Noise Type	2	35.630	34.769	**$< 10^{-8}$***
	Age Group	1	33.918	17.155	**< 0.001***
	Noise Type × Age Group	2	35.630	0.656	0.525
$PDC_{Subcort \rightarrow Cort}$	Noise Type	2	54	5.302	**0.008**
	Age Group	1	27	0.023	0.881
	Noise Type × Age Group	2	54	0.618	0.543
$PDC_{Cort \rightarrow Subcort}$	Noise Type	2	54	4.948	**0.012***
	Age Group	1	27	0.226	0.638
	Noise Type × Age Group	2	54	0.497	0.611

Table 2.3. Statistical result of linear mixed-effect regression for FFR_{ENV_F0} with PTA_{Low} as the covariate. DV, df, F, and *p* refer to the dependent variable, degrees of freedom, F values, and *p* values (uncorrected), respectively. Significant *p* values (< 0.05) are indicated in bold. **p < 0.01; ***p < 0.001.

DV	Fixed-effect factors/covariate	*df1*	*df2*	*F*	*p*
FFR_{ENV_F0}	Noise Type	2	26.059	41.418	**$< 10^{-8}$***
	Age Group	1	28.214	8.038	**0.008****
	PTA_{Low}	1	28.214	11.765	**0.002****
	Noise Type × Age Group	2	26.059	2.182	0.133
	Noise Type × PTA_{Low}	2	26.059	3.015	0.066

2.3.3 Behavioural-neural relationship

Behavioural-neural relationships were assessed by linear regressions in which SRT in SpN was predicted by EEG signatures obtained in SpN, whilst SRT in BbN was predicted by EEG signatures obtained in BbN. SRTs were further predicted by EEG signatures obtained in Quiet. PTAs (PTA_{Low}, PTA_{High} or PTA_{Wide}) were also included as predictors provided that including them improved the statistical capacity of EEG signatures to predict SRTs in the Best-Subsets. Similar to the procedure in *3.2*, participants' data obtained under a particular noise type (Quiet, SpN or BbN) were not included if artefact-free sweeps in the high arousal periods were < 1450 for that noise type (see *2.4.5*). This resulted in 31 (17) participants retained (i.e., 10 (6) were excluded; rejection rate 24%) for analyses in SpN, 34 (18) participants retained (i.e., 7 (5) were excluded; rejection rate 17%) for analyses in BbN, and 32 (18) participants retained (9 (5) were excluded; rejection rate 22%) for analyses in Quiet (the numbers in brackets represent the numbers of young adults).

Regression results including data of both young and older adults

Statistics for the Best-Subset Regressions are shown as in **Tables 4 and 5**. Results that included data of both young and older adults were analysed first. When SRTs were predicted by EEG signatures obtained in the respective noise types, SRTs were significantly correlated with Logit-theta-band PLV (SpN, $t_{(28)}$ = -3.104, *p* = 0.004; BbN, $t_{(31)}$ = -2.508, *p* = 0.018; greater Logit-theta-band PLV correlated with better SiN perception) after PTA was controlled for (PTA_{High} for SpN and PTA_{Wide} for BbN) (**Table 2.4**; **Figure 2.8A**). When SRTs were predicted by EEG signatures obtained in Quiet, a significant correlation was found between FFR_{ENV_F0} magnitude and SRT in BbN ($t_{(29)}$ = -2.698, *p* = 0.012; greater FFR_{ENV_F0} magnitude correlated with better SiN perception) after PTA_{Low} was controlled for (**Table 2.5**; **Figure 2.8B**).

Table 2.4. Results for the Best-Subset Regressions in which SRTs were predicted by EEG signatures obtained in the corresponding noise types (i.e., SRT_{SpN} was predicted by EEGs obtained in SpN; SRT_{BbN} was predicted by EEGs obtained in BbN). DVs refers to the dependent variables; β, CI, *T*, *p*, VIF refer to standardized β-coefficient, 95% confidence interval for standardized β, *t* values, *p* values (uncorrected) and variance inflation factors, respectively. *N* denotes the numbers of participants. *F* denotes the F values of the models (with corresponding *p* values in the brackets). Significant *p* values (< 0.05) are in bold. *p < 0.05; **p < 0.01; ***p < 0.001.

DVs	Participants	*N*	Predictors	β	CI	*T*	*p*	VIF	*F*
SRT_{SpN}	Young + older	31	Logit-theta PLV_{SpN}	-0.499	[-0.828, -0.170]	-3.104	**0.004****	1.040	6.104 (**0.006****)
			PTA_{High}	0.351	[0.022, 0.680]	2.183	**0.038***	1.040	
	Young	17	PTA_{High}	0.445	[-0.048, 0.938]	1.922	0.074	1.000	3.695 (0.074)
	Older	14	$FFR_{ENV\_F0\_SpN}$	-0.475	[-0.903, -0.048]	-2.480	**0.033***	1.042	6.120 (**0.012***)
			Logit-theta PLV_{SpN}	-0.475	[-0.903, -0.047]	-2.475	**0.033***	1.045	
			PTA_{Wide}	0.363	[-0.064, 0.791]	1.892	0.088	1.045	
SRT_{BbN}	Young + older	34	Logit-theta PLV_{BbN}	-0.377	[-0.684, -0.070]	-2.508	**0.018***	1.365	14.738 (**$<10^{-4}$*****)
			PTA_{Wide}	0.814	[0.506, 1.122]	5.418	**$< 10^{-5}$*****	1.365	
	Young	18	PTA_{Wide}	0.318	[-0.184, 0.820]	1.343	0.198	1.000	1.804 (0.198)
	Older	16	$FFR_{TFS\_H2\_BbN}$	-0.351	[-0.578, -0.124]	-3.368	**0.006****	1.047	28.054 (**$<10^{-4}$*****)
			Logit-theta PLV_{BbN}	-0.769	[-1.005, -0.533]	-7.103	**$< 10^{-4}$*****	1.128	
			PTA_{Wide}	0.696	[0.456, 0.936]	6.350	**$< 10^{-4}$*****	1.157	

Table 2.5. Results for the Best-Subset Regression in which SRTs (both in SpN and BbN) were predicted by EEG signatures obtained in Quiet. DV refers to the dependent variables; β, CI, *T*, *p*, VIF refers to standardized β-coefficient, 95% confidence interval for standardized β, *t* values, *p* values (uncorrected) and variance inflation factors, respectively. *N* denotes the numbers of participants. *F* denotes the F values of the models (with corresponding *p* values in the brackets). Significant *p* values (< 0.05) are in bold. *p < 0.05; **p < 0.01.

DVs	Participants	*N*	Predictors	*B*	CI	*T*	*p*	VIF	*F*
SRT_{SpN}	Young + older	32	$FFR_{ENV\_F0\_Quiet}$	-0.307	[-0.661, 0.048]	-1.765	0.088	1.000	3.116 (0.088)
	Young	18	$FFR_{ENV\_F0\_Quiet}$	-0.508	[-1.001, -0.014]	-2.192	**0.045***	1.161	3.304 (0.065)
			PTA_{Wide}	0.478	[-0.016, 0.973]	2.062	0.057	1.161	

	Older	14	FFR$_{ENV\_F0\_Quiet}$	-0.471	[-1.010, 0.069]	-1.919	0.081	1.068	3.375 (0.072)
			FFR$_{TFS\_H2\_Quiet}$	-0.534	[-1.074, 0.005]	-2.179	0.052	1.068	
SRT$_{BbN}$	Young + older	32	FFR$_{ENV\_F0\_Quiet}$	-0.410	[-0.720, -0.099]	-2.698	**0.012***	1.012	7.425 (**0.002***)
			PTA$_{Low}$	0.460	[0.150, 0.771]	3.030	**0.005****	1.012	
	Young	18	PTA$_{Wide}$	0.323	[-0.181, 0.829]	1.363	0.192	1.000	1.858 (0.192)
	Older	14	FFR$_{ENV\_F0\_Quiet}$	-0.455	[-0.857, -0.054]	-2.525	**0.030***	1.069	7.622 (**0.006***)
			FFR$_{TFS\_H2\_Quiet}$	-0.551	[-0.969, -0.133]	-2.941	**0.015***	1.153	
			PTA$_{Wide}$	0.441	[0.037, 0.845]	2.434	**0.035***	1.081	

(A) Relation between SRT and Logit-θ-PLV obtained in corresponding noise types
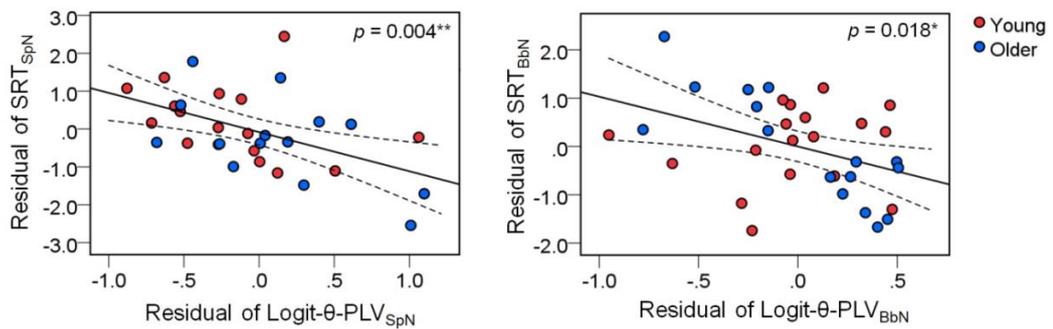
(B) Relation between SRT$_{BbN}$ and FFR$_{ENV\_F0}$ obtained in Quiet
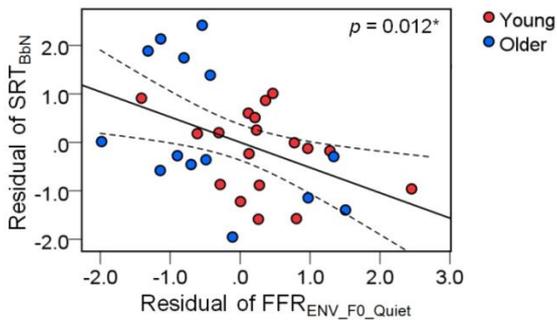
Figure 2.8. Scatter plots (young + older) for significant partial correlations after controlling for PTAs. *The plots visualize the relation between **(A)** SRT and Logit-theta-band PLV obtained in the corresponding noise types (see statistics in **Table 2.4**); and **(B)** SRT$_{BbN}$ and FFR$_{ENV\_F0}$ magnitude obtained in Quiet (see statistics in **Table 2.5**). Red and blue dots represent young and older participants, respectively.*

Regression results in the young and the older group separately

Best-Subset Regressions were then conducted separately for the young and older groups. In the young group, SRT in SpN was significantly correlated with FFR$_{ENV\_F0}$ magnitude

obtained in Quiet after PTA_{Low} was controlled for ($t_{(15)}$ = -2.195, p = 0.045; greater FFR_{ENV_F0} magnitude correlated with better SiN perception; **Table 2.5**). No significant correlations were found between SRTs and EEG signatures in the corresponding noise types nor between SRT in BbN and in EEG signatures obtained in Quiet (**Tables 4** and **5**). In the older group, when SRTs were predicted by EEG signatures in the respective noise types, SRTs were significantly correlated with FFRs (FFR_{ENV_F0} for SpN: $t_{(10)}$ = -2.480, p = 0.033; FFR_{TFS_H2} for BbN: $t_{(12)}$ = -3.368, p = 0.006) and Logit-theta-band PLV (SpN: $t_{(10)}$ = -2.475, p = 0.033; BbN: $t_{(12)}$ = -7.103, p < 10^{-4}) (greater FFR magnitudes and Logit-theta-band PLV correlated with better SiN perception) after PTA_{Wide} was controlled for (**Table 2.4**). When SRTs were predicted by EEG signatures in Quiet, SRT in BbN was significantly correlated with FFR_{ENV_F0} and FFR_{TFS_H2} (FFR_{ENV_F0}: $t_{(10)}$ = -2.941, p = 0.030; FFR_{TFS_H2}: $t_{(10)}$ = -2.525, p = 0.015; greater FFR magnitudes correlated with better SiN perception) after PTA_{Wide} was controlled for (**Table 2.5**).

Taken together, the regression analyses showed that: (1) when combining data from both young and older adults, by controlling for the degree of hearing loss (PTAs), SRTs can be predicted by cortical phase-locked responses (Logit-theta-band PLV) to speech obtained in noise (greater Logit-theta-band PLV associated with better SiN perception) and by subcortical phase-locked responses to speech F_0 obtained in Quiet (greater FFR_{ENV_F0} magnitude associated with better SiN perception); (2) SRTs are predicted by subcortical and cortical responses (FFR_{ENV_F0}, FFR_{TFS_H2} and Logit-theta-band PLV) obtained in noise in the older group, not in the young group, and by subcortical responses obtained in Quiet (FFR_{ENV_F0}/FFR_{TFS_H2}) in the young (SRT_{SpN}) and the older group (SRT_{BbN}).

2.4 Summary of results and brief discussions

Fuller discussions of this study will be in Chapter 5 (*5.1*).

The present study found that theta-band PLV increased with age. Further analyses showed that, after PTAs (PTA_{Low} and PTA_{High}) were controlled for, the statistical effect of age was maintained. No statistical correlations occurred between PTAs and theta-band PLV, indicating that the age-related increase cannot be explained by hearing loss. This is consistent with previous studies showing greater theta-band PLV to amplitude-modulated tones (Tlumak et al., 2015; Goossens et al., 2016) and that older adults have larger magnitudes of cortical auditory-evoked responses (Alain et al., 2014; Herrmann et al., 2013, 2016). Consequently, the results may be attributable to hyperexcitiblity of the central auditory system during the aging process (Caspary et al., 2008).

No significant differences were found between young and older adults for magnitudes of speech-evoked FFRs. However, it was shown that FFR_{ENV_F0} magnitude was significantly smaller in older than young adults after PTA_{Low} (0.25–1 kHz) was controlled for. Also, PTA_{Low} correlated positively with the FFR_{ENV_F0} magnitude (i.e., FFR_{ENV_F0} magnitude increased with

low-frequency hearing loss). Therefore, these results are in line with findings that encoding of envelopes at high-gamma frequencies corresponding to the F_0 range declines during aging when peripheral hearing is normal but increases when there is hearing loss (Goossens et al., 2016, 2019), indicating the age-related neural declines in phase-locking and reduced neural inhibition related to hearing loss.

The relationships between age-related phase-locked responses to speech in the auditory sensory systems and SiN perception were then investigated. The older adult group had peripheral hearing at high frequencies that ranged from normal to mild/moderate hearing loss that reflect the typical demographics in normal aging populations (Gopinath et al., 2009; Humes et al., 2010). Furthermore, regressions were conducted with neural signatures and SiN perception under the same types of background noise; neural data obtained in quiet were additionally used as predictors, as FFRs in quiet could be associated with SiN perception (Anderson et al., 2011). Different patterns of age effects on auditory phase-locked responses were revealed at the cortical and subcortical levels: aging is related to increases in cortical responses (theta-band PLV) and decreases in subcortical responses (FFRs). Relationships between behavioural and neural performance showed that cortical responses obtained in noise and subcortical responses obtained in quiet had significant positive associations with SiN perception, indicating that effects of aging on cortical (increase with aging) and subcortical (decrease with aging) activities make different impacts on SiN perception.

Limitations of the current study should be addressed. One of the most important factors is the lack of measurements of higher-level cognitive functions. Older adults suffer from declines in cognitive functions related to not only aging itself, but also hearing loss (Lin et al., 2013). The most important cognitive functions are working memory and selective-attention that have been shown to influence SiN perception in older adults (Schoof and Rosen, 2014; Rimmele et al., 2015). A relevant issue which needs attention is that perception of a repeatedly presented single syllable used for neural measurements should be very different from comprehension of sentences used for behavioural measurements. Compared to the former, the latter requires additional top-down strategies which would engage higher-level cognitive functions as mentioned above (Davis and Jonhsrude, 2007). Besides, Mild Cognitive Impairment (MCI) is another factor that could happen in some of the older participants (Petersen et al., 1999). Furthermore, speech-evoked responses can be affected by MCI as well (Anderson et al., 2013; Bidelman et al., 2017).

Another important consideration is that different types of hearing loss may need to be disentangled. The audiograms (tested via air-conduction) for older adults with hearing loss in the present study showed clear patterns consistent with the sensorineural hearing loss due to presbycusis (**Figure 2.1**). However, we cannot exclude that conductive hearing loss may occur in some of these participants. Future work is needed to further include bone-conduction to test for conductive hearing loss in order to better clarify the roles of different types of hearing loss.

A further concern may be the relatively small sample size. The present study recruited older participants who covered a wide range of peripheral hearing from normal to mild/moderate hearing loss. Although the degree of hearing loss was controlled for by using PTAs as covariates during the analyses, a more direct and better approach would be to compare normal-hearing older adults with those who had hearing loss to clarify the respective effects of aging and hearing loss. However, sample size in the present study (<17 older participants after data rejection) made it difficult to use this approach. Future research will need to recruit participants with bigger sample sizes with better control in hearing loss for older adults. In addition, sample sizes were relatively small when conducting separate regression analyses for the young and older groups (both groups had <20 participants after data rejection; see **Table 2.4** and **Table 2.5**). Small sample sizes may limit the power to detect an effect and this could be a reason for the lack of significant neural-behavioural relations especially in the young group. Future work may need to recruit more participants to obtain greater effect sizes and statistical powers so that the results are more reliable.

Furthermore, although arousal effects were controlled in the present study by restricting the data used in analyses to those when participants were in the high arousal states, amount of attention to the acoustic stimuli was not controlled during the EEG recording. While selective auditory attention can modulate auditory cortical (Choi et al., 2013; Kong et al., 2014) and subcortical (Galbraith et al., 2003; Hairston et al., 2013; Lehmann and Schönwiesner, 2014) electrophysiological responses, it is not totally clear how they are affected by unpredictable changes in attention during passive listening as in the present study. Therefore, additional tasks of active listening to target speech stimuli under the corresponding noise types need to be conducted in the future to investigate whether the current neural-behavioural relationships is replicable.

Taken together, the present study hypothesized that effects of age on these activities should be associated with SiN perception. Compared to young adults, it was found that older adults have greater theta-band PLV and smaller FFR magnitude (when low-frequency hearing loss was controlled for), illustrating distinct mechanisms of age effects at the subcortical and cortical levels. Greater theta-band PLV reflects the neural hyperexcitability in the auditory cortex during aging and was associated with increased SiN perception whilst smaller FFR magnitude reflects declines in subcortical phase-locking during aging and was associated with decreased SiN perception. The current study thus provided evidence for different mechanisms at the sensory cortical and subcortical levels by which age affects speech-evoked phase-locked activities and SiN perception. Future work need to be conducted by combining cognitive assessments to study how higher-level cognitive functions influence such mechanisms and contribute to SiN perception together with sensory processing during aging.

Chapter 3

Modulation of speech-evoked phase-locked neural responses during different arousal states in young and older adults

3.1 Introduction

Study 2 of this thesis studied how speech-evoked responses are affected by arousal and how age may modulate this process. Auditory signals are processed by the sleeping brain (Issa and Wang, 2008; Nir et al., 2015). However, neural responses to speech in the cortical (Czisch et al., 2002, 2004; Wilf et al., 2016) and subcortical (Portas et al., 2000) auditory regions reduce during sleep compared to wakefulness. Phase-locked responses to complex auditory signals change in different arousal states. For example, Makov et al. (2017) examined relationships between episodes designated as wakefulness, nREM and REM and processing at different linguistic levels. EEG phase-locked responses at rates corresponding to higher-order linguistic units (words, phrases and sentences) were statistically greater in wakefulness than in sleep, but not at the rates corresponding to those of lower-order units (syllables). It is unclear how arousal affects neural phase-locked responses to fine-grained speech acoustic properties, such as speech envelopes (i.e., Slow-ENV) and F_0. The former corresponds to speech-evoked theta-band PLV at the cortical level which reflects the neural tracking of Slow-ENV and/or evoked responses to amplitude variations of speech, while the latter corresponds to FFR_{ENV_F0} at the subcortical level the encodes the F_0 information.

An important property to describe arousal is sleep spindles that can be used to locate episodes where arousal is low and to indicate whether and when arousal state changes within EEG sessions. These bursts of oscillatory neural activity occur at frequencies of 12–16 Hz (Warby et al., 2014) and are transmitted to the cortex from thalamus. Auditory responses have been shown to be affected by this activity during sleep (Dang-Vu et al., 2011; Schabus et al., 2012). Sleep spindle properties, including magnitude, duration and density during nREM sleep, decrease with age (Martin et al., 2013; Mander et al., 2017). Hence, spindles are expected to neuromodulate speech-evoked phase-locked responses and the neuromodulation could differ across ages. Spindle activity within entire recording sessions, such as spindle density, has been used previously to indicate arousal state and sleep stability (Kim et al., 2012). However, such activity has not been used to explore the effects on phase-locked responses in speech perception across age groups. Due to the influence of age on spindle properties, it is plausible to hypothesize that the age would modulate the arousal effects.

Here the links between arousal, sleep spindle density and speech-evoked phase-locked activity (theta-band PLV and FFR_{ENV_F0}) were assessed in human adults across a wide age range (19–75 years old). The present study hypothesized that: (1) State of arousal can affect theta-band PLV and the magnitude of FFR_{ENV_F0}; and (2) Arousal effects may covary with spindle density within different age groups, reflecting the modulation of age.

3.2 Materials and Methods

3.2.1 Participants
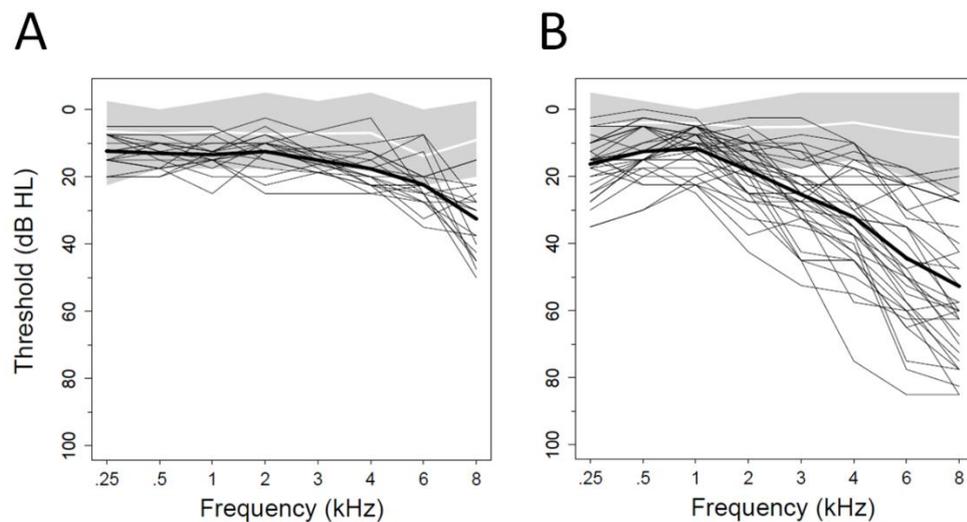
Figure 3.1 Individual audiograms for (A) Exp 1 and (B) Exp 2. *Individual participant data were averaged across ears. The grey shaded areas and white lines represent the ranges and average thresholds for the young adults. The grey and bold lines represent the individual and average thresholds of the older adults. Thresholds at 3 kHz were only obtained in Exp 2. In Exp 2, thresholds of one older participant at 6 kHz (both ears) and six older participants at 8 kHz (two on both ears and four on either left or right ear) were > 85 dB and were not measureable at these frequencies. For these points, thresholds were entered as the highest possible value (85 dB) (Mai et al., 2018).*

Data from Schoof and Rosen (2016) and Mai et al. (2018) (Exp 1 and Exp 2) were used. Participants in both experiments were native English speakers who had no history of language or neurological disorders. In Exp 1 there were 20 young (19–29 yrs; Mean ± SD = 23.7 ± 2.9 yrs; 10 males) and 20 older adults (60–72 yrs; Mean ± SD = 64.1 ± 3.3 yrs; 3 males). They all had near-normal hearing defined as pure-tone thresholds ≤ 25 dB between 0.25 and 4 kHz in both ears and at 6 kHz in at least one ear. In Exp 2 there were 23 young (19–42 yrs; Mean ± SD

= 26.3 ± 5.5 yrs; 15 males) and 35 older adults[7] (53–75 yrs; Mean ± SD = 67.6 ± 5.1 yrs; 12 males). Pure-tone audiometric thresholds (PTA) were measured via a MAICO MA41 Audiometer (MAICO Diagnostics, Germany) in a sound-attenuating booth in both experiments. All young participants had normal hearing (thresholds ≤ 25 dB) from 0.25 to 8 kHz in both ears except for one whose pure-tone thresholds on the left ear were 35 and 45 dB at 6 and 8 kHz. For older adults, 27 out of 35 had normal hearing at low frequencies (≤ 1 kHz) but PTAs ranged from normal hearing to severe hearing loss at high frequencies (2 to 8 kHz). **Figure 3.1** gives individual PTAs separately for each experiment.

3.2.2 Stimuli

In Exp 1, EEG was recorded in response to repeated presentations of a 100 ms /a/ vowel presented with an inter-stimulus interval (ISI) of 100 ms (5 syllables per second). The vowel had a flat fundamental frequency (F_0) at 160 Hz and F1, F2, F3 and F4 were at 710, 1200, 2900 and 3400 Hz, respectively (**Figure 3.2A**). The stimulus was ramped on/off for 6.25 ms with a cosine window. The syllables were presented at 80 dB SPL binaurally via electrically shielded ER-3 insert earphones (Intelligent Hearing Systems, Miami, FL).

In Exp 2, the stimuli were repeated presentations of a 120 ms /i/ and ISIs varied randomly between 60 and 120 ms (~ 4.8 syllables per second). The vowel had an F_0 contour that dropped from 160 to 110 Hz. F1, F2 and F3 were approximately 280, 2400 and 3100 Hz, respectively (**Figure 3.2B**). The vowel was ramped on and off with a 5 ms cosine window. The stimuli were presented over a Rogers LS3/5A loudspeaker (Falcon Acoustics, UK). The intensity at a distance of 1 metre from the loudspeaker at 0 degrees azimuth, which corresponded to where participants' heads were located during measurements, was 77 dB SPL.

Syllables were presented in quiet and when different types of background noise were present in both experiments (steady-state and amplitude-modulated speech-shaped noise in Exp 1; steady-state speech-shaped and 16-talker babble noise in Exp 2). Syllables were presented with positive and negative polarities. In Exp 1 syllables with different polarities were presented sequentially in separate blocks (positive followed by negative polarity) whilst in Exp 2 they were temporally intermixed., See Schoof and Rosen (2016) and Mai et al. (2018) for detailed information of the paradigms. In the present paper, only EEG responses to syllables in the quiet background were used. There were 6000 and 3200 sweeps for each polarity in Exp 1 and 2, respectively.

---

[7] There were 47 older participants in total in Exp 2. This included 12 hearing aid users and 35 participants who did not use hearing aids (Mai et al., 2018). Hearing aid users were excluded from the present study to avoid possible additional effects of hearing aids. In addition, the 23 young participants and 18 older participants (out of the 35 older participants) in Exp 2 were those whose data were also used for Chapter 2. Data for the other 17 out of the 35 older participants were not used in Chapter 2 because a different SNR in noisy backgrounds was used for these participants (7 dB, see Mai et al., 2018) from the SNR that was reported in Chapter 2 (-1 dB).
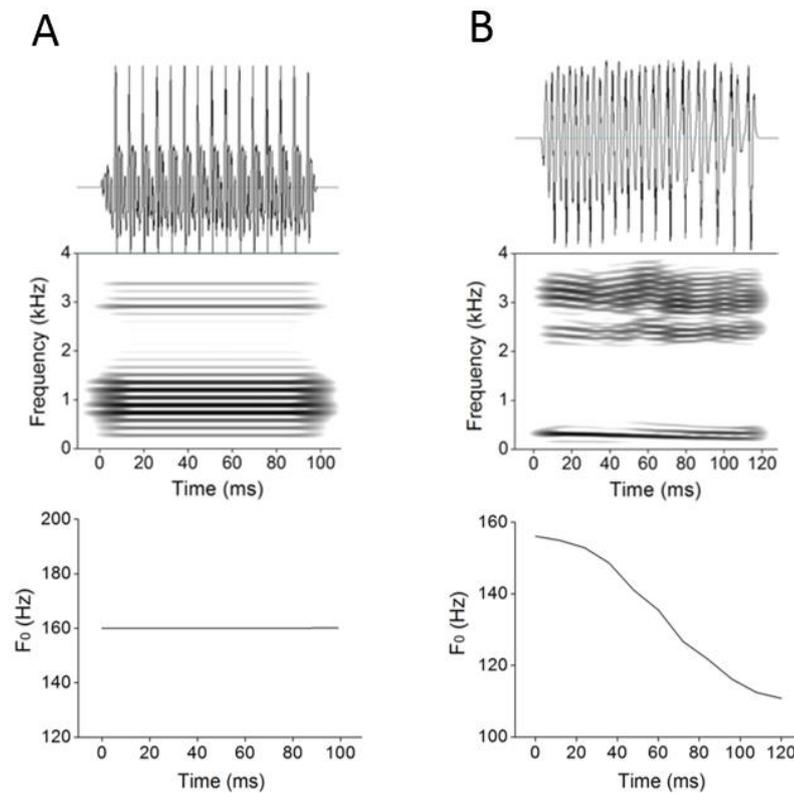
Figure 3.2 Vowel stimuli for Exp 1 and 2. (A) *Exp 1 (/a/ with an F_0 at 160 Hz).* ***(B)*** *Exp 2 (/i/ with an F_0 falling from 160 to 110 Hz). Top, middle and bottom panels show the acoustic waveforms, narrow-band spectrograms, and the F_0 contours respectively.*

3.2.3 EEG recording procedure

In both experiments, participants sat in a reclining chair in a sound-attenuating, electromagnetically-shielded booth. Participants were instructed to relax, close their eyes and keep as still as possible. They were allowed to fall asleep during stimulus presentation in both experiments. Movements were monitored by a webcam in both experiments and no significant changes in head or body position were observed.

EEG was recorded using an ActiveTwo BioSemi system (Biosemi, The Netherlands) at a sampling rate of 16384 Hz. Three active electrodes positioned at Cz (vertex), C3 and C4 according to the 10/20 configuration were used for analyses. Cz was used to obtain FFRs (Skoe and Kraus, 2010) and to classify arousal states (Martin et al., 2013). Cortical responses were measured via C3 and C4, representing activity in the auditory cortex (Carpenter and Shahin, 2013; Noguchi et al., 2015). Bilateral earlobes were used as reference. Ground electrodes were CMS/DRL. Electrode impedance was always below 40 mV.

3.2.4 Classification of arousal states

Subsequent analyses of EEG signals were conducted using Matlab R2014a (Mathworks, USA). Sleep spindles in the EEG sigma frequency band (12–16 Hz) were used as signatures of

Stage 2 nREM sleep (Warby et al., 2014) using a method adapted from Martin et al. (2013). EEGs at Cz were filtered into alpha (8–11 Hz), sigma (12–16 Hz) and beta (17–20 Hz) bands using a 2nd-order zero-phase Butterworth filter. Then the filtered signals were divided into 250-ms-long successive segments (without temporal overlaps between segments). A spindle was labelled when the following criteria were met: (1) root-mean-square (RMS) voltage in the sigma band in a given segment exceeded the threshold of the 95th percentile of the sigma RMS of all segments; (2) RMS of the sigma band was higher than both alpha and beta RMS in the current segment; (3) two successive segments met both criteria (1) and (2). (1) and (2) were invoked because dominance of the sigma-band in the spectrum is the major characteristic of sleep spindles (Martin et al., 2013; Warby et al., 2014). The requirement to extend across two segments was included because sleep spindles usually last for at least 500 ms (De Gennaro and Ferrara, 2003).

After the spindles were detected, the entire EEG recordings were segmented into epochs of 21- and 20-second lengths (in Exp 1 and Exp 2, respectively). These lengths were chosen so that each epoch contained responses to 100 vowel repetitions No participant reported deep sleep during the tests. Consistent with this, high-amplitude delta (1 ~ 4 Hz) activity (Hilbert envelope > 60 µV) that lasted for 25% of the time within an epoch was not detected for any epoch for any participant showing that they were not in Stage 3 or 4 of nREM sleep (i.e., Slow-Wave Sleep). Hence participants were either awake, or in Stage 1 or 2 of nREM sleep (Brown et al., 2012).

The epochs were then classified into high arousal, low arousal, and transition between high and low arousal states. Low arousal epochs were those that contained at least one sleep spindle. High arousal epochs were those that contained no spindles and were not adjacent to an epoch with a sleep spindle. Transition epochs were those that were neither high arousal nor low arousal epochs. High arousal epochs approximate to wakefulness or nREM Stage 1, whilst low arousal epochs approximate to nREM Stage 2. Transition epochs were discarded.
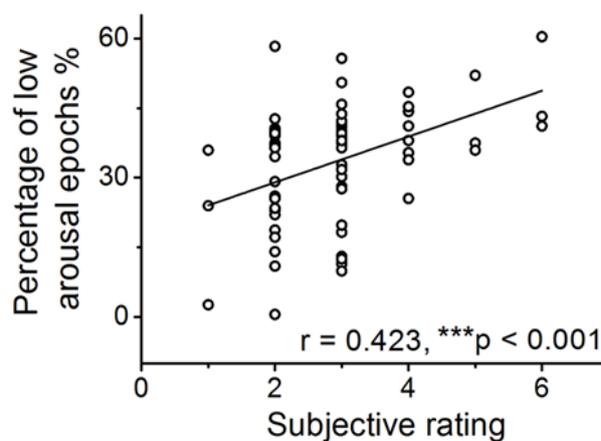
Figure 3.3. Correlations of the subjective rating of sleepiness with the percentage of epochs classified as low arousal in Exp 2. Higher ratings indicate higher levels of sleepiness.

Participants in Exp 2 rated how much they slept after each session (scale points from 1 to 7, each of which had a written description; see *Appendix 3* for the detailed written descriptions). Subjective ratings correlated significantly with the percentage of epochs classified as 'low arousal' (Pearson's r = 0.423, p < 0.001; **Figure 3.3**). This confirmed the validity of the spindle detection and classification steps.

3.2.5 Sleep spindle parameters

Density, magnitude and duration of sleep spindles were calculated. Spindle density was the number of spindles per minute averaged across the low arousal states[8]. Magnitude of each spindle was quantified as the maximum power value in the Hilbert envelope during spindle activity. Spindle duration was the time between the start- and end-point values at half spindle magnitude in the amplitude envelope. Participants who had fewer than five epochs classified as the low arousal states (those that contained spindles) were excluded from these analyses. This left 91 participants (38 young and 53 older adults, i.e., 5 young and 2 older adults were removed).

3.2.6 Frequency-following responses (FFRs)

Baseline was adjusted using the 40-ms pre-stimulus period. EEGs were re-referenced to the average of bilateral earlobes and bandpass filtered between 70 and 2000 Hz[9] using a zero-phase 2nd-order Butterworth filter. Sweeps exceeding ±25 µV were rejected to reduce the incidence of movement artefacts (Schoof and Rosen, 2016; Mai et al., 2018). FFRs with positive (FFR$_{pos}$) and negative (FFR$_{neg}$) polarities were obtained by averaging across sweeps with their respective polarities. In Exp 1, FFR magnitudes were quantified as the magnitude along the F_0 trajectory of the /a/ vowel (160 Hz) based on either FFR$_{pos}$ or FFR$_{neg}$. In Exp 2, FFR magnitudes were quantified as the magnitude along the F_0 trajectory of the /i/ vowel (160–110 Hz) using the waveform resulting from addition of FFR$_{pos}$ and FFR$_{neg}$ that was then divided by 2 (Aiken and Picton, 2008). The procedure that added alternate polarities minimized fine structure temporal information at the auditory periphery (i.e. the cochlear microphonic) and emphasized the processing of envelope cues at the brainstem (Aiken and Picton, 2008). The addition step was not conducted in Exp 1 because the sequential order of the polarities (positive polarity was followed by negative polarity) led to different temporal distributions of the two polarities. This raises the possibility that the magnitudes of FFR$_{pos}$ and FFR$_{neg}$ may differ because neural

---

[8] Spindle density measures were based on the periods classified as low arousal states (when spindle occurred), not on the entire measurement period. This was because the purpose of using spindle density was to help characterize participants' status during the low arousal states, rather than their average status across the entire experiment.

[9] The cutoff frequencies of 70 and 2000 Hz were different from Chapter 2 which used 70 and 4000 Hz. This is due to different vowels and frequency ranges of interest between Chapter 2 and 3. Chapter 2 used a syllable /i/ (which covered F2-F3 range up to 4000 Hz), while this current chapter (Chapter 3) used syllables /a/ in Exp 1 (which covered formants <2000 Hz) and /i/ in Exp 2. This chapter focused only on FFRs at F_0, while Chapter 2 also measured FFRs at F2-F3 range that led to greater filtered range up to 4000 Hz. Furthermore, because the 70–2000 Hz range was used in Schoof and Rosen (2016) (Exp 1) whose data were used in this chapter, such range was thus adapted here.

adaptation effects differ across the two polarities[10]. See *Minimizing adaptation effects and normalization of sweep numbers* for the procedures that checked for adaptation effects.

A set of 40-ms sliding windows at 1-ms steps was applied to the FFR waveforms across the stimulus period (100 ms for Exp 1 and 120 ms for Exp 2). Each 40-ms waveform was Hanning-windowed and zero-padded to 1 second. The spectral magnitude was measured at the frequency that corresponded to the F_0 value of the vowel at that step. Magnitudes were then averaged across all steps. As neural transmission from the cochlea to the auditory brainstem for FFRs takes between 5 and 10 ms (Chandrasekaran and Kraus, 2010; Skoe and Kraus, 2010), the maximum magnitude for time lags in the range 5 to 10 ms was used as the final FFR magnitude. An additional 3 ms was required in Exp 2 to account for air transmission from the loudspeaker to the cochlea (1-ms steps between 8 and 13 ms were used) (see Mai et al., 2018).

3.2.7 Theta-band phase-locked responses

Phase-locking values (PLV) at theta-band frequencies were measured (4–6 Hz, corresponding to stimulus repetition rates of ~5 syllables per second; see *Part 2.2*). This followed the same procedure as described in Chapter 2 for calculating the theta-band PLV. EEGs were decimated to 1024 Hz, re-referenced to the average of the bilateral earlobes and bandpass filtered (4–6 Hz) using a 2nd-order zero-phase Butterworth filter. Sweeps exceeding ± 15 µV on either electrode were rejected (Mai et al., 2018). A lower rejection threshold was used for theta-band PLV (± 15 µV) compared to FFRs (± 25 µV) because the theta-band signal normally does not have excessively high amplitude since a relatively narrow frequency range (4–6 Hz) was used. More than 80% of the sweeps were retained in all participants after artefact rejection. PLV time series (*PLV(t)*) were calculated (Morillon et al., 2012) as follows:

$$PLV(t) = \frac{1}{n}\left|\sum_{i=1}^{n} e^{j\phi_i(t)}\right|$$

where *n* denotes the total number of sweeps, $\phi_i(t)$ denotes the Hilbert phase series of the filtered EEG of the *i*th sweep time-locked to the syllable onset and *j* refers to $\sqrt{-1}$. Hilbert phase was used as it reflects phase-locking to stimuli even when EEG amplitude variation due to relaxation and eye closure occurred (Thatcher, 2012). The decision to measure theta-band PLV was the desire to examine the degree of EEG phase coherence relative to syllable onset. The perfect scenario is that theta-band EEG will be reset to the same phase value at each onset. This requires similar lengths of one cycle of the theta-band EEG and the stimulus onset asynchrony (SOA) of the stimuli, in order that the same phase value of EEG can appear around

---

[10] As the addition step was not conducted in Exp 1, the cochlear microphonic would also exist at the frequency of the first harmonics. However, because Exp 1 used the /a/ vowel which has relatively low energy at the first harmonics compared to other frequencies (see **Figure 3.2(A)**), the cochlear microphonic would not significantly influence the FFR magnitude, as demonstrated in previous studies (c.f. Skoe and Kraus, 2010).

each onset. Here, both one cycle of theta-band EEG (4–6 Hz) and SOA of the stimuli (~5 syllables per second) were at ~200 ms, which met this requirement.

As PLV is restricted to values between 0 and 1, it was logit-transformed to bound it between -∞ and +∞, making it appropriate for statistical analysis using linear regression (Waschke et al., 2017):

$$Logit\text{-}\ theta\text{-}band\ PLV\ (t) = ln\frac{PLV(t)}{1-PLV(t)}$$

Logit-theta-band PLV(t) values were then averaged across the stimulus period (100 ms for Exp 1and 120 ms for Exp 2). As neural transmission from cochlea to auditory cortex takes 10 to 30 ms in primates (Lakatos et al., 2007), the final Logit-theta-band PLV was taken as the maximum value for time lags between 10 and 30 ms at 1-ms steps (13 to 33 ms for Exp 2 with the added 3-ms for air transmission).

The θ-band phase-locked responses obtained using the current method correlated significantly with the behavioural performances (SiN perception) (Mai et al., 2018), which also supports the validity of the claim that cortical phase-locked sensory processing was estimated.

3.2.8 Statistical analyses

All statistical analyses used SPSS 23 (IBM, USA).

Minimizing adaptation effects and normalization of sweep numbers

As well as arousal states, two other factors that potentially may affect FFRs/Logit-theta-band PLVs were considered. First, since magnitudes of phase-locked activities are sensitive to the number of sweeps (Aviyente et al., 2011), problems can arise during statistical analyses if numbers of sweeps differ between the two arousal states. Second, neural adaptation could affect FFRs (Pérez-Gonzalez and Malmierca, 2014) and Logit-theta-band PLV since phase-locked adaptation has been reported in auditory cortex (Noda et al., 2014). Difference in temporal distributions of the high and low arousal epochs could lead to different adaptation between the two arousal states. Therefore, such adaptation differences may be confused with the arousal effects on FFRs and Logit-theta-band PLV.

To tackle the first issue, the number of sweeps was normalized to around 1500 for FFRs and around 500 for Logit-theta-band PLV for both types of arousal period for each participant (c.f., Dajani et al., 2005; Wong et al., 2007). To ensure the data quality was adequate with respect to number of syllable repetitions, participants whose artefact-free sweeps were fewer than 1,450 (for FFR) or 450 (for Logit-theta-band PLV) in either low or high arousal states were not included in subsequent analyses. This gave 58 and 91 participants for FFRs and Logit-theta-band PLV, respectively (see *Part 2.8.2 Combining data sets* for more details). Normalization was then conducted by randomly selecting epochs which contained the requisite

numbers of artefact-free sweeps between 1,450 and 1,550 for FFRs, and between 450 and 550 for Logit-theta-band PLV.

To tackle the second issue, two "adaptation indices" (AI) were defined for the 30-s blocks used in both experiments: (1) Within-Block AI (AI_{Within_Block}), and (2) Across-Block AI (AI_{Across_Block}) AI_{Within_Block} was defined as:

$$P_{High} = \frac{\sum_k \sum_i P_{i,k,High}}{\sum_k N_{k,High}}$$

$$P_{Low} = \frac{\sum_k \sum_i P_{i,k,Low}}{\sum_k N_{k,Low}}$$

$$AI_{Within\_Block} = P_{Low} - P_{High}$$

where $P_{i,k}$ denotes the position of the ith classified high/low arousal epoch in the kth block (where epoch hereafter refers specifically to those where the stimuli were presented in a quiet background); N_k denotes the number of the high/low arousal epochs in the kth block. P_{High} and P_{Low} thus represent the average within-block positions of the high and low arousal epochs, respectively. As such, $AI_{Within\_Block} > 0$ means that, on average, high arousal epochs were in earlier temporal positions than were low arousal epochs within blocks.

AI_{Across_Block} was defined as:

$$P_{High} = \frac{\sum_i P_{i,High}}{N_{High}}$$

$$P_{Low} = \frac{\sum_i P_{i,Low}}{N_{Low}}$$

$$AI_{Across\_Block} = P_{Low} - P_{High}$$

where P_i denotes the position of the ith high/low arousal epoch; N denotes the total number of high/low epochs across all blocks. P_{High} and P_{Low} thus represent the average across-block positions of the high and low arousal epochs, respectively. Therefore, $AI_{Across\_Block} > 0$ means that, on average, high arousal epochs were in earlier position than low arousal epochs across all blocks. Greater FFR/Logit-theta-band PLV magnitudes in high arousal periods may be due to less neural adaptation, rather than being ascribed to the effect of arousal itself. To avoid this situation, AI_{Within_Block} and AI_{Across_Block} both needed to be ≤ 0 at the group level.

To combine approaches that normalize sweep numbers and minimize adaptation, the signals were processed for each participant as follows: (1) The normalization procedure using 1450-1550 (FFR) or 450-550 (Logit-theta-band PLV) artefact-free sweeps was conducted 1000 times to generate 1000 sets of high and low arousal epochs. Within these 1000 sets, only those where $AI_{Within\_Block} \leq 0$ were retained, unless AI_{Within_Block} was above 0 for all sets. (2) The set of epochs with the minimum absolute value of AI_{Within_Block} was chosen for FFR/Logit-theta-band

PLV measurements. (3) Steps (1) and (2) were repeated 500 times giving 500 estimates of *AI* values, density of sleep spindles in low arousal periods, and FFR/Logit-theta-band PLV magnitudes in both arousal periods. Measures averaged over these 500 estimates were used in the final statistical analyses. The reason for refining AI_{Within_Block} rather than AI_{Across_Block} was because subsequent analyses found that both *AI*s were ≤ 0 at the group level when AI_{Within_Block} was ≤ 0 but not when the AI_{Across_Block} was ≤ 0.

For both FFRs and Logit-theta-band PLV, all *AI*s had mean values below zero (mean ± SD: $AI_{Within\_Block\_FFRs}$ = -0.018 ± 0.129; $AI_{Across\_Block\_FFRs}$ = -0.055 ± 0.272; $AI_{Within\_Block\_PLV}$ = -0.001 ± 0.028; $AI_{Across\_Block\_PLV}$ = -0.031 ± 0.237). None of the *AI*s differed statistically from zero (all p > 0.1). *AI*s being lower than zero reflected a later temporal position for the high than for the low arousal epochs. The results therefore indicated that, if any adaptation occurred, it should result in greater suppression on magnitudes of both FFRs and Logit-theta-band PLV in the high than in the low arousal state. Thus any effects of arousal that are found cannot be explained by adaptation.

Combining data sets

Stimuli were presented at similar sound intensities in both experiments (80 and 77 dB, respectively, see *Stimuli*). Gama et al., (2017) showed that FFRs generated via free-field acoustic stimulation (loudspeaker) are comparable to those measured in close field (inserted earphones) with the same sound intensity. To further confirm the validity of combining the data from the two experiments, three-way mixed ANOVAs were conducted for magnitudes of FFRs and Logit-theta-band PLV with the within-subject factor of Arousal (high vs. low) and the between-subject factors of Age Group (young vs. older) and Data Set (Exp 1 vs. Exp 2; **Table 3.1)**. Data from Exp 1 and Exp 2 were combined in subsequent analyses since there were no significant main effects or interactions involving Data Set. This resulted in data for 58 participants (25 young and 33 older) for FFRs and 91 participants (38 young and 53 older) for Logit-theta-band PLV.

It may be considered that FFRs would differ across data sets because the pitch contours of the stimulus differed (static in Exp 1 and falling in Exp 2). The falling contour used here corresponds to that used in some tonal languages (e.g., Mandarin). Non-tonal language speakers may be less sensitive to this linguistic-related feature compared to static pitch (e.g., Krishnan et al., 2005). However, lack of effects of Data Set indicates that the pitch contour did not affect the results.

Table 3.1 Summary statistics for the three-way ANOVA with factors of Arousal, Age Group and Data Set. The top and bottom panels are for FFRs and Logit-theta-band PLV respectively. DVs, *Df, F,* and *p* refer to dependent variables, degrees of freedom, *F*-values and *p*-values, respectively. Significant *p*-values are in bold. *p < 0.05; ***p < 0.001.

DVs	Factors	df1	df2	F	p
FFRs	Arousal	1	54	6.357	**0.015***
	Age Group	1	54	0.816	0.370
	Data Set	1	54	0.034	0.853
	Arousal × Age Group × Data Set	1	54	1.722	0.195
	Arousal × Age Group	1	54	1.176	0.283
	Arousal × Data Set	1	54	0.004	0.948
	Age Group × Data Set	1	54	0.296	0.588
Logit-theta-band PLV	Arousal	1	87	5.289	**0.024***
	Age Group	1	87	22.217	**< 0.001***
	Data Set	1	87	2.379	0.127
	Arousal × Age Group × Data Set	1	87	2.209	0.141
	Arousal × Age Group	1	87	0.241	0.624
	Arousal × Data Set	1	87	0.025	0.875
	Age Group × Data Set	1	87	0.445	0.506

Effects of arousal, age and sleep spindle density

To address the question whether state of arousal affected the phase-locked responses cortically and subcortically, and whether such effects change with age and sleep spindle density, linear mixed-effect regressions were conducted for FFRs and Logit-theta-band PLV. In these analyses arousal was the fixed-effect factor, Age and the spindle density in the low arousal states were fixed-effect covariates, and Participant was a random-effect factor. Spindle Density was included as a fixed-effect covariate because spindle density can reflect the sleep stability (Kim et al., 2012) and the degree of sensory deafferentation (Spoormaker et al., 2010, 2011; Picchioni et al., 2014) during the low arousal states. Due to the individual differences in spindle density in low arousal epochs, measuring the differences in phase-locked responses between the high and low arousal epochs may only crudely reflect the neuro-regulatory effects on these responses. Including Spindle Density in the low arousal states as a fixed-effect covariate should thus more accurately quantify the neuro-regulations when states of arousal are altered. Both Age and Spindle Density were mean-centred. The type of covariance matrix that

was chosen was the one that generated the smallest BIC value. Age (a continuous variable), rather than Age Group (with categorical levels), was used in the model here so that the effects of age itself and age-related variables could be compared. The extra age-related variables examined were pure-tone audiometric threshold (PTA) averaged across 0.25 and 4 kHz over both ears, and sleep spindle duration. The results of the analyses that used the age-related variables as fixed effect covariates instead of Age are given in the *Appendix*.

3.3 Results

3.3.1 Age effects on sleep spindles

Sleep spindle density, magnitude and duration in the low arousal epochs were compared between young and older adults using independent sample t-tests. Equal variances were not assumed during these t-tests, as Levene's test showed that variances differed significantly between the two age groups (all $p < 0.02$). There were no significant differences between young and older adults for spindle density (**Figure 3.4A,** t-test: $t(63.772) = 1.221$, $p = 0.227$) or spindle magnitude (**Figure 3.4B** , t-test: $t(54.723) = 0.767$, $p = 0.447$). Spindle duration was significantly longer in young than in older adults (**Figure 3.4C**, t-test: $t(58.756) = 3.006$, $p = 0.004$).
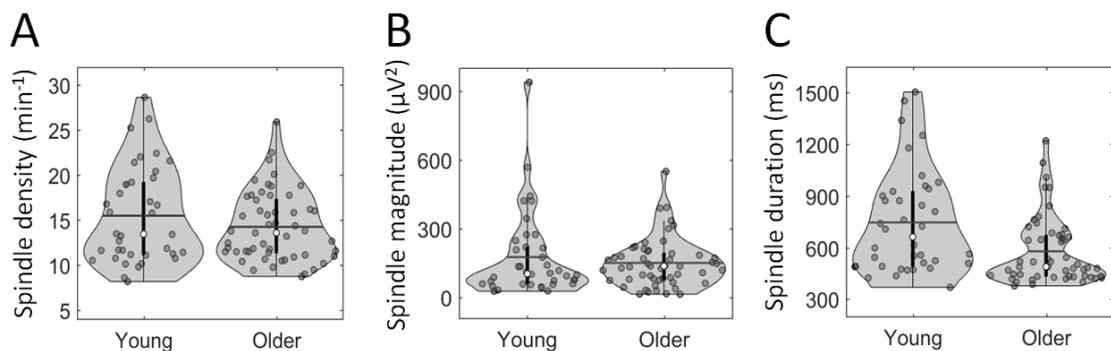
Figure 3.4 Violin plots of sleep spindle parameters for the two age groups. (A) *Spindle density.* ***(B)*** *Spindle magnitude.* ***(C)*** *Spindle duration.*

3.3.2 Effects of arousal, age and spindle density on FFRs and Logit-theta-band PLV

Figure 3.5A shows waveforms and spectra of FFR responses for one participant and **Figure 3.5B** shows changes of Logit-theta-band PLV across time for another participant. These participants were selected so that the differences in the respective measures between the two arousal states were closest to the group averages.
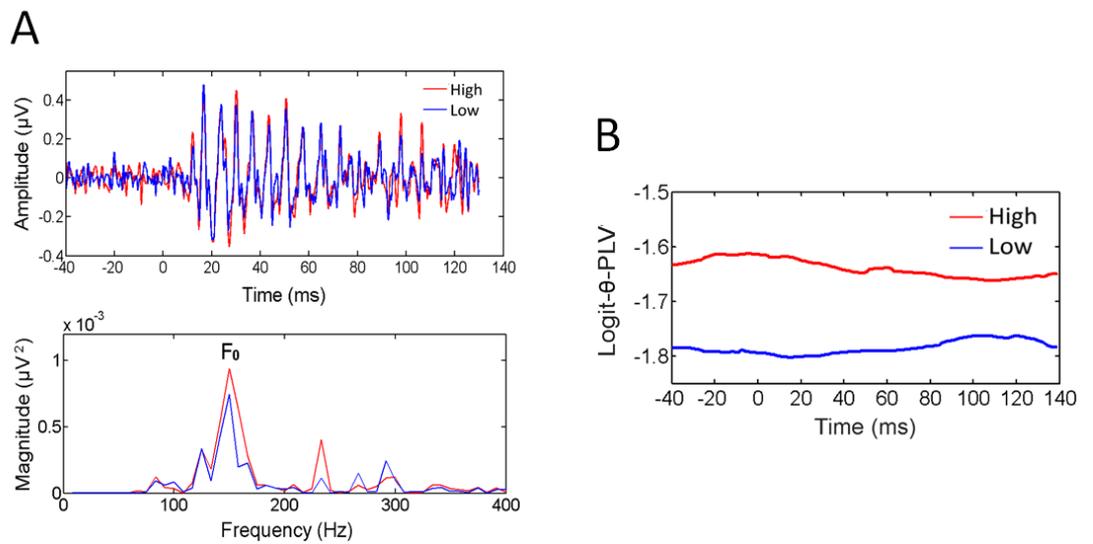
Figure 3.5 FFRs and Logit-theta-band PLV for selected individual participants. *Participants were chosen whose differences between the high and low arousal states were closest to the group averages for that measure. The red and blue lines indicate the high and the low arousal states, respectively.* ***(A)*** *Top: FFR waveforms across time; Bottom: the spectra for the sections between 0–120 ms. The spectra at the bottom of* ***(A)*** *peak at around F_0 frequency (labelled).* ***(B)*** *Changes of Logit-theta-band PLV across time.*

Distributions of FFR magnitudes and Logit-theta-band PLV across the arousal states and age groups are shown as violin plots in **Figure 3.6** and **Figure 3.7**, respectively.
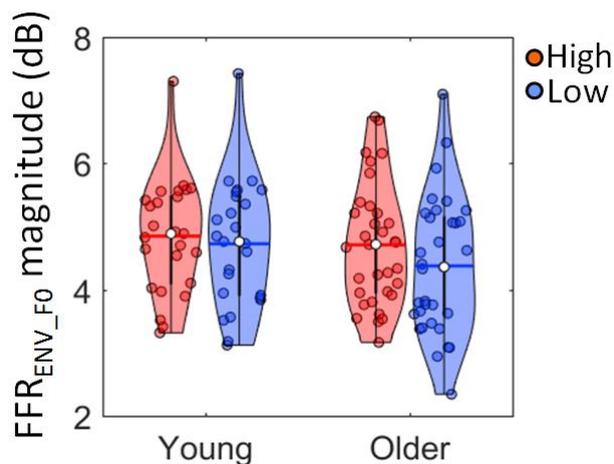
Figure 3.6. Violin plots of FFR magnitudes for the combined data sets. *Magnitudes are shown for the high and low arousal states across the two age groups. Red and blue bars indicate the high and the low arousal states, respectively.*
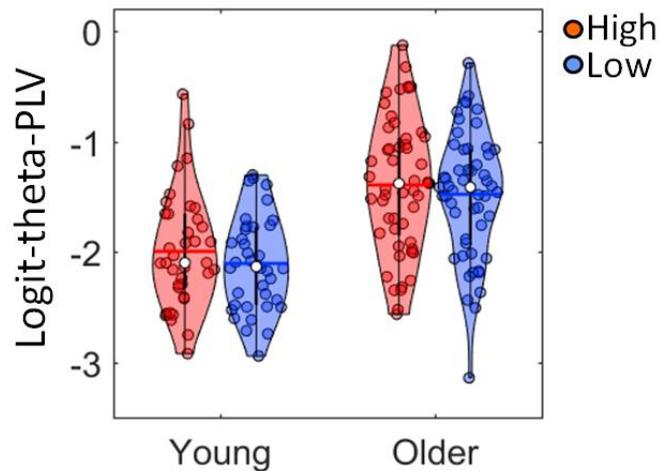
***Figure 3.7 Violin plots of Logit-theta-band PLV for the high and low arousal states across
the two age groups**. Red and blue bars indicate the high and the low arousal states,
respectively.*

Linear mixed-effect regressions were conducted using Arousal as the fixed-effect factor,
Age and Spindle Density in the low arousal states as the fixed-effect covariates, and Participant
as the random-effect factor. Statistics for the linear mixed-effect regressions are summarized in
Table 3.2. Significant main effects of Arousal were found for both FFRs ($F(1, 54) = 6.263$, $p =$
0.015) and Logit-theta-band PLV ($F(1, 87) = 5.520$, $p = 0.021$), with greater FFR magnitude and
Logit-theta-band PLV in the high than in the low arousal state. There was a significant main
effect of Age for Logit-theta-band PLV (Logit-theta-band PLV increased with age; $F(1, 87) =$
32.076, $p < 0.001$) but not for FFRs.

The [Arousal × Age] interactions were not significant either for FFRs or for Logit-theta-
band PLV. However, there was a significant three-way [Arousal × Age × Spindle Density]
interaction for Logit-theta-band PLV ($p = 0.010$; see **Table 3.2**) that suggests an interplay
between age and the effect of arousal on the cortical phase-locked processing. To follow this
up, a post-hoc analysis was conducted to examine how the [Arousal × Spindle Density]
interaction differed across ages. **Figure 3.8** shows the [Arousal × Spindle Density] interaction
(i.e., correlation between the effect of arousal on Logit-theta-band PLV (Logit-theta-band
$PLV_{High\_vs\_Low}$) and spindle density) in the young and older adults. The interaction was significant
for the young adults ($r = 0.402$, $p = 0.012$, Logit-theta-band $PLV_{High\_vs\_Low}$ increased with spindle
density; **Figure 3.8** left panel), but not for the older adults ($r = -0.062$, $p = 0.661$; **Figure 3.8**
right panel). Age-related variables (mean-centred PTA and spindle duration) were used
respectively as covariates that replaced Age in the model to test whether the aging effect could
result from age-related changes in peripheral hearing loss (PTA) or a spindle property (spindle
duration). No significant three-way interactions relevant to PTA or spindle duration were found
(see *Appendix*).

In summary, both FFR and Logit-theta-band PLV magnitudes were significantly affected by arousal, with greater magnitudes in the high than in the low arousal state. FFR magnitude did not show a significant decline with age as was predicted. Logit-theta-band PLV, as predicted, increased significantly with age. The significant three-way [Arousal × Age × Spindle Density] interaction for Logit-theta-band PLV showed that age interplays with the effect of arousal on the cortical phase-locked processing. Post-hoc analysis showed that the effect of arousal on Logit-theta-band PLV increased significantly with spindle density only in the young adults. Furthermore, no evidence was found for aging effects when the age-related factors of PTA or spindle density were used as covariates.

Table 3.2. Summary statistics for linear mixed-effect regressions, using Arousal as the fixed-effect factor, Age and Spindle Density as fixed-effect covariates, and Participant as the random-effect factor. DVs, *Df, F,* and *p* refer to dependent variables, degrees of freedom, *F*-values and *p*-values, respectively. Significant *p*-values are in bold and *p* < 0.05, and ***p* < 0.001.

DVs	Fixed-effect factors/covariates	*df1*	*df2*	*F*	*p*
FFRs	Arousal	1	54	6.263	**0.015***
	Age	1	54	1.401	0.242
	Spindle Density	1	54	0.004	0.947
	Arousal × Age × Spindle Density	1	54	0.242	0.625
	Arousal × Age	1	54	0.352	0.555
	Arousal × Spindle Density	1	54	0.207	0.651
	Age × Spindle Density	1	54	1.101	0.299
Logit-theta-band PLV	Arousal	1	87	5.520	**0.021***
	Age	1	87	32.076	**< 0.001***
	Spindle Density	1	87	0.976	0.326
	Arousal × Age × Spindle Density	1	87	6.848	**0.010***
	Arousal × Age	1	87	0.093	0.762
	Arousal × Spindle Density	1	87	1.754	0.189
	Age × Spindle Density	1	87	0.033	0.856
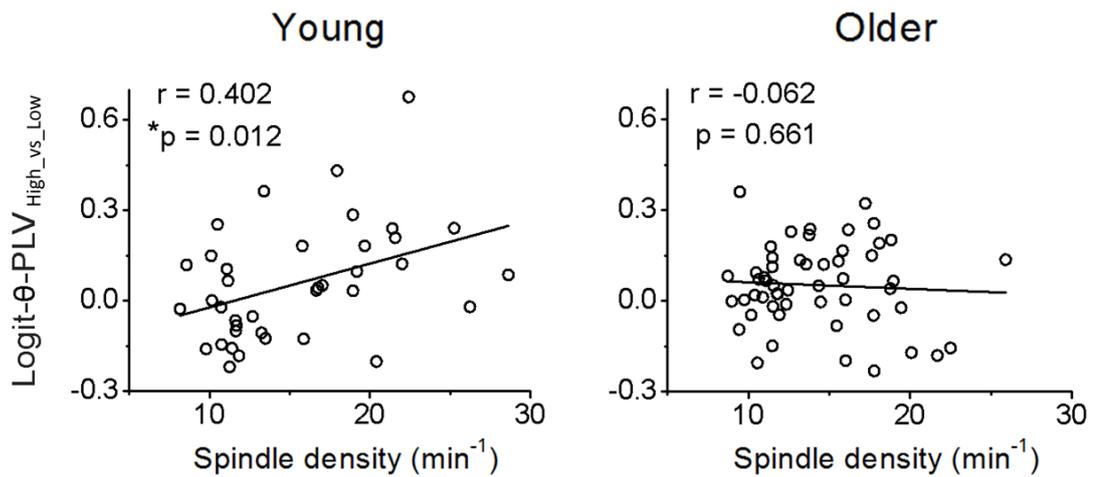
Figure 3.8. Interaction between Arousal and Spindle Density for Logit-theta-PLV (i.e. the correlation between the effect of arousal (Logit-theta-PLV_{High_vs_Low}) and spindle density for separate age groups. A significant [Arousal × Spindle Density] correlation occurred for the young (left), but not for the older adults (right).

3.4 Summary of results and brief discussions

Fuller discussions of this study will be in Chapter 5 (*5.2*).

The present study showed that arousal affects both speech-evoked theta-band PLV and FFR_{ENV_F0}. Both types of responses were statistically greater under high (no sleep spindles) than under low arousal states (with spindles) after potential neural adaptation had been ruled out. The effects were statistically significant and the effect sizes were in the medium range (Cohen, 1988; see **Table 3.2**). These results show that arousal affects the neuro-temporal precision of responses to speech at early sensory levels in the auditory system[11]. Age effects were found on theta-band PLV and sleep spindle duration. As predicted, theta-band PLV increased with age as spindle duration decreased. Furthermore, age interacted significantly with arousal and sleep spindle density in the low arousal states for theta-band PLV. The arousal effect on theta-band PLV increased significantly as spindle density increased in the young, but not the older, adults. Thus, incidence of sleep spindles during nREM sleep affects auditory processing differentially across ages.

The effect of arousal on theta-band PLV is consistent with the previous studies showing cortical responses to speech decrease with decreased level of arousal (Czisch et al., 2002, 2004; Davis et al., 2007; Wilf et al., 2016). The effect of arousal on the speech-evoked FFRs, however, was not in line with the previous studies showing that decreases in magnitudes of

---

[11] It is also noteworthy that FFR here was quantified as spectral magnitude. So mathematically it is not merely determined by temporal precision/synchrony of phase, but also by single-trial spectral magnitudes. Therefore, it is possible that arousal level affects FFRs by changing single-trial spectral magnitudes as well as temporal precision.

auditory steady-state responses (ASSRs) which, during sleep, occurred only when the modulation rates was below 70 Hz (Cohen et al., 1991; Lins et al., 1995; Picton et al., 2003). Here, the F_0 frequencies were over 100 Hz and FFRs were significantly reduced with decreased level of arousal. It is possible that the stimuli to elicit ASSRs were acoustically much simpler (amplitude-modulation with a pure-tone as the carrier, see e.g., Picton et al., 2003) compared to stimuli used in the present study (speech sounds with complex harmonic carriers).

Effects of age were found in the present study. The results showed significantly higher theta-band PLV and shorter spindle duration in older adults compared to young adults. The finding of greater theta-band PLV in older adults is consistent with Study 1 as well as previous studies that have shown age-related increases in theta-band ASSRs (Tlumak et al., 2015; Goossens et al., 2016) reflecting the hyperexcitability of the central auditory system as a result of aging (Caspary et al., 2008). The observation that spindle duration is shorter in older adults is also consistent with a previous report (Martin et al., 2013). Although no significant interaction was found between Age and Arousal for magnitude of either FFRs or theta-band PLVs, aging could still interplay with the effect of arousal. A significant three-way [Arousal × Age × Spindle Density] interaction was found for Logit-theta-band PLV. Post-hoc analysis showed that the effect of arousal on Logit-theta-band PLV increased statistically with sleep spindle density for the young adults alone. Furthermore, it was shown that this discrepancy between the young and older adults was attributable to age itself, rather than age-related variables such as hearing loss (PTA) or shorter spindle duration (see *Appendix 4*). The lack of the two-way [Arousal × Age] interaction indicate that Arousal can only be a crude proxy for characterizing the degree of arousal, as sleep spindle density in the low arousal states differed across participants. It is imperative to include Spindle Density into the model. This is because spindle density should more accurately describe the stability of sleep status (Kim et al., 2012) and degree of sensory deafferentation (Spoormaker et al., 2010, 2011; Picchioni et al., 2014) during nREM sleep and thus should more accurately measure the neuro-regulation of the auditory responses. This is evidenced by the observed correlation between the effect of arousal and spindle density in young adults. In older adults, such correlation was lacking, which indicates that speech spindles play less of a role in regulating auditory responses as people age. The present results found no evidence that the reduced regulatory role of sleep spindles was due to age-related changes in peripheral hearing or particular spindle properties (spindle duration here). As such, the neural mechanisms underlying this phenomenon (i.e., neuro-regulatory role of spindle density in young but not in older adults) need further clarification/investigation in the future.

The present study was the first to investigate the effect of arousal on phase-locked neural responses to speech signals and to examine how aging interplays with these effects. The results highlight the significant role arousal plays in assisting processing of fine-grained acoustic properties of F_0 and envelope modulations at the sensory level. A possible regulatory role of sleep spindles for phase-locked responses in the auditory cortex was revealed and it was further found that aging reduced the role of spindle regulation.

Chapter 4

Causal relationship between the right auditory cortex and frequency-following responses: A combined tDCS and EEG study

4.1 Introduction

As discussed in the previous chapters, speech-evoked FFRs are closely related to both fundamental auditory processes and are proposed to be clinical biomarkers for various speech and language processing disorders. It is argued that the fundamental and clinical importance of FFRs is linked to the neural fidelity of speech signals in the inferior colliculus at the brainstem, which has been proposed as the main neural origin of FFRs (Chandrasekaran and Kraus, 2010; Bidelman, 2018). However, recent studies have shown an additional source of FFRs in the right auditory cortex that is associated with musical experience, pitch discrimination ability and speech-in-noise perception (Coffey et al., 2016, 2017a). FFR strength was further shown to be associated with right-lateralized hemodynamic activity in the auditory cortex (Coffey et al., 2017b), consistent with the relative specialization of right auditory cortex for pitch and tonal processing (Zatorre and Berlin, 2001; Patterson et al., 2002; Hyde et al., 2008; Albouy et al., 2013; Cha et al., 2016).

Despite findings that show the potential cortical contributions to FFRs, it is unclear whether such contributions are causal. The aim of the present study (Study 3 of this thesis) was to determine whether there is a causal relationship between auditory cortex and the FFRs. Here, neuro-stimulation by transcranial direct current stimulation (tDCS) was applied to alter neural excitability in the left and right auditory cortex. I tested for the after-effects of tDCS on the speech-evoked FFRs using electroencephalography (EEG). tDCS is a non-invasive neuro-stimulation technique that modulates cortical excitability (Jacobson et al., 2012). By applying direct currents over the scalp, tDCS causes depolarization (anodal) and hyperpolarization (cathodal) of neurons that leads, respectively, to neural excitation and inhibition in proximal parts of the cortex that last for up to 90 minutes post-stimulation (Nitsche and Paulus, 2001). Previous studies showed that applying tDCS over the right, compared to the left, auditory cortex can significantly change pitch discrimination performances (Mathys et al., 2010; Matsushita et al., 2015). Thus, this supports the causal role of the right auditory cortex for pitch perception. However, such causality has not been established for neurophysiological signatures like FFRs.

The present study tested the hypothesis that tDCS over the right auditory cortex should change the strength of FFRs. Furthermore, I predicted that such after-effects should occur

particularly along the contralateral auditory pathway (i.e., from the left ear to the right auditory cortex).

4.2 Materials and Methods

4.2.1 Participants

Ninety participants (18-40 years old; 45 females) were recruited and completed the entire experiment. Two other participants dropped out during the tDCS phase because they felt uncomfortable with the skin sensation when stimulation was applied. All participants had normal-hearing (pure-tone audiometric thresholds <25 dB HL within frequency range of 0.25–6 kHz for both ears) tested using a MAICO MA41 Audiometer (MAICO Diagnostics, Germany). Participants were non-tonal language speakers, had no long-term musical training and reported no history of neurological or speech/language disorders. They had not participated in any brain stimulation experiments in the two weeks prior to the present experiment.

All participants were right-handed (Handedness Index (HI) > 40; Oldfield, 1971). Participants were assigned at random to one of five groups, each of which received different types of tDCS (detailed in *Experimental design*). HI did not differ significantly between the five groups (all p > 0.4, uncorrected), indicating that the degree of handedness was well matched across stimulation types. The absence of HI differences across groups is important because handedness has been argued to influence functional hemispheric specialization (Carey et al., 2014; Willems et al., 2014). Hence matching the HI across groups ensured that any effects of tDCS were not confounded with handedness.

4.2.2 Syllable stimulus for the FFR recording
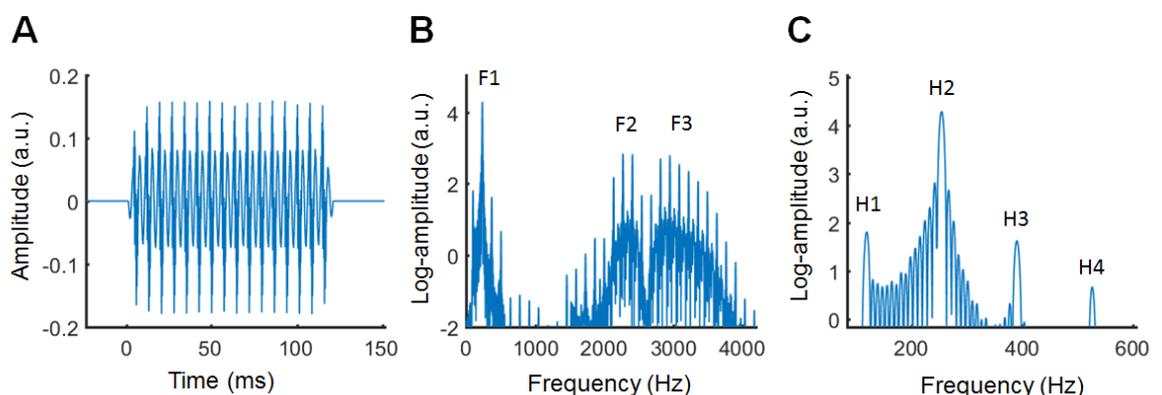
Figure 4.1. The syllable stimulus for FFR recordings. (A) *Temporal waveform of the syllable /i/.* ***(B)*** *Spectrum of the syllable (0–4000 Hz) showing the formant distributions (F1, F2 and F3).* ***(C)*** *The same spectrum as* ***(B)*** *that shows the first four harmonics with F_0 at 136 Hz. N.B., the spectrum was obtained via Fast Fourier Transform (FFT) after zero-padding the temporal waveform to 1 second.*

A 120-ms-long syllable /i/ spoken by a male with a static fundamental frequency (F_0) at 136 Hz was used for the FFR recordings. The waveform and spectrum of the syllable are shown as **Figure 4.1**. The syllable has three formants (F1, F2 and F3 at ~280, 2400 and 3100 Hz, respectively). It has a stable amplitude profile across the syllable period except for the 5-ms rising and falling cosine ramps applied at the onset and offset to avoid transients.

4.2.3 Experimental design

The experimental procedure is summarized in **Figure 4.2**. FFRs were recorded pre- and post-tDCS during monaural listening to the syllable stimulus to test for any after-effects of tDCS.

FFR recording

EEG were recorded over participants' scalps (Biosemi ActiView, The Netherlands) whilst they listened to the repeatedly-presented syllable /i/ (see *Syllable stimulus for the FFR recording*) both pre- and post-tDCS. The recording site was at the vertex (Cz localized using a standard Biosemi cap, which is the conventional site used for obtaining FFRs, (Skoe and Kraus, 2010). Bilateral earlobes served as the reference and the sampling rate was 16,384 Hz. The auditory stimulus were presented at ~4 syllables per second (inter-stimulus interval (ISI) of 120 ms). The stimulus was played monaurally via electrically shielded inserted earphone (ER-3 insert earphones, Intelligent Hearing Systems, Miami, FL) at 85 dB SL (excluding ISIs) in each ear (e.g. left-ear listening followed by right-ear listening or vice versa with order of ear presentation counterbalanced across participants). Monaural listening ensured that after-effects of ipsilateral and contralateral tDCS (relative to the listening ears) could be tested separately (see *Statistical analyses*). For each ear, there were 1,500 sweeps for the positive and 1,500 sweeps for the negative polarity presented in an intermixed order (3,000 syllables in total).

Participants were seated comfortably in an armchair in an electromagnetically- and sound-shielded booth. They listened passively to the stimulus sequence whilst keeping their eyes on a fixation cross on the centre of a computer screen. The 3,000 syllable sweeps in each ear were broken into six 2-minute-long blocks (500 sweeps each) with ~40 second breaks between blocks. Participants were required to keep awake and refrain from body and head movements whilst they were listening to sounds. The FFR recording lasted for ~30 minutes for both pre- and post-tDCS. The post-tDCS recording was completed within 45 minutes post-tDCS for all participants to ensure that any after-effects of tDCS on FFRs were sustained (Nitsche and Paulus, 2001).

tDCS

tDCS was applied over the scalp using a battery-driven direct current stimulator (Magstim HDCStim, UK) with a pair of rubber-surface electrodes (5×5 cm) contained in saline-soaked cotton pads. Participants were assigned at random to one of the five groups (18 participants (9 females) per group; single-blinded). The five groups received the following different types of

tDCS: (1) anodal stimulation on the left auditory cortex (AC) (Left-Anod); (2) cathodal stimulation on the left AC (Left-Cathod); (3) anodal stimulation on the right AC (Right-Anod); (4) cathodal stimulation on the right AC (Right-Cathod); and (5) Sham, with electrode configurations randomly chosen from (1)–(4) for each participant (in this group, the active electrode was put on the left AC for half of the participants and on the right AC for the other half). Centre position of the active electrode was on T7/T8 (according to the 10/20 EEG system) for the left/right AC. Reference electrode was placed on the forehead above the eyebrow contralateral to the active electrode (see Matsushita et al., 2015; also see **Figure 4.2**). For groups (1)–(4), tDCS was applied at 1 mA for 25 minutes with the currents ramping up/down for 15 seconds at the stimulation onset/offset. Sham applied tDCS only for 30 seconds in total (15 seconds ramping up and down respectively) at the onset of stimulation. This created the usual sensations associated with tDCS in Sham but without actual stimulation during the remainder of the 25-minute run. All experimental sessions were conducted during the day time (mornings or early afternoons) and all participants had enough sleep (at least 6 hrs) the night before (based on self-report prior to the experiment) to ensure adequate cortical plasticity triggered by tDCS (Salehinejad et al., 2019).
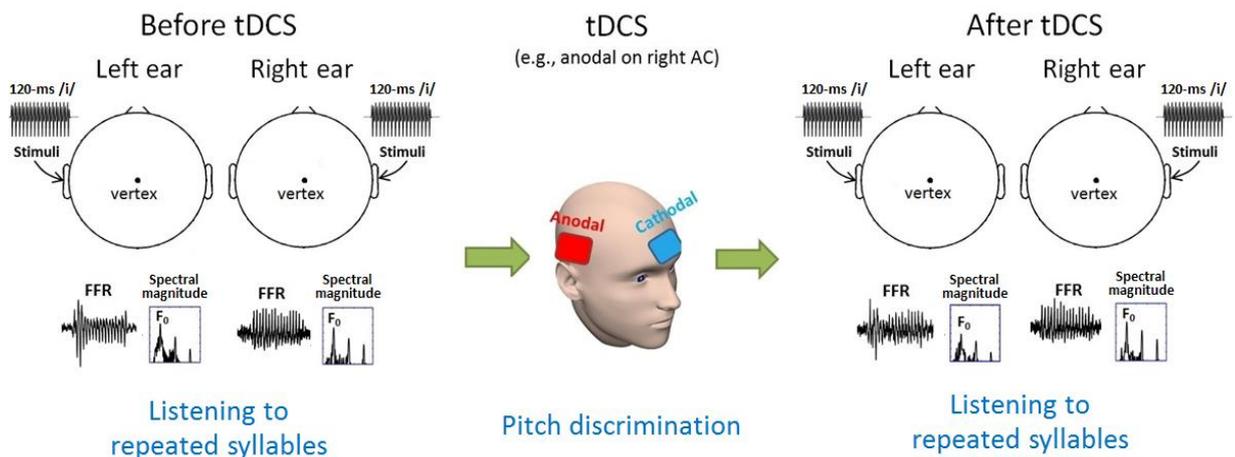
Figure 4.2. Illustrations for the experiment design. *Participants first listened to a repeated syllable /i/ monaurally while FFRs were recorded over scalp-EEGs at Cz. tDCS was then applied over the auditory cortex (AC) along with a pitch discrimination task. The same syllable listening task as in the first step was finally performed following tDCS to detect any after-effects of neuro-stimulation.*

During neuro-stimulation, participants completed a pitch discrimination task while they listened to sound stimuli over a loudspeaker 1 metre in front of them in the same sound-shielded booth used for the FFR recordings. Three short complex tones (400 ms long with frequencies spanning from the first harmonic to ~4000 Hz) were presented on each trial at a calibrated level of 75 dB SL at the 1 metre position. The task was an 'ABX' task. In each trial, two tones 'A' and 'B' with different fundamental frequencies (F_0) were played consecutively

followed by a third tone 'X' randomly selected from 'A' or 'B'. The 'standard' F_0 (F_0 of either 'A' or 'B') was at 136 Hz which was the same F_0 as in the syllable stimulus used for FFR recording. The initial F_0 difference between 'A' and 'B' was set at 16 Hz. Participants had to identify whether 'X' was the same as 'A' or 'B'. They gave their best guess when they were unsure of the answer. The process followed a '2-down, 1-up' adaptive procedure, in which the F_0 difference between 'A' and 'B' decreased by $\sqrt{2}$ times following two consecutive correct trials and increased by $\sqrt{2}$ times following an incorrect trial. No feedback about response accuracy was provided. Half-minute breaks were taken every 4 minutes. This task was included during tDCS because tDCS modulates neural networks that are currently active (Reato et al., 2010; Ranieri et al., 2012; Bikson and Rahman, 2013). Concurrent tDCS and the pitch discrimination task could therefore maintain auditory cortical activity during neuro-stimulation, hence maximizing the effect of tDCS on neural excitability.

4.2.4 EEG Signal processing

All EEG signal processing was conducted via Matlab R2017a (The Mathworks).

Pre-processing

As mentioned, FFRs were captured from Cz. The EEG signals were first re-referenced to the bilateral earlobes and bandpass-filtered between 90 and 4000 Hz using a 2nd-order zero-phase Butterworth filter. The filtered signals were then segmented for each sweep (-50 to 150 ms relative the syllable onset). Each segment was baseline-corrected by subtracting the average of the pre-stimulus (-50–0 ms) period. Segments that exceeded ±25 mV were rejected to minimize movement artefacts. The resultant rejection rates were < 2.5% averaged across participants for all cases (pre- and post-tDCS for the five stimulation groups for both left and right ear conditions).

FFR magnitudes

FFRs for the positive and negative polarities (FFR_{Pos} and FFR_{Neg}) were first obtained by temporally averaging the pre-processed signals across sweeps with the respective polarities. FFRs for envelopes of F_0 and its harmonics (i.e., periodicity; FFR_{ENV}) and temporal fine structures (TFS; FFR_{TFS}) were obtained by adding and subtracting FFR_{Pos} and FFR_{Neg}, respectively (Aiken and Picton, 2008). The addition and subtraction minimized the responses to TFS in FFR_{ENV} and to envelopes in FFR_{TFS}, so that purer FFRs to envelopes and TFS were obtained separately (Aiken and Picton, 2008). Spectral magnitudes of FFR_{ENV} and FFR_{TFS} were then calculated.
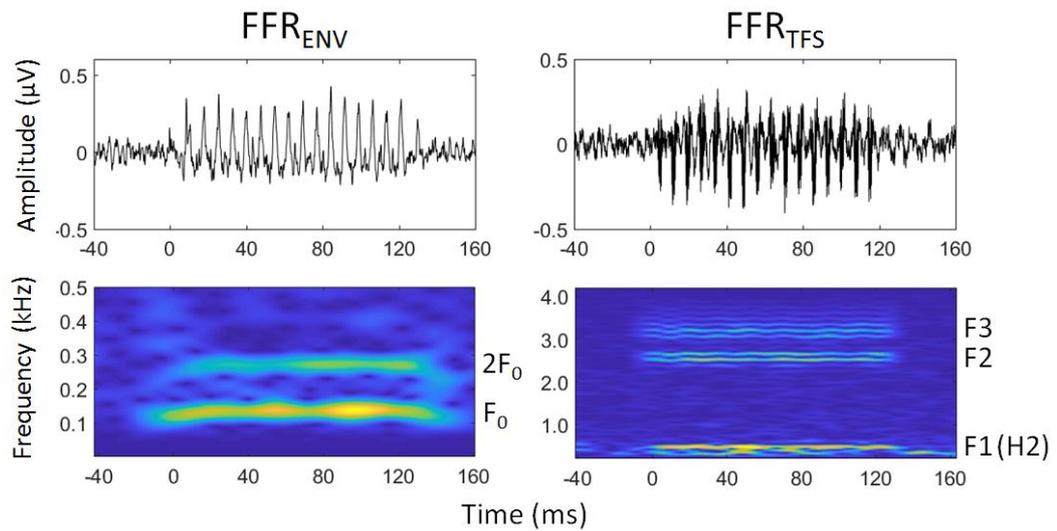
Figure 4.3 A representative sample of FFRs. *Sample waveforms (top panels) and the corresponding spectrograms (lower panels) of FFR_{ENV} (left) and FFR_{TFS} (right) were obtained from a single participant in the left ear listening condition before tDCS was applied. The first two harmonics of F_0 (F_0 and $2F_0$) dominate the power of FFR_{ENV} as indicated in the FFR_{ENV} spectrogram (lower left). The three formants (F1, F2 and F3) in FFR_{TFS} are shown and indicated in the FFR_{TFS} spectrogram (lower right); F1 occurs at H2 for this vowel (the 2nd harmonic).*

For FFR_{ENV}, FFR_{ENV_F0} and FFR_{ENV_2F0} (FFR_{ENV} at F_0 and its 2nd harmonic, $2F_0$) that dominate the power of FFR_{ENV} (see **Figure 4.3** left panel) were focused on. Whereas FFR_{ENV_F0} and FFR_{ENV_2F0} reflect neural phase-locking to the stimulus envelope periodicity in the central auditory systems, higher harmonics (≥ 3) of FFR_{ENV} may reflect distortion products resulting from non-linearities in response to acoustic stimuli on the basilar membrane (Smalt et al., 2012). Whilst it is expected that FFR_{ENV_F0} plays the major role in phase-locking to speech periodicity, FFR_{ENV_2F0} may also contribute (e.g., Aiken and Picton, 2008) because of the non-sinusoidal characteristics of speech periodicity (Holmberg et al., 1988; also see discussions in Smalt et al., 2012). The procedure for measuring the magnitudes of FFR_{ENV_F0} and FFR_{ENV_2F0} was as follows: a 120 ms (same length as the stimulus syllable) window with a 5-ms rising/falling cosine ramp at the onset/offset, was applied to the FFR_{ENV} waveform. Furthermore, as FFR_{ENV} occurs at the auditory brainstem (Chandrasekaran and Kraus, 2010; Bidelman, 2015, 2018) and/or primary auditory cortex (Coffey et al., 2016), the neural transmission delays were set at 5–20 ms. Therefore, the window was applied with an onset between 6 and 21 ms (allowing for an additional ~1 ms sound transmission through the plastic tube of the earphone to the cochlea) after the syllable onset. The window then slid within this time range (1 ms per step) to find the optimal onset/neural delay for the power measurement (see below). The windowed FFR_{ENV} waveform in each step was then zero-padded to 1 second to allow for a frequency resolution of 1 Hz and the log-transformed FFT-powers ($10*\log_{10}$[power]) centred at F_0 and $2F_0$ were measured (averaged across 136 ± 2 Hz and 272 ± 2 Hz, respectively). Finally, the FFR_{ENV_F0} and FFR_{ENV_2F0} magnitudes were taken as the powers at the optimal neural delays (i.e., when

powers are maximal for F_0 and $2F_0$, respectively). Such procedure is different from that in Study 1 and 2 (see Chapter 2 and 3, respectively) in which a set of short (40 ms) sliding windows was applied before considering the neural transmission delays. The reason is that Study 1 and 2 involved syllable stimuli with a dynamic pitch (a falling tone in the syllable /i/). Measuring the FFR magnitudes thus requires applying short sliding windows so that the magnitude of responses to pitch frequency at each step was correctly measured. On the other hand, syllable stimulus in the current study had a static pitch. In this case, a longer window that covered the whole syllable duration (120 ms) is adequate to accurately measure the FFR magnitudes.

For FFR_{TFS}, FFR_{TFS_H2} and FFR_{TFS_F2F3} (FFR_{TFS} at the 2nd harmonic that represents F1 for this vowel, and at F2 and F3, respectively; see **Figure 4.3** right panel) were focused on. FFR_{TFS_H2} reflects FFRs to TFS at the resolved-harmonic region while FFR_{TFS_F2F3} reflects FFRs to TFS at the unresolved-harmonic region. The same procedure was followed and the same 120 ms window was used when measuring magnitudes of FFR_{ENV_F0} and FFR_{ENV_2F0}, except that: (1) the procedure was applied on FFR_{TFS} at H2 (for FFR_{TFS_H2}) and at H16–H27 (the 16th to 27th harmonics corresponding to the range of F2 and F3 for FFR_{TFS_F2F3}; the final magnitude was taken as the mean magnitude across all harmonics in this range); (2) the neural delays during analyses were set at 1–6 ms (0–5 ms delays allowing an additional 1 ms sound transmission through the plastic tube of the earphone) as FFR_{TFS} arises at earlier stages of auditory processing in the periphery (Aiken and Picton, 2008).

Because of the different neural origins of FFR_{ENV} (brainstem/auditory cortex) and FFR_{TFS} (periphery), the present study thus allows us to confirm whether tDCS applied to auditory cortex affects FFRs that arise at different levels of the auditory systems.

4.2.5 Statistical analyses

Before testing the after-effects of tDCS, analyses were first conducted to check whether baseline (pre-tDCS) characteristics were matched across stimulation. Linear mixed-effect regressions were conducted using the baseline magnitudes and optimal neural delays of FFRs as dependent variables, Stimulation (Left-Anode, Left-Cathode, Right-Anode, Right-Cathode and Sham) and Ear (left vs. right) as the fixed-effect factors and Participant as the random-effect factor. Post-hoc analyses were conducted following significant interactions or main effects.

After-effects of tDCS (differences in FFR magnitudes between post- and pre-tDCS) were tested also using linear mixed-effect regressions. These were conducted using after-effects as dependent variables, Stimulation and Ear as fixed-effect factors and Participant as the random-effect factor. Post-hoc analyses were conducted following significant interactions or main effects.

Furthermore, regardless of whether interaction effects occurred between Stimulation and Ear, planned comparisons for the after-effects were conducted between different stimulation types in the left and right ear conditions, respectively. This was because collapsing the left and

right ears would smear the distinctions between any after-effects along the contralateral pathway (ears with tDCS on the opposite side) and those along the ipsilateral pathway (ears with tDCS on the same side), which was one of the aspects addressed in the present study. As multiple comparisons were conducted for each ear (5 stimulation types leading to 10 comparisons), the critical α value for detecting significance was adjusted at 0.005 for such planned comparisons. It was predicted that, compared to Sham, significantly greater after-effects of tDCS over the right auditory cortex (Right-Anode and Right-Cathode), but not the left auditory cortex (Left-Anode or Left-Cathode), should be found. Particularly, the after-effects are predicted to occur in the left ear listening condition (stimulus presentation side contralateral to the stimulation over the right auditory cortex), consistent with the current hypothesis that the right auditory cortex makes specific contributions to FFRs along the contralateral pathway.

FFR magnitudes were magnitudes of FFR_{ENV} (FFR_{ENV_F0} and FFR_{ENV_2F0}) and FFR_{TFS} (FFR_{TFS_H2} and FFR_{TFS_F2F3}) (see *EEG signal processing*). For FFR_{ENV}, the present study combined the magnitudes of FFR_{ENV_F0} and FFR_{ENV_2F0}, rather than use them as separate dependent variables. The reason was that, it was observed that the summed FFR_{ENV_F0} and FFR_{ENV_2F0} magnitude yielded greater effect sizes during planned comparisons where statistical significance ($p < 0.05$, uncorrected) was detected using FFR_{ENV_F0} or FFR_{ENV_2F0} magnitude alone: Cohen's d = 0.752 and 1.001 for FFR_{ENV_F0} and for the summed FFR_{ENV_F0} and FFR_{ENV_2F0} magnitude, respectively, when Right-Anode was compared with Sham in the left ear listening condition; Cohen's d = 0.934 and 1.140 for FFR_{ENV_F0} and for combined FFR_{ENV_F0} and FFR_{ENV_2F0} magnitude, respectively, when Right-Cathode was compared with Sham in the left ear listening condition (see *Results* for further details).

4.3 Results

4.3.1 Baseline characteristics

Table 4.1 and **4.2** shows the baseline magnitudes and neural delays for FFR_{ENV}, FFR_{TFS_H2} and FFR_{TFS_F2F3} in both the left and right ear conditions. ANOVAs were conducted for baseline magnitudes and optimal neural delays of FFR_{ENV}, FFR_{TFS_H2} and FFR_{TFS_F2F3}.

For FFR_{ENV}, a significant main effect of Ear was found for the magnitude ($F(1, 85)$ = 12.318, $p < 0.001$; greater magnitude in the left than in the right ear condition) but not for the neural delay ($F(1, 85)$ = 0.055, $p = 0.815$); no main effects of Stimulation (magnitude: $F(4, 85)$ = 0.932, $p = 0.450$; neural delay: $F(4, 85)$ = 0.799, $p = 0.529$) or [Stimulation × Ear] interactions were found (magnitude: $F(4, 85)$ = 0.541, $p = 0.706$; neural delay: $F(4, 85)$ = 0.046, $p = 0.996$). Furthermore, no significant differences were found between any stimulation type in either ear condition (magnitude: all $p > 0.07$; neural delay: all $p > 0.1$). **Figure 4.4** illustrates the comparison of baseline magnitudes for FFR_{ENV} between the left and right ear conditions after

collapsing across stimulation types (due to the significant main effect of Ear but no main effect of Stimulation).

For FFR_{TFS_H2}, there were no significant main effects of Stimulation (magnitude: $F(4, 85) = 0.692$, $p = 0.600$; neural delay: $F(4, 85) = 1.421$, $p = 0.234$) or Ear (magnitude: $F(1, 85) = 3.483$, $p = 0.065$; neural delay: $F(1, 85) = 1.842$, $p = 0.178$), or [Stimulation $\times$ Ear] interactions (magnitude: $F(4, 85) = 0.744$, $p = 0.565$; neural delay: $F(4, 85) = 0.587$, $p = 0.673$). No significant differences were found between any stimulation type in either ear condition (magnitude: all $p > 0.1$; neural delay: all $p > 0.05$).

For FFR_{TFS_F2F3}, significant main effects of Stimulation ($F(4, 85) = 40.872$, $p < 0.001$) and Ear ($F(1, 85) = 4.225$, $p = 0.002$; greater in the right than the left ear condition) were found for the magnitude, but not for the neural delay (Stimulation: $F(4, 85) = 1.504$; $p = 0.208$; Ear: $F(1, 85) = 0.324$, $p = 0.571$). A significant [Stimulation $\times$ Ear] interaction was found for the neural delay ($F(4, 85) = 2.549$, $p = 0.045$), but not for the magnitude ($F(4, 85) = 1.763$, $p = 0.144$). Post-hoc analyses found significant differences in magnitudes between several stimulation types (collapsing the left and right ears: Left-Anode vs. Right-Anode, $t(34) = -2.110$, $p = 0.042$; Left-Anode vs. Sham, $t(34) = -2.713$, $p = 0.010$; Left-Cathode vs. Right-Anode, $t(34) = -2.796$, $p = 0.008$; Left-Cathode vs. Right-Cathode, $t(34) = -2.566$, $p = 0.015$; Left-Cathode vs. Sham, $t(34) = -3.498$, $p = 0.001$). Significant differences were found between stimulation types for the neural delay in both the left ear (Left-Anode vs. Right-Cathode, $t(34) = -2.703$, $p = 0.011$) and the right ear condition (Right-Anode vs. Right-Cathode, $t(34) = 2.279$, $p = 0.029$; Left-Anode vs. Right-Anode, $t(34) = -2.240$, $p = 0.032$; Right-Anode vs. Sham, $t(34) = 2.629$, $p = 0.013$). All p-values here are reported without correction.

The results thus indicate that the baseline characteristics of FFR_{ENV} and FFR_{TFS_H2}, but not FFR_{TFS_F2F3}, were well matched across stimulation types. As such, although after-effects were tested for all three FFR signatures, FFR_{ENV} and FFR_{TFS_H2} are focused on. In addition, the main effects of Ear for FFR_{ENV} and FFR_{TFS_F2F3} magnitudes may reflect the laterality of speech encoding at the subcortical (Chandrasekaran and Kraus, 2010; Bidelman, 2015, 2018) and/or cortical levels (Coffey et al., 2016, 2017b), which will be discussed further (see *4.4* and *5.3*).

Table 4.1. Baseline magnitudes (standard deviations shown in the brackets; in *dB*) for FFR_{ENV}, FFR_{TFS_H2} and FFR_{TFS_F2F3} across stimulation types in the left and right ear conditions.

FFRs	Ear	Left-Anode	Left-Cathode	Right-Anode	Right-Cathode	Sham
FFR_{ENV}	Left	76.28 (5.27)	78.84 (7.03)	76.05 (6.96)	76.24 (8.56)	73.45 (10.05)
	Right	75.05 (5.79)	75.42 (4.72)	72.82 (6.62)	74.24 (7.91)	72.17 (10.92)
FFR_{TFS_H2}	Left	30.35 (5.70)	30.70 (7.71)	32.52 (7.12)	31.68 (6.24)	33.37 (6.86)

	Right	32.71 (3.88)	30.63 (6.98)	32.59 (7.97)	33.40 (7.51)	34.36 (5.66)
FFR_{TFS_F2F3}	Left	15.31 (7.26)	13.21 (7.24)	19.45 (7.16)	20.14 (6.58)	20.46 (5.75)
	Right	17.13 (7.07)	16.28 (6.58)	22.97 (7.57)	20.90 (7.82)	23.58 (6.09)

Table 4.2. Baseline neural delays (standard deviations shown in the brackets; in *ms*) for FFR_{ENV}, FFR_{TFS_H2} and FFR_{TFS_F2F3} across stimulation types in the left and right ear conditions.

FFRs	Ear	Left-Anode	Left-Cathode	Right-Anode	Right-Cathode	Sham
FFR_{ENV}	Left	8.75 (2.45)	9.42 (2.44)	9.67 (2.70)	8.56 (2.81)	8.81 (2.71)
	Right	8.78 (2.02)	9.47 (3.49)	9.50 (3.25)	8.58 (1.69)	9.08 (2.33)
FFR_{TFS_H2}	Left	3.50 (2.28)	4.50 (1.82)	3.50 (1.82)	3.67 (2.06)	4.28 (1.60)
	Right	3.61 (2.30)	4.94 (1.59)	4.44 (1.95)	4.11 (2.00)	4.06 (1.92)
FFR_{TFS_F2F3}	Left	2.90 (0.36)	3.04 (0.27)	3.03 (0.44)	3.20 (0.31)	3.05 (0.53)
	Right	2.93 (0.48)	3.03 (0.48)	3.28 (0.47)	2.97 (0.34)	2.87 (0.48)
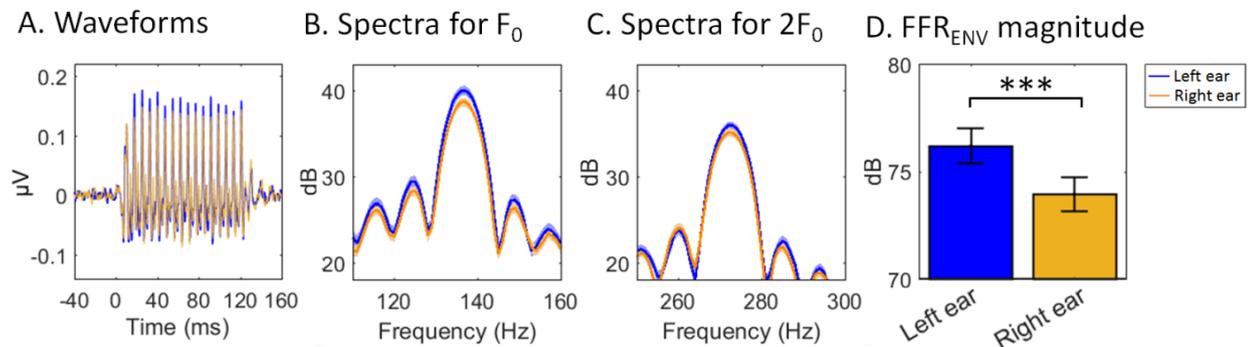
Figure 4.4. Comparison of baseline magnitude for FFR_ENV between the left and the right ear conditions. The comparison was conducted by collapsing the stimulation types following the ANOVA results which showed a significant main effect of Ear, but no significant main effect of Stimulation or [Stimulation $\times$ Ear] interaction for the baseline FFR_ENV magnitude. The left and the right ear conditions are indicated as blue and orange, respectively. (A) Waveforms of FFR_ENV averaged across stimulation types. (B)(C) FFT-power spectra averaged across stimulation types, obtained using the individual optimal neural delays for (B) FFR_ENV_F0 (showing 110–160 Hz peaking at F_0 of 136 Hz) and (C) FFR_ENV_2F0 (showing 250–300 Hz peaking at 2F_0 272 Hz) (shaded areas in the spectra cover the ranges of ±1 standard errors (SEs)). (D) FFR_ENV magnitude (summed magnitude of FFR_ENV_F0 and FFR_ENV_2F0). Significant greater FFR_ENV magnitude was found in the left than in the right ear condition (***p < 0.001, uncorrected). Error bars indicate the SEs.

4.3.2 After-effects on FFR$_{ENV}$

A. Left ear
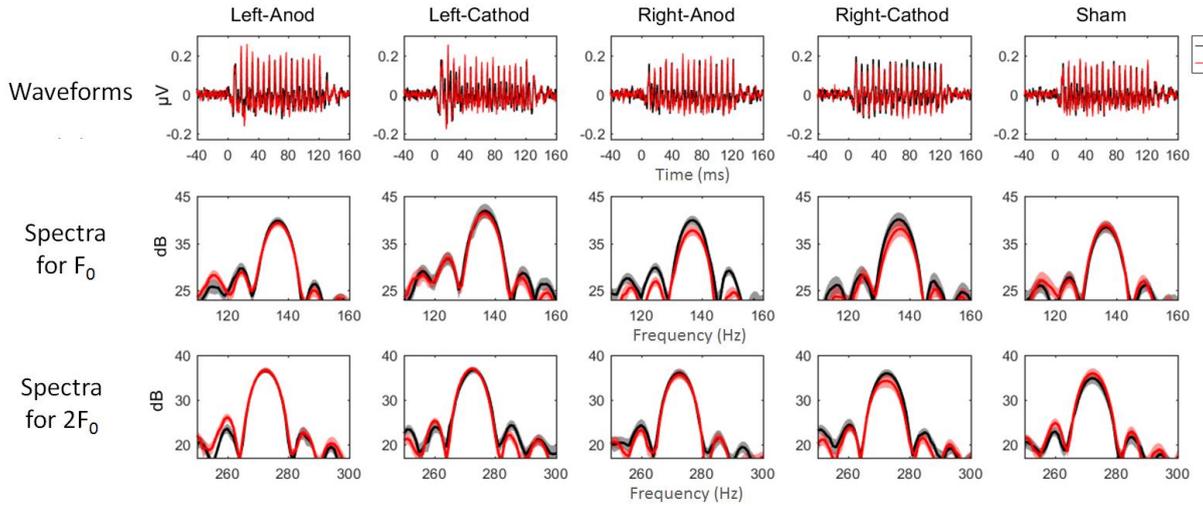
B. Right ear
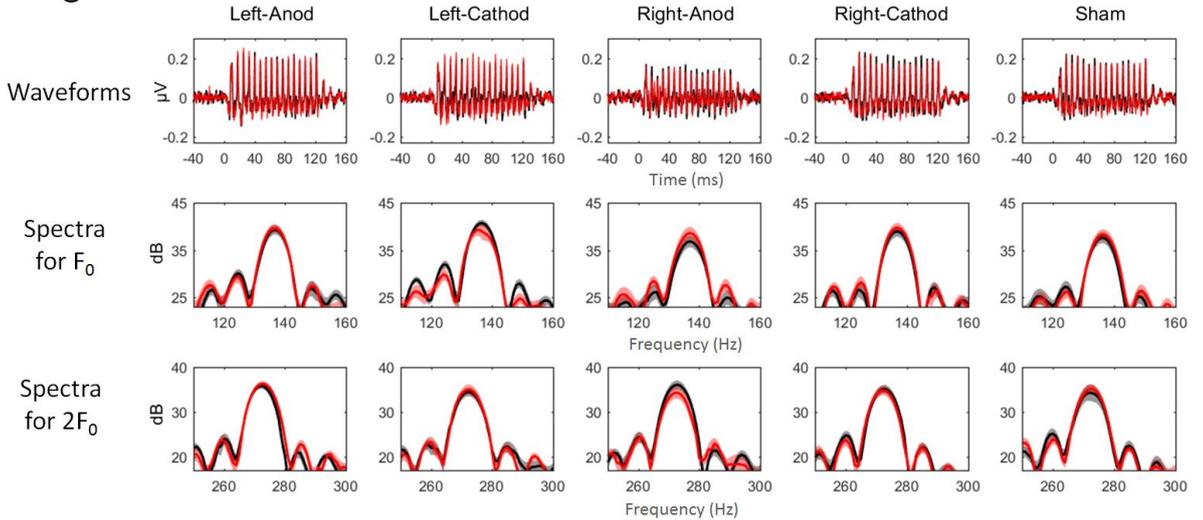
Figure 4.5. Waveforms and power spectra for FFR$_{ENV}$ averaged across participants. ***(A)****
and* ***(B)*** *show the waveforms and FFT-power spectra in the left and right ear condition,
respectively. Pre- and post-tDCS were indicated as black and red, respectively (shaded areas
in the spectra cover the ranges of ±1 SEs from the means). From left to right are different
stimulation types (Left-Anode, Left-Cathode, Right-Anode, Right-Cathode and Sham). Upper
panels: waveforms of FFR$_{ENV}$; mid and lower panels: power spectra obtained using the
individual optimal neural delays for FFR$_{ENV\_F0}$ (mid; showing 110–160 Hz peaking at F$_0$ of 136
Hz) and FFR$_{ENV\_2F0}$ (lower; showing 250–300 Hz peaking at 2F$_0$ 272 Hz).*

FFR$_{ENV}$ magnitude refers to the summed FFR$_{ENV\_F0}$ and FFR$_{ENV\_2F0}$ magnitudes (see
Statistical analyses). **Figure 4.5** shows the waveforms and FFT-power spectra for FFR$_{ENV}$
across participants. Linear mixed-effect regression showed a significant main effect of
Stimulation (F(4, 85) = 2.549, p = 0.045). No main effect of Ear (F(1, 85) = 0.784, p = 0.378) or

[Stimulation × Ear] interaction (F(4, 85) = 1.309, p = 0.273) was found. Post-hoc independent-sample t-tests were thus conducted between different stimulation types following the main effect of Stimulation (collapsing the left and right ear due to the lack of [Stimulation × Ear] interaction). After-effects of tDCS over the right AC were significantly lower than that of Sham (Right-Anode vs. Sham, t(34) = -2.569, p = 0.015 (uncorrected), Cohen's d = 0.856; Right-Cathode vs. Sham, t(34) = -2.219, p = 0.033 (uncorrected), Cohen's d = 0.740) (**Figure 4.6**).
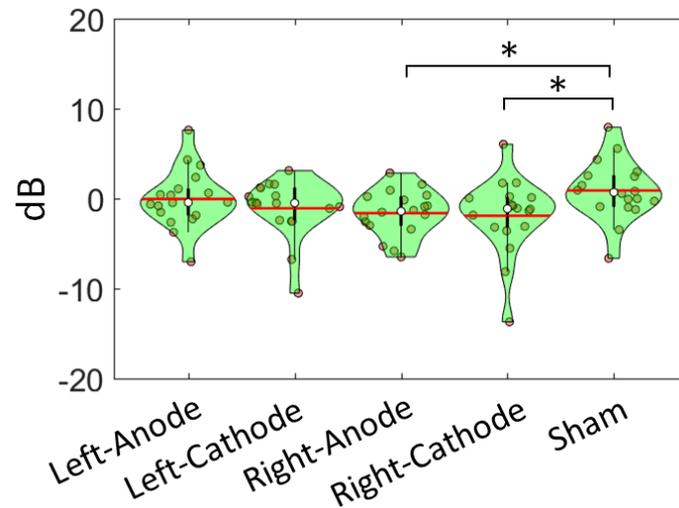
Figure 4.6. After-effects of tDCS on FFR$_{ENV}$ magnitudes comparing across stimulation types after collapsing the left and right ears. Collapsing the left and right ears was conducted following a significant main effect of Stimulation but no significant main effect of Ear or [Stimulation × Ear] interaction. Red circles indicate individual data for the corresponding stimulation types (Left-Anode, Left-Cathode, Right-Anode, Right-Cathode and Sham). Post-hoc paired comparisons showed significant differences between tDCS over the right AC (Right-Anode and Right-Cathode) and Sham (*p < 0.05, uncorrected). Error bars indicate the SEs.

Planned comparisons between different stimulation types were subsequently conducted for the left and right ear listening conditions to determine whether tDCS has effects along the contralateral or ipsilateral pathway. The critical α value for detecting significance was adjusted to 0.005 (there were 10 pairs of comparisons in each ear condition). The results are illustrated in **Figure 4.7** (upper panels). In the left ear condition, significant differences were found between tDCS over the right AC and Sham (Right-Anod vs. Sham, t(34) = -3.024, p < 0.005, Cohen's d = 1.001; Right-Cathod vs. Sham, t(34) = -3.420, p = 0.002, Cohen's d = 1.140). No significant effects were found for any other comparison (all p > 0.2). In the right ear condition, no significant effects were found for any pair of comparison (all p > 0.2). All p-values shown here are reported without correction.
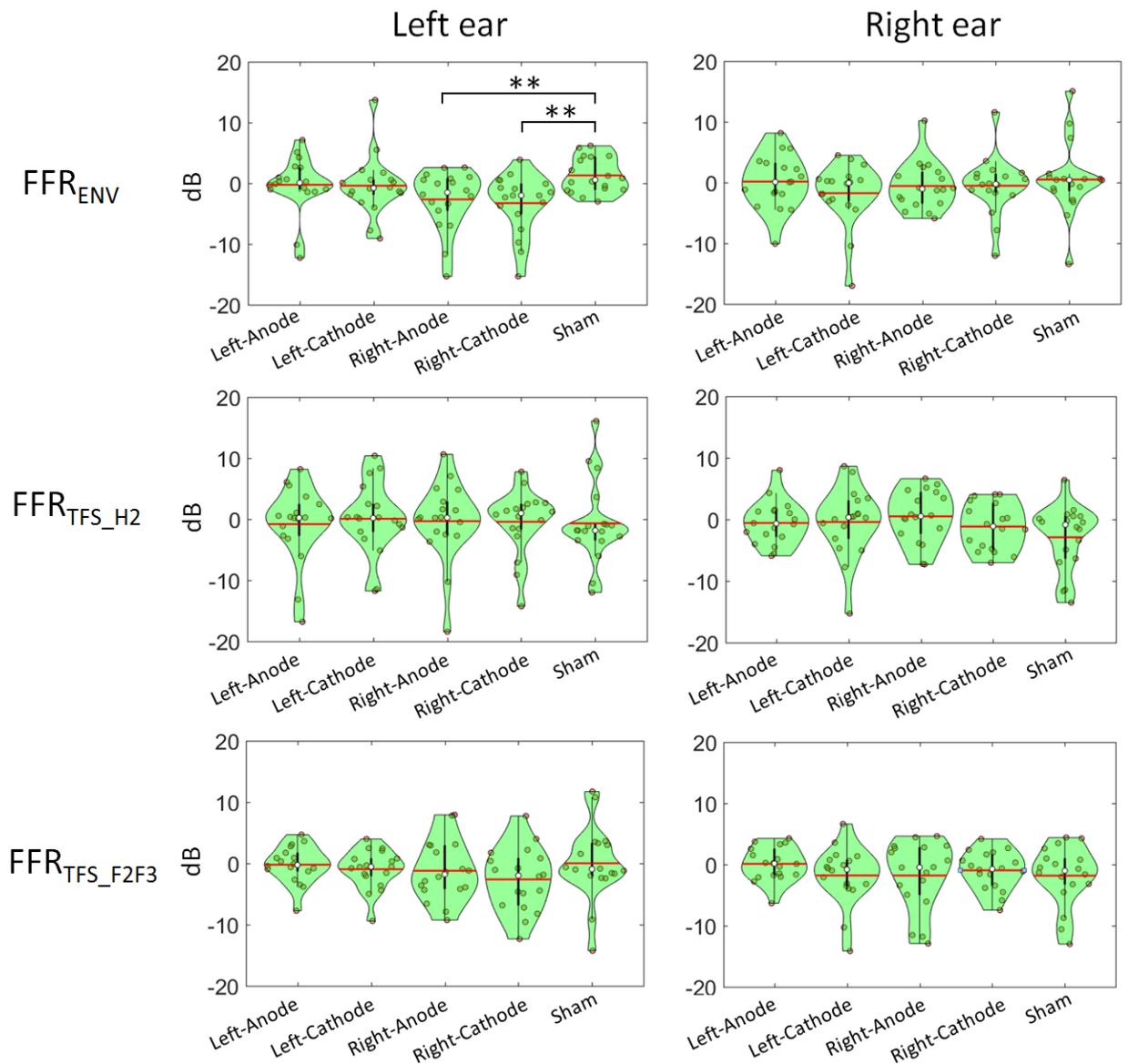
Figure 4.7. After-effects of tDCS on FFR magnitudes. *Upper, mid and lower panels indicate the after-effects on magnitudes of FFR$_{ENV}$, FFR$_{TFS\_H2}$ and FFR$_{TFS\_F2F3}$, respectively. Planned comparisons were conducted between different stimulation types in both the left and right ear conditions, with the critical α value set at 0.005 according to multiple comparisons. Significant differences were found between tDCS over the right auditory cortex (Right-Anod and Right-Cathod) and Sham in the left ear condition for FFR$_{ENV}$. (**p < 0.005, uncorrected; p < 0.05 after correction based on multiple comparisons) Red circles indicate individual data for the corresponding stimulation types. Error bars indicate the SEs.*

4.3.3 After-effects on FFR$_{TFS}$

Equivalent analyses to those conducted for the FFR$_{ENV}$ magnitude were conducted for the magnitudes of FFR$_{TFS\_H2}$ and FFR$_{TFS\_F2F3}$. The linear mixed-effect regressions did not show significant main effects of Stimulation (FFR$_{TFS\_H2}$: $F(4, 85) = 0.528$, $p = 0.715$; FFR$_{TFS\_F2F3}$: $F(4,$

85) = 0.613, p = 0.655) or Ear (FFR$_{TFS\_H2}$: F(1, 85) = 0.496, p = 0.467; FFR$_{TFS\_F2F3}$: F(1, 85) = 0.213, p = 0.646), or significant [Stimulation × Ear] interactions (FFR$_{TFS\_H2}$: F(4, 85) = 0.530, p = 0.714; FFR$_{TFS\_F2F3}$: F(4, 85) = 1.189, p = 0.322).

Planned comparisons did not find significant after-effects between different stimulation types in the left or right ear condition (FFR$_{TFS\_H2}$: all p > 0.6 in the left ear condition and all p > 0.1 in the right ear condition; FFR$_{TFS\_F2F3}$: all p > 0.09 in the left ear condition and all p > 0.1 in the right ear condition; see **Figure 4.7**, mid and lower panels). All p-values are reported without correction.

4.4 Summary of results and brief discussions

Fuller discussions of this study will be in Chapter 5 (*5.3*).

The current study established a causal relationship between the right auditory cortex and FFRs using combined tDCS and EEG measurements. The results showed that tDCS, both anodal and cathodal, over the right auditory cortex, generated significantly greater after-effects on FFR$_{ENV}$ magnitude compared to Sham. Specifically, such effects were present only in the left, but not the right ear listening condition, indicating that the changes in processing of speech periodicity information occur along the contralateral pathway (i.e., from the left ear to the right auditory cortex). This should be the first causal evidence to validate the contribution of the right auditory cortex to FFRs. After-effects found for FFR$_{ENV}$ indicate that tDCS had impacts on the responses at the subcortical and/or cortical levels above the auditory periphery. Lack of after-effects on FFR$_{TFS}$ may be because of the impacts of tDCS on the cortex had not reached the peripheral level; alternatively, tDCS could have impacts on the auditory periphery, but better signal quality or signal-to-noise ratio of FFR$_{TFS}$ may be required in order to detect such effects.

The present results can be linked to the relationship between FFRs and pitch perception, the neural processing of which is thought to be right-lateralized in the brain (Zatorre and Berlin, 2001; Patterson et al., 2002; Hyde et al., 2008; Mathys et al., 2010; Albouy et al., 2013). FFR strength is closely related to pitch and music perception (Carcagno and Plack, 2011; Musacchia et al., 2007; Wong et al., 2007; Strait et al., 2009; Bidelman et al., 2011) as well as perception of linguistic pitch (Krishnan et al., 2004, 2005, 2009). While strong links have long been established between FFRs and pitch perception (Musacchia et al., 2007; Wong et al., 2007; Strait et al., 2009; Bidelman et al., 2011; Carcagno and Plack, 2011), relationship between FFRs and the right auditory cortex was only established recently (Coffey et al., 2016, 2017a, 2017b). It is possible that FFRs reflect the neural pitch representation at the subcortical level before the process of pitch extraction that takes place in the auditory cortex (Bendor and Wang, 2005; Puschmann et al., 2010). Alternatively, FFRs may be directly involved in the pitch processing that occurs in the auditory cortex. The limitation in the present study is that, despite the finding of the causal relationship, the current approach is not able to validate the neural

sources of FFRs (i.e., whether the sources are from subcortical or auditory cortex). Scalp recording tools that better localize the FFR sources, like high-density EEG (Bidelman, 2018) or MEG (Coffey et al., 2016), should help in the future research to figure out where the neurophysiological changes come from.

Current effects of tDCS on FFRs were found compared between stimulation over the right auditory cortex and sham, despite the diffuse property of the tDCS. Direct currents can spread widely through various different regions that the electrodes are not placed at (Faria et al., 2011; Bai et al., 2014; Unal and Bikson, 2018). This may be the reason why no significant after-effects of tDCS were found between stimulations over the right and over the left auditory cortex. Future studies could use more focal stimulation like transcranial magnetic stimulation (TMS) to alleviate this problem.

The current results validate the previous findings that right auditory cortex makes significant contributions to speech-evoked FFRs (Coffey et al., 2016, 2017a, 2017b) by establishing a causal relationship between the two. To our knowledge, this is the first evidence for this causality and it could be essential due to the fundamental and clinical importance of FFRs. Thus, these findings should advance our understanding of how speech periodicity and pitch information are processed along the central auditory pathways in the human brain.

Chapter 5

General Discussion

Three experiments were reported in this thesis that investigated how speech-evoked phase-locked responses interact with the following factors: background noise, aging and hearing loss (Study 1 described in Chapter 2); changes in arousal and aging (Study 2 described in Chapter 3); and change in cortical neural excitability (Study 3 described in Chapter 4). These studies provide further evidence for how phase-locked responses can change as a function of factors across the lifespan and how these changes may affect our auditory and speech perception abilities during everyday life.

5.1 Effects of background noise, aging and hearing loss on speech-evoked phase-locked responses

Study 1 (Chapter 2) investigated how speech-evoked phase-locked responses in the auditory sensory systems are associated with SiN perception in participants across a wide age range (19–75 years old). Groups were divided by age and the older adult group had peripheral hearing at high frequencies that ranged from normal to mild/moderate hearing loss. Thus, this mimicked the ecological demographics in normal aging populations (Gopinath et al., 2009; Humes et al., 2010). Crucially, effects of background noise, aging and hearing loss on speech-evoked phase-locked responses were investigated. Furthermore, the study also demonstrated how the effects of age on neural activities (i.e., speech phase-locked responses) are associated with behavioural performances (i.e., speech reception threshold, or SRT).

5.1.1 Effects of background noise

Investigating the effects of background noise on neural processing of speech can provide information on how speech may be perceived at the behavioural level (Wong et al., 2009). For speech-evoked phase-locked responses, the influence of background noise has been extensively investigated in previous studies (e.g., Russo et al., 2004; Ding and Simon, 2013; Presacco et al., 2016, 2019). However, the following two questions have not been examined extensively: (1) How does background noise affect phase-locked responses at different stages (peripheral, subcortical and cortical levels) in the sensory auditory systems? (2) How do phase-locked responses change according to different types of noise, and do any such changes predict the SiN performances? Study 1 in this thesis attempted to answer these questions.

For question (1), Study 1 used a passive listening paradigm in which participants listened to a repeatedly presented syllable /i/ and was not required to perform a task. The syllable does not have semantic or syntactic information which effectively rules out the effects of higher-level (semantic/syntactic) neural processing. It can thus be argued that any phase-locked responses occurred at early-stage of auditory processing. The results showed that FFR_{ENV_F0} magnitudes, FFR_{PLV_F0} and theta-band PLV decreased when background noise was present compared to quiet background (see **Figure 2.6**, **Table 2.2**). This is consistent with previous studies that have shown that low-frequency phase-locked responses to speech Slow-ENV (Ding and Simon, 2013) and FFR magnitude at F_0 decreases in noisy environments (Russo et al., 2004; Presacco et al., 2016). In contrast, FFR_{TFS_H2} and FFR_{TFS_F2F3} magnitudes were not influenced by the presence of background noise (see **Figure 2.6**, **Table 2.2**). Thus, the findings imply that background noise affects phase-locked responses at the cortical and subcortical levels (where theta-band PLV and FFR_{ENV_F0}/FFR_{PLV_F0} originate; Ng et al., 2013; Aiken and Picton, 2008) but not at the peripheral level (where FFR_{TFS} originates; Aiken and Picton, 2008). A possible explanation may be that, compared to phase-locked responses to the acoustic signals per se (i.e. TFS), responses to acoustically modulated signals (i.e., envelope modulations at low-frequencies such as theta-band or at F_0) are more susceptible to background noise. However, Henry and Heinz (2012) found that, in chinchillas, reduced phase-locked responses to pure-tones at the TFS frequencies (0.6–2.5 kHz) when background noise was present. Importantly, this was found at the nerve fibres recorded through inserted invasive sensors placed at the auditory periphery (Henry and Heinz, 2012). This is not consistent with the present finding that FFR_{TFS} were not affected by noise. An alternative explanation for this result may be that since the neural phase-locked ability in the auditory systems is relatively weak at high acoustic frequencies as in TFS (Verschooten et al., 2019), precise quantification of this ability may require measurements with high temporal precision as in the study of Henry and Heinz (2012). The FFR_{TFS}, which was recorded over the scalp in the present thesis, may therefore not be able to accurately reflect the phase-locked responses to TFS that occur at the peripheral level. It is also noteworthy that possible contributions of electrical artefacts to FFR_{TFS} (*esp.* FFR_{TFS_F2F3} which represents phase-locked responses to TFS at very high frequencies) should not be ruled out (see more detailed explanations in *2.2.4*). Further efforts are needed to exclude such possibility in the future.

For question (2), Study 1 used two different types of noise (steady-state speech-shaped noise (SpN) and 16-talker babble noise (BbN)) for both behavioural (word recognition) tasks and the EEG recording. The behavioural performances showed that the speech reception threshold (SRT) was significantly lower (i.e., better SiN performances) in SpN than in BbN. This is consistent with previous research (Rosen et al., 2013). The reason for the differences in performances under different types of noise was argued to be due to the different properties of SpN and BbN. Whilst SpN serves as a steady-state energetic masker (EM), BbN is an envelope-modulated EM with more similar temporal acoustic properties to target speech as compared to SpN; furthermore, SpN has no linguistic structure whilst BbN also serves as an

informational masker (IM) that leads to phonetic interference[12] (Rosen et al., 2013). Therefore, the more similar acoustic properties to speech and the informational interference may lead BbN to have a greater masking effect than SpN. For the neural measurements, although the results showed that FFR_{ENV_F0} and FFR_{PLV_F0} were susceptible to background noise, they were not differentially susceptible to different types of noise as in the behavioural performances. In contrast to $FFR_{ENV\_F0}/FFR_{PLV\_F0}$, variations of theta-band phase-locking as a function of noise type (NoN or Quiet, SpN and BbN) followed the pattern of the behavioural results: theta-band PLV was significantly greater in NoN than in noise and significantly greater in SpN than in BbN (see **Figure 2.4**). This indicated that, unlike FFRs to F_0-ENV and TFS that mainly originate from the subcortical/peripheral level, cortical phase-locked responses to Slow-ENV (i.e., theta-band phase-locking) can predict the behavioural performances of SiN perception that varies a function of noise type. Furthermore, regression analyses also showed that, compared to FFRs, theta-band phase-locking can better predict the behavioural performances for the corresponding noise types (see **Table 2.4** and **2.5**).

Besides measuring cortical (theta-band PLV) and subcortical/peripheral (FFRs) activities, Study 1 also assessed the functional connectivity between subcortex and cortex. It used Partial Directed Coherence (PDC) which can characterize the directed connectivity between cortex and subcortex (Baccala and Sameshima, 2001; Schelter et al., 2005). The results showed that the afferent connectivity ($PDC_{Subcort \rightarrow Cort}$) was greater in quiet than in noise (but no difference was found between SpN and BbN), whilst efferent connectivity ($PDC_{Cort \rightarrow Subcort}$) was greater in BbN than in quiet and SpN. Thus this showed that afferent information flows were reduced by the interference of external noise, but they were not susceptible to different noise types; on the other hand, increased efferent flows in BbN may be explained by a compensatory top-down mechanism due to decreased afferent connectivity. These results thus showed that although cortico-subcortical connectivity was affected by noise, the pattern was not consistent with the behavioural results. Again, this further implies that low-frequency cortical phase-locking (theta-band PLV) may be a better neural signature to predict speech-in-noise performances as a function of noise type compared to subcortical signatures and cortico-subcortical connectivity.

5.1.2 Effects of aging and hearing loss

Study 1 showed that aging and hearing loss play different roles for speech-evoked phase-locked responses. At the cortical level, theta-band phase-locking increased with age (i.e., was greater in older than in young adults). Linear mixed-effect regressions using PTA (that reflects the degree of peripheral hearing loss) as a covariate did not show significant effects of PTA on theta-band phase-locking; however, the significant effects of Age were maintained as in the results where PTAs were not included as covariates. This thus implies that that the effects of Age were due to aging itself rather than age-related hearing loss. At the subcortical/peripheral level, no significant differences were found between young and older

---

[12] N.B., BbN should had no lexical interference here due to the high number of talkers (16) in the babble (Hoen et al., 2007).

adults for FFRs (FFR_{ENV_F0}, FFR_{PLV_F0}, FFR_{TFS_H2}, or FFR_{TFS_F2F3}). However, using PTA_{Low}, (0.25–1 kHz) as a covariate showed that FFR_{ENV_F0} magnitude decreased with age but increased with greater degree of hearing loss in the low frequency range (PTA_{Low}).

Theta-band PLV reflects the cortical tracking of speech envelopes (Slow-ENV) (Luo and Poeppel, 2007; Howard and Poeppel, 2010; Peelle et al., 2013). The present finding that theta-band PLV increased with age is thus in line with a previous MEG study that showed greater cortical tracking of speech envelopes in older than in young adults (Presacco et al., 2016a). It is also consistent with results showing that theta-band phase-locking in response to amplitude-modulated tones increases with age (Tlumak et al., 2015; Goossens et al., 2016) and findings that older adults have larger magnitudes of cortical auditory-evoked responses (Alain et al., 2014; Herrmann et al., 2013, 2016). As theta-band phase-locking has been shown to predict neural excitability in response to acoustic stimuli in the auditory cortex (Ng et al., 2013), the present result may thus be attributable to hyperexcitablity of the central auditory system during the aging process (Caspary et al., 2008).

At the subcortical level, the present finding in terms of FFR_{ENV_F0} magnitude (decreased with age but increased with hearing loss) is in line with the previous studies showing that FFRs magnitudes were smaller in older than young adults when both groups had normal audiometric hearing (Anderson et al., 2012; Presacco et al., 2016a, 2019). The present result is also in line with previous findings showing that encoding of envelopes at high-gamma frequencies corresponding to the F_0 range declines during aging when peripheral hearing is normal, but increases when there is hearing loss (Goossens et al., 2016, 2019). The effect of hearing loss may be due to the mechanism of reduced neural inhibition. There have been studies in animals (Kale and Heinz, 2010; Henry et al., 2014; Zhong et al., 2014) and humans (Anderson et al., 2013) which provide evidence to show that hearing loss can reduce neural inhibition, resulting in increased magnitudes of F_0-rate envelope modulations.

However, caution is required when interpreting the present results in terms of FFRs. FFR magnitudes are associated with participants' hearing sensitivity. Greater audibility could lead to greater FFR magnitudes (Ananthakrishnan et al., 2016). Therefore, the relation between PTA and FFR magnitudes probably depends on individual's degree of hearing loss. This could possibly explain why PTA_{Low} (PTA averaged across 0.25–1 kHz) but not PTA_{High} (PTA averaged across 2–4 kHz), significantly correlated with FFR_{ENV_F0} magnitude. This may be because older adults in the present study had a much smaller range of PTA_{Low} (roughly 0–30 dB HL) than PTA_{High} (varying from normal hearing to mild/moderate hearing loss; roughly 0–50 dB HL) (see **Figure 2.1**) and the wider range of PTA_{High} might lead to a greater mix of reduced audibility and neural inhibition on FFR_{ENV_F0}. Taken together, effects of aging on subcortical phase-locked processing may be due to a combined effect of age-related hearing loss and the aging process that is independent of peripheral hearing.

Effect of age and hearing loss were also assessed for cortico-subcortical connectivity (PDC) and no significant effect of either factor was found. Recent research showed that older

adults have poorer afferent functional connectivity between brainstem and auditory cortex during SiN perception (Bidelman et al., 2019). However, this was not replicated in Study 1. The possible reason is that the listening statuses differed across the studies: Bidelman et al. (2019) used a syllable identification task that reinforced participants' attention to the target syllable stimuli, whereas participants passively listened to the sound in the present study. Furthermore, Bidelman et al. (2019) applied source localization before assessment of connectivity (phase transfer entropy between FFRs and ERPs which are grand-averaged responses across trials and are in distinct frequency ranges), whereas Study 1 applied single-trial-based PDC analyses using continuous cortical and subcortical signals obtained from specific electrodes of interest. The continuous signals here ensured that cortical and subcortical signals were in the same frequency range (with the subcortical signal represented as its envelope profile) and allowed for the characterization of PDC at the presentation rate of the syllable stimuli (see 'Cortico-subcortical connectivity' in *2.2.4* for details). Future studies using high-density EEGs are needed to localize the sources of subcortical and cortical signals so that connectivity analyses can be better characterized.

In sum, the findings imply that the mechanisms by which aging and peripheral hearing loss affect the phase-locked neural activities differ at subcortical and cortical levels. Also, note that in Study 1, participants listened passively to a repeatedly-presented syllable that lacked higher-order semantic/syntactic information, indicating that the present effects of age and hearing loss occurred at the early stage of neural processing in the auditory systems. The findings thus show the complexity of the neurophysiological consequences of the aging process in the auditory domain. Accordingly, it should be emphasized that when considering SiN perception, and/or auditory perception in general, impacts of aging and hearing loss should be considered separately.

5.1.3 Contributions of age effects on phase-locked activities to SiN perception

Study 1 also addressed how the effects of age on phase-locked activities contribute to impaired SiN perception. Neural-behavioural relationships (relations between phase-locked responses and SiN performances) were studied using linear regressions using data that combined the young and older adult groups, as well as data from the young and older adult groups separately. The present study used an approach that differs from previous studies (Presacco et al., 2016a; Goossens et al., 2018). Presacco et al. (2016a) studied the relation between speech-evoked phase-locked responses and SiN perception in both young and older adults, but did not find any significant neural-behavioural correlations. A limitation of this study was that it used different types of background noise for the neural recording (single-talker background) and for the SiN perception tasks (four-talker babble). Therefore the neural-behavioural relationship may not be appropriately assessed. Goossens et al., (2018) investigated the relation between subcortical/cortical auditory steady-state responses (ASSRs) and SiN perception in both normal-hearing and hearing impaired adults across ages and included age itself as an additional predictor when modelling the neural-behavioural

relationship. This did not fulfil the aim of testing how age-related neural factors contribute to SiN perception as including age itself in the model would mathematically partial out the effect of age. In the present study, regressions were conducted with neural signatures and SiN perception under the same types of background noise. That is, SiN perception in SpN and 16-talker BbN were predicted by neural data obtained in SpN and 16-talker BbN, respectively; neural data obtained in quiet were additionally used as predictors, as FFRs in quiet could be associated with SiN perception (Anderson et al., 2011). Also, age itself was not included as a predictor.

Theta-band PLV, which was shown to increase with age, was associated with increased SiN perception. Thus this argues against the view that increased tracking of speech envelopes (Slow-ENV) or auditory cortical excitability reflects a diminished excitation-inhibition balance in the neural network which results in impaired SiN perception (Presacco et al., 2016a). An alternative explanation may be that, cortical hyperexcitability *per se* does not impair SiN perception, but influences SiN perception in an indirect way: increased cortical excitability is significantly related to decreased attention and inhibitory control in older adults which could, in turn, impair SiN perception (Presacco et al., 2016b). At the subcortical level, previous studies have argued that declines in subcortical speech-evoked FFRs are an important determinant of SiN perception difficulty in older adults (Anderson et al., 2011, 2012; Presacco et al., 2016a). Consistent with this argument, the current results showed that the FFR_{ENV_F0} declined with age and that decreased FFR_{ENV_F0} obtained in quiet was associated with decreased SiN perception. Taken together, the current results thus showed that impacts of age effects on SiN perception are different at cortical (increased θ-PLV with age associated with increased SiN perception) and subcortical (decreased FFR_{ENV_F0} with age associated with decreased SiN perception) levels.

The results also showed potential different mechanisms of neural-behavioural relationships across ages, in which SiN perception is better predicted by cortical and subcortical signatures obtained under noisy backgrounds in the older compared to the young group. This could be because higher-level cognitive functions, such as working memory and selective-attention capacity (not tested in the present study) that decline with age. These factors could play greater roles in modelling the individual variability of SiN perception ability in young adults. The results could also arise because individual variability of SiN perception was relatively small in young adults (especially SRT_{BbN}, which had significantly lower variability in the young, than the older, group according to Levene's test, see *3.1*). This would make SiN perception more difficult to predict using linear regression.

A further issue related to age, phase-locked responses and SiN perception needs further discussion. The current results found an interaction between noise types and age for the SiN performances (SRT), where older adults had more difficulty in speech perception in babble noise (BbN) than young adults, but not in speech-shaped noise (SpN). However, this interaction was not found for phase-locked responses, either theta-band PLV or FFRs. The current results thus did not provide evidence to explain this interaction at the sensory neural level. Future work should be conducted to investigate this phenomenon. Specifically, while both SpN and BbN can

cause energetic masking, BbN additionally causes informational masking such as phonetic interferences that are higher-order interferences beyond sensory processing (Rosen et al., 2013). Informational masking may interfere with participants' cognitive processing, especially attention during SiN perception (Carlile and Corkhill, 2015), which declines following aging (Humes et al., 2006; Anderson et al., 2013) and hearing loss (Lin et al., 2013) (see further discussions in *5.4 Limitations and future work*).

5.1.4 Study 1 summary

Study 1 tested the effects of age and hearing loss on speech-evoked phase-locked responses (theta-band PLV to Slow-ENV and FFRs to F_0-ENV and TFS) and tested the relationship between the phase-locked responses and SiN perception in young and older adult groups. The results showed that older adults have smaller FFR_{ENV_F0} magnitude after low-frequency PTA was controlled for and greater theta-band PLV. Furthermore, greater low-frequency PTA (greater degree of hearing loss) was associated with increased FFR_{ENV_F0} magnitude. Age-related effects (increased theta-band PLV and decreased FFR_{ENV_F0}) were associated with SiN perception, where SiN performances increased with greater theta-band PLV and decreased with decreased FFR_{ENV_F0}. These results illustrate distinct mechanisms by which age and hearing loss affect phase-locked responses at the subcortical and cortical levels and the SiN perception. Future work needs to be conducted that includes cognitive assessments to study how higher-level cognitive functions influence such mechanisms and contribute to SiN perception together with sensory processing during the aging processes.

5.2 Effects of arousal on speech-evoked phase-locked responses and the age-dependent modulation

Study 2 (Chapter 3) investigated the effects of arousal on speech-evoked phase-locked responses and how such effects are modulated by age. It combined two data sets that used the same paradigm (passive listening to a repeatedly presented syllable) when participants' EEG signals were recorded over the scalp and when they were either awake or fell asleep. The age of participants ranged from 19 to 75 years old. Besides assessing the effect of arousal, the study particularly focused on how age modulated the neuro-regulation of sleep spindles on the phase-locked responses.

5.2.1 Effects of arousal

The results showed that both theta-band PLV and FFR_{ENV_F0} were statistically greater in high (wakefulness) than in low arousal (Stage 2 nREM sleep) states. Furthermore, potential effects of neural adaptation that may cause these results were ruled out.

Neural deactivation during sleep is a way to protect the brain from arousing stimuli (Czisch et al., 2002, 2004). The cortical changes (i.e., significantly decreased theta-band PLV) with arousal state could be explained by findings of sensory deafferentation as a result of reduced thalamocortical connectivity during the transition from wakefulness to nREM sleep (Spoormaker et al., 2010, 2011; Picchioni et al., 2014). This is supposed to be achieved by the thalamic 'gating mechanism' by which most sensory information must pass through the thalamus before they reach the sensory cortices (Steriade et al., 1993; McCormick and Bal, 1994; 1997). This gating process can selectively control the information flow through the thalamus. The thalamic neurons fire in bursts when the brain's arousal level decreases, which creates membrane hyperpolarization that results in increases in the potential threshold for postsynaptic neurons to fire (Sherman, 2001). As a result, the incoming postsynaptic potentials are less likely to pass the threshold, leading to a partial blockage of sensory information ascending from thalamus to cortex (Coenen, 2010). This could serve the mechanism underlying the effects of arousal on neural responses to auditory stimuli. The current result is, however, not in line with Makov et al., (2017) which showed that phase-locked responses to low-level linguistic (i.e., syllable) rhythms were not affected by the level of arousal. This could be because the statistical results of the Makov et al. *(2017) study were based on a relatively small sample size (17 participants when comparing wakefulness vs. nREM sleep). In the present Study 2, a much larger sample size was used (91 participants including 38 younger and 53 older adults) which would thus surpass the study by Makov et al., (2017) in its ability to find any effect of arousal on cortical phase-locked responses.

The effect of arousal was also found for the speech-evoked FFR_{ENV_F0}. FFR_{ENV_F0} magnitude significantly decreased during nREM sleep compared to wakefulness. Previous studies have reported that magnitudes of auditory steady-state responses (ASSRs), which are phase-locked responses to amplitude-modulated (AM) tones, decreased significantly in sleep compared to wakefulness; however, this occurred only at AM rates <70 Hz (Cohen et al., 1991; Lins et al., 1995; Picton et al., 2003). In the present study, the modulated frequency (F_0s, corresponding to AM rates) for FFR_{ENV_F0} were >100 Hz. The ASSRs in previous reports were responses to AM tones which were carried by a pure tone with a fixed frequency. In contrast, the FFR_{ENV_F0} were responses to periodicity carried by complex harmonic structures with formants that spanned across a wide range of frequencies. It is hence possible that the FFR_{ENV_F0} decreased according to arousal because of the stimuli used here were speech signals with complex harmonic carriers rather than amplitude-modulated tones carried by simple pure tone in previous studies. This is the first study to show that speech-evoked FFR magnitude is reduced following decreases in the level of arousal. It is important for future studies to further clarify whether this effect is related to complexity of the acoustic stimuli.

The paradigm used in Study 2 was the same as in Study 1: participants listened passively to a repeatedly presented syllable that lacked higher-level linguistic features (semantic or syntactic information). Therefore, the current results reflect the effects of arousal on early-stage brain processing of speech. It was shown in Study 1 that both theta-band PLV and FFR_{ENV_F0} were associated with speech-in-noise perception indicating their important roles for speech

processing. The current findings thus have important theoretical implications about the impact of arousal on neural processing. Such impact on speech perception starts from the early sensory stages rather than from the higher-order linguistic processing stages (i.e., stages of perceiving words, phrases or sentences argued by Makov et al., (2017)).

A further issue concerning the effect of arousal lies on whether the effect started at the cortical or subcortical level. Low-frequency phase-locked responses (theta-band PLV here) have their neural origin in the auditory cortex (Luo and Poeppel, 2007; Howard and Poeppel, 2010; Peelle et al., 2013) and FFR_{ENV_F0} originate primarily in the auditory brainstem (Chandrasekaran and Kraus, 2010; Bidelman, 2018). However, recent studies have shown that FFR also has origins in the auditory cortex (Coffey et al., 2016, 2017). It is thus not certain in which locations the current observed effects started. It is possible that arousal takes effect at both the cortical and subcortical levels, or at the cortical level only. Previous fMRI studies gave mixed findings on this issue. Portas et al. (2000), Wilf et al. (2016) and Czisch et al. (2002, 2004) studied the effect of arousal on speech and auditory processing at both subcortical and cortical levels. Portas et al. (2000) showed reduced BOLD activity during nREM sleep than wakefulness at the thalamus, but studies by Wilf et al. (2016) and Czisch et al. (2002, 2004) did not find this effect. Subcortical BOLD responses can be highly contaminated by systemic physiological signals (Brooks et al., 2013), which may be the reason that makes subcortical BOLD insensitive to auditory processing changes due to varying arousal levels. Also, it is unclear how speech-evoked phase-locked responses (FFRs) are related to subcortical BOLD, which needs further investigation.

5.2.2 Age-dependent modulation

Another goal that Study 2 pursued was to examine how aging may take a role in the neuro-regulation of arousal. Previous studies have focused on how the speech-evoked phase-locked responses are modulated by aging (Anderson et al., 2012; Tlumak et al., 2015; Goossens et al., 2016, 2018; Presacco et al., 2016) and arousal (Makov et al., 2017). However, no studies have tested how aging and arousal might interact. It is reasonable to suspect that age would modulate the process of sensory gating, because (1) age can affect speech-evoked phase-locked responses (Anderson et al., 2012; Presacco et al., 2016a, 2019; also as revealed by Study 1); and (2) age can modulate properties of the sleep spindles (Martin et al., 2013; Mander et al., 2017) such as magnitude, density and duration. Sleep spindles play an important role in the sensory gating process (Dang-Vu et al., 2011; Schabus et al., 2012). Furthermore, sleep spindle properties are also related to hippocampal activity involved memory consolidation (Schabus et al., 2007; Andrade et al., 2011; Bergmann et al., 2012) that is affected by aging (Mander et al., 2017).

Effects of age on speech-evoked responses and spindle properties were found in Study 2. First, theta-band PLV was significantly greater in older than in young adults, which is in line with the results in Study 1. This is in agreement with previous findings showing age-related increases in theta-band ASSRs (Tlumak et al., 2015; Goossens et al., 2016) and cortical

tracking of speech envelopes (Presacco et al., 2016a, 2019) which can be explained by hyperexcitability of central auditory systems following aging (Caspary et al., 2008). Second, duration of sleep spindles was significantly shorter in older than in young adults. This is consistent with previous studies showing that spindle properties change following aging where magnitude, density and duration of spindles decrease with age (Martin et al., 2013). However, no significant differences in magnitude or density of spindles were found between the older and the young group in Study 2. This might be because the measurements were made over different time scales between this study and others. While sleep spindle properties were typically collected across the entire nights of sleep (e.g., Martin et al., 2013), experiments in Study 2 here lasted for much shorter time frames (~2 hours for Experiment 1 and ~70 minutes for Experiment 2) Also, the spindles were measured only for the parts the experiments when participants were exposed to sounds in the quiet background (which means the actual time frames for the measurements were even much shorter than 2 hours or 70 minutes).

Interplay of age on the arousal effect was found in Study 2. Specifically, although no significant interaction between age and the level of arousal was found, a significant three-way interaction between age, the level of arousal and spindle density was found for the theta-band PLV. Post-hoc analyses showed a significant interaction between the level of arousal and spindle density in the young, but not in the older, adult group, i.e., the effect of arousal (high arousal/wakefulness vs. low arousal/sleep) on theta-band PLV showed significant positive correlation with the spindle density in the young, but not older, adults. This correlation echoes previous research which found reduction of neural responses to external acoustic inputs when sleep spindles occur during nREM sleep (Dang-Vu et al., 2011; Schabus et al., 2012). Greater spindle density could reflect a greater degree of sensory deafferentiation in sleep (Spoormaker et al., 2010, 2011; Picchioni et al., 2014) because spindle density reflects tolerance to auditory disturbance (Dang-Vu et al., 2010) and sleep stability (Kim et al., 2012) during nREM sleep. The current results stress that age plays a significant role in the regulation of this process. The lack of the interaction between the level of arousal and age may be because the level of arousal itself would only be appropriate to characterize whether participants were in the state of wakefulness or nREM sleep. However, this cannot accurately estimate the degree or stability of sleep because sleep spindle density, which can better estimate the degree of arousal during nREM sleep, differed across participants. Therefore, including spindle density in the statistical model can more precisely assess the interplay between age and arousal.

The absence of correlation between the effect of arousal and sleep spindle density in older adults indicates lower degree of neuro-regulation of auditory processing as people age. It is noteworthy that the discrepancy between the older and young adults was attributable to age itself, rather than age-related variables such as PTA (indicating the degree of hearing loss) or shorter duration of sleep spindle. One possible explanation is that the generation of sleep spindles does not only involve changes in thalamocortical circuits that influence the processing of auditory inputs, but also involves changes in hippocampal activity that is influential in memory consolidation (Schabus et al., 2007; Andrade et al., 2011; Bergmann et al., 2012). Aging could change hippocampal activity in terms of atrophy of the hippocampus and such changes could

be associated with functional changes in sleep spindles (Mander et al., 2017). If this is the case, non-auditory processes (e.g., changes in hippocampal functions) that are related to changes in sleep spindles may play a role in regulation of auditory processing during aging. Such complexity in older adults may then be a confounder that deteriorates the neuro-regulation of spindles on cortical phase-locked responses. Future research may be conducted to test this hypothesis, for example, using combined fMRI technique to look into hippocampal functions in relation to sensory gating in older adults.

5.2.3 Practical issues regarding the effect of arousal

Besides the theoretical values of the current finding, a practical value would be for experiment designs in research of phase-locked responses, especially FFRs. It has been argued that brainstem responses, like click-evoked ABRs and speech-evoked FFRs, are not susceptible to sleepiness (Skoe and Kraus, 2010). The current results, however, indicate that scalp-recorded FFRs using EEG can be affected by the level of arousal. When using FFRs as a clinical tool in the future, it could be important to attend to the effect of arousal, especially in populations like neonates, whose FFRs are not always easy to obtain during wakefulness (e.g., Ribas-Prats et al., 2019).

Furthermore, the relationship between speech-evoked FFRs and behavioural functions were investigated (e.g., speech perception in reverberation, Fujihira and Shiraishi, 2015; pitch perception, Krishnan et al., 2005; Bidelman et al., 2011). In these studies, participants were allowed, or even encouraged, to fall asleep during EEG recording without controlling for levels of arousal. The neural-behavioural relationship was studied using between-subject analyses (regression analyses as conducted in Study 1) were conducted. The approach of allowing participants to fall sleep may need to be revised by including procedures that control for each individual's state of arousal or by processing the data as here. This is because different levels of arousal across participants are potentially between-subject confounds that could possibly bias the observed neural-behavioural relationship. Disregarding the influence of arousal might affect the validity of neural phase-locked activities as neuro-markers for behavioural performances. On the other hand, the sizes of the arousal effects observed in the present thesis was only in the medium range which is relatively small compared to the between-subject variability of the neural and behavioural data, Therefore, It is important to carefully investigate in future studies how much additional variance of the arousal effects would explain when analysing the neural-behavioural relationship.

Another issue that is related to the neural-behavioural relationship is that Study 2 only analysed the neural responses to speech in a quiet background which may raise additional concerns for the influence of arousal on relationship between phase-locked responses and SiN perception. The reason that the data in noisy backgrounds were not analysed was because different types of background noise and signal-to-noise ratios (SNRs) were used in the two data sets (Experiment 1 and Experiment 2, see details in *3.2.2*). The only background type shared across all participants was the quiet background that is appropriate to focus on so that within-

subject analyses can be conducted. Therefore, it may be worth in the future to study whether the observed effects can be also detected for phase-locked responses to speech in noisy backgrounds and how the effects occur in different types of noise and at different SNRs. This could better estimate how arousal might cause biases to the neural-behavioural relationship for SiN perception.

5.2.4 Study 2 summary

The current findings provide evidence that arousal can affect neural phase-locked responses to speech. Furthermore, the arousal effect on the cortical responses (theta-band PLV) were modulated by age through different regulatory roles of sleep spindles in the older and young adult groups. Such findings thus emphasized the importance of arousal and the regulatory role of spindles for early-stage brain processing of speech. The age-dependency found here lays the ground for studying how cognitive states, such as arousal, anaesthesia and attention, affects auditory neural activities across the lifespan. The mechanisms behind the role of age merit further investigations in the future using multimodal imaging techniques such as combined EEG and fMRI that can detect auditory and non-auditory processes that could both interact with sleep spindle properties during aging.

5.3 Effects of cortical excitability on FFRs

Study 3 (Chapter 4) was designed to address the question whether FFRs have a close relationship with the auditory cortex. It used a combined tDCS and EEG approach to test for a causal contribution of auditory cortex to speech-evoked FFRs in healthy right-handed participants. Left and right auditory cortices were neuro-stimulated using tDCS that changed the cortical excitability and the after-effects of tDCS on FFRs were examined during monaural listening to a repeatedly-presented speech syllable. This was the first study to provide evidence to validate the contribution of the auditory cortex to FFRs.

5.3.1 Laterality for FFR$_{ENV}$

Before tDCS was applied, analyses for the FFRs were conducted to assess the laterality at the baseline. Ear laterality for FFR$_{ENV}$ and FFR$_{TFS\_F2F3}$ magnitudes was respectively found. Here, the discussion only focuses on FFR$_{ENV}$ alone because (1) there was a significant main effect of Stimulation for the baseline FFR$_{TFS\_F2F3}$ magnitude meaning that the baseline was not well matched across stimulation types; (2) it lacked a significant after-effect of tDCS on FFR$_{TFS\_F2F3}$ magnitude; and (3) because of the high frequency range at F2 and F3 of the stimuli (syllable /i/), the actual FFR$_{TFS\_F2F3}$ magnitudes may not be easily obtained in an accurate manner (see relevant discussions in *5.1.1*).

Baseline FFR$_{ENV}$ had significantly greater magnitude in the left than in the right ear condition. This supports the laterality of speech periodicity encoding along the contralateral

auditory pathway from the left ear to the right auditory cortex. This result echoes a previous finding that showed the right-hemispheric lateralization of auditory steady-state response (ASSR which is envelope-following responses as FFR_{ENV}, Dimitrijevic et al., 2004) at 40 Hz (Ross et al., 2005; Luke et al., 2017) and 80 Hz (Vanvooren et al., 2014). Whilst the neural origin of 40 Hz ASSR is at the cortical level (Herdman et al., 2002; Ross et al., 2002, 2005), 80 Hz ASSR has predominant activity that occurs at the subcortical level (Herdman et al., 2002) which is also the location where speech-evoked FFR_{ENV} mainly originates from (Chandrasekaran and Kraus, 2010; Bidelman, 2015, 2018). As discussed above, FFRs have been argued to have additional neural sources in the auditory cortex (Coffey et al., 2016, 2017a). Therefore, it is unclear whether the observed laterality of FFR_{ENV} occurs at the cortical or subcortical level or, even more equivocally, whether auditory cortex contributes to this laterality. As such, the current combined tDCS and EEG approach showed how altering neural excitability of auditory cortex in the left or right hemisphere can lead to changes in FFRs which therefore provides confirmatory evidence of causal cortical contributions.

5.3.2 Causal role of the right auditory cortex for FFR_{ENV}

After-effects of tDCS were found for FFR_{ENV} but not for FFR_{TFS}. This indicates that tDCS may affect the phase-locked responses at the cortical/subcortical level above the auditory periphery. Alternatively, tDCS could have impacts on the auditory periphery, but better signal-to-noise ratio of FFR_{TFS} may be required to detect the effects. The findings argue for a causal role of the right auditory cortex in processing speech periodicity along the contralateral pathway in the central auditory systems.

From a theoretical stance, the present study thus advances our understanding of the relationship between FFRs and pitch processing in the right auditory cortex. Previous studies have shown that FFRs are closely related to pitch perception. It has been found that FFR strength is enhanced by both short-term perceptual training of pitch discrimination (Carcagno and Plack, 2011) as well as long-term musical experience (Musacchia et al., 2007; Wong et al., 2007; Strait et al., 2009; Bidelman et al., 2011). Furthermore, FFRs are often used as indices of neural fidelity of linguistic pitch and the fidelity has been reported to be greater in tonal than in non-tonal language speakers (Krishnan et al., 2004, 2005, 2009). Despite this, however, rather than reflecting the result of pitch extraction, FFRs are suggested to reflect subcortical responses to monaural temporal information (e.g., periodicity cues) that are important for extracting pitch of complex sounds (i.e., 'pitch-bearing' information; Gockel et al., 2011). On the other hand, the process of pitch extraction itself takes place in the auditory cortex (Penagos et al., 2004; Bendor and Wang, 2005; Puschmann et al., 2010) with a right hemispheric specialization (Zatorre and Berlin, 2001; Patterson et al., 2002; Hyde et al., 2008; Mathys et al., 2010; Albouy et al., 2013). In this respect, the current after-effects of tDCS may reflect a top-down corticofugal modulation process in which right auditory cortex affects the processing of pitch-bearing information that occurs at the subcortical level. Alternatively, as FFRs have cortical sources dominated in the right hemisphere (Coffey et al., 2016; 2017b), tDCS may directly affect the FFR magnitude at

the cortical level. It is noteworthy that the current finding could not disentangle whether the effects emerge at the subcortical or cortical level, or both.

From a clinical perspective, these findings are important because FFRs are biomarkers for various speech, language processing and cognitive disorders. These include hearing deficits such as cochlear synaptopathy (Encina-Llamas et al., 2019) and auditory processing disorders (e.g., Schochat et al., 2017), learning disorders and cognitive impairments in children, such as learning difficulties in literacy (Cunningham et al., 2001; Banai et al., 2007; White-Schwoch et al., 2015), dyslexia (Hornickel et al., 2013) and autism (Russo et al., 2008). These disorders are often indicated by abnormal FFRs in populations over a wide age range. The finding in Study 3 may thus lay the ground for further research to investigate and/or develop interventions like neuro-stimulation or combined neuro-stimulation and training based on the short-term plastic features of FFRs (Song et al., 2008; Carcagno and Plack, 2011; Skoe et al., 2014). Also, future studies would need to use more focal stimulation, such as high-definition tDCS (HD-tDCS) and transcranial magnetic stimulation (TMS) compared to the current tDCS method which could generate diffuse currents flowing across wide regions of the brain (Faria et al., 2011; Bai et al., 2014; Unal and Bikson, 2018).

5.3.3 Neurophysiological consequences of tDCS

An intriguing finding of the Study 3 is that different directions of polarity in the neuro-stimulation, i.e., anodal and cathodal tDCS, resulted in the same direction of changes, both causing decreases in FFR_{ENV} magnitude. Conventionally, anodal and cathodal stimulations reflect depolarization and hyperpolarization of neurons, respectively, which should lead to opposite directions of after-effects (Jacobson et al., 2012). However, it is not unusual that tDCS has polarity-independent effects due to the underlying complexity of its neurophysiological consequences. For example, several studies have shown that anodal and cathodal tDCS have the same effects on excitability of motor cortex (Antal et al., 2007), motor learning (de Xivry et al., 2011), cerebellar functions for working memory (Ferrucci et al., 2008) and visuomotor learning (Shah et al., 2013). The first possibility would be the non-linear effects depend on the current intensity. It has been shown that cathodal tDCS with an electrode size of 35 cm^2 can lead to inhibition in the motor cortex at 1 mA but excitation at 2 mA (Batsikadze et al., 2013). The present study used a current intensity at 1 mA but with smaller electrode size (25 cm^2; hence greater current density). It could be that this current density led to similar non-linear effects in the auditory cortex as in the motor cortex. Second, it is possible that similar changes in concentrations of relevant neurotransmitters are caused by anodal and cathodal tDCS. It was found that with 1 mA currents, anodal tDCS caused decreases in GABA (gamma-Aminobutyric acid, an chief inhibitory neurotransmitter in the mammalian nervous systems) concentration that lead to cortical excitation; cathodal tDCS also causes decreases in GABA, but with greater concurrent decreases in glutamate that lead to cortical inhibition (Stagg et al., 2009). It is possible that GABA concentrations, which decrease following both anodal and cathodal tDCS, play an important role for changes in FFR_{ENV} magnitude.

Also, another issue relevant to the properties of tDCS also arose. In the present study, stronger evidence would be provided for the specific contributions of right auditory cortex to FFRs if significant differences in after-effects were further found between tDCS over the right and the left auditory cortex. However, such difference were not observed. A possible explanation is that tDCS not only alters excitability of regions in which electrodes are located but can yield widespread changes across the brain (for review: Filmer et al., 2014). This could be due to the diffuse nature of the tDCS where currents do not only flow between electrodes, but also spread widely through various other regions (Faria et al., 2011; Bai et al., 2014; Unal and Bikson, 2018). tDCS also changes functional connectivity (Sehm et al., 2012; Kunze et al., 2016) by which interactions of auditory cortices between the two hemispheres may be further activated. Therefore, tDCS over the left auditory cortex could also cause some changes in the right side that yield similar (but smaller) after-effects as direct stimulation over the right auditory cortex.

5.3.4 Study 3 summary

The current results validate that the right auditory cortex makes significant contributions to speech-evoked FFRs (Coffey et al., 2016, 2017a, 2017b) by establishing a causal relationship between the two. This provides the first evidence for such a causal link and it could be essentially due to the theoretical and clinical importance of FFRs. Thus, these findings should advance our understanding of how the brain processes speech periodicity along the central auditory pathways.

Future research is needed to further clarify the locations where this causality may emerge, i.e., to disentangle whether the observed effects are realized through top-down corticofugal modulations on the subcortical level, or through modulations of neural excitability directly in the auditory cortex (see *5.4* and *5.5* for more discussions). Other advanced neuro-stimulation method like TMS which provide more focal stimulation and MEG and/or high-density EEG which may provide source localisation could be used in the future. Also, it is worthwhile to further study how this causality is associated with changes in concentrations of neurotransmitters triggered by neuro-stimulation. A possible approach could be, besides using combined EEG and neuro-stimulation, involving imaging tools like magnetic resonance spectroscopy (MRS) that can detect changes in transmitter concentration in particular brain regions (Stagg et al., 2009). Future endeavours should provide us with better understanding of the mechanisms that underlie the contributions of auditory cortex to FFRs.

5.4 Overarching summary on the effects of aging

This thesis has shown how speech acoustic properties are neurally processed in settings of our everyday listening circumstances. Specifically, the overarching findings have illustrated the multifaceted way in which the aging process influences speech perception.

Aging is a process that is associated with various neural degenerations and abnormalities. The present thesis specifically looked into the age-related factors that can impact speech perception and the neural phase-locking to speech. These factors include age-related peripheral hearing loss, age-related declines in the central auditory systems and the age-dependent neuro-regulations during transitions between different arousal states. These factors are likely to play roles in speech perception through different underlying mechanisms. As summarized above, Study 1 showed that age-related hearing loss and aging (which was disentangled from hearing loss) play distinct roles for subcortical and cortical phase-locking to speech. Study 2 showed that aging modulates the neuro-regulatory role of sleep spindles on cortical phase-locking to speech during different arousal states. Study 3 revealed a causal relationship between auditory cortex and subcortical/cortical phase-locking to speech (i.e., FFRs), Note that Study 3 did not look into the effect of age, however, the role of aging/age-related factors merits future investigations as it can reveal how the cortex-FFR causality is modulated by age that may further impact speech perception. This may reveal an important mechanism by which speech is perceived along the central auditory system across different ages.

The current results thus reflect the multifaceted manners in which the aging process influences speech and auditory perception in the brain. This echoes the accumulated evidence showing that aging is a process that changes brain functions with multiple distinct mechanisms (e.g., Smith et al., 2020). The work of this thesis thus advances our understanding of this process and lays the ground for future studies. While it is noteworthy that this thesis has focused on sensory processing in the auditory systems, the multifaceted impacts by the age-related factors would have occurred prior to higher-level brain dysfunctions (such as cognitive declines). The complexity of the aging process for the everyday challenges of speech perception in older adults would need further investigations in the future. For example, a typical question is when and where in the brain these different mechanisms take effects and whether higher-level brain regions are further involved. This has not been clearly answered in this thesis due to the insufficient ability of source localisation of the EEG method. Future research using more sophisticated methods like MEG should be able to better localise brain networks that encompass both low-level sensory and high-level cognitive regions to reveal how aging modulates neural phase-locking to speech (see *5.5* for further discussions).

5.5 Limitations and future work

5.5.1 Possible role of higher-level functions for SiN perception

Phase-locked activities obtained in the present thesis were recorded with a paradigm where participants listened passively to repeatedly presented syllables without high-level linguistic features (semantic or syntactic information). Thus, the responses reflected early-stage sensory processing of speech in the auditory brainstem and cortex. Therefore, studies in this

thesis focussed mainly on how various factors influence neural phase-locked responses without considering the higher-level cognitive processing.

Lack of measurements of higher-level cognitive functions would lead to limitations for detecting the effects of age and hearing loss on SiN perception. As discussed above, Study 1 showed that older adults had significantly worse SiN perception than young adults in BbN, but not in SpN, consistent with previous findings (Helfer and Freyman, 2008; Schoof and Rosen, 2014). However, no significant [Noise Type × Age] interactions were found for phase-locked responses, hence no sensori-neural evidence was provided to explain this behavioural phenomenon. This leaves the alternative explanation through age-related declines in higher-level cognitive functions. Aging and age-related hearing loss are associated with declines in cognitive functions, such as working memory (WM) and attention (Lin et al., 2013). WM capacity is related to SiN perception in older adults (Schoof and Rosen, 2014), while attention is critical for suppressing neural processing of background noise in SiN perception (Rimmele et al., 2015) and such ability deteriorates during aging (Andres et al., 2006; Presacco et al., 2019). WM and attention may be particularly important for SiN perception in BbN as they contribute to resisting the informational masking caused by BbN (Schneider et al., 2007; Shinn-Cunningham and Best, 2008). Furthermore, recognition of words in sentences during behavioural tasks should involve top-down processing strategies in which these higher-level cognitive functions play important roles (Davis and Johnsrude, 2007). This processing had not been reflected in the neural measurements using single syllables as stimuli. Future work would need to use sentence stimuli in addition to the current paradigm for neural measurements to compensate such caveat.

As well as WM and attention capacity during SiN perception, Mild Cognitive Impairment (MCI) could occur in some participants during normal aging (Petersen et al., 1999) and speech-evoked cortical and subcortical activities have been reported to be related to MCI that may have further affected SiN perception (Anderson et al., 2013; Bidelman et al., 2017). For example, Bidelman et al. (2017) found that people with MCI showed hyperexcitability of subcortical and cortical responses to speech compared to controls. It is thus not clear how MCI contributes to the neural-behavioural relation found the in the present study. Future studies should include measurements of cognitive functions and screening for cognitive impairment in the model to better predict SiN perception. Furthermore, additional tasks of active listening to speech stimuli under the corresponding noise types need to be conducted in the future to investigate whether the current neural-behavioural relationships is replicable.

5.5.2 Source locations of speech-evoked phase-locked responses

It is disputable whether FFRs have neural sources at the cortical level. FFRs have been argued to have main sources in the subcortical level (Bidelman, 2015; 2018), but recent MEG studies using source localization have shown additional FFR sources in the auditory cortex (Coffey et al., 2016, 2017a). FFR$_{ENV}$ in the present thesis was obtained by scalp EEG recording at the vertex, which thus cannot disentangle whether the current observed effects were applied at the cortical or subcortical level.

Despite the current finding of the causal relationship between FFRs and the right auditory cortex (Study 3), the present thesis is not able to validate the neural sources of FFRs (i.e., whether the sources are from brainstem/subcortex or auditory cortex). Scalp recording tools that better localize the FFR sources, like high-density EEG (Bidelman, 2018) or MEG (Coffey et al., 2016), should help in future research to figure out where the FFRs or changes in FFRs originate from. This could be especially important because FFRs that originate from different sources may play distinct roles in auditory perception. For example, it was shown that FFR strength at the auditory cortex had higher correlation with SiN perception than FFR strength at the subcortical level (Coffey et al., 2017a); furthermore, the correlation was found to higher in the right than the left hemisphere in the auditory cortex (Coffey et al., 2017a). Also, FFR strength at the right, but not left, auditory cortex is significantly associated with pitch discrimination ability (Coffey et al., 2016).

5.5.3 Challenges to disentangle age and hearing loss, and different types of hearing loss

Age and hearing loss were one of the main factors assessed the present thesis (Study 1 and 2). Statistical methods were used to disentangle these effects. It should be acknowledged, however, that although there are great variations between aging and hearing loss (i.e., some older adults at younger age have bad hearing while some older adults with older age have good hearing), these two factors are highly correlated (Humes et al., 2010). Future work is needed to recruit different groups of participants (e.g., three groups of younger normal-hearing, older normal-hearing and older hearing-impaired participants) that can better disentangle the effects of aging and hearing loss (e.g., Presacco et al., 2019).

Also, different types of hearing loss may have made impacts on the findings shown in the present thesis. Here, degree of hearing loss was tested via air-conduction. The audiogram for majority of the older participants showed hearing loss at high frequencies (≥2 kHz) (e.g., **Figure 2.1**), consistent with the pattern of presbycusis which is usually sensorineural hearing loss (SNHL) caused by age-related functional declines in the inner ear. Despite this, it cannot totally be excluded that some of the participants may also suffer from other types of hearing loss such as conductive hearing loss (CHL). CHL is a type of hearing loss caused by sound-blockage in the outer/middle ear without damaging the auditory nervous systems (British Society of Audiology, 2015). CHL can be a confounding factor as it influences SiN perception with different mechanisms compared to SNHL (Beales, 1997). Future work should further combine use of air- and bone-conduction to test for conductive hearing loss in order to better clarify the roles of different types of hearing loss, or to partial out the effects of CHL and only focus on those who have SNHL.

5.6 Summary

This thesis has addressed questions on how neural phase-locked responses to speech signals, particularly at the early-stage subcortical and cortical levels, are regulated by different factors. This topic is important because the phase-locked responses investigated here reflect neuro-temporal processing of fine-grained speech acoustic properties of low-frequency envelope modulations and F_0-ENV at a millisecond-scale precision that are essential for audition in general and speech perception in particular. The factors studied by this thesis include internal factors of aging (Study 1 and 2), hearing loss (Study 1), changes in level of arousal (Study 3) and cortical neural excitability (Study 3), and external factors of background noise (Study 1). The findings advance our understanding of neural processing of speech in the auditory sensory systems during our daily life and across our lifespan. The findings are summarized as follows:

(1) Background noise reduces speech-evoked phase-locked responses at both cortical (theta-band PLV) and subcortical (FFRs) levels. However, only cortical responses were evidenced to preferentially change as a function of noise types that corresponded to SiN performances. Aging and hearing loss can affect phase-locked responses in quiet and noisy backgrounds through different mechanisms and they should be considered separately during SiN perception.

(2) Speech-evoked phase-locked responses decrease following decreases in level of arousal. Furthermore, changes in theta-band PLV across arousal states are regulated by occurrence of sleep spindles. However, this neuro-regulation was found only in young, but not in older adults, indicating an age-dependent effect.

(3) Through combined neuro-stimulation and neurophysiological measures, a causal relationship between the right auditory cortex and FFRs is established for the first time, validating the right-hemispheric cortical contributions to the neural encoding of speech periodicity.

References

Ahissar, E., Nagarajan, S., Ahissar, M., Protopapas, A., Mahncke, H., & Merzenich, M. M. (2001). Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proceedings of the National Academy of Sciences*, *98*(23), 13367-13372.

Aiken, S. J., & Picton, T. W. (2008). Envelope and spectral frequency-following responses to vowel sounds. *Hearing Research*, *245*(1-2), 35-47.

Alain, C. (2014). Effects of age-related hearing loss and background noise on neuromagnetic activity from auditory cortex. *Frontiers in Systems Neuroscience*, *8*, 8.

Albouy, P., Mattout, J., Bouet, R., Maby, E., Sanchez, G., Aguera, P. E., et al. (2013). Impaired pitch perception and memory in congenital amusia: the deficit starts in the auditory cortex. *Brain*, *136*(5), 1639-1661.

Ananthakrishnan, S., Krishnan, A., & Bartlett, E. (2016). Human frequency following response: neural representation of envelope and temporal fine structure in listeners with normal hearing and sensorineural hearing loss. *Ear and Hearing*, *37*(2), e91.

Anderson S, Parbery-Clark, White-Schwoch, T., Drehobl, S., & Kraus, N. (2013). Effects of hearing loss on the subcortical representation of speech cues. *Journal of Acoustical Society of America* 133:3030-3038.

Anderson, S., Parbery-Clark, A., White-Schwoch, T., & Kraus, N. (2012). Aging affects neural precision of speech encoding. *Journal of Neuroscience*, *32*(41), 14156-14164.

Anderson, S., Parbery-Clark, A., Yi, H. G., & Kraus, N. (2011). A neural basis of speech-in-noise perception in older adults. *Ear and Hearing*, *32*(6), 750.

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.

Andrade, K. C., Spoormaker, V. I., Dresler, M., Wehrle, R., Holsboer, F., Sämann, P. G., & Czisch, M. (2011). Sleep spindles and hippocampal functional connectivity in human NREM sleep. *Journal of Neuroscience*, *31*(28), 10331-10339.

Andrés, P., Parmentier, F. B., & Escera, C. (2006). The effect of age on involuntary capture of attention by irrelevant sounds: a test of the frontal hypothesis of aging. *Neuropsychologia*, *44*(12), 2564-2568.

Antal, A., Terney, D., Poreisz, C., & Paulus, W. (2007). Towards unravelling task-related modulations of neuroplastic changes induced in the human motor cortex. *European Journal of Neuroscience*, *26*(9), 2687-2691.

Arehart, K. H., King, C. A., & McLean-Mudgett, K. S. (1997). Role of fundamental frequency differences in the perceptual separation of competing vowel sounds by listeners with normal hearing and listeners with hearing loss. *Journal of Speech, Language, and Hearing Research*, *40*(6), 1434-1444.

Aviyente, S., Bernat, E. M., Evans, W. S., & Sponheim, S. R. (2011). A phase synchrony measure for quantifying dynamic functional integration in the brain. *Human Brain Mapping*, *32*(1), 80-93.

Baccalá, L. A., & Sameshima, K. (2001). Partial directed coherence: a new concept in neural structure determination. *Biological Cybernetics*, *84*(6), 463-474.

Bai, S., Dokos, S., Ho, K. A., & Loo, C. (2014). A computational modelling study of transcranial direct current stimulation montages used in depression. *NeuroImage*, *87*, 332-344.

Banai, K., Abrams, D., & Kraus, N. (2007). Sensory-based learning disability: Insights from brainstem processing of speech sounds. *International Journal of Audiology*, *46*(9), 524-532.

Batsikadze, G., Moliadze, V., Paulus, W., Kuo, M. F., & Nitsche, M. A. (2013). Partially non-linear stimulation intensity-dependent effects of direct current stimulation on motor cortex excitability in humans. *The Journal of Physiology*, *591*(7), 1987-2000.

Beales, P. H. (1997). Otosclerosis, in Scott-Brown's Otolaryngology (6th edition) Kerr A. G. (Ed.), Butterworths, London. pp. 306-307

Bench, J., Kowal, Å., & Bamford, J. (1979). The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *British Journal of Audiology*, *13*(3), 108-112.

Bendor, D., & Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature*, *436*(7054), 1161-1165.

Bergmann, T. O., Mölle, M., Diedrichs, J., Born, J., & Siebner, H. R. (2012). Sleep spindle-related reactivation of category-specific cortical regions after learning face-scene associations. *Neuroimage*, *59*(3), 2733-2742.

Bidelman, G. M. (2015). Multichannel recordings of the human brainstem frequency-following response: scalp topography, source generators, and distinctions from the transient ABR. *Hearing Research*, *323*, 68-80.

Bidelman, G. M. (2018). Subcortical sources dominate the neuroelectric auditory frequency-following response to speech. *NeuroImage*, *175*, 56-69.

Bidelman, G. M., Davis, M. K., & Pridgen, M. H. (2018). Brainstem-cortical functional connectivity for speech is differentially challenged by noise and reverberation. *Hearing Research*, *367*, 149-160.

Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2011). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *Journal of Cognitive Neuroscience*, *23*(2), 425-434.

Bidelman, G. M., Krishnan, A., & Gandour, J. T. (2011). Enhanced brainstem encoding predicts musicians' perceptual advantages with pitch. *European Journal of Neuroscience*, *33*(3), 530-538.

Bidelman, G. M., Lowther, J. E., Tak, S. H., & Alain, C. (2017). Mild cognitive impairment is characterized by deficient brainstem and cortical representations of speech. *Journal of Neuroscience*, *37*(13), 3610-3620.

Bidelman, G. M., Price, C. N., Shen, D., Arnott, S. R., & Alain, C. (2019). Afferent-efferent connectivity between auditory brainstem and cortex accounts for poorer speech-in-noise comprehension in older adults. *Hearing Research*, *382*, 107795.

Billings, C. J., Penman, T. M., McMillan, G. P., & Ellis, E. (2015). Electrophysiology and perception of speech in noise in older listeners: effects of hearing impairment & age. *Ear and Hearing*, *36*(6), 710.

Binns, C., & Culling, J. F. (2005). The role of fundamental frequency (F 0) contours in the perception of speech against interfering speech. *The Journal of the Acoustical Society of America*, *117*(4), 2606-2607.

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.

Bird, J., & Darwin, C. J. (1998). Effects of a difference in fundamental frequency in separating two sentences. *Psychophysical and physiological advances in hearing*, 263-269.

Boersma, P., & Weenink, D. (2013). Praat: doing phonetics by computer [Computer program]. Version 5.3.51. *Online: http://www.praat.org*.

Bregman, A. S. (1994). Auditory Scene Analysis: The Perceptual Organization of Sound (MIT Press, Cambridge, MA).

British Society of Audiology (2015). Recommended Procedure: Pure-tone air-conduction and bone-conduction threshold audiometry with and without masking. Retrieved from: http://www.thebsa.org.uk/wpcontent/uploads/2011/04/BSA_PTA_Dec_15_minor_ammen dments.pdf

Brooks, J. C. W., Faull, O. K., Pattinson, K. T., & Jenkinson, M. (2013). Physiological noise in brainstem FMRI. *Frontiers in Human Neuroscience*, *7*, 623.

Brown, R. E., Basheer, R., McKenna, J. T., Strecker, R. E., & McCarley, R. W. (2012). Control of sleep and wakefulness. *Physiological Reviews*, *92*(3), 1087-1187.

Burnham, K. P. dr AnderSon. 2002. Model selection and multimodel inference: a practical information-theoretic approach. *Ecological Modelling. Springer Science & Business Media, New York, New York, USA*.

Carcagno, S., & Plack, C. J. (2011). Subcortical plasticity following perceptual learning in a pitch discrimination task. *Journal of the Association for Research in Otolaryngology*, *12*(1), 89-100.

Carey, D., & Johnstone, L. T. (2014). Quantifying cerebral asymmetries for language in dextrals and adextrals with random-effects meta analysis. *Frontiers in Psychology*, *5*, 1128.

Carlile, S., & Corkhill, C. (2015). Selective spatial attention modulates bottom-up informational masking of speech. *Scientific Reports*, *5*(1), 1-7.

Carpenter, A. L., & Shahin, A. J. (2013). Development of the N1–P2 auditory evoked response to amplitude rise time and rate of formant transition of speech sounds. *Neuroscience Letters*, *544*, 56-61.

Caspary, D. M., Ling, L., Turner, J. G., & Hughes, L. F. (2008). Inhibitory neurotransmission, plasticity and aging in the mammalian central auditory system. *Journal of Experimental Biology*, *211*(11), 1781-1791.

Cha, K., Zatorre, R. J., & Schönwiesner, M. (2016). Frequency selectivity of voxel-by-voxel functional connectivity in human auditory cortex. *Cerebral Cortex*, *26*(1), 211-224.

Chandrasekaran, B., & Kraus, N. (2010). The scalp-recorded brainstem response to speech: Neural origins and plasticity. *Psychophysiology*, *47*(2), 236-246.

Choi, I., Rajaram, S., Varghese, L. A., & Shinn-Cunningham, B. G. (2013). Quantifying attentional modulation of auditory-evoked cortical responses from single-trial electroencephalography. *Frontiers in Human Neuroscience*, *7*, 115.

Coenen, A. (2010). Subconscious stimulus recognition and processing during sleep. PSYCHE: An Interdisciplinary Journal of Research on Consciousness 16:90-97

Coffey, E. B., Musacchia, G., & Zatorre, R. J. (2017a). Cortical correlates of the auditory frequency-following and onset responses: EEG and fMRI evidence. *Journal of Neuroscience*, *37*(4), 830-838.

Coffey, E. B., Chepesiuk, A. M., Herholz, S. C., Baillet, S., & Zatorre, R. J. (2017b). Neural correlates of early sound encoding and their relationship to speech-in-noise perception. *Frontiers in Neuroscience*, *11*, 479.

Coffey, E. B., Herholz, S. C., Chepesiuk, A. M., Baillet, S., & Zatorre, R. J. (2016). Cortical contributions to the auditory frequency-following response revealed by MEG. *Nature Communications*, *7*(1), 1-11.

Coffey, E. B., Nicol, T., White-Schwoch, T., Chandrasekaran, B., Krizman, J., Skoe, E., Zotorre, R. J. & Kraus, N. (2019). Evolving perspectives on the sources of the frequency-following response. *Nature Communications*, *10*(1), 1-10.

Cohen, J. (1988). Statistical power analysis for the behavioural sciences (2nd Ed.). Mahwah, NJ: Erlbaum.

Cohen, L. T., Rickards, F. W., & Clark, G. M. (1991). A comparison of steady-state evoked potentials to modulated tones in awake and sleeping humans. *The Journal of the Acoustical Society of America*, *90*(5), 2467-2479.

Cunningham, J., Nicol, T., Zecker, S. G., Bradlow, A., & Kraus, N. (2001). Neurobiologic responses to speech in noise in children with learning problems: deficits and strategies for improvement. *Clinical Neurophysiology*, *112*(5), 758-767.

Czisch, M., Wehrle, R., Kaufmann, C., Wetter, T. C., Holsboer, F., Pollmächer, T., & Auer, D. P. (2004). Functional MRI during sleep: BOLD signal decreases and their electrophysiological correlates. *European Journal of Neuroscience*, *20*(2), 566-574.

Czisch, M., Wetter, T. C., Kaufmann, C., Pollmächer, T., Holsboer, F., & Auer, D. P. (2002). Altered processing of acoustic stimuli during sleep: reduced auditory activation and visual deactivation detected by a combined fMRI/EEG study. *NeuroImage*, *16*(1), 251-258.

Dajani, H. R., Purcell, D., Wong, W., Kunov, H., & Picton, T. W. (2005). Recording human evoked potentials that follow the pitch contour of a natural vowel. *IEEE Transactions on Biomedical Engineering*, *52*(9), 1614-1618.

Dang-Vu, T. T., Bonjean, M., Schabus, M., Boly, M., Darsaud, A., Desseilles, M., ... & Sejnowski, T. J. (2011). Interplay between spontaneous and induced brain activity during human non-rapid eye movement sleep. *Proceedings of the National Academy of Sciences*, *108*(37), 15438-15443.

Dang-Vu, T. T., McKinney, S. M., Buxton, O. M., Solet, J. M., & Ellenbogen, J. M. (2010). Spontaneous brain rhythms predict sleep stability in the face of noise. *Current Biology*, *20*(15), R626-R627.

Davis, M. H., Coleman, M. R., Absalom, A. R., Rodd, J. M., Johnsrude, I. S., Matta, B. F. et al. (2007). Dissociating speech perception and comprehension at reduced levels of awareness. *Proceedings of the National Academy of Sciences*, 104(41), 16032-16037.

Davis, M. H., & Johnsrude, I. S. (2007). Hearing speech sounds: top-down influences on the interface between audition and speech perception. *Hearing Research*, *229*(1-2), 132-147.

De Gennaro, L., & Ferrara, M. (2003). Sleep spindles: an overview. *Sleep Medicine Reviews*, *7*(5), 423-440.

de Xivry, J. J. O., Marko, M. K., Pekny, S. E., Pastor, D., Izawa, J., Celnik, P., & Shadmehr, R. (2011). Stimulation of the human motor cortex alters generalization patterns of motor learning. *Journal of Neuroscience*, *31*(19), 7102-7110.

Dimitrijevic, A., John, M. S., & Picton, T. W. (2004). Auditory steady-state responses and word recognition scores in normal-hearing and hearing-impaired adults. *Ear and Hearing*, *25*(1), 68-84.

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.

Doelling, K. B., Arnal, L. H., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage, 85*, 761-768.

Eaves, J. M., Quentin Summerfield, A., & Kitterick, P. T. (2011). Benefit of temporal fine structure to speech perception in noise measured with controlled temporal envelopes. *The Journal of the Acoustical Society of America*, *130*(1), 501-507.

Encina-Llamas, G., Harte, J. M., Dau, T., Shinn-Cunningham, B., & Epp, B. (2019). Investigating the effect of cochlear synaptopathy on envelope following responses using a model of the auditory nerve. *Journal of the Association for Research in Otolaryngology*, *20*(4), 363-382.

Faria, P., Hallett, M., & Miranda, P. C. (2011). A finite element analysis of the effect of electrode area and inter-electrode distance on the spatial distribution of the current density in tDCS. *Journal of Neural Engineering*, *8*(6), 066017.

Ferrucci, R., Marceglia, S., Vergari, M., Cogiamanian, F., Mrakic-Sposta, S., Mameli, F., et al. (2008). Cerebellar transcranial direct current stimulation impairs the practice-dependent proficiency increase in working memory. *Journal of Cognitive Neuroscience*, *20*(9), 1687-1697.

Filmer, H. L., Dux, P. E., & Mattingley, J. B. (2014). Applications of transcranial direct current stimulation for understanding brain function. *Trends In Neurosciences*, *37*(12), 742-753.

Fujihira, H., & Shiraishi, K. (2015). Correlations between word intelligibility under reverberation and speech auditory brainstem responses in elderly listeners. *Clinical Neurophysiology*, *126*(1), 96-102.

Füllgrabe, C., Moore, B. C., & 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*, 347.

Galbraith, G. C., & Doan, B. Q. (1995). Brainstem frequency-following and behavioral responses during selective attention to pure tone and missing fundamental stimuli. *International Journal of Psychophysiology*, *19*(3), 203-214.

Galbraith, G. C., Olfman, D. M., & Huffman, T. M. (2003). Selective attention affects human brain stem frequency-following response. *NeuroReport*, *14*(5), 735-738.

Gama, N., Peretz, I., & Lehmann, A. (2017). Recording the human brainstem frequency-following-response in the free-field. *Journal of Neuroscience Methods*, *280*, 47-53.

Gockel, H. E., Carlyon, R. P., Mehta, A., & Plack, C. J. (2011). The frequency following response (FFR) may reflect pitch-bearing information but is not a direct representation of pitch. *Journal of the Association for Research in Otolaryngology*, *12*(6), 767-782.

Goossens, T., Vercammen, C., Wouters, J., & van Wieringen, A. (2018). Neural envelope encoding predicts speech perception performance for normal-hearing and hearing-impaired adults. *Hearing Research*, *370*, 189-200.

Goossens, T., Vercammen, C., Wouters, J., & van Wieringen, A. (2019). The association between hearing impairment and neural envelope encoding at different ages. *Neurobiology of Aging*, *74*, 202-212.

Goossens, T., Vercammen, C., Wouters, J., & Wieringen, A. V. (2016). Aging affects neural synchronization to speech-related acoustic modulations. *Frontiers in Aging Neuroscience*, *8*, 133.

Gopinath, B., Rochtchina, E., Wang, J. J., Schneider, J., Leeder, S. R., & Mitchell, P. (2009). Prevalence of age-related hearing loss in older adults: Blue Mountains Study. *Archives of Internal Medicine*, *169*(4), 415-418.

Greenberg, S., Carvey, H., Hitchcock, L., & Chang, S. (2003). Temporal properties of spontaneous speech—a syllable-centric perspective. *Journal of Phonetics*, *31*(3-4), 465-485.

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., & Garrod, S. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biol*, *11*(12), e1001752.

Hairston, W. D., Letowski, T. R., & McDowell, K. (2013). Task-related suppression of the brainstem frequency following response. *PLoS One*, *8*(2).

Harris, K. C., Eckert, M. A., Ahlstrom, J. B., & Dubno, J. R. (2010). Age-related differences in gap detection: Effects of task difficulty and cognitive ability. *Hearing Research*, *264*(1-2), 21-29.

Helfer, K. S., & Freyman, R. L. (2008). Aging and speech-on-speech masking. *Ear and hearing*, *29*(1), 87.Henry, K. S., Kale, S., & Heinz, M. G. (2014). Noise-induced hearing loss increases the temporal precision of complex envelope coding by auditory-nerve fibers. *Frontiers in Systems Neuroscience*, *8*, 20.

Henry, K. S., Kale, S., & Heinz, M. G. (2014). Noise-induced hearing loss increases the temporal precision of complex envelope coding by auditory-nerve fibers. *Frontiers in Systems Neuroscience*, *8*, 20.

Henry, K. S., & Heinz, M. G. (2012). Diminished temporal coding with sensorineural hearing loss emerges in background noise. *Nature Neuroscience*, *15*(10), 1362.

Henry, M. J., & Obleser, J. (2012). Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proceedings of the National Academy of Sciences*, *109*(49), 20095-20100.

Herdman, A. T., Lins, O., Van Roon, P., Stapells, D. R., Scherg, M., & Picton, T. W. (2002). Intracerebral sources of human auditory steady-state responses. *Brain Topography*, *15*(2), 69-86.

Herrmann, B., Henry, M. J., Johnsrude, I. S., & Obleser, J. (2016). Altered temporal dynamics of neural adaptation in the aging human auditory cortex. *Neurobiology of Aging*, *45*, 10-22.

Herrmann, B., Henry, M. J., Scharinger, M., & Obleser, J. (2013). Auditory filter width affects response magnitude but not frequency specificity in auditory cortex. *Hearing Research*, *304*, 128-136.

Hoen, M., Meunier, F., Grataloup, C. L., Pellegrino, F., Grimault, N., Perrin, F., et al (2007). Phonetic and lexical interferences in informational masking during speech-in-speech comprehension. Speech communication, 49(12), 905-916.

Hopkins, K., & Moore, B. C. (2011). The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise. *The Journal of the Acoustical Society of America*, *130*(1), 334-349.

Howard, M. F., & Poeppel, D. (2010). Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. *Journal of Neurophysiology*, *104*(5), 2500-2511.

Humes, L. E., & Dubno, J. R. (2010). Factors affecting speech understanding in older adults. In *The aging auditory system* (pp. 211-257). Springer, New York, NY.

Humes, L. E., Busey, T. A., Craig, J. C., & Kewley-Port, D. (2009). The effects of age on sensory thresholds and temporal gap detection in hearing, vision, and touch. *Attention, Perception, & Psychophysics*, *71*(4), 860-871.

Humes, L. E., Kewley-Port, D., Fogerty, D., & Kinney, D. (2010). Measures of hearing threshold and temporal processing across the adult lifespan. *Hearing research*, *264*(1-2), 30-40.

Humes, L. E., Lee, J. H., & Coughlin, M. P. (2006). Auditory measures of selective and divided attention in young and older adults using single-talker competition. *The Journal of the Acoustical Society of America*, *120*(5), 2926-2937.

Hunter, L. L., & Sanford, C. A. (2015). Tympanometry and wideband acoustic immitance. *Katz J seventh ed: Physiologic Principles and Measures, Section*, *2*, 137-163.

Hyde, K. L., Peretz, I., & Zatorre, R. J. (2008). Evidence for the role of the right auditory cortex in fine pitch resolution. *Neuropsychologia*, *46*(2), 632-639.

Issa, E. B., & Wang, X. (2008). Sensory responses during sleep in primate primary and secondary auditory cortex. *Journal of Neuroscience*, *28*(53), 14467-14480.

Jacobson, L., Koslowsky, M., & Lavidor, M. (2012). tDCS polarity effects in motor and cognitive domains: a meta-analytical review. *Experimental Brain Research*, *216*(1), 1-10.

Kale, S., & Heinz, M. G. (2010). Envelope coding in auditory nerve fibers following noise-induced hearing loss. *Journal of the Association for Research in Otolaryngology*, *11*(4), 657-673.

Keshavarzi, M., Kegler, M., Kadir, S., & Reichenbach, T. (2020). Transcranial alternating current stimulation in the theta band but not in the delta band modulates the comprehension of naturalistic speech in noise. *NeuroImage, 210*, 116557.

Keshavarzi, M., & Reichenbach, T. (2020). Transcranial alternating current stimulation with the theta-band portion of the temporally-aligned speech envelope improves speech-in-noise comprehension. *Frontiers in Human Neuroscience*.

Kim, A., Latchoumane, C., Lee, S., Kim, G. B., Cheong, E., Augustine, G. J., & Shin, H. S. (2012). Optogenetically induced sleep spindle rhythms alter sleep architectures in mice. *Proceedings of the National Academy of Sciences*, *109*(50), 20673-20678.

Kong, Y. Y., Mullangi, A., & Ding, N. (2014). Differential modulation of auditory responses to attended and unattended speech in different listening conditions. *Hearing Research*, *316*, 73-81.

Krishnan, A., Gandour, J. T., Bidelman, G. M., & Swaminathan, J. (2009). Experience dependent neural representation of dynamic pitch in the brainstem. *Neuroreport*, *20*(4), 408.

Krishnan, A., Xu, Y., Gandour, J. T., & Cariani, P. A. (2004). Human frequency-following response: representation of pitch contours in Chinese tones. *Hearing research*, *189*(1-2), 1-12.

Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research*, *25*(1), 161-168.

Kunze, T., Hunold, A., Haueisen, J., Jirsa, V., & Spiegler, A. (2016). Transcranial direct current stimulation changes resting state functional connectivity: a large-scale brain network modeling study. *NeuroImage*, *140*, 174-187.

Lakatos, P., Chen, C. M., O'Connell, M. N., Mills, A., & Schroeder, C. E. (2007). Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron*, *53*(2), 279-292.

Langner, G., & Schreiner, C. E. (1988). Periodicity coding in the inferior colliculus of the cat. I. Neuronal mechanisms. Journal of Neurophysiology, 60(6), 1799-1822.

Lehmann, A., & Schönwiesner, M. (2014). Selective attention modulates human auditory brainstem responses: relative contributions of frequency and spatial cues. *PloS One*, *9*(1).

Lin, F. R., Yaffe, K., Xia, J., Xue, Q. L., Harris, T. B., Purchase-Helzner, E., et al. (2013). Hearing loss and cognitive decline in older adults. *JAMA Internal Medicine*, *173*(4), 293-299.

Lins, O. G., Picton, P. E., Picton, T. W., Champagne, S. C., & Durieux-Smith, A. (1995). Auditory steady-state responses to tones amplitude modulated at 80–110 Hz. *The Journal of the Acoustical Society of America*, *97*(5), 3051-3063.

Luke, R., De Vos, A., & Wouters, J. (2017). Source analysis of auditory steady-state responses in acoustic and electric hearing. *NeuroImage*, *147*, 568-576.

Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron*, *54*(6), 1001-1010.

Mai, G., Minett, J. W., & Wang, W. S. Y. (2016). Delta, theta, beta, and gamma brain oscillations index levels of auditory sentence processing. *NeuroImage*, *133*, 516-528.

Mai, G., Tuomainen, J., & Howell, P. (2018). Relationship between speech-evoked neural responses and perception of speech in noise in older adults. *The Journal of the Acoustical Society of America*, *143*(3), 1333-1345.

Makov, S., Sharon, O., Ding, N., Ben-Shachar, M., Nir, Y., & Golumbic, E. Z. (2017). Sleep disrupts high-level speech parsing despite significant basic auditory processing. *Journal of Neuroscience*, *37*(32), 7772-7781.

Mander, B. A., Winer, J. R., & Walker, M. P. (2017). Sleep and human aging. *Neuron*, *94*(1), 19-36.

Martin, N., Lafortune, M., Godbout, J., Barakat, M., Robillard, R., Poirier, G., ... & Carrier, J. (2013). Topography of age-related changes in sleep spindles. *Neurobiology of aging*, *34*(2), 468-476.

Mathys, C., Loui, P., Zheng, X., & Schlaug, G. (2010). Non-invasive brain stimulation applied to Heschl's gyrus modulates pitch discrimination. *Frontiers in Psychology*, *1*, 193.

Matsushita, R., Andoh, J., & Zatorre, R. J. (2015). Polarity-specific transcranial direct current stimulation disrupts auditory pitch learning. *Frontiers in Neuroscience*, *9*, 174.

McCormick, D. A., & Bal, T. (1994). Sensory gating mechanisms of the thalamus. *Current Opinion in Neurobiology*, *4*(4), 550-556.

McCormick, D. A., & Bal, T. (1997). Sleep and arousal: thalamocortical mechanisms. *Annual Review of Neuroscience*, *20*(1), 185-215.

Moore, B. C. (2008). The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people. *Journal of the Association for Research in Otolaryngology*, *9*(4), 399-406.

Moore B.C.J. (2014). *Auditory processing of temporal fine structure: Effects of age and hearing loss.* World Scientific, Sigapore.

Morillon, B., Liégeois-Chauvel, C., Arnal, L. H., Bénar, C. G., & Giraud, A. L. (2012). Asymmetric function of theta and gamma activity in syllable processing: an intra-cortical study. *Frontiers in Psychology*, *3*, 248.

Moushegian, G., Rupert, A. L., & Stillman, R. D. (1973). Scalp-recorded early responses in man to frequencies in the speech range. *Electroencephalography and Clinical Neurophysiology*, *35*(6), 665-667.

Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences*, *104*(40), 15894-15898.

Ng, B. S. W., Logothetis, N. K., & Kayser, C. (2013). EEG phase patterns reflect the selectivity of neural firing. *Cerebral Cortex*, *23*(2), 389-398

Nir, Y., Vyazovskiy, V. V., Cirelli, C., Banks, M. I., & Tononi, G. (2015). Auditory responses and stimulus-specific adaptation in rat auditory cortex are preserved across NREM and REM sleep. *Cerebral Cortex*, *25*(5), 1362-1378.

Nitsche, M. A., & Paulus, W. (2001). Sustained excitability elevations induced by transcranial DC motor cortex stimulation in humans. *Neurology*, *57*(10), 1899-1901.

Noda, T., Kanzaki, R., & Takahashi, H. (2014). Amplitude and phase-locking adaptation of neural oscillation in the rat auditory cortex in response to tone sequence. *Neuroscience Research*, *79*, 52-60.

Noguchi, Y., Fujiwara, M., & Hamano, S. (2015). Temporal evolution of neural activity underlying auditory discrimination of frequency increase and decrease. *Brain Topography*, *28*(3), 437-444.

Oya, H., Gander, P. E., Petkov, C. I., Adolphs, R., Nourski, K. V., Kawasaki, H., ... & Griffiths, T. D. (2018). Neural phase locking predicts BOLD response in human auditory cortex. *NeuroImage*, *169*, 286-301.

Parbery-Clark, A., Marmel, F., Bair, J., & Kraus, N. (2011). What subcortical–cortical relationships tell us about processing speech in noise. *European Journal of Neuroscience*, *33*(3), 549-557.

Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron*, *36*(4), 767-776.

Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology*, *3*, 320.

Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex*, *23*(6), 1378-1387.

Penagos, H., Melcher, J. R., & Oxenham, A. J. (2004). A neural representation of pitch salience in nonprimary human auditory cortex revealed with functional magnetic resonance imaging. *Journal of Neuroscience*, *24*(30), 6810-6815.

Pérez-González, D., & Malmierca, M. S. (2014). Adaptation in the auditory system: an overview. *Frontiers in Integrative Neuroscience*, *8*, 19.

Petersen, R. C., Smith, G. E., Waring, S. C., Ivnik, R. J., Tangalos, E. G., & Kokmen, E. (1999). Mild cognitive impairment: clinical characterization and outcome. *Archives of neurology*, *56*(3), 303-308.

Picchioni, D., Pixa, M. L., Fukunaga, M., Carr, W. S., Horovitz, S. G., Braun, A. R., & Duyn, J. H. (2014). Decreased connectivity between the thalamus and the neocortex during human nonrapid eye movement sleep. *Sleep*, *37*(2), 387-397.

Picton, T. W., John, M. S., Dimitrijevic, A., & Purcell, D. (2003). Human auditory steady-state responses: Respuestas auditivas de estado estable en humanos. *International Journal of Audiology*, *42*(4), 177-219.

Plomp, R., & Mimpen, A. M. (1979). Speech-reception threshold for sentences as a function of age and noise level. *The Journal of the Acoustical Society of America*, *66*(5), 1333-1342.

Portas, C. M., Krakow, K., Allen, P., Josephs, O., Armony, J. L., & Frith, C. D. (2000). Auditory processing across the sleep-wake cycle: simultaneous EEG and fMRI monitoring in humans. *Neuron*, *28*(3), 991-999.

Presacco, A., Simon, J. Z., & Anderson, S. (2016a). Evidence of degraded representation of speech in noise, in the aging midbrain and cortex. *Journal of Neurophysiology*, *116*(5), 2346-2355.

Presacco, A., Simon, J. Z., & Anderson, S. (2016b). Effect of informational content of noise on speech representation in the aging midbrain and cortex. *Journal of Neurophysiology*, *116*(5), 2356-2367.

Presacco, A., Simon, J. Z., & Anderson, S. (2019). Speech-in-noise representation in the aging midbrain and cortex: Effects of hearing loss. *PloS One*, *14*(3).

Puschmann, S., Uppenkamp, S., Kollmeier, B., & Thiel, C. M. (2010). Dichotic pitch activates pitch processing centre in Heschl's gyrus. *NeuroImage*, *49*(2), 1641-1649.

Ribas-Prats, T., Almeida, L., Costa-Faidella, J., Plana, M., Corral, M. J., Gómez-Roig, M. D., & Escera, C. (2019). The frequency-following response (FFR) to speech stimuli: a normative dataset in healthy newborns. *Hearing Research*, *371*, 28-39.

Riecke, L., Formisano, E., Herrmann, C. S., & Sack, A. T. (2015). 4-Hz transcranial alternating current stimulation phase modulates hearing. *Brain Stimulation*, *8*(4), 777-783.

Rimmele, J. M., Golumbic, E. Z., Schröger, E., & Poeppel, D. (2015). The effects of selective attention and speech acoustics on neural speech-tracking in a multi-talker scene. *Cortex*, *68*, 144-154.

Roß, B., Picton, T. W., & Pantev, C. (2002). Temporal integration in the human auditory cortex as represented by the development of the steady-state magnetic field. *Hearing Research*, *165*(1-2), 68-84.

Rosen, S. (1992). Temporal information in speech: acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *336*(1278), 367-373.

Rosen, S., Souza, P., Ekelund, C., & Majeed, A. A. (2013). Listening to speech in a background of other talkers: Effects of talker number and noise vocoding. *The Journal of the Acoustical Society of America*, *133*(4), 2431-2443.

Ross, B., Herdman, A. T., & Pantev, C. (2005). Right hemispheric laterality of human 40 Hz auditory steady-state responses. *Cerebral Cortex*, *15*(12), 2029-2039.

Russo, N. M., Skoe, E., Trommer, B., Nicol, T., Zecker, S., Bradlow, A., & Kraus, N. (2008). Deficient brainstem encoding of pitch in children with autism spectrum disorders. *Clinical Neurophysiology*, *119*(8), 1720-1731.

Russo, N., Nicol, T., Musacchia, G., & Kraus, N. (2004). Brainstem responses to speech syllables. *Clinical Neurophysiology*, *115*(9), 2021-2030.

Salehinejad, M., Kuo, M., & Nitsche, M. (2019). The impact of chronotypes and time of the day on tDCS-induced motor cortex plasticity and cortical excitability. *Brain Stimulation: Basic, Translational, and Clinical Research in Neuromodulation*, *12*(2), 421.

Schabus, M. D., Dang-Vu, T. T., Heib, D. P. J., Boly, M., Desseilles, M., Vandewalle, G., et al. (2012). The fate of incoming stimuli during NREM sleep is determined by spindles and the phase of the slow oscillation. *Frontiers in Neurology*, *3*, 40.

Schabus, M., Dang-Vu, T. T., Albouy, G., Balteau, E., Boly, M., Carrier, J., et al. (2007). Hemodynamic cerebral correlates of sleep spindles during human non-rapid eye movement sleep. *Proceedings of the National Academy of Sciences*, *104*(32), 13164-13169.

Schelter, B., Timmer, J., & Eichler, M. (2009). Assessing the strength of directed influences among neural signals using renormalized partial directed coherence. *Journal of Neuroscience Methods*, *179*(1), 121-130.

Schneider, B. A., & Hamstra, S. J. (1999). Gap detection thresholds as a function of tonal duration for younger and older listeners. *The Journal of the Acoustical Society of America*, *106*(1), 371-380.

Schneider, B. A., Li, L., & Daneman, M. (2007). How competing speech interferes with speech comprehension in everyday listening situations. *Journal of the American Academy of Audiology*, *18*(7), 559-572.

Schochat, E., Rocha-Muniz, C. N., & Filippini, R. (2017). Understanding Auditory Processing Disorder Through the FFR. In *The Frequency-Following Response* (pp. 225-250). Springer, Cham.

Schoof, T., & Rosen, S. (2014). The role of auditory and cognitive factors in understanding speech in noise by normal-hearing older listeners. *Frontiers in Aging Neuroscience*, *6*, 307.

Schoof, T., & Rosen, S. (2016). The role of age-related declines in subcortical auditory processing in speech perception in noise. *Journal of the Association for Research in Otolaryngology*, *17*(5), 441-460.

Schroeder, C. E., & Lakatos, P. (2009). Low-frequency neuronal oscillations as instruments of sensory selection. *Trends in Neurosciences*, *32*(1), 9-18.

Sehm, B., Schäfer, A., Kipping, J., Margulies, D., Conde, V., Taubert, M., et al. (2012). Dynamic modulation of intrinsic functional connectivity by transcranial direct current stimulation. *Journal of Neurophysiology*, *108*(12), 3253-3263.

Shah, B., Nguyen, T. T., & Madhavan, S. (2013). Polarity independent effects of cerebellar tDCS on short term ankle visuomotor learning. *Brain Stimulation*, *6*(6), 966-968.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*(5234), 303-304.

Sherman, S. M. (2001). Tonic and burst firing: dual modes of thalamocortical relay. *Trends in Neurosciences*, *24*(2), 122-126.

Shinn-Cunningham, B. G., & Best, V. (2008). Selective attention in normal and impaired hearing. *Trends in Amplification*, *12*(4), 283-299.

Skoe, E., & Kraus, N. (2010). Auditory brainstem response to complex sounds: a tutorial. *Ear and Hearing*, *31*(3), 302.

Skoe, E., Chandrasekaran, B., Spitzer, E. R., Wong, P. C., & Kraus, N. (2014). Human brainstem plasticity: the interaction of stimulus probability and auditory learning. *Neurobiology of Learning and Memory*, *109*, 82-93.

Smith, J. C., Marsh, J. T., & Brown, W. S. (1975). Far-field recorded frequency-following responses: evidence for the locus of brainstem sources. *Electroencephalography and Clinical Neurophysiology*, 39(5), 465-472.

Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002). Chimaeric sounds reveal dichotomies in auditory perception. *Nature*, *416*(6876), 87-90.

Smith, S. M., Elliott, L. T., Alfaro-Almagro, F., McCarthy, P., Nichols, T. E., Douaud, G., & Miller, K. L. (2020). Brain aging comprises many modes of structural and functional change with distinct genetic and biophysical associations. *Elife*, *9*, e52677.

Sohmer, H., & Pratt, H. (1977). Identification and separation of acoustic frequency following responses (FFRs) in man. Electroencephalography and Clinical Neurophysiology, 42(4), 493-500.

Sommers, M. S., & Gehr, S. E. (1998). Auditory suppression and frequency selectivity in older and younger adults. *The Journal of the Acoustical Society of America*, *103*(2), 1067-1074.

Song, J. H., Skoe, E., Wong, P. C., & Kraus, N. (2008). Plasticity in the adult human auditory brainstem following short-term linguistic training. *Journal of Cognitive Neuroscience*, *20*(10), 1892-1902.

Song, J. H., Skoe, E., Banai, K., & Kraus, N. (2011). Perception of speech in noise: neural correlates. *Journal of Cognitive Neuroscience*, *23*(9), 2268-2279.

Souza, P., & Rosen, S. (2008). Factors affecting recognition of vocoded speech: Effect of envelope cutoff frequency and carrier type. *The Journal of the Acoustical Society of America*, *123*(5), 3864-3864.

Spoormaker, V. I., Czisch, M., Maquet, P., & Jäncke, L. (2011). Large-scale functional brain networks in human non-rapid eye movement sleep: insights from combined electroencephalographic/functional magnetic resonance imaging studies. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *369*(1952), 3708-3729.

Spoormaker, V. I., Schröter, M. S., Gleiser, P. M., Andrade, K. C., Dresler, M., Wehrle, R., et al.. (2010). Development of a large-scale functional brain network during human non-rapid eye movement sleep. *Journal of Neuroscience*, *30*(34), 11379-11387.

Stagg, C. J., Best, J. G., Stephenson, M. C., O'Shea, J., Wylezinska, M., Kincses, Z. T., et al. (2009). Polarity-sensitive modulation of cortical neurotransmitters by transcranial stimulation. *Journal of Neuroscience*, *29*(16), 5202-5206.

Steriade, M., McCormick, D. A., & Sejnowski, T. J. (1993). Thalamocortical oscillations in the sleeping and aroused brain. *Science*, *262*(5134), 679-685.

Stickney, G. S., Assmann, P. F., Chang, J., & Zeng, F. G. (2007). Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentences. *The Journal of the Acoustical Society of America*, *122*(2), 1069-1078.

Stine, R. A. (1995). Graphical interpretation of variance inflation factors. *The American Statistician*, *49*(1), 53-56.

Strait, D. L., Kraus, N., Skoe, E. & Ashley, R. Musical experience and neural efficiency: effects of training on subcortical processing of vocal expressions of emotion. *Eur. J. Neurosci.* **29**, 661–668 (2009).

Thatcher, R. W. (2012). Coherence, phase differences, phase shift, and phase lock in EEG/ERP analyses. *Developmental Neuropsychology*, *37*(6), 476-496.

Tlumak, A. I., Durrant, J. D., & Delgado, R. E. (2015). The effect of advancing age on auditory middle-and long-latency evoked potentials using a steady-state-response approach. *American Journal of Audiology*, *24*(4), 494-507.

Tun, P. A., McCoy, S., & Wingfield, A. (2009). Aging, hearing acuity, and the attentional costs of effortful listening. *Psychology and Aging*, *24*(3), 761.

Tun, P. A., O'Kane, G., & Wingfield, A. (2002). Distraction by competing speech in young and older adult listeners. *Psychology and Aging*, *17*(3), 453.

Unal, G., & Bikson, M. (2018). Transcranial Direct Current Stimulation (tDCS). In *Neuromodulation* (pp. 1589-1610). Academic Press.

Vanvooren, S., Hofmann, M., Poelmans, H., Ghesquière, P., & Wouters, J. (2015). Theta, beta and gamma rate modulations in the developing auditory system. *Hearing Research*, *327*, 153-162.

Verschooten, E., Shamma, S., Oxenham, A. J., Moore, B. C., Joris, P. X., Heinz, M. G., & Plack, C. J. (2019). The upper frequency limit for the use of phase locking to code temporal fine structure in humans: A compilation of viewpoints. Hearing research, 377, 109-121.

Wang, H., Li, R., & Tsai, C. L. (2007). Tuning parameter selectors for the smoothly clipped absolute deviation method. *Biometrika*, *94*(3), 553-568.

Warby, S. C., Wendt, S. L., Welinder, P., Munk, E. G., Carrillo, O., Sorensen, H. B., Jennum, P., Peppard, P. E., Perona, P., & Mignot, E. (2014). Sleep-spindle detection: crowdsourcing

and evaluating performance of experts, non-experts and automated methods. *Nature Methods*, *11*(4), 385.

Waschke, L., Wöstmann, M., & Obleser, J. (2017). States and traits of neural irregularity in the age-varying human brain. *Scientific Reports*, *7*(1), 1-12.

White-Schwoch, T., Carr, K. W., Thompson, E. C., Anderson, S., Nicol, T., Bradlow, A. R., Zecker, S. G., & Kraus, N. (2015). Auditory processing in noise: A preschool biomarker for literacy. *PLoS Biology*, *13*(7).

Wilf, M., Ramot, M., Furman-Haran, E., Arzi, A., Levkovitz, Y., & Malach, R. (2016). Diminished auditory responses during NREM sleep correlate with the hierarchy of language processing. *PloS One*, *11*(6).

Willems, R. M., Van der Haegen, L., Fisher, S. E., & Francks, C. (2014). On the other hand: including left-handers in cognitive neuroscience and neurogenetics. *Nature Reviews Neuroscience*, *15*(3), 193-201.

Wilsch, A., Neuling, T., Obleser, J., & Herrmann, C. S. (2018). Transcranial alternating current stimulation with speech envelopes modulates speech comprehension. *NeuroImage*, *172*, 766-774.

Wong, P. C., Jin, J. X., Gunasekera, G. M., Abel, R., Lee, E. R., & Dhar, S. (2009). Aging and cortical mechanisms of speech perception in noise. Neuropsychologia, 47(3), 693-703.

Wong, P. C., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, *10*(4), 420-422.

Worden, F. G., & Marsh, J. T. (1968). Frequency-following (microphonic-like) neural responses evoked by sound. *Electroencephalography and Clinical Neurophysiology*, *25*(1), 42-52.

Zatorre, R. J. (1988). Pitch perception of complex tones and human temporal-lobe function. *The Journal of the Acoustical Society of America*, *84*(2), 566-572.

Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex*, *11*(10), 946-953.

Zeng, F. G., Nie, K., Stickney, G. S., Kong, Y. Y., Vongphoe, M., Bhargave, A. et al. (2005). Speech recognition with amplitude and frequency modulations. *Proceedings of the National Academy of Sciences*, *102*(7), 2293-2298.

Zhong, Z., Henry, K. S., & Heinz, M. G. (2014). Sensorineural hearing loss amplifies neural coding of envelope information in the central auditory system of chinchillas. *Hearing Research*, *309*, 55-62.

Zoefel, B., Allard, I., Anil, M., & Davis, M. H. (2020). Perception of rhythmic speech is modulated by focal bilateral transcranial alternating current stimulation. *Journal of Cognitive Neuroscience*, *32*(2), 226-240.

Zoefel, B., Archer-Boyd, A., & Davis, M. H. (2018). Phase entrainment of brain oscillations causally modulates neural responses to intelligible speech. *Current Biology*, *28*(3), 401-408.

Zoefel, B., & VanRullen, R. (2017). Oscillatory mechanisms of stimulus processing and selection in the visual and auditory systems: state-of-the-art, speculations and suggestions. *Frontiers in Neuroscience*, *11*, 296.

Appendices

Appendix 1. Results for linear mixed-effect regressions with PTAs as covariates (Chapter 2)

Linear mixed-effect regressions were conducted for the EEG signatures (FFR_{ENV_F0}, FFR_{PLV_F0}, FFR_{TFS_H2}, FFR_{TFS_F2F3}, Logit-theta PLV and PDCs) using Noise Type (Quiet, SpN and BbN) and Age Group (young vs. older) as fixed-effect factors, Participant as random-effect factors and PTA (PTA_{Low}, PTA_{High} or PTA_{Wide}) as fixed-effect covariates. **Tables A1.1**, **A1.2** and **A1.3** summarize the statistics using PTA_{Low}, PTA_{High} and PTA_{Wide} as the covariate, respectively. The type of covariance matrix that was chosen was the one that generated the smallest BIC value (Wang et al., 2007).

Table A1.1 Statistical results of linear mixed-effect regressions with PTA_{Low} as the fixed-effect covariate. DVs, df, F, and p refer to the dependent variables, degrees of freedom, F values, and p values, respectively. Significant p values (< 0.05) are indicated in bold. *p < 0.05; **p < 0.01; ***p < 0.001.

DVs	Fixed-effect factors/covariate	*df1*	*df2*	*F*	*p*
FFR_{ENV_F0}	Noise Type	2	26.059	41.418	**< 10⁻⁸*****
	Age Group	1	28.214	8.038	**0.008****
	PTA_{Low}	1	28.214	11.765	**0.002****
	Noise Type × Age Group	2	26.059	2.182	0.133
	Noise Type × PTA_{Low}	2	26.059	3.015	0.066
FFR_{PLV_F0}	Noise Type	2	27.473	24.986	**< 10⁻⁶*****
	Age Group	1	30.284	1.235	0.275
	PTA_{Low}	1	30.284	0.005	0.942
	Noise Type × Age Group	2	27.473	0.905	0.416
	Noise Type × PTA_{Low}	2	27.473	0.838	0.443
FFR_{TFS_H2}	Noise Type	2	52	0.477	0.624
	Age Group	1	26	0.041	0.840
	PTA_{Low}	1	26	0.050	0.824
	Noise Type × Age Group	2	52	2.698	0.077
	Noise Type × PTA_{Low}	2	52	2.839	0.068

FFR_{TFS_F2F3}	Noise Type	2	52	1.493	0.234
	Age Group	1	26	2.035	0.166
	PTA_{Low}	1	26	0.192	0.665
	Noise Type × Age Group	2	52	0.636	0.533
	Noise Type × PTA_{Low}	2	52	1.954	0.152
Logit-theta PLV	Noise Type	2	34.660	33.107	**$< 10^{-8}$*****
	Age Group	1	32.467	4.793	**0.036***
	PTA_{Low}	1	32.467	1.445	0.238
	Noise Type × Age Group	2	34.660	0.497	0.612
	Noise Type × PTA_{Low}	2	34.660	0.208	0.813
$PDC_{Subcort \to Cort}$	Noise Type	2	26.430	4.827	**0.016***
	Age Group	1	26.569	0.126	0.726
	PTA_{Low}	1	26.569	0.451	0.508
	Noise Type × Age Group	2	26.430	2.456	0.105
	Noise Type × PTA_{Low}	2	26.430	3.678	**0.039***
$PDC_{Cort \to Subcort}$	Noise Type	2	29.650	3.619	**0.039***
	Age Group	1	26.100	0.603	0.444
	PTA_{Low}	1	26.100	2.719	0.111
	Noise Type × Age Group	2	29.650	0.569	0.572
	Noise Type × PTA_{Low}	2	29.650	2.395	0.109

^Although [Noise Type × Age Group] and [Noise Type × PTA_{Low}] interactions were found, post-hoc analyses did not show significant effects of Age Group or PTA_{Low} in any level of Noise Type (all $p > 0.2$).

Table A1.2 Statistical results of linear mixed-effect regressions with PTA_{High} as the fixed-effect covariate. DVs, df, F, and p refer to the dependent variables, degrees of freedom, F values, and p values, respectively. Significant p values (< 0.05) are indicated in bold. *$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

DVs	Fixed-effect factors/covariate	*df1*	*df2*	*F*	*p*
FFR_{ENV_F0}	Noise Type	2	52	52.028	**$< 10^{-12}$*****

	Age Group	1	26	< 0.001	0.978
	PTA_{High}	1	26	0.186	0.670
	Noise Type × Age Group	2	52	0.524	0.595
	Noise Type × PTA_{High}	2	52	1.704	0.192
FFR_{PLV_F0}	Noise Type	2	27.946	24.563	**$< 10^{-6}$*****
	Age Group	1	30.421	0.472	0.497
	PTA_{High}	1	30.421	0.197	0.660
	Noise Type × Age Group	2	27.946	0.844	0.440
	Noise Type × PTA_{High}	2	27.946	0.253	0.778
FFR_{TFS_H2}	Noise Type	2	52	0.381	0.685
	Age Group	1	26	0.247	0.623
	PTA_{High}	1	26	0.348	0.560
	Noise Type × Age Group	2	52	1.852	0.167
	Noise Type × PTA_{High}	2	52	0.492	0.614
FFR_{TFS_F2F3}	Noise Type	2	52	1.564	0.219
	Age Group	1	26	2.678	0.114
	PTA_{High}	1	26	1.019	0.322
	Noise Type × Age Group	2	52	0.059	0.943
	Noise Type × PTA_{High}	2	52	0.030	0.970
Logit-theta PLV	Noise Type	2	35.513	35.525	**$< 10^{-8}$*****
	Age Group	1	32.812	6.520	**0.016***
	PTA_{High}	1	32.812	0.019	0.892
	Noise Type × Age Group	2	35.513	0.695	0.506
	Noise Type × PTA_{High}	2	35.513	3.042	0.060
$PDC_{Subcort \rightarrow Cort}$	Noise Type	2	52	4.740	**0.013***
	Age Group	1	26	0.022	0.883
	PTA_{High}	1	26	0.005	0.945
	Noise Type × Age Group	2	52	0.252	0.778
	Noise Type × PTA_{High}	2	52	1.724	0.188
$PDC_{Cort \rightarrow Subcort}$	Noise Type	2	52	4.726	**0.013***

	Age Group	1	26	0.534	0.472
	PTA_{High}	1	26	0.313	0.580
	Noise Type × Age Group	2	52	0.112	0.894
	Noise Type × PTA_{High}	2	52	0.304	0.739

Table A1.3 Statistical results of linear mixed-effect regressoins with PTA_{Wide} as the fixed-effect covariate. DVs, df, F, and *p* refer to the dependent variables, degrees of freedom, F values, and *p* values, respectively. Significant *p* values (< 0.05) are indicated in bold. *p < 0.05; **p < 0.01; ***p < 0.001.

DVs	Fixed-effect factors/covariate	*df1*	*df2*	*F*	*p*
FFR_{ENV_F0}	Noise Type	2	52	52.205	**< 10^{-12}***
	Age Group	1	26	1.278	0.269
	PTA_{Wide}	1	26	0.967	0.334
	Noise Type × Age Group	2	52	0.899	0.413
	Noise Type × PTA_{Wide}	2	52	1.960	0.151
FFR_{PLV_F0}	Noise Type	2	28	24.661	**< 10^{-6}***
	Age Group	1	30.232	0.420	0.522
	PTA_{Wide}	1	30.232	0.144	0.707
	Noise Type × Age Group	2	28	0.424	0.658
	Noise Type × PTA_{Wide}	2	28	0.046	0.955
FFR_{TFS_H2}	Noise Type	2	52	0.488	0.617
	Age Group	1	26	0.242	0.627
	PTA_{Wide}	1	26	0.315	0.580
	Noise Type × Age Group	2	52	2.313	0.109
	Noise Type × PTA_{Wide}	2	52	1.421	0.251
FFR_{TFS_F2F3}	Noise Type	2	52	1.450	0.244
	Age Group	1	26	2.777	0.108
	PTA_{Wide}	1	26	0.970	0.334
	Noise Type × Age Group	2	52	0.080	0.923
	Noise Type × PTA_{Wide}	2	52	0.265	0.768
Logit-theta PLV	Noise Type	2	35.491	34.257	**< 10^{-8}***
	Age Group	1	32.368	3.753	0.061

	PTA_{Wide}	1	32.368	0.402	0.530
	Noise Type × Age Group	2	35.491	0.268	0.766
	Noise Type × PTA_{Wide}	2	35.491	1.542	0.228
PDC_{Subcort→Cort}	Noise Type	2	26.002	3.998	**0.031***
	Age Group	1	27.018	0.011	0.916
	PTA_{Wide}	1	27.018	0.058	0.812
	Noise Type × Age Group	2	26.002	0.465	0.633
	Noise Type × PTA_{Wide}	2	26.002	1.550	0.231
PDC_{Cort→Subcort}	Noise Type	2	52	5.018	**0.010***
	Age Group	1	26	0.005	0.944
	PTA_{Wide}	1	26	0.065	0.801
	Noise Type × Age Group	2	52	0.321	0.727
	Noise Type × PTA_{Wide}	2	52	1.050	0.357

Appendix 2. Simulations that test relative reliability of FFR measurements (Chapter 2)

Simulations were conducted to test which measurements – the FFR magnitudes, or response SNRs (difference in magnitudes between FFRs and the EEG noise floors) – can more reliably quantify FFRs.

Appendix 2.1 Methods

Simulated FFRs and EEG background noise were created. FFRs were created based on the stimulus syllable /i/ used in the present study: FFR_{ENV} used the Hilbert Envelope of the syllable, while FFR_{TFS} used the syllable *per se*. EEG background noise was pink noise (with random phases) that fits the 1/f power law of EEG. This artificially created 'real' FFRs (FFRs before adding EEG noise) and 'observed' FFRs (FFRs after adding EEG noise). The 'observed' FFR magnitudes and 'observed' SNRs (difference in magnitudes between 'observed' FFRs and EEG noise floors) were then measured in order to investigate which one can better reflect the 'real' FFR magnitudes. Measurements of the 'real' and 'observed' FFR magnitudes followed the same procedure as described in *Chapter 2*. EEG noise floors were measured as EEG noise magnitudes at the corresponding frequency range of FFRs (110 ~ 160 Hz (F_0) for FFR_{ENV_F0}; 220 ~ 320 Hz (H2) for FFR_{TFS_H2}; 150-Hz bandwidth centred at 2400 Hz (F2) and 300-Hz

bandwidth centred at 3100 Hz (F3) for FFR_{TFS_F2F3}) at the 50-ms FFR pre-stimulus period (see *Chapter 2*; also see Schoof & Rosen, 2016; Mai et al., 2018).

100 pairs of FFRs and EEG noise with different magnitudes were created and were made sure that they were in line with the characteristics of actual data in the present experiment: (i) the 'observed' FFRs had significantly greater magnitudes than the noise floors (see *2.4.5*); (ii) average magnitudes of the 'observed' FFRs and the noise floors approximated those of the actual data in the experiment (averaged across all participants) (see **Table A2.1**). The 'real' FFR magnitudes were then correlated with the 'observed' FFR magnitudes and 'observed' SNRs, where higher correlation values indicate better reliability of measurements. Such simulations were repeated 20 times and pair-wise t-tests were finally conducted to compare the two correlations (correlations between the 'real' FFR magnitudes and 'observed' FFR magnitudes vs. correlations between the 'real' FFR magnitudes and 'observed' SNRs; Fisher-transformed) for each of the three FFR signatures (FFR_{ENV_F0}, FFR_{TFS_H2} and FFR_{TFS_F2F3}) under each noise type (Quiet, SpN and BbN).

Table A2.1. Average magnitudes of the observed FFRs, EEG noise floors and SNRs of the actual and simulated data (numbers in the brackets indicate the simulated data) for the three FFR signatures (FFR_{ENV_F0}, FFR_{TFS_H2} and FFR_{TFS_F2F3}) under different noise types (Quiet, SpN and BbN). As shown, the average simulated magnitudes of the 'observed' FFRs, noise floors and SNRs approximate those of the actual data in the experiment.

FFR signatures	Noise types	Observed FFR magnitudes	Magnitudes of EEG noise floors	SNRs
FFR_{ENV_F0}	Quiet	4.771 (4.770)	3.022 (3.041)	1.749 (1.729)
	SpN	3.697 (3.679)	3.062 (3.055)	0.635 (0.624)
	BbN	3.752 (3.762)	3.222 (3.220)	0.530 (0.542)
FFR_{TFS_H2}	Quiet	4.164 (4.160)	2.362 (2.327)	1.802 (1.834)
	SpN	4.170 (4.180)	2.270 (2.313)	1.901 (1.867)
	BbN	4.127 (4.132)	2.329 (2.353)	1.797 (1.779)
FFR_{TFS_F2F3}	Quiet	-30.632 (-30.694)	-37.884 (-37.841)	7.252 (7.147)
	SpN	-30.341 (-30.512)	-38.192 (-38.143)	7.852 (7.631)
	BbN	-30.389 (-30.313)	-37.730 (-37.784)	7.341 (7.471)

Appendix 2.2 Results

Results showed that correlations between the 'real' FFR magnitudes and 'observed' FFR magnitudes were significantly higher than correlations between the 'real' FFR magnitudes and

'observed' SNRs for all three FFR signatures in all noise types ($FFR_{ENV\_F0\_Quiet}$, $p < 10^{-10}$; $FFR_{ENV\_F0\_SpN}$, $p < 10^{-5}$; $FFR_{ENV\_F0\_BbN}$, $p < 10^{-7}$; $FFR_{TFS\_H2\_Quiet}$, $p < 10^{-12}$; $FFR_{TFS\_H2\_SpN}$, $p < 10^{-9}$; $FFR_{TFS\_H2\_BbN}$, $p < 10^{-10}$; $FFR_{TFS\_F2F3\_Quiet}$, $p < 10^{-11}$; $FFR_{TFS\_F2F3\_SpN}$, $p < 10^{-15}$; $FFR_{TFS\_F2F3\_BbN}$, $p < 10^{-13}$) (**Figure A1.1**). This therefore indicates that, compared to the response SNR, the observed FFR magnitude should more reliably quantify the real FFR magnitude in the present study.
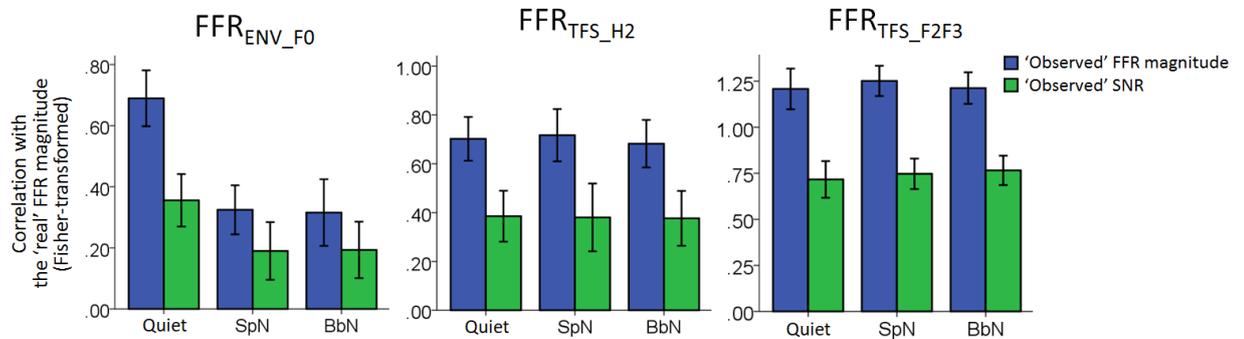
Figure A1.1. Relative reliability of FFR measurements. *The 'real' FFR magnitudes were correlated (Fisher-transformed) with the 'observed' FFR magnitudes (blue) and the 'observed' SNRs (green) for the three FFR signatures (FFR_{ENV_F0}, FFR_{TFS_H2} and FFR_{TFS_F2F3}) under each noise type (Quiet, SpN and BbN). Error bars denote the standard deviations across the 20 simulations.*

Appendix 3. Written descriptions for scales of sleepiness (Chapter 3)

There were 7 subjective scales of sleepiness with the following written descriptions:

[Scale 1]: I was always awake across the entire experiment.

[Scale 2]: I fell asleep occasionally, but was awake most of the time during the experiment.

[Scale 3]: I slept for substantial amount of time during the experiment, but was awake for more than half of the time.

[Scale 4]: I slept for about half of the time during the experiment.

[Scale 5]: I slept for more than half of the time during the experiment.

[Scale 6]: I was sleeping most of the time during the experiment and only awake occasionally.

[Scale 7]: I have been deeply sleeping across the entire experiment.

Appendix 4. Results for linear mixed-effect regressions with age-related factors as covariates (Chapter 3)

Age was used as one covariate and Spindle Density (number of sleep spindles per minute across the epochs of low arousal state) as the other covariate in the linear mixed regression analyses for Logit-theta PLV in the report. Similar analyses were conducted using other age-related factors as covariates that replaced the Age covariate. The two replacement age-related covariates were pure-tone audiometric threshold (PTA) averaged across 0.25 to 4 kHz over both ears, and sleep spindle duration; both covariates were mean-centred. Together with the analysis that used Age as a covariate, these analyses tested whether the effect of Age resulted from age-related changes in peripheral hearing (PTA) or spindle duration.

The [Arousal × PTA × Spindle Density] (**Table A3.1**) and the [Arousal × Spindle Duration × Spindle Density] (**Table A3.2**) interactions were not significant. Hence there was no evidence that the interaction between Age, Arousal and Spindle Density resulted from age-related changes in PTA or Spindle Duration. This contrasted with the analysis which used age as covariate where there was a significant three-way [Arousal × Age × Spindle Density] interaction for Logit-theta PLV (see *Chapter 3*).

Table A3.1 Linear mixed regression for Logit-theta PLV, using Arousal as the fixed-effect factor, PTA and Spindle Density as fixed-effect covariates, and Participant as the random-effect factor. *DV*, *Df, F,* and *p* refer to dependent variable, degrees of freedom, *F*-values, and *p*-values, respectively. Significant *p*-values are in bold. $^*p < 0.05$, $^{***}p < 0.001$.

DV	Fixed-effect factors/covariate	df1	df2	F	p
Logit-theta PLV	Arousal	1	87	5.740	**0.019***
	PTA	1	87	16.199	**< 0.001***
	Spindle Density	1	87	0.126	0.723
	Arousal × PTA × Spindle Density	1	87	2.206	0.141
	Arousal × PTA	1	87	0.003	0.957
	Arousal × Spindle Density	1	87	1.647	0.203
	PTA × Spindle Density	1	87	0.862	0.356

Table A3.2 Linear mixed regression for Logit-θ-PLV, using Arousal as the fixed-effect factor, Spindle Duration and Spindle Density as fixed-effect covariates, and Participant as the random-effect factor. *DV*, *Df, F,* and *p* refer to dependent variable, degrees of freedom, *F*-values, and *p*-values, respectively. Significant *p*-values are in bold. *$p < 0.05$, **$p < 0.01$.

DV	Fixed-effect factors/covariate	df1	df2	F	p
Logit-theta PLV	Arousal	1	87	5.593	**0.020***
	Spindle Duration	1	87	7.402	**0.008****
	Spindle Density	1	87	3.209	0.077
	Arousal × Spindle Duration × Spindle Density	1	87	0.037	0.847
	Arousal × Spindle Duration	1	87	2.118	0.149
	Arousal × Spindle Density	1	87	6.209	**0.015***
	Spindle Duration × Spindle Density	1	87	0.810	0.370