



# The possible role of early-stage phase-locked neural activities in speech-in-noise perception in human adults across age and hearing loss



Guangting Mai <sup>a,b,c,\*</sup>, Peter Howell <sup>c</sup>

<sup>a</sup> National Institute for Health Research Nottingham Biomedical Research Centre, Nottingham NG1 5DU, UK

<sup>b</sup> Academic Unit of Mental Health and Clinical Neurosciences, School of Medicine, University of Nottingham, Nottingham NG7 2UH, UK

<sup>c</sup> Department of Experimental Psychology, University College London, London WC1H 0AP, UK

## ARTICLE INFO

### Article history:

Received 21 March 2022

Revised 26 October 2022

Accepted 4 November 2022

Available online 10 November 2022

### Keywords:

Speech-in-noise (SiN) perception

Frequency-following response (FFR)

Theta-band phase-locking values ( $\theta$ -PLV)

Age and hearing loss

## ABSTRACT

Ageing affects auditory neural phase-locked activities which could increase the challenges experienced during speech-in-noise (SiN) perception by older adults. However, evidence for how ageing affects SiN perception *through* these phase-locked activities is still lacking. It is also unclear whether influences of ageing on phase-locked activities in response to different acoustic properties have similar or different mechanisms to affect SiN perception. The present study addressed these issues by measuring early-stage phase-locked encoding of speech under quiet and noisy backgrounds (speech-shaped noise (SSN) and multi-talker babbles) in adults across a wide age range (19–75 years old). Participants passively listened to a repeated vowel whilst the frequency-following response (FFR) to fundamental frequency that has primary subcortical sources and cortical phase-locked response to slowly-fluctuating acoustic envelopes were recorded. We studied how these activities are affected by age and age-related hearing loss and how they are related to SiN performances (word recognition in sentences in noise). First, we found that the effects of age and hearing loss differ for the FFR and slow-envelope phase-locking. FFR was significantly decreased with age and high-frequency ( $\geq 2$  kHz) hearing loss but increased with low-frequency ( $< 2$  kHz) hearing loss, whilst the slow-envelope phase-locking was significantly increased with age and hearing loss across frequencies. Second, potential relationships between the types of phase-locked activities and SiN perception performances were also different. We found that the FFR and slow-envelope phase-locking positively corresponded to SiN performance under multi-talker babbles and SSN, respectively. Finally, we investigated how age and hearing loss affected SiN perception through phase-locked activities via mediation analyses. We showed that both types of activities significantly mediated the relation between age/hearing loss and SiN perception but in distinct manners. Specifically, FFR decreased with age and high-frequency hearing loss which in turn contributed to poorer SiN performance but increased with low-frequency hearing loss which in turn contributed to better SiN performance under multi-talker babbles. Slow-envelope phase-locking increased with age and hearing loss which in turn contributed to better SiN performance under both SSN and multi-talker babbles. Taken together, the present study provided evidence for distinct neural mechanisms of early-stage auditory phase-locked encoding of different acoustic properties through which ageing affects SiN perception.

© 2022 The Author(s). Published by Elsevier B.V.  
This is an open access article under the CC BY-NC-ND license  
(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

## 1. Introduction

Older adults often experience increased difficulty with speech-in-noise (SiN) perception (Hume and Dubno, 2010). It has been suggested that the degraded neuro-temporal encoding of speech sounds in older adults could be a reason for such difficulty

**Abbreviations:** EEG, electroencephalography; SiN, speech-in-noise; FFR, frequency-following response;  $\theta$ -PLV, theta-band phase-locking value; PTAT, pure-tone audiometric threshold; PCA, principal component analysis; SRT, speech reception threshold; SSN, speech-shaped noise; SNR, signal-to-noise ratio; nREM, non-rapid eye movement.

\* Corresponding author at: Hearing Theme, NIHR Nottingham Biomedical Research Centre, 113 The Ropewalk, Ropewalk House, Nottingham NG1 5DU, United Kingdom.

E-mail address: [guangting.mai@nottingham.ac.uk](mailto:guangting.mai@nottingham.ac.uk) (G. Mai).

(Anderson et al., 2012). Phase-locked activity, which reflects the brain's ability to synchronize its neural activity with input stimuli, informs how neuro-temporal encoding is achieved (Henry and Heinz, 2012; Gransier et al., 2021). Therefore, an important research question is how phase-locked activity plays a role to affect SiN perception according to ageing. Currently, there are two important types of phase-locked activities in response to speech: (1) phase-locked response to speech fundamental frequency ( $F_0$ ) termed frequency-following response (FFR) that has primary sources at the subcortex (Coffey et al., 2019); and (2) cortical phase-locked response to slowly-fluctuating speech envelopes (with rates especially at theta range) (Howard and Poeppel, 2010; Peelle et al., 2013).

FFR is an important neural signature for temporal encoding of  $F_0$  cues of speech that originates primarily from the subcortex (Aiken and Picton, 2008; Chandrasekaran and Kraus, 2010; Bidelman, 2018; Coffey et al., 2019) as well as the primary auditory cortex (Coffey et al., 2016, 2019; Hartmann and Weisz, 2019; Ross et al., 2020). Previous studies showed that greater FFR magnitudes correspond to better SiN perception in older adults (Anderson et al., 2011; Fujihira and Shiraishi, 2015; Mai et al., 2018). It has been also shown that older adults have smaller FFR magnitudes than young adults (Anderson et al., 2012; Presacco et al., 2016). These findings thus suggest that degraded FFR may be an indicator to impaired SiN perception during ageing. In addition to FFR, phase-locked response to speech at theta rates (4–8 Hz), which reflects cortical tracking of slowly-fluctuating envelopes, is also an important neural index for speech perception (Luo and Poeppel, 2007; Peelle et al., 2013). In contrast to FFR, this cortical response increases with age (Presacco et al., 2016, 2019), consistent with findings showing enhanced cortical phase-locking to amplitude-modulated tones following ageing (Tlumak et al., 2015; Goossens et al., 2016). Auditory cortical phase-locking is related to neural excitability in the auditory cortex (Ng et al., 2013; Oya et al., 2018), indicating that such increased phase-locking could reflect hyperexcitability possibly related to reduction in neural inhibition in the auditory cortex following ageing (Goossens et al., 2016). It was also argued that such excessive cortical phase-locking in older adults may reflect a loss of balance between inhibitory and excitatory neural processes (Presacco et al., 2016). This could impair network connectivity and over-represent speech envelopes relative to other speech features which may in turn impair SiN perception (Presacco et al., 2016).

Despite the reported effects of ageing on phase-locked activities and arguments for how these effects may influence SiN perception, there has not been direct evidence that links such effects with behavioral data during SiN perception, i.e., how ageing may affect SiN perception through phase-locked activities. Recent studies had attempted to address this issue. For example, Presacco et al. (2016) and Schoof and Rosen (2016) tested the relationship between FFR and SiN performances (word recognition in sentences in noise) in normal-hearing (audiometric thresholds  $\leq 25$  dB at frequencies  $\leq 4$  kHz) adults across different ages. Both studies showed declined FFR and SiN performances in older ( $> 60$  years) compared to young ( $< 30$  years) adults. However, neither showed statistical association between FFR and SiN perception. Presacco et al. (2016) further investigated cortical phase-locking to speech envelopes which increased in older compared to young adults. However, the authors had not found a correlation between cortical phase-locking and SiN perception. This may be because different types of background noise were used when measuring neural data (single-talker background) and SiN perception tasks (four-talker babbles) (Presacco et al., 2016). Moreover, older participants in these studies were normal-hearing, so their audiometric status had not reflected the real-world scenario in typical ageing populations who have great inter-individual variability of hearing

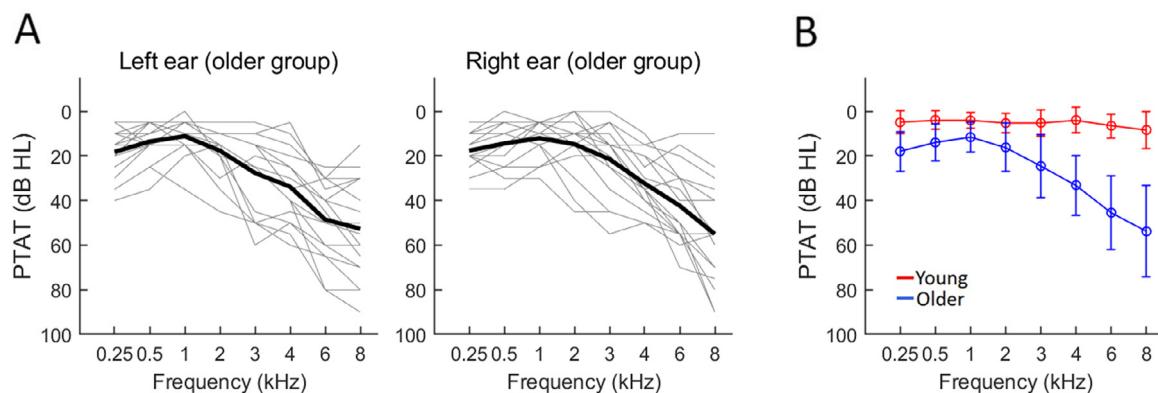
(Gopinath et al., 2009; Humes et al., 2010).

A more recent study by Goossens et al. (2018) investigated the relationship between neural phase-locked responses to non-speech amplitude modulations (subcortical and cortical auditory steady-state responses (ASSR)) and SiN performances in normal-hearing and hearing-impaired adults across a wide age range (20–80 years old). Subcortical and cortical ASSRs were responses to amplitude modulations at 80 Hz that approximated the  $F_0$  range at 4 Hz that approximated the range for slowly-fluctuating envelopes, respectively. It was found that greater ASSRs predicted poorer SiN performance across hearing and age groups. Greater subcortical ASSR may reflect excessive encoding of envelope cues relative to other important acoustic features like temporal fine structures. On the other hand, because the background noise during the SiN perception tasks had similar acoustic features as the stimuli used for cortical measurements, greater cortical ASSR may correspond to greater encoding of background noise during SiN perception associated with poorer performances. Therefore, this study had not elucidated how measurements that reflect encoding of target speech are related to SiN perception. More importantly, the question for how effects of ageing on phase-locked activities contribute to SiN perception is still unclear.

In summary of the discussions above, the following issues are yet to be addressed: (1) how effects of ageing on phase-locked activities may contribute to SiN perception; in other words, how ageing affects SiN perception through these activities; and (2) whether mechanisms by which ageing affects phase-locked activities as well as affects SiN perception through these activities are similar or different for encoding of  $F_0$  and slowly-fluctuating envelope cues. The aim of the present study was thus to address these issues. More specifically, we aimed to study how ageing affects SiN perception through the early, sensory stages of subcortical and cortical phase-locked encoding of speech. This was because whilst it is already noticeable that declines in high-level cognitive processing following ageing deteriorate SiN perception (see a review by Slade et al., 2020), we wished to focus on elucidating how relatively low-level, early-stage neural encoding may contribute to this process. It is noteworthy that 'ageing' here refers to the effects that include not only the effect of increased age itself, but also hearing loss accompanied by ageing. The present study recruited adult participants who spanned a wide age range (19–75 years old) without deliberately controlling for hearing loss. Participants who were  $> 50$  years old had audiometric thresholds ranging from normal hearing to various degrees of hearing loss (see *Methods* for details). This reflected the great inter-individual variability of peripheral hearing in the real-life, ecological scenarios during normal ageing (Gopinath et al., 2009; Humes et al., 2010)<sup>1</sup> as opposed to some previous studies that recruited older participants with normal hearing (Anderson et al., 2011; Fujihira and Shiraishi, 2015; Presacco et al., 2016; Schoof and Rosen, 2016).

Both neural and behavioral assessments were conducted in the present study. For the neural assessments, participants' phase-locked activities (FFR and the cortical phase-locked response) were measured over the scalp using electroencephalography (EEG) when they passively listened to a single vowel /i/ repeated at  $\sim 5$  Hz (hence resembling the rate of slowly-fluctuating envelopes in natural speech) under quiet and two types of noisy backgrounds (speech-shaped Gaussian noise (SSN) and multi-talker babbles).

<sup>1</sup> Our subsequent analyses employed principal component analysis (PCA) on age and hearing loss that largely separated the effects of age/high-frequency ( $\geq 2$  kHz) hearing loss and low-frequency ( $< 2$  kHz) hearing loss (but could not separate age from high-frequency hearing loss; see *Methods* and *Results* for details). As such, we acknowledge that 'ageing' is a process with combined effects of increased age and hearing loss and the present study had not specifically tried to distinguish the two from each other in a strict manner.



**Fig. 1.** Pure-tone audiograms. (A) The thin lines represent individual audiograms for the older group in both ears ranging from 0.25 to 8 kHz. The bold lines are the grand averages across participants. Three older participants had thresholds 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 90 dB. (B) Audiograms for the young (red) and older (blue) group averaged across participants. Error bars represent the standard deviations.

The lack of semantic/syntactic information of the stimuli and the passive listening paradigm meant that the phase-locked activities largely reflect the early stages of processing, in line with our current emphasis as mentioned above. For the behavioral assessments, participants completed word recognition tasks when they listened to sentences under the same types of background noise as in the neural assessments. Consequently, we investigated how phase-locked activities differed across age and hearing loss. We then tested how these activities were correlated with behavioral performances of SiN perception. Finally, to further elucidate the possible role of these phase-locked activities played for SiN perception during ageing, we studied whether and how age and hearing loss affect SiN performances *through* phase-locked activities via mediation analyses. Taken together, we expected to provide evidence for the early-stage neural phase-locked mechanisms that underlie how ageing affects SiN perception.

## 2. Methods

The present study was approved by the University College London (UCL) Research Ethics Committee and informed consents were obtained from all participants. The study followed the same experimental procedure and used parts of the data from Mai et al. (2018) (see 2.1 for details).

### 2.1. Participants

Participants comprised 41 adult adults (22 males) aged between 19 and 75 years old. One female participant's data were excluded due to technical errors detected retrospectively through examinations of the EEG signals (the reference electrodes were unintentionally detached during neural assessments). Thus, 40 participants were eventually included. All participants were native British English speakers with no reports of neurological diseases, language-related or psychiatric problems. For the illustrative purpose, we split the participants into a young (23 participants aged 19–42 years) and an older group (17 participants aged 53–75 years) (see Figs. 1, 5 and 6). For the subsequent statistical analyses, however, we used age as a continuous rather than diatomic parameter (see 2.5). Data of the older group were from our previous study (the non-hearing-aid participants in Group 2 of Mai et al., 2018). These data were used as we found significant correlations between phase-locked activities and SiN performances in this group of participants (see Mai et al., 2018). The young participants were subsequently recruited and they completed the same neural and behavioural tasks with the same experimental setups as for the older group.

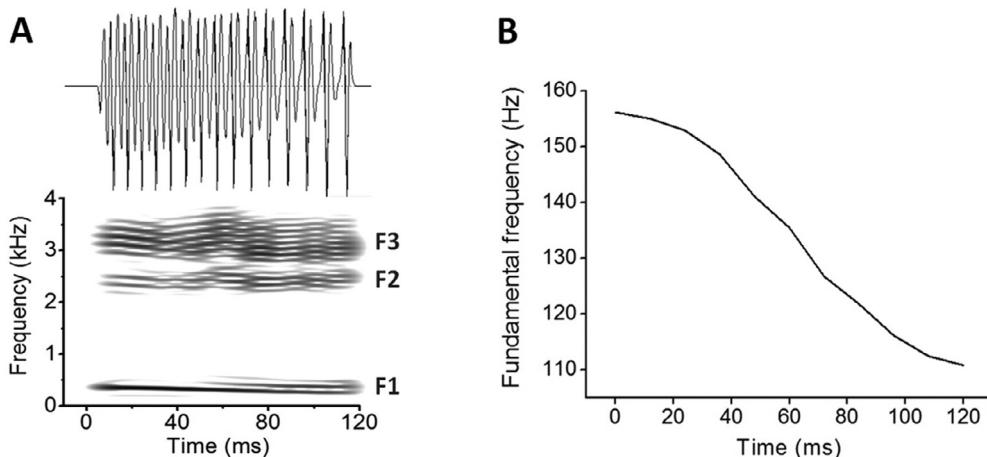
Fig. 1 shows pure-tone audiometric thresholds (PTAT) for frequencies 0.25–8 kHz measured via air-conduction using an MA41 Audiometer (MAICO Diagnostics, Germany). In the young group, all participants had normal hearing ( $PTAT \leq 25$  dB HL; see the grand-averaged curve in Fig. 1B). In the older group, inter-individual variability was high particularly at high frequencies ( $\geq 2$  kHz; ranging from normal hearing to mild-to-severe hearing loss, see individual curves in Fig. 1A), consistent with the ecological distribution pattern of older populations during normal ageing (Gopinath et al., 2009; Humes et al., 2010).

## 2.2. Neural experiment

### 2.2.1. Acoustic stimuli

Participants listened to a repeatedly-presented, 120-ms-long vowel /i/ produced by a male speaker (Fig. 2A). The  $F_0$  contour of the vowel fell from  $\sim 160$  to  $\sim 110$  Hz (Fig. 2B). The  $F_0$  contour covered a similar frequency range and direction of change as those in the  $F_0$ s 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; see 2.3). The three formants in the vowel were at  $\sim 280$  Hz (F1),  $\sim 2400$  Hz (F2) and  $\sim 3100$  Hz (F3). The amplitude envelope profile was stable across the stimulus period with 5-ms-long rising and falling cosine windows applied at the onset and offset to avoid transients.

The vowel was presented repeatedly at both original and inverted (i.e., all positive values become negative and all negative values become positive) polarities in a random order with inter-stimulus intervals that varied randomly between 60 and 120 ms (hence repetition rate was approximately 5 vowels per second). The stimuli were presented under quiet and two types of background noise: steady-state speech-shaped noise (SSN) and 16-talker babble noise. The babble noise was a mixture of utterances spoken by 16 male British English speakers with similar  $F_0$  range as the target stimulus (ranging from  $\sim 80$  to  $\sim 200$  Hz). SSN was formed by randomizing the phases of the long-term spectrum of the babble noise and transforming the spectrum back to the time domain. As a result, SSN has the same long-term power spectrum as the babble noise and stable time-domain properties (Rosen et al., 2013). The signal-to-noise ratios (SNRs) were set at  $-1$  dB, which was shown to lead to neural responses that correlated significantly with SiN performances 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



**Fig. 2.** The vowel /i/ used for the neural measurement. (A) The temporal waveform (top) and spectrogram (bottom) of the vowel. F1, F2 and F3 frequencies are around 280, 2400 and 3100 Hz, respectively. (B) The falling F0 contour ranging from around 160 to 110 Hz obtained by autocorrelation. The waveform, spectrogram and F0 were generated using PRAAT (Boersma and Weenink, 2013).

in total with 400 sweeps per segment. The segments were played in succession in an intermixed order.

#### 2.2.2. EEG data acquisition

EEGs were recorded over participants' scalps via an ActiveTwo system (Biosemi, The Netherlands) at a sampling rate of 16,384 Hz. Three active electrodes were placed at Cz (vertex), C3 and C4 according to the 10/20 configuration with the help of a standard Biosemi EEG cap<sup>2</sup>. FFRs were obtained at Cz which is the conventional site to obtain robust brainstem signals (Skoe and Kraus, 2010). Cortical activities were measured at C3 and C4, the sites of which are used to obtain activity in the auditory cortex (Carpenter and Shahin, 2013; Noguchi et al., 2015). Furthermore, our previous study has already shown that reliable cortical phase-locked activity can be recorded on these sites and the activity is significantly associated with SiN performances (Mai et al., 2018). Bilateral earlobes were used as the reference. Ground electrodes were CMS/DRL. All electrode offsets were 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). A loudspeaker instead of inserted earphones (also for the behavioral experiment, see 2.3) was used to present stimuli because the present study used part of the data in Mai et al. (2018) (see 2.1) which recruited some hearing-aid users (whose data were not included here) who could not use inserted earphones. The stimulus level (measured across time including inter-stimulus intervals) at the distance between the loudspeaker and participants' ears (constant at 1 meter) was calibrated at 74.5 dB before background noise was added (i.e., under quiet background). The stimulus level was at 79.5 dB after either SSN or babble noise was added.

Participants were instructed to relax, close their eyes and keep still in order to attenuate movement artefacts. They did not have to make any response to the stimuli (i.e., passive listening). A web-cam monitored the participants throughout the test and no signif-

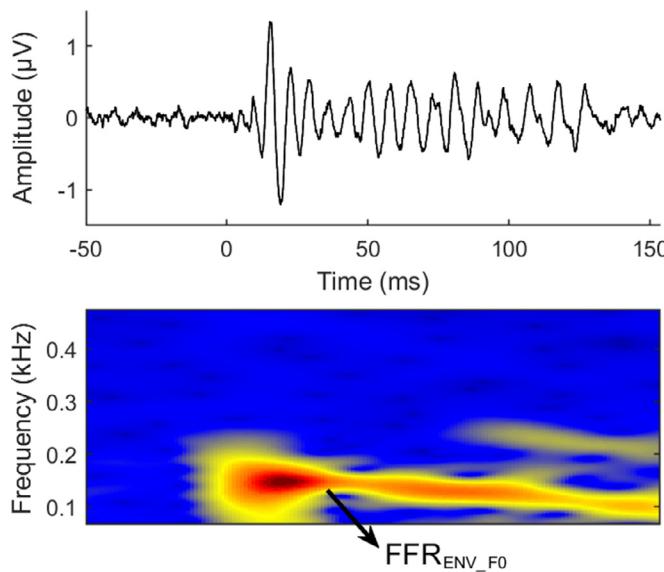
icant changes in head or body position were observed. The whole recording process lasted for ~70 min.

#### 2.3. Behavioral experiment

The behavioral experiments involved participants listening to BKB sentences (Bench et al., 1979) and recognizing words in the sentences under the types of background noise same as in the neural measurements (SSN and 16-talker babble noise). All sentences were pre-recorded utterances spoken by a male British English speaker whose F0 ranged from 80 to 200 Hz. Each sentence included three key (content) words, e.g., "The clown has a funny face" with key words of "clown", "funny" and "face". According to the features of SSN and the babble noise as described previously, there was an important difference between the tasks under these two types of background noise. That is, while both SSN and babble noise led to energetic masking, babble noise had an additional informational masking (Rosen et al., 2013). Such information masking mainly consisted of acoustic-phonetic interference due to the high number of talkers in the babble noise (i.e., 16) (hence there should be no lexical interference; see Hoen et al., 2007).

Participants were seated comfortably in the same sound-treated booth as in the neural experiment facing a Fostex 6301B loudspeaker (Canford Group Ltd.) at zero-degree azimuth. Same as in the neural experiment, distance between the loudspeaker and participants' ears was constant at 1 meter. All sentences (target sentence plus noise) were played at an intensity calibrated at 70 dB SPL at this distance. Participants listened to two different sets of 30 sentences (for the backgrounds of SSN and babble noise, respectively) after presentations of 8 practice sentences for each set. Participants were asked to orally report as many words as they could from each sentence. Sentences were presented via Matlab (Mathwork, USA) and SNR varied adaptively to track for the speech reception threshold (SRT, Plomp and Mimpel, 1979) at which 50% of key words were correctly reported. The 'loose key word scoring' approach was followed, meaning that a reported word was considered as correct as long as it matched the root of a key word (e.g., 'stand' was considered as correct for the key word 'stood') (MacLeod and Summerfield, 1990). For each background type, the first sentence was played at a relatively high SNR (8 and 10 dB for SSN and babble noise, 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

<sup>2</sup> More electrodes were not used because the study had not got consent from the recruited older participants to place the whole-head electrode set due to their general unwillingness to allow us to do so. The young participants were recruited subsequently and the same electrode configuration as for the older participants was followed.



**Fig. 3.** FFR<sub>ENV</sub> waveform (top) and spectrogram (bottom) of a participant recorded under quiet background. FFR<sub>ENV\_F0</sub> (at F<sub>0</sub> with a falling direction from 160 to 110 Hz) is indicated by the arrow. '0' in the x axis corresponds to the vowel onset.

sentences. The SRT was calculated as the mean SNR across the reversals at the step size of 2 dB (Schoof and Rosen, 2014). Lower SRT therefore represents better SiN performance.

#### 2.4. Signal processing for EEG data

The signal processing was conducted using Matlab 2019b (Mathwork, USA).

##### 2.4.1. 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  $\pm 25 \mu\text{V}$  were rejected to exclude movement artefacts. FFRs with original (FFR<sub>original</sub>) and inverted (FFR<sub>inverted</sub>) polarities were obtained by averaging across sweeps with their respective polarities. Here, we focused on FFRs that represent the neural responses to the envelope modulations (FFR<sub>ENV</sub>) by addition of FFR<sub>original</sub> and FFR<sub>inverted</sub> (which were then divided by 2; Aiken and Picton, 2008) but not temporal fine structures (TFS) (FFR<sub>TFS</sub>) via subtraction between FFR<sub>original</sub> and FFR<sub>inverted</sub>). We looked into FFR<sub>ENV</sub> rather than FFR<sub>TFS</sub>, because FFR<sub>ENV</sub> originates primarily from the subcortex (brainstem) that reflects neural encoding in the central auditory system that we focused on, whilst FFR<sub>TFS</sub> mainly reflects cochlear microphonics at the peripheral level (Aiken and Picton, 2008). In addition, subtraction of FFR<sub>original</sub> and FFR<sub>inverted</sub> (the way in which FFR<sub>TFS</sub> was obtained) could lead to electrical artefacts, while FFR<sub>ENV</sub> was obtained by adding the two polarities such that these artefacts are minimized (Skoe and Kraus, 2010). Fig. 3 shows an example of FFR<sub>ENV</sub> obtained in the present study (from a single participant recorded under the quiet background).

The FFR magnitude was then measured as the spectral magnitude along the F<sub>0</sub> trajectory (FFR<sub>ENV\_F0</sub>) that represents neural encoding of envelope modulations at F<sub>0</sub> (Fig. 3). The procedures followed our previous study (Mai et al., 2018), except that a more reliable approach for estimation of neural delay was used (Krizman and Kraus, 2019; Mai and Howell, 2022). The procedures were as follows:

- (1) Determine the neural delay. Specifically, both the vowel stimulus and FFR<sub>ENV</sub> waveforms were bandpass filtered at 110–160

Hz (the F<sub>0</sub> range) using a zero-phase 2nd-order Butterworth filter. The 100 ms pre- and post-stimulus periods were included during the filtering to prevent filtering-induced boundary artifacts from contaminating the stimulus period. Cross-correlations between the filtered stimulus and FFR<sub>ENV</sub> were then conducted over a range of time delays (FFR<sub>ENV</sub> lagged behind the stimulus) for the stimulus period. This range was set at 8–13 ms, for which 5–10 ms are latencies when FFR<sub>ENV</sub> starts to occur in the brainstem (Skoe and Kraus, 2010) with an additional 3 ms for air transmission of sound from the loudspeaker to cochlea (1 meter apart). The time delay which corresponded to the maximum absolute correlation value was taken as the neural delay (Krizman and Kraus, 2019; Mai and Howell, 2022).

- (2) Measure FFR<sub>ENV\_F0</sub> magnitude along the F<sub>0</sub> trajectory (Mai et al., 2018). Note that the vowel stimulus and FFR<sub>ENV</sub> in this procedure were waveforms *prior* to the bandpass filtering mentioned in the previous procedure, i.e., bandpass filtering in procedure (1) was only for determining the neural delay, not for measuring the FFR<sub>ENV\_F0</sub> magnitude here. The stimulus' F<sub>0</sub> trajectory was first obtained using a set of 40-ms sliding windows (1-ms per step) applied to the vowel'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 was chosen as the F<sub>0</sub> value at each step. The same set of 40-ms sliding windows was applied to FFR<sub>ENV</sub> with onset of the first window set at the neural delay determined in procedure (1). The same Hanning-windowing, zero-padding and Fourier-transforms were performed for each step. For each given step, the mean log-magnitude was measured across a 20 Hz bandwidth centered at the frequency of the F<sub>0</sub> trajectory at that step. The magnitudes were then averaged across all steps along the F<sub>0</sub> trajectory across the entire 120-ms period as the final FFR<sub>ENV\_F0</sub> magnitude.

As FFR magnitudes are small, their robustness was further confirmed by statistically comparing the FFR<sub>ENV\_F0</sub> magnitudes with the EEG noise floors. The noise floors were quantified as the EEG magnitudes at the 50-ms pre-stimulus period (Mai et al., 2018). The quantification procedure was similar to that used in calculating the FFR<sub>ENV\_F0</sub> magnitudes, in which a set of 40-ms sliding windows (with 1-ms steps) was applied over the pre-stimulus period. The same Hanning-windowing and zero-padding as in calculating FFR<sub>ENV\_F0</sub> magnitudes were performed for each step. Magnitudes of noise floors were then measured as the spectral magnitudes across the corresponding F<sub>0</sub> frequency range (i.e., 110–160 Hz) averaged across all steps. The results confirmed such robustness that FFR<sub>ENV\_F0</sub> magnitudes were all significantly greater than the noise floor under all background types (see *Results* for details).

##### 2.4.2. Cortical phase-locked activities

Cortical phase-locked activity was measured as theta-band (4–6 Hz, to correspond to the stimulus presentation rate of  $\sim 5$  vowels per second) phase-locking value ( $\theta$ -PLV) at C3 and C4. The frequency at 4–6 Hz, whose range is narrower than the conventional definition of theta (4–8 Hz), was used because it better approximated the stimulus presentation rate to quantify the neural phase-locking to the stimuli (Mai et al., 2018). 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  $\pm 15 \mu\text{V}$  on either electrode were rejected (Mai et al., 2018). This lower rejection threshold was used than with FFRs ( $\pm 25 \mu\text{V}$ ) because the theta-band signal usually does not have excessively high amplitude since it occupies a relatively narrow frequency range and  $> 80\%$  of the sweeps were retained in all participants after artefact rejection.  $\theta$ -PLV was mea-

sured as the inter-trial phase coherence (Morillon et al., 2012). In the current case, it was the phase coherence across vowel sweeps time-locked to the onset of each vowel. The  $\theta$ -PLV time series ( $PLV(t)$ ) were thus calculated as:

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

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 vowel onset and  $j$  is imaginary unit  $\sqrt{-1}$ . Hilbert phase was used because, thanks to its relative independence of the signal amplitude, PLV based on Hilbert phase effectively reflects phase-locking to stimuli even when EEG amplitude variation occurs due to relaxation and eye closure (Thatcher, 2012).

As PLV is restricted to values between 0 and 1, it was further logit-transformed to have the range of  $[-\infty, +\infty]$ , which made it appropriate to be used for linear regression analyses (Waschke et al., 2017):

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

As neural transmission from cochlea to auditory cortex takes 10 to 30 ms in primates (Lakatos et al., 2007),  $\text{Logit} - \theta - PLV(t)$  values were averaged across the stimulus period (120 ms) with neural delays between 13 and 33 ms (including the 3 ms for air transmission of sound; at 1-ms steps). The maximal value of the time-averaged Logit- $\theta$ -PLV among these neural delays was chosen for each electrode site. The final Logit- $\theta$ -PLV was taken as the average across the two electrodes.

## 2.5. Statistical analyses

Analyses in 2.5.1 and 2.5.2 (PCA, mixed-effect linear regressions and multiple linear regressions) were conducted using SPSS 26.0 (SPSS Inc., USA). Analyses in 2.5.3 (bootstrapping) were conducted using Matlab 2019b (Mathwork, USA).

### 2.5.1. Effects of age and hearing loss on SRT and the neural signatures

We first analyzed how age and hearing loss affected the SiN performances (SRT) and the neural signatures (FFR<sub>ENV\_F0</sub> and Logit- $\theta$ -PLV). As age and hearing loss are highly correlated, we applied Principal Component Analysis (PCA) to transform the age and PTATs (which reflect the degree of hearing loss) into sets of uncorrelated principal components (PCs). Specifically, PCA was applied on the gathering of age and PTATs at all frequencies (0.25, 0.5, 1, 2, 3, 4, 6 and 8 kHz). PCs with eigenvalues  $> 1$  were retained and transformed into Anderson-Rubin scores via varimax rotation (Yong and Pearce, 2013). Two PCs were obtained following this approach (see 3.1 and Table 1).

Following the PCA, linear mixed effect regressions were conducted using SRT, FFR<sub>ENV\_F0</sub> and Logit- $\theta$ -PLV as dependent variables, Background Type (SSN and babble noise for SRT; quiet, SSN and babble noise for FFR<sub>ENV\_F0</sub> and Logit- $\theta$ -PLV) as the fixed-effect factor and PCs obtained from the PCA as fixed-effect covariates. Participant was used as the random-effect factor, i.e., a random intercept was fitted for each participant. For the fixed-effect configuration, we did not find significant interactions between PCs (including three-way interactions between Background Type and the two PCs) for any of the models. We thus did not include the interaction terms between PCs due to a tradeoff between using a full-scale factorial configuration and avoiding model overfitting. For the random-effect configuration, a previous study has shown that further fitting a random slope for each of the within-subject factor (Background Type in the current case) can lead to better model

**Table 1**

Loadings of principal components (PCs) obtained via PCA on the age and PTATs after they were transformed into Anderson-Rubin scores via varimax rotation. Two PCs ( $PC1_{Age\_PTAT}$  and  $PC2_{Age\_PTAT}$ ) with eigenvalues  $> 1$  were retained. Loadings highlighted in bold ( $> 0.7$ ) indicate the relatively high contributions of certain age/PTAT parameters to a given PC.

Measures	$PC1_{Age\_PTAT}$	$PC2_{Age\_PTAT}$
Age	<b>0.783</b>	0.456
PTAT <sub>0.25kHz</sub>	0.324	<b>0.880</b>
PTAT <sub>0.5kHz</sub>	0.289	<b>0.917</b>
PTAT <sub>1kHz</sub>	0.438	<b>0.801</b>
PTAT <sub>2kHz</sub>	<b>0.726</b>	0.322
PTAT <sub>3kHz</sub>	<b>0.904</b>	0.257
PTAT <sub>4kHz</sub>	<b>0.933</b>	0.292
PTAT <sub>6kHz</sub>	<b>0.873</b>	0.403
PTAT <sub>8kHz</sub>	<b>0.849</b>	0.382

fitting (Barr et al., 2013). However, when such approach was employed, either the models failed to converge or the Hessian matrices failed to be positively definite (i.e., matrices cannot be inverted and standard errors were thus untrustworthy, see McNeish and Bauer, 2020). Thus, eventually we did not include random slopes for Background Type, but only include random intercepts for participants. The linear mixed-effect model for a given dependent variable was therefore fitted via the formula (1) as below:

$$DV_{ij} = \beta_0 + \beta_1 BT_i + \beta_2 PC1_j + \beta_3 PC2_j + \beta_4 (BT_i \times PC1_j) + \beta_5 (BT_i \times PC2_j) + \beta_{random\_j} + \varepsilon_{ij} \quad (1)$$

where  $DV_{ij}$  denotes the value of the dependent variable at the  $i$ th level of Background Type for the  $j$ th participant.  $\beta_0$  denotes the global intercept of the model.  $\beta_1$  to  $\beta_5$  denote the coefficients for the fixed-effect factors/interactions.  $BT_i$  is the value of the  $i$ th level of Background Type based on sum coding for F-contrasts (-1 and 1 for SSN and babble noise, respectively, when SRT was the dependent variable; -1, 0 and 1 for quiet, SSN and babble noise, respectively, when FFR<sub>ENV\_F0</sub> or Logit- $\theta$ -PLV was the dependent variable).  $PC1$  and  $PC2$  denote the two PCs obtained via PCA applied on the age and PTATs ( $PC1_{Age\_PTAT}$  and  $PC2_{Age\_PTAT}$ , see Table 1).  $PC1_j$  and  $PC2_j$  are values of  $PC1$  and  $PC2$  for the  $j$ th participant. Note that as mentioned above, the terms involving interactions between the two PCs were not included in the formula.  $\beta_{random\_j}$  denotes the random intercept fitted for the  $j$ th participant.  $\varepsilon_{ij}$  denotes the residual. The models were fitted via restricted maximum likelihood. Structure types of the covariance matrices for the residuals were chosen among the commonly used structures (first-order autoregression, diagonal, compound symmetry, scaled identity, Toeplitz and unstructured) which generated the lowest Bayesian Information Criterion (BIC) value for each model (i.e., best goodness of fit). Following this approach, scaled identity was chosen as the covariance type for all three models. The degrees of freedom were estimated via Satterthwaite approximation.

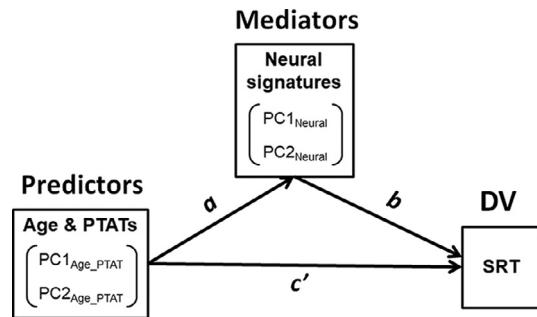
Post-hoc analyses were further conducted if: (i) a significant main effect of Background Type occurred to see how neural signatures differed between different background types (via pairwise T-tests; but not when SRT was the dependent variable because in this case there were only two levels of Background Type); or (ii) a significant [Background Type  $\times$  PC1] or [Background Type  $\times$  PC2] interaction occurred to see whether there was a significant main effect of PC1/PC2 under each individual background type (via Pearson correlations between PC1/PC2 and data under each background type).  $P$  values for the post-hoc analyses were corrected via Bonferroni correction according to the number of levels within the Background Type.

A further issue we considered involved a potential confounding effect of arousal for the neural signatures. In the present study, participants were instructed to relax, close their eyes and keep still during the neural assessments (see 2.2.2). They were thus not prevented from falling asleep which may lead to changes in arousal states that could in turn affect subcortical and cortical phase-locked activities (Mai et al., 2019). It is unclear whether such changes had confounding impacts on quantifying how age and hearing loss affected these activities. To test such confounding impacts, we conducted two additional analyses. First, we extracted the sleep spindles that can quantify the degree of arousal during wakefulness and the stage 1 and 2 of non-Rapid Eye Movement (nREM) sleep in which participants had experienced during the neural assessments<sup>3</sup> (Kim et al., 2012) (for detailed procedure to extract sleep spindles, see Appendices; also see Mai et al., 2019). We conducted further analyses using the density of spindles (number of spindles per minute) as an additional fixed-effect covariate in the linear mixed-effect regressions for the neural signatures. These were conducted for the purpose of controlling for the effect of arousal and testing whether this resulted in changes in statistical significance from the original results. It was shown that the original statistical results that argued for the effects of age and hearing loss were maintained (see Results and Appendices). Second, we fitted separate linear mixed-effect models for the neural signatures only based on a subset of data where participants were awake or at high arousal states before real sleepiness occurred. The periods of wakefulness/high arousal states were determined by the occurrence of sleep spindles (i.e., periods without spindles, see Appendices for the detailed procedure of determination). We found that the statistical significances remained the same as the original results (see Results and Appendices). Therefore, changes from high arousal to partial sleepiness in some participants had not resulted in significant impacts on measuring the effects of age and hearing loss on the neural signatures, indicating such potential confounds were negligible. To keep the data integrity and quality of the neural signatures (esp. for FFR<sub>ENV\_F0</sub> which requires a large number of sweeps to ensure decent signal quality, see Skoe and Kraus, 2010), we will interpret the results based on the whole dataset rather than the subset of high arousal data as described above.

### 2.5.2. Neural-behavioral relationship

Multiple linear regressions were then conducted to test for neural-behavioral relationship by linearly regressing SRTs on neural signatures (FFR<sub>ENV\_F0</sub> and Logit- $\theta$ -PLV) under the corresponding types of noise. This avoided problems that could arise due to behavioral and neural measurements being associated under different types of noise as in previous work (Presacco et al., 2016). Additionally, since FFRs in quiet have been suggested to be associated with SiN performances (Anderson et al., 2011), we further included neural signatures obtained under the quiet background to predict SRTs. Namely, neural signatures obtained under quiet and SSN were used to predict SRT under SSN, while neural signatures obtained under quiet and babble noise were used to predict SRT under babble noise.

As different elements within the predictors may be highly correlated with each other, we applied PCA on the predictors prior to the multiple linear regressions. This has been shown to effectively



**Fig. 4.** The mediation model that tested how age and hearing loss (measured as PTATs) affected SiN performances (measured as SRTs) through the neural signatures. Age and PTATs, neural signatures and SRT were the predictors, mediators and dependent variable (DV), respectively. SRTs under SSN and babble noise were predicted in separate models, each with neural signatures obtained under quiet and the corresponding noise type for SRT. Both the predictors and mediators had two variables (PC1<sub>Age\_PTAT</sub> and PC2<sub>Age\_PTAT</sub> for predictors and PC1<sub>Neural</sub> and PC2<sub>Neural</sub> for mediators). Path *a* and *b* indicate the indirect/mediation effect through the neural signatures while path *c'* indicates the direct effect of age and PTATs on SRT.

extract key components of the neural signatures and meanwhile avoid concerns of collinearity between predictors during multiple regressions (Schoof and Rosen, 2016). Same as in 2.5.1, PCs with eigenvalues  $> 1$  were retained and transformed into Anderson-Rubin scores via varimax rotation. Multiple linear regressions were then conducted using PCs as predictors (independent variables) and SRTs as dependent variables.

### 2.5.3. Mediation effects of neural signatures

While testing the neural-behavioral relationships could reveal how neural signatures may be related to SiN performances, it is still not clear how the neural signatures play a role for SiN perception during the ageing process. To address this issue, we further conducted mediation analyses (MacKinnon, 2007) to show whether and how age and hearing loss affect SRTs through FFR<sub>ENV\_F0</sub> and Logit- $\theta$ -PLV (i.e., how FFR<sub>ENV\_F0</sub> and Logit- $\theta$ -PLV mediate the effects of age and hearing loss on SRTs). The mediation model is illustrated as Fig. 4. Age and PTATs, neural signatures and SRTs were predictors, mediators and dependent variables, respectively. Similar to 2.5.2, SRTs were predicted by age and PTATs with neural signatures as mediators under quiet and the corresponding types of background noise. Namely, neural signatures obtained under quiet and SSN were used as mediators when predicting SRT under SSN, whilst neural signatures obtained under quiet and babble noise were used as mediators when predicting SRT under babble noise.

Furthermore, we used corresponding PCs obtained via PCA to represent variables of the age and PTATs (as described in 2.5.1) and neural signatures (as described in 2.5.2), respectively. As such, for each model (model for SRT under SSN or babble noise), there were two predictor variables which were the two PCs retained via PCA applied on the age and PTATs (PC1<sub>Age\_PTAT</sub> and PC2<sub>Age\_PTAT</sub> that largely reflected age/high-frequency ( $\geq 2$  kHz) PTATs and low-frequency ( $< 2$  kHz) PTATs, respectively; see 3.1 and Table 1). Similarly, there were two mediator variables which were the two PCs retained via PCA applied on the neural signatures (PC1<sub>Neural</sub> and PC2<sub>Neural</sub> that largely reflected FFR<sub>ENV\_F0</sub> and Logit- $\theta$ -PLV, respectively; see 3.1 and Table 4).

The following effects were quantified: (i) indirect/mediation effects that reflect the degree to which neural signatures mediated the effects of age and PTATs on SRT (put in another way, the effect of age and PTATs on SRT through the neural signatures); (ii) direct effects which are effects of age and PTATs on SRT by controlling for the effects of neural signatures; and (iii) total effects which are combinations of the indirect/mediation and direct effects. Mathematically, these effects were quantified via the following formulas:

<sup>3</sup> Subsequent analyses segmented the neural recordings into multiple epochs (each ~20 s). We did not find high-amplitude delta-band (1–4 Hz) activities (magnitude of Hilbert envelope  $> 60 \mu\text{V}$ ) that lasted for  $> 25\%$  of the time in any epoch in any single participant, showing that no participants were in the stage 3 or 4 of nREM sleep (i.e., 'deep sleep', Brown et al., 2012) (see detailed analyses in Mai et al., 2019). Also, no participants reported that they were in a deep sleep during the recording. Therefore, participants should be either awake or in the stage 1 or 2 of nREM sleep (i.e., 'light sleep') where sleep spindles can be used to quantify the degree of arousal (Kim et al., 2012).

$$a = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad (2)$$

$$b = \begin{bmatrix} b_1 & b_2 \end{bmatrix} \quad (3)$$

$$c' = \begin{bmatrix} c'_1 & c'_2 \end{bmatrix} \quad (4)$$

$$P = \begin{bmatrix} PC1_{Age\_PTAT} \\ PC2_{Age\_PTAT} \end{bmatrix} \quad (5)$$

$$M = \begin{bmatrix} PC1_{Neural} \\ PC2_{Neural} \end{bmatrix} \quad (6)$$

$$M = aP + \beta_{01} + e_1 \quad (7)$$

$$SRT = bM + c'P + \beta_{02} + e_2 \quad (8)$$

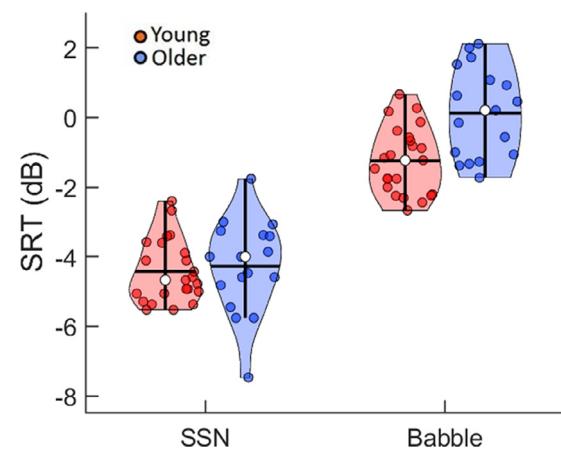
where  $a$  (formula (2)) refers to the coefficient matrix when linearly regressing  $M$  (the mediator matrix, formula (6)) on  $P$  (the predictor matrix, formula (5)) (formula (7) and path  $a$  as in Fig. 4).  $b$  (formula (3)) and  $c'$  (formula (4)) refer to the coefficient matrices when linearly regressing SRT (the dependent variable) on  $M$  and  $P$  (formula (8) and path  $b$  and  $c'$  as in Fig. 4).  $\beta_{01}$  and  $\beta_{02}$  refer to the intercept, and  $e_1$  and  $e_2$  refer to the residual matrix/scalar during regressions. The mediation effects are multiplications between coefficients in matrices  $a$  and  $b$  that link to specific mediator and predictor variables (i.e.,  $a_{ij}b_i$  denotes the degree to which the  $i$ th mediator variable mediates the effect of the  $j$ th predictor variable on SRT). The direct effects are coefficients in matrix  $c'$  (i.e.,  $c'_j$  denotes the direct effect of the  $j$ th predictor variable on SRT). The total effects  $c$  are the sum of the indirect/mediation effects and the direct effects (i.e.,  $c_j = \sum_i a_{ij}b_i + c'_j$  denotes the total effect of the  $j$ th predictor variable on SRT). Note that all regression coefficients were calculated as standardized coefficients when they were used to measure each effect.

We then used a bootstrapping approach to quantify the statistical significance of these effects (Shrout and Bolger et al., 2002). Bootstrapping is an empirical, non-parametric method that avoids requirement for assumptions of specific data distributions (e.g., normality) as in the parametric methods. Specifically, we resampled data with replacement in each repetition and then obtained a bootstrap distribution for each effect. We used the bias-corrected and accelerated (BCa) bootstrap (Efron, 1987) to measure the confidence intervals for each distribution. The BCa approach corrected the confidence limits by accounting for the deviations of the bootstrapped mean from the sample mean and skewness of the distribution (Efron, 1987; Efron and Tibshirani, 1994). The replication was set at 10,000 times and confidence intervals were measured for each distribution (using the Matlab function 'bootci'). An effect was considered as statistically significant if the value of zero fell outside the 95% confidence interval of the corresponding distribution.

### 3. Results

#### 3.1. Effects of age and hearing loss on SRT and the neural signatures

First of all, robustness of FFR signals was confirmed by comparing  $FFR_{ENV\_F0}$  magnitudes with their corresponding EEG noise floors.  $FFR_{ENV\_F0}$  magnitudes were all significantly greater than noise floors under all background types ( $FFR_{ENV\_F0\_Quiet}$ :  $t_{(39)} = 10.385$ ,  $p < 10^{-12}$ ;  $FFR_{ENV\_F0\_SSN}$ :  $t_{(39)} = 6.021$ ,  $p < 10^{-6}$ ;  $FFR_{ENV\_F0\_Babble}$ :  $t_{(39)} = 4.788$ ,  $p < 10^{-4}$ ).



**Fig. 5.** Data distributions of SRT illustrated as violin plots grouped by the background types (SSN and babble noise) and the young/older group (red/blue). The horizontal and vertical lines within the violins indicate the grand-averaged values and 1.5 interquartile ranges. The white dots indicate the median values. N.B., lower SRT indicates better SiN performance.

Prior to testing the effects of age and hearing loss, PCA was first applied on the age and PTATs. The first two PCs (which explained 85.8% of the total variance) had eigenvalues  $> 1$  and were thus retained and transformed to Anderson-Rubin scores via varimax rotation. The loadings for the PCs are shown in Table 1. For  $PC1_{Age\_PTAT}$ , age and high-frequency PTATs ( $\geq 2$  kHz) had relatively high loadings ( $> 0.7$ ) compared to low-frequency PTATs ( $< 2$  kHz) (loadings  $< 0.5$ ). For  $PC2_{Age\_PTAT}$ , on the other hand, low-frequency PTATs had relatively high loadings ( $> 0.8$ ) compared to age and high-frequency PTATs (loadings  $< 0.5$ ) (see Table 1). Therefore,  $PC1_{Age\_PTAT}$  largely reflected age and the degree of high-frequency hearing loss (however unable to separate these two) while  $PC2_{Age\_PTAT}$  largely reflected the degree of low-frequency hearing loss.

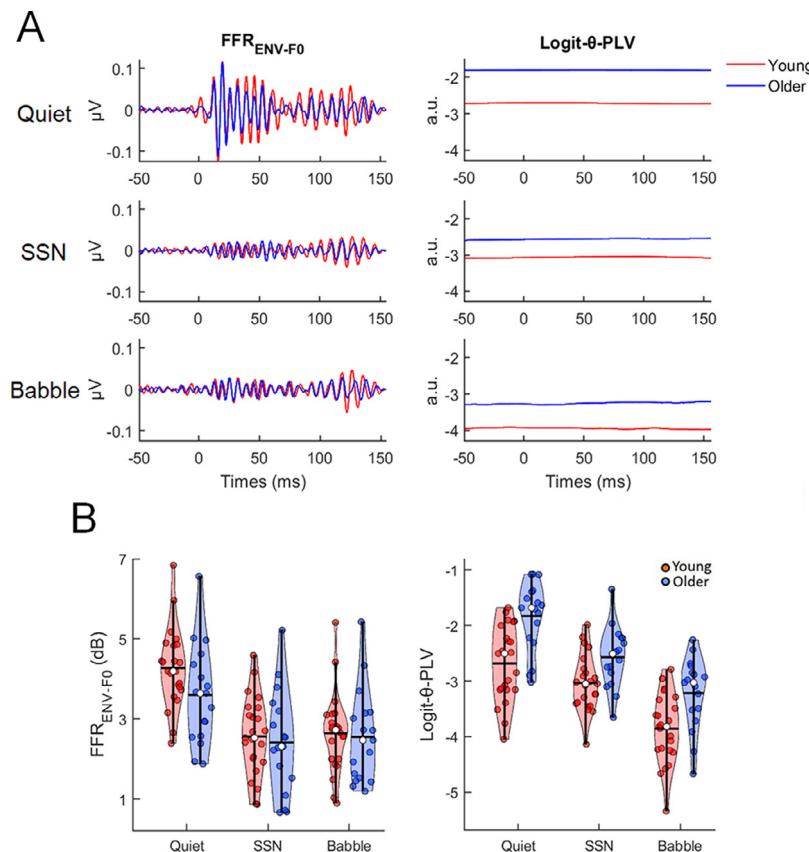
Effects of age and hearing loss were then tested. Linear mixed-effect regressions were conducted for SRT,  $FFR_{ENV\_F0}$  and Logit- $\theta$ -PLV using Background Type as the fixed-effect factor, the two PCs as fixed-effect covariates and Participant as the random-effect factor. The statistical results are summarized in Table 2. Fig. 5 illustrates the data distribution of SRT grouped by the background types and age groups (the young and older groups which were for illustrative purpose only rather than being used as a statistical factor). Fig. 6 illustrates the time series and data distributions of  $FFR_{ENV\_F0}$  and Logit- $\theta$ -PLV grouped by the background types and age groups. Descriptions of significant main effects and interactions for each dependent variable were as follows (note that  $p$  values for the post-hoc analysis following significant main effects/interactions were Bonferroni-corrected according to the number of levels within Background Type):

For SRT, there were significant effects of Background Type ( $F_{(1, 37)} = 428.950$ ,  $p < 10^{-21}$ ;  $SRT_{SSN} < SRT_{Babble}$ ) and  $PC1_{Age\_PTAT}$  ( $F_{(1, 37)} = 8.779$ ,  $p = 0.005$ ; greater  $PC1_{Age\_PTAT}$  correlated with higher SRT). There were also significant interactions of [Background Type  $\times$   $PC1_{Age\_PTAT}$ ] ( $F_{(1, 37)} = 4.800$ ,  $p = 0.035$ ) and [Background Type  $\times$   $PC2_{Age\_PTAT}$ ] ( $F_{(1, 37)} = 10.112$ ,  $p = 0.003$ ). Post-hoc analyses following these interactions showed that  $PC1_{Age\_PTAT}$  and  $PC2_{Age\_PTAT}$  were significantly correlated with SRT under babble noise (greater  $PC1_{Age\_PTAT}$  correlated with higher  $SRT_{Babble}$ ,  $t_{(39)} = 3.402$ ,  $p = 0.003$ ; greater  $PC2_{Age\_PTAT}$  correlated with higher  $SRT_{Babble}$ ,  $t_{(39)} = 2.817$ ,  $p = 0.015$ ) but not under SSN ( $PC1_{Age\_PTAT}$  with  $SRT_{SSN}$ :  $t_{(39)} = 1.321$ ,  $p = 0.389$ ;  $PC1_{Age\_PTAT}$  with  $SRT_{Babble}$ :  $t_{(39)} = -0.203$ ,  $p > 1$ ). Note that lower SRT indicates better SiN perception.

**Table 2**

Statistical results for the linear mixed-effect regression for the SRT,  $FFR_{ENV\_FO}$  and Logit- $\theta$ -PLV.  $PC1_{Age\_PTAT}$  and  $PC2_{Age\_PTAT}$  largely reflected age/high-frequency PTATs and low-frequency PTATs, respectively (indicated in the brackets with Italics). DV, df, F, p refer to the dependent variable, degrees of freedom, F values, and p values, respectively. Significant p values ( $< 0.05$ ) are indicated in bold. \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .

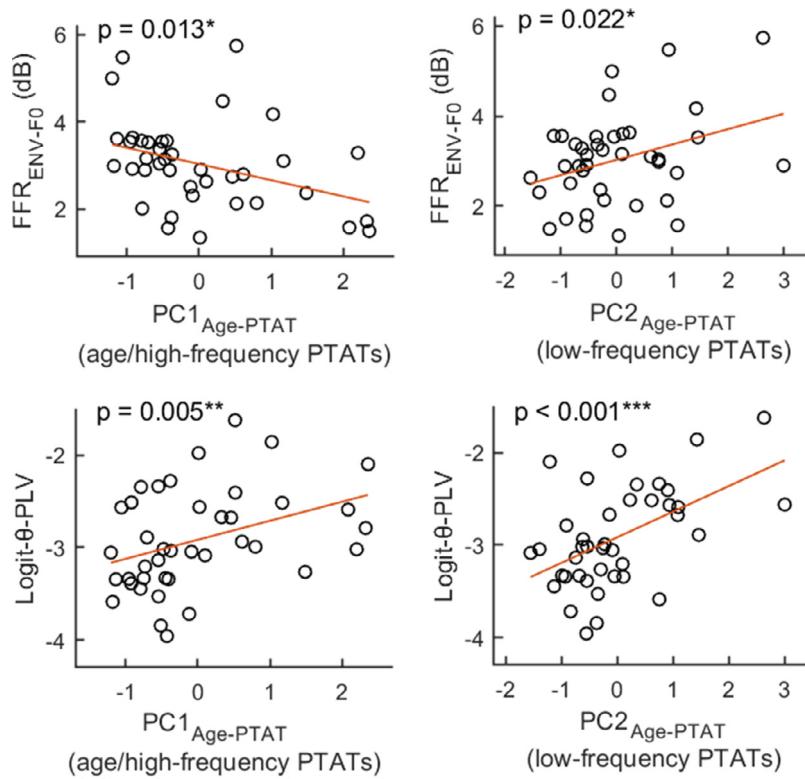
DV	Fixed effects	df1	df2	F	p
SRT	Background Type	1	37	428.950	<b>&lt; 10<sup>-21</sup>***</b>
	$PC1_{Age\_PTAT}$ (age/high-frequency PTATs)	1	37	8.779	<b>0.005**</b>
	$PC2_{Age\_PTAT}$ (low-frequency PTATs)	1	37	3.024	0.090
	Background Type $\times$ $PC1_{Age\_PTAT}$	1	37	4.800	<b>0.035*</b>
	Background Type $\times$ $PC2_{Age\_PTAT}$	1	37	10.112	<b>0.003**</b>
$FFR_{ENV\_FO}$	Background Type	2	74	106.299	<b>&lt; 10<sup>-21</sup>***</b>
	$PC1_{Age\_PTAT}$ (age/high-frequency PTATs)	1	37	6.775	<b>0.013*</b>
	$PC2_{Age\_PTAT}$ (low-frequency PTATs)	1	37	5.699	<b>0.022*</b>
	Background Type $\times$ $PC1_{Age\_PTAT}$	2	74	3.871	<b>0.025*</b>
	Background Type $\times$ $PC2_{Age\_PTAT}$	2	74	2.098	0.130
Logit- $\theta$ -PLV	Background Type	2	74	63.642	<b>&lt; 10<sup>-16</sup>***</b>
	$PC1_{Age\_PTAT}$ (age/high-frequency PTATs)	1	37	9.010	<b>0.005**</b>
	$PC2_{Age\_PTAT}$ (low-frequency PTATs)	1	37	16.396	<b>&lt; 0.001***</b>
	Background Type $\times$ $PC1_{Age\_PTAT}$	2	74	1.424	0.247
	Background Type $\times$ $PC2_{Age\_PTAT}$	2	74	0.232	0.794



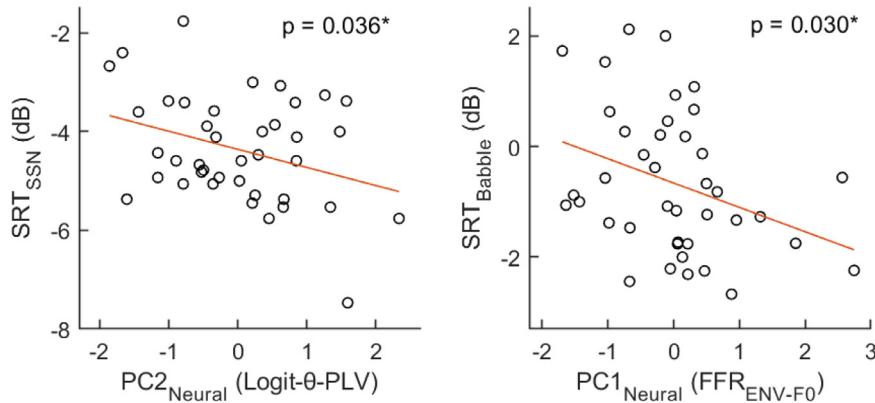
**Fig. 6.** Time series and data distributions of  $FFR_{ENV\_FO}$  and Logit- $\theta$ -PLV. (A) Time series of  $FFR_{ENV\_FO}$  and Logit- $\theta$ -PLV under quiet (upper), SSN (mid) and babble noise (lower). The series were averages across participants within the young and older group, respectively. Series of  $FFR_{ENV\_FO}$  are shown as  $FFR_{ENV}$  bandpass filtered at 90–180 Hz (corresponding to the  $F_0$  range). '0' corresponds to the stimulus onset. (B) Data distributions of  $FFR_{ENV\_FO}$  and Logit- $\theta$ -PLV illustrated as violin plots grouped by the three background types and the young/older group. The horizontal and vertical lines within the violins indicate the grand-averaged values and 1.5 interquartile ranges. The white dots indicate the median values.

For  $FFR_{ENV\_FO}$ , there were significant main effects of Background Type ( $F_{(2, 74)} = 106.299, p < 10^{-21}$ ),  $PC1_{Age\_PTAT}$  (greater  $PC1_{Age\_PTAT}$  correlated with smaller  $FFR_{ENV\_FO}$  magnitude,  $F_{(1, 37)} = 6.775, p = 0.013$ ; Fig. 7 upper left) and  $PC2_{Age\_PTAT}$  (greater  $PC2_{Age\_PTAT}$  correlated with greater  $FFR_{ENV\_FO}$  magnitude,  $F_{(1, 37)} = 5.699, p = 0.022$ ; Fig. 7 upper right). Post-hoc analyses following the effect of Background Type showed that  $FFR_{ENV\_FO}$  had significant greater magnitude under quiet than noisy backgrounds

( $FFR_{ENV\_FO\_Quiet} > FFR_{ENV\_FO\_SSN}, t_{(39)} = 11.455, p < 10^{-12}$ ;  $FFR_{ENV\_FO\_Quiet} > FFR_{ENV\_FO\_Babble}, t_{(39)} = 10.997, p < 10^{-12}$ ), but did not differ significantly between the two types of noise ( $t_{(39)} = -1.032, p = 0.926$ ). There was also a significant [Background Type  $\times$   $PC1_{Age\_PTAT}$ ] interaction ( $F_{(2, 74)} = 3.871, p = 0.025$ ). Post-hoc analyses following this interaction showed that  $PC1_{Age\_PTAT}$  was significantly correlated with  $FFR_{ENV\_FO}$  under quiet (greater  $PC1_{Age\_PTAT}$  correlated with smaller  $FFR_{ENV\_FO\_Quiet}$  magnitude,



**Fig. 7.** Scatter plots that illustrate the effects of age and hearing loss on  $FFR_{ENV\_F0}$  and  $\text{Logit-}\theta\text{-PLV}$ . The red curves are fitted lines via linear regressions. X axes denote  $PC1_{Age\_PTAT}$  and  $PC2_{Age\_PTAT}$  which largely reflected age/high-frequency PTAT and low-frequency PTAT, respectively (indicated in the brackets). Y axes denote the magnitudes of  $FFR_{ENV\_F0}$ /Logit- $\theta$ -PLV (averaged across background types).  $P$  values are indicated. \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .



**Fig. 8.** Scatter plots that illustrate the significant neural-behavioral relationships. The red curves are fitted lines via linear regressions. X axes denote  $PC1_{Neural}/PC2_{Neural}$  which largely reflected magnitudes of  $FFR_{ENV\_F0}/\text{Logit-}\theta\text{-PLV}$  (indicated in the brackets). Y axes denote the SRTs. Significant correlations shown in **Table 6** were illustrated (left: between  $PC2_{Neural}$  and  $SRT_{SSN}$ ; right: between  $PC1_{Neural}$  and  $SRT_{Babble}$ ).  $P$  values are indicated. \* $p < 0.05$ .

$t_{(39)} = -3.470$ ;  $p = 0.004$ ), but not under noisy backgrounds (with  $FFR_{ENV\_F0\_SSN}$ :  $t_{(39)} = -1.665$ ,  $p = 0.312$ ; with  $FFR_{ENV\_F0\_Babble}$ :  $t_{(39)} = -1.661$ ,  $p = 0.314$ ).

For Logit- $\theta$ -PLV, there were significant main effects of Background Type ( $F_{(2, 74)} = 63.642$ ,  $p < 10^{-16}$ ),  $PC1_{Age\_PTAT}$  (greater  $PC1_{Age\_PTAT}$  correlated with greater Logit- $\theta$ -PLV,  $F_{(1, 37)} = 9.010$ ,  $p = 0.005$ ) and  $PC2_{Age\_PTAT}$  (greater  $PC2_{Age\_PTAT}$  correlated with greater Logit- $\theta$ -PLV,  $PC1$ :  $F_{(1, 37)} = 16.396$ ,  $p < 0.001$ ). Post-hoc analyses following the effect of Background Type showed that Logit- $\theta$ -PLV was significantly greater under quiet than noisy backgrounds (Logit- $\theta$ -PLV<sub>Quiet</sub> > Logit- $\theta$ -PLV<sub>SSN</sub>,  $t_{(39)} = 5.107$ ,  $p < 10^{-4}$ ; Logit- $\theta$ -PLV<sub>Quiet</sub> > Logit- $\theta$ -PLV<sub>Babble</sub>,  $t_{(39)} = 9.772$ ,  $p < 10^{-10}$ ) and significantly greater under SSN than under babble noise

( $t_{(39)} = 7.205$ ,  $p < 10^{-7}$ ). No significant interaction effects were observed.

We also tested whether arousal during neural measurements influenced the effects of age and hearing loss on the neural signatures. Additional analyses were conducted by including the density of sleep spindles as fixed-effect covariates in the linear mixed-effect regressions for  $FFR_{ENV\_F0}$  and Logit- $\theta$ -PLV. The statistical results are described in the *Appendices A1*, which showed that adding the density of spindles in the regression models did not change the statistical significance of the effects of age and hearing loss (see **Table A1**). We also fitted separate linear mixed-effect models for the neural signatures only based on the subset of data where participants were awake or at high arousal states. The results showed

**Table 3**

Loadings of principal components (PCs) obtained via PCA on the neural signatures (obtained in quiet and SSN and in quiet and babble noise, respectively) after they were transformed into Anderson-Rubin scores via varimax rotation. Two PCs (PC1<sub>Neural</sub> and PC2<sub>Neural</sub>) with eigenvalues  $> 1$  were retained. Loadings highlighted in bold ( $> 0.8$ ) indicate the relatively high contributions of certain neural signatures to a given PC.

Background types	Measures	PC1 <sub>Neural</sub>	PC2 <sub>Neural</sub>
Quiet + SSN	FFR <sub>ENV_F0_Quiet</sub>	<b>0.927</b>	0.083
	FFR <sub>ENV_F0_SSN</sub>	<b>0.933</b>	-0.055
	Logit- $\theta$ -PLV <sub>Quiet</sub>	-0.047	<b>0.892</b>
	Logit- $\theta$ -PLV <sub>SSN</sub>	0.074	<b>0.884</b>
Quiet + Babble	FFR <sub>ENV_F0_Quiet</sub>	<b>0.939</b>	-0.074
	FFR <sub>ENV_F0_Babble</sub>	<b>0.925</b>	0.157
	Logit- $\theta$ -PLV <sub>Quiet</sub>	0.039	<b>0.810</b>
	Logit- $\theta$ -PLV <sub>Babble</sub>	0.028	<b>0.849</b>

that statistical significances remain the same as the original results (see Appendices A2, Table A2). These thus indicate that changes in arousal may have affected the magnitudes of FFR<sub>ENV\_F0</sub> and Logit- $\theta$ -PLV (Mai et al., 2019), however, such changes had not caused significant impacts on how age and hearing loss affected these neural signatures.

In sum, results of the linear mixed-effect regressions showed that: (1) SiN performance was better under SSN than under babble noise; age and hearing loss contributed to poorer performances, however, the effects were present only under babble noise but not under SSN; (2) FFR<sub>ENV\_F0</sub> magnitude was susceptible to changes from quiet to noisy backgrounds (but not to changes between SSN and babble noise); it had smaller magnitude in participants with older age and higher degree of high-frequency hearing loss but greater magnitude in participants with higher degree of low-frequency hearing loss; (3) Logit- $\theta$ -PLV was susceptible to changes from quiet to noisy backgrounds and to changes between different noise types (SSN vs. babble noise); it was greater in participants with older age and higher degree of hearing loss; (4) no evidence was found that changes in arousal during the neural measurements significantly influenced the effects of age and hearing loss on the neural signatures.

### 3.2. Neural-behavioral relationship

Neural-behavioral relationships were assessed by linearly regressing SRTs on the neural signatures obtained in quiet and the corresponding types of noise. PCAs were first applied on the predictors (neural signatures). The first two PCs (PC1<sub>Neural</sub> and PC2<sub>Neural</sub>) whose eigenvalues  $> 1$  were retained for both predictors for SRT<sub>SSN</sub> (explaining 83.2% of the total variance) and predictors for SRT<sub>Babble</sub> (explaining 78.7% of the total variance). Table 3 shows the loadings of PCs. For PC1<sub>Neural</sub>, FFR<sub>ENV\_F0</sub> had relatively high loadings ( $> 0.9$ ) compared to relatively low loadings for Logit- $\theta$ -PLV ( $< 0.1$ ). For PC2<sub>Neural</sub>, Logit- $\theta$ -PLV had relatively high loadings ( $> 0.8$ ) compared to relatively low loadings for FFR<sub>ENV\_F0</sub> ( $< 0.2$ ). Therefore, PC1<sub>Neural</sub> largely reflected FFR<sub>ENV\_F0</sub> magnitudes, while PC2<sub>Neural</sub> largely reflected Logit- $\theta$ -PLV.

Multiple linear regressions were then conducted using SRTs as dependent variables and the corresponding PCs as the independent variables. The statistical results are shown in Table 4. For predicting SRT<sub>SSN</sub>, we found that SRT<sub>SSN</sub> was significantly correlated with PC2<sub>Neural</sub> ( $\beta = -0.334$ ,  $p = 0.036$ ) but not with PC1<sub>Neural</sub> ( $\beta = -0.145$ ,  $p = 0.350$ ). For predicting SRT<sub>Babble</sub>, we found that SRT<sub>Babble</sub> was significantly correlated with PC1<sub>Neural</sub> ( $\beta = -0.340$ ,  $p = 0.030$ ) but not with PC2<sub>Neural</sub> ( $\beta = 0.210$ ,  $p = 0.171$ ). The results therefore indicate that better SiN performance corresponded to greater PC2<sub>Neural</sub> (which largely reflected Logit- $\theta$ -PLV) under SSN and with

greater PC1<sub>Neural</sub> (which largely reflected FFR<sub>ENV\_F0</sub> magnitude) under babble noise.

### 3.3. Mediation effects of neural signatures

In the mediation models, SRTs (SiN performances) were predicted by the PCs that represented age and PTATs (PC1<sub>Age\_PTAT</sub> and PC2<sub>Age\_PTAT</sub>) with PCs that represented the neural signatures as mediators (PC1<sub>Neural</sub> and PC2<sub>Neural</sub>) obtained under quiet and the corresponding types of noise (quiet and SSN for SRT<sub>SSN</sub>; quiet and babble noise for SRT<sub>Babble</sub>). Table 5 shows the statistical results of the indirect/mediation, direct and total effects. An effect was considered as statistically significant when zero fell outside the 95% confidence interval of the corresponding bootstrap distribution. Note that higher SRT stands for poorer SiN performance.

We found significant mediation/indirect effects of PC1<sub>Neural</sub>, in which PC1<sub>Age\_PTAT</sub> and PC2<sub>Age\_PTAT</sub> respectively affected SRT through PC1<sub>Neural</sub> under babble noise (effect sizes: former = 0.146; latter = -0.108). These effects were the combined effects of a negative/positive correlation between PC1<sub>Age\_PTAT</sub>/PC2<sub>Age\_PTAT</sub> and PC1<sub>Neural</sub> and a negative correlation between PC1<sub>Neural</sub> and SRT<sub>Babble</sub>. We also found significant mediation effects of PC2<sub>Neural</sub>, in which PC1<sub>Age\_PTAT</sub> and PC2<sub>Age\_PTAT</sub> respectively affected SRT through PC2<sub>Neural</sub> under both SSN (effect sizes: former = -0.155; latter = -0.242) and babble noise (effect size: former = -0.170; latter = -0.183). These effects were the combined effects of positive correlations between PC1<sub>Age\_PTAT</sub>/PC2<sub>Age\_PTAT</sub> and PC2<sub>Neural</sub> and negative correlations between PC2<sub>Neural</sub> and SRT<sub>SSN</sub>/SRT<sub>Babble</sub>. Empirically, values to designate effect sizes for regressions as small, medium and large are 0.1, 0.3 and 0.5, respectively (Cohen, 1992). These values were then squared for mediation (i.e., 0.01, 0.09 and 0.25, respectively) because each mediation effect is the product of two regression coefficients (see 2.5.3) (Preacher and Kelley, 2011). Therefore, the significant mediation effects currently found had medium to large effect sizes.

Significant direct effects of PC1<sub>Age\_PTAT</sub> on SRT were found under both SSN (effect size = 0.337) and babble noise (effect size = 0.507). There was also a significant direct effect of PC2<sub>Age\_PTAT</sub> on SRT under babble noise (effect size = 0.706). There were significant total effects of PC1<sub>Age\_PTAT</sub> (effect size = 0.483) and PC2<sub>Age\_PTAT</sub> (effect size = 0.416) on SRT under babble noise. No significant total effects of PC1<sub>Age\_PTAT</sub>/PC2<sub>Age\_PTAT</sub> on SRT were found under SSN.

In summary, the results indicate that both subcortical/cortical FFR to F<sub>0</sub> (i.e., FFR<sub>ENV\_F0</sub> largely reflected by PC1<sub>Neural</sub>) and cortical slow-envelope phase-locking (i.e., Logit- $\theta$ -PLV largely reflected by PC2<sub>Neural</sub>) significantly mediate the effects of age and hearing loss on SiN performances. Specifically, the FFR decreased with age/high-frequency hearing loss (largely reflected by PC1<sub>Age\_PTAT</sub>) which in turn contributed to poorer SiN performance, but increased with low-frequency hearing loss (largely reflected by PC2<sub>Age\_PTAT</sub>) which in turn contributed to better SiN performance. These mediations occurred only under babble noise. Note that the mediations had the same sign as the direct effect of PC1<sub>Age\_PTAT</sub> but opposite sign to the direct effect of PC2<sub>Age\_PTAT</sub>. This indicates that the FFR facilitated the effect of age/high-frequency hearing loss (i.e., made SiN performances even worse), but suppressed the effect of low-frequency hearing loss (i.e., helped to resist worsening of performances) (MacKinnon et al., 2000). On the other hand, the slow-envelope phase-locking increased with both age/high-frequency hearing loss and low-frequency hearing loss which in turn contributed to better SiN performance. These effects occurred under both SSN and babble noise. These mediations had the opposite sign to the direct effects of PC1<sub>Age\_PTAT</sub> and PC2<sub>Age\_PTAT</sub>, indicating that the slow-envelope phase-locking suppressed the effects of age and hearing loss (i.e., helped to resist worsening of SiN per-

**Table 4**

Statistical results for the neural-behavioral relationships using multiple linear regressions. SRT<sub>SSN</sub> was predicted by neural signatures obtained under quiet and SSN, whilst SRT<sub>Babble</sub> was predicted by neural signatures obtained under quiet and babble noise. Predictors were the first two PCs (PC1<sub>Neural</sub> and PC2<sub>Neural</sub>) retained via PCAs applied on neural signatures (see loadings of corresponding PCs in **Table 3**). FFR<sub>ENV\_F0</sub> and Logit- $\theta$ -PLV were largely reflected by the PC1<sub>Neural</sub> and PC2<sub>Neural</sub>, respectively and are indicated correspondingly in the brackets with Italics. DV,  $\beta$ , CI and  $p$  refer to dependent variables, standardized regression coefficients, confidence intervals (95%) and  $p$  values, respectively. Significant  $p$  values ( $< 0.05$ ) are indicated in bold. \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ .

DV	Background types of neural predictors	Predictors	$\beta$	95% CI	$p$
SRT <sub>SSN</sub>	Quiet + SSN	PC1 <sub>Neural</sub> ( <i>FFR<sub>ENV_F0</sub></i> )	-0.145	[-0.455, 0.165]	0.350
		PC2 <sub>Neural</sub> ( <i>Logit-<math>\theta</math>-PLV</i> )	-0.334	[-0.644, -0.024]	<b>0.036*</b>
SRT <sub>Babble</sub>	Quiet + Babble	PC1 <sub>Neural</sub> ( <i>FFR<sub>ENV_F0</sub></i> )	-0.340	[-0.645, -0.034]	<b>0.030*</b>
		PC2 <sub>Neural</sub> ( <i>Logit-<math>\theta</math>-PLV</i> )	0.210	[-0.095, 0.516]	0.171

**Table 5**

Statistical results for the mediation analyses. The predictors were the first two PCs (PC1<sub>Age\_PTAT</sub> and PC2<sub>Age\_PTAT</sub>) retained via PCAs applied on the age and PTATs. The mediators were the first two PCs (PC1<sub>Neural</sub> and PC2<sub>Neural</sub>) retained via PCAs applied on the neural signatures (obtained under quiet and SSN when predicting SRT<sub>SSN</sub> and under quiet and babble noise when predicting SRT<sub>Babble</sub>). Age/high-frequency PTATs and low-frequency PTATs were largely reflected by PC1<sub>Age\_PTAT</sub> and PC2<sub>Age\_PTAT</sub>, respectively. FFR<sub>ENV\_F0</sub> and Logit- $\theta$ -PLV were largely reflected by PC1<sub>Neural</sub> and PC2<sub>Neural</sub>, respectively. They are indicated correspondingly in the brackets with Italics. DV refers to the dependent variables. All effect sizes were obtained based on the standardized regression coefficients. The corresponding 95% confidence intervals (CI) obtained via bias-corrected and accelerated (BCa) bootstrapping are shown underneath each effect size with square brackets. Significant effects are indicated in bold. \*zero fell outside 95% CI; \*\*zero fell outside 99% CI; \*\*\*zero fell outside 99.9% CI.

DV	Background types of neural signatures	Predictors	Mediation/indirect effects		Direct effects	Total effects
			Mediator 1	Mediator 2		
SRT <sub>SSN</sub>	Quiet + SSN	PC1 <sub>Age_PTAT</sub> (age/high-frequency PTATs)	0.028	<b>-0.155*</b>	<b>0.337*</b>	0.210
		PC2 <sub>Age_PTAT</sub> (low-frequency PTATs)	-0.018	<b>-0.242**</b>	0.227	-0.033
	Quiet + Babble	PC1 <sub>Age_PTAT</sub> (age/high-frequency PTATs)	<b>0.146*</b>	<b>-0.170**</b>	<b>0.507**</b>	<b>0.483***</b>
		PC2 <sub>Age_PTAT</sub> (low-frequency PTATs)	<b>-0.108*</b>	<b>-0.183**</b>	<b>0.706***</b>	<b>0.416**</b>

formances). Taken together, these results imply distinct mediation effects of neural encoding of F<sub>0</sub> and slowly-fluctuating envelopes.

#### 4. Discussion

We studied neural phase-locked encoding of speech with the aims of assessing how early-stage phase-locked activities are affected by ageing and how ageing affects SiN perception *through* these activities. We also aimed to assess whether such processes involve similar or different mechanisms for the encoding of different acoustic properties. These activities include phase-locked responses to F<sub>0</sub> (FFR<sub>ENV\_F0</sub>) that primarily originates from the subcortical level, [Chandrasekaran and Kraus, 2010](#); [Bidelman, 2018](#); [Coffey et al., 2019](#)) and to slowly-fluctuating acoustic envelopes (Logit- $\theta$ -PLV at the cortical level, [Howard and Poeppel, 2010](#)). It is necessary to stress that these activities we focused on reflect the relatively low-level, early-stage phase-locked encoding as participants passively listened to stimuli of a repeated single vowel during the neural assessments. It is also noteworthy that we recruited participants in which older adults had great inter-individual audiometric variability ranging from relatively normal hearing to mild-to-severe hearing loss that resembled the hearing status of typical ageing populations in the real world ([Gopinath et al., 2009](#); [Humes et al., 2010](#)). Based on our results, we will discuss the effects of ageing elements collected from this study (i.e., age and age-related hearing loss) on the phase-locked activities and the potential relationships between these activities and SiN perception. We will then interpret their possible mediation roles in terms of how age and hearing loss affect SiN perception *through* these activities. Finally, limitations of the present study will be discussed for future research.

#### 4.1. Effects of age and hearing loss on phase-locked encoding of speech

The first step of the present study was to address how age and hearing loss affect the phase-locked encoding of speech. As age and hearing loss are highly correlated, it is difficult to dissociate them in participants who were currently recruited to reflect real-world scenarios. We thus applied PCA which previous studies have not applied on the age and hearing loss. We largely separated age/high-frequency ( $\geq 2$  kHz) hearing loss from low-frequency ( $< 2$  kHz) hearing loss and were able to test the different effects of these two components on the phase-locked activities.

First, FFR<sub>ENV\_F0</sub> had smaller magnitudes in participants who had older age and higher degree of high-frequency hearing loss, but greater magnitudes in those with higher degree of low-frequency hearing loss. To our knowledge, this is a finding that has not been reported by previous studies. Previous studies showed declined FFRs in normal-hearing older adults compared to young adults ([Anderson et al., 2012](#); [Presacco et al., 2016, 2019](#)), whilst hearing loss may result in reduced neural inhibition that leads to increased neural responses to F<sub>0</sub>-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., 2016, 2019](#)). The reason that low-, not high-, frequency hearing loss was associated with greater FFR<sub>ENV\_F0</sub> may be that the PCA applied on the age and PTATs was able to largely separate low-frequency hearing loss from age, rather than high-frequency hearing loss from age. The effect of hearing loss that was dissociated from increased age could thus be reflected by the low- rather than high-frequency hearing loss. Therefore, these results indicate that the impact of the ageing process on phase-locked encoding of F<sub>0</sub> could be a combined effect of age-related hearing loss

and the ageing process that is independent of peripheral hearing. We argue that, if this is the case, such ageing process would be a functional brain ageing process. Specifically, timing precision of neural firing could have reduced in the subcortex that would lead to damped phase-locking to  $F_0$  cues following brain ageing (Anderson et al., 2012). In addition, although  $FFR_{ENV\_F0}$  possesses the neural sources that primarily originate from the subcortex (Chandrasekaran and Kraus, 2010; Bidelman, 2018; Coffey et al., 2019), recent research has demonstrated additional sources in the primary auditory cortex (Coffey et al., 2016, 2019; Hartmann and Weisz, 2019; Ross et al., 2020). Therefore, it is possible that contributions of age and hearing loss observed here might also partly come from the cortical level.

Second, Logit- $\theta$ -PLV was greater in participants who had older age and higher degree of hearing loss. This is consistent with previous studies showing that, compared to young adults, older adults have greater cortical auditory-evoked responses (Alain et al., 2014; Herrmann et al., 2013, 2016) and theta-band auditory steady-state responses (ASSR) (Tlumak et al., 2015; Goossens et al., 2016). We argue that the result could reflect the neural hyperexcitability during ageing as phase-locked responses to auditory and speech stimuli at theta rates reflects the neural excitability (in terms of neural firing) in the auditory cortex (Ng et al., 2013). It is suggested that this hyperexcitability may be due to age-related down-regulation in synthesis and release of GABA neurotransmitters leading to reduced cortical inhibition (Goossens et al., 2016). Alternatively, auditory compensation in the cortex for hearing loss may have occurred. It has been reported that, when adults with hearing loss passively listened to a meaningless syllable (a similar paradigm to the present study to test early-stage speech processing), auditory-evoked responses were increased in higher-level cortical regions (frontal cortices) as a possible compensation of decreased responses in the temporal cortex (Campbell and Sharmer, 2013). In this sense, it is necessary for future work to dissociate low- and high-level regions by including more EEG electrodes so that phase-locked activities in these regions can be localized (see 4.4 for further discussions). Also, theta-band phase-locked responses to speech can reflect neural tracking of slowly-fluctuating speech envelopes (Luo and Poeppel, 2007; Peelle et al., 2013). As the stimulus presentation rate used during the current neural recording ( $\sim 5$  vowels per second) approximated the syllable rate (theta rates) in normal speech conversations (Greenberg et al., 2003), Logit- $\theta$ -PLV can thus largely reflect neural processing of speech envelopes at the syllable level as in previous studies (Luo and Poeppel, 2007; Howard and Poeppel, 2010; Peelle et al., 2013). The current result is thus consistent with the finding showing greater cortical tracking of speech envelopes in older adults (Presacco et al., 2016, 2019) and adults with hearing loss (Decruy et al., 2020). It is important to note that, however, in these studies (Presacco et al., 2016, 2019; Decruy et al., 2020), participants' tasks were to actively pay attention to the target speech, while in the present study, participants listened passively to a repeated vowel without higher-order linguistic information. This indicates that the hyperexcitability or compensatory effect occurred at early stages of speech processing. Also, it is unlikely that this was due to increased listening effort in older adults, because the greater Logit- $\theta$ -PLV was found even under quiet background that should not require effortful listening.

In sum, these results imply that the mechanisms by which age and hearing loss affect early-stage phase-locked activities differ between encoding of  $F_0$  and slowly-fluctuating envelopes.

#### 4.2. Potential neural-behavioral relationships

Potential neural-behavioral relationships were demonstrated by testing how the phase-locked activities were correlated with SiN performance under the two types of background noise (SSN and

multi-talker babble noise). We showed that greater  $FFR_{ENV\_F0}$  magnitude was correlated with better SiN performance under babble noise, while greater Logit- $\theta$ -PLV was correlated with better SiN performance under SSN.

These results are in line with previous studies showing that greater  $FFR_{ENV\_F0}$  magnitudes correspond with better SiN performance (Anderson et al., 2011; Fujihira and Shiraishi, 2015; Mai et al., 2018) and that neural tracking of slowly-fluctuating speech envelopes are critical for speech perception (Ahissar et al., 2001; Peelle et al., 2013; Doelling et al., 2014). Our results stresses that the potential relationships for the two types of phase-locked activities differ across noise types. The results indicates that neural encoding of  $F_0$  cues might be more important for SiN perception under babble noise than under SSN. Compared to SSN, babble noise causes an additional informational masking that mainly included acoustic-phonetic interference (N.B., the number of talkers of babble noise was 16, hence no lexical interference, see Hoen et al., 2007). Therefore, neural encoding of  $F_0$  may contribute to extracting target speech information against such interference. On the other hand, cortical tracking of slowly-fluctuating envelopes could be more important for SiN perception under SSN than under babble noise. This echoes another finding in the present study where Logit- $\theta$ -PLV was significantly greater under SSN than babble noise and correspondingly SiN performance was significantly better under SSN than babble noise. These results thus emphasize that Logit- $\theta$ -PLV could be an important neural indicator for SiN perception, especially under SSN.

Despite such findings, it is equally important to note that regression analyses only revealed potential or indirect relationships, i.e., the contributions made by phase-locked encoding to SiN perception are deductive not causal.

#### 4.3. The mediation roles of phase-locked encoding

Although we had demonstrated how age and hearing loss affected neural phase-locked activities and the relationships between these activities and SiN perception, it had still not been clarified how age and hearing loss may affect SiN perception through these activities. Understanding this should further elucidate the contributions that the effects of age/hearing loss on phase-locked activities made to SiN perception. Therefore, we further conducted mediation analyses to address this issue.

The results showed significant mediation effects (with medium to large effect sizes) of both  $FFR_{ENV\_F0}$  and Logit- $\theta$ -PLV but with different mechanisms. Specifically,  $FFR_{ENV\_F0}$  decreased with age/high-frequency hearing loss which in turn contributed to poorer SiN performances (hence facilitating the effect of age/high-frequency hearing loss that made the performances even worse) but increased with low-frequency hearing loss which in turn contributed to better SiN performances (hence suppressing the effect of low-frequency hearing loss that resisted worsening of the performances) under babble noise. Logit- $\theta$ -PLV, on the other hand, increased with age and hearing loss which in turn contributed to better SiN performances (hence suppressing the effect of age and hearing loss that resisted worsening of the performances) under both SSN and babble noise. These results are in line with the effects of age and hearing loss on the neural signatures (see 3.1 and 4.1) and the potential neural-behavioral relationships where  $FFR_{ENV\_F0}$  and Logit- $\theta$ -PLV positively correlated with SiN performances (see 3.2 and 4.2). Note that although significant correlation between Logit- $\theta$ -PLV and SiN performance was present only under SSN, significant mediation of Logit- $\theta$ -PLV was found under both SSN and babble noise. This may be because the previous neural-behavioral analyses were conducted to show the overall relationship between Logit- $\theta$ -PLV and SiN perception without considering age and hearing loss, while the mediation analyses considered how

Logit- $\theta$ -PLV mediated the effect of age and hearing loss on SiN perception.

The facilitative mediation of FFR<sub>ENV\_F0</sub> for the effect of age/high-frequency hearing loss is consistent with the arguments that decreased FFR following ageing may be a cause for difficulties during SiN perception in older adults (Anderson et al., 2012). Such effect occurred only under babble noise probably due to the possible role of neural processing of F<sub>0</sub> cues in resisting the acoustic-phonetic informational masking under babble noise as mentioned in the previous discussions (see 4.2). On the other hand, the suppressive mediation of FFR<sub>ENV\_F0</sub> and Logit- $\theta$ -PLV indicates that neural changes following age/hearing loss may not necessarily contribute to impaired SiN performances, but could also help alleviate SiN difficulties. Our results thus had not provided evidence for the claim that increased cortical phase-locking to speech envelopes following ageing can lead to imbalance between neural excitation and inhibition that in turn impairs SiN perception (Presacco et al., 2016). An argument by Presacco et al. (2016) was that the excitation-inhibition imbalance may impair SiN perception through the decreased neural resources of higher-level cognitive functions. It would thus be worthwhile for future work to study how such increased cortical activities relate to changes in higher-level cognitive functions that are important for SiN perception (data of which the present study had not collected, see 4.4 for discussions on limitations).

The mediation results thus indicate, for the first time according to our knowledge, the distinct mechanisms that underlie how age and hearing loss affect SiN perception through early-stage neural phase-locked encoding of different speech acoustic properties.

#### 4.4. Limitations and future work

First, participants listened to a repeated single vowel during the neural assessments, which is a conventional paradigm to capture FFR in human participants (Aiken and Picton, 2008; Skoe and Kraus, 2010). Because of the passive listening and the lack of higher-level linguistic features (semantic/syntactic information) in the stimuli, the phase-locked activities should therefore mainly reflect early-stage processing, consistent with our current focus. It is noteworthy that this may reflect processing of speech-specific properties (phonetic and phonological cues of the vowel /i/) and/or non-speech-specific phase-locked encoding of complex acoustic properties (F<sub>0</sub> of the harmonic complexes and the theta-band amplitude envelope modulations). Due to the focus on the early-stage phase-locked activities, participants' higher-level cognitive functions had not been included in the current screening plan. However, age-related cognitive declines would be important for further explaining the impacts of ageing on SiN perception (Slade et al., 2020). This is reflected by the results of the mediation analyses showing significant direct effects of age and hearing loss (see 3.3 and Table 5), indicating that age and hearing loss may have also affected SiN perception through other age-related factors besides early-stage phase-locked activities, such as higher-level cognitive functions. Therefore, mediation of phase-locked activities should not dominate the explanation for the way in which age and hearing loss affect SiN perception. The cognitive functions related to age/hearing loss and SiN perception include working memory and attention (Lin et al., 2013). Working memory capacity is related to SiN perception in older adults (Schoof and Rosen, 2014), while attention is critical for suppressing neural sensitivity to background noise (Rimmele et al., 2015) and such ability deteriorates during ageing (Andrés et al., 2006). Working memory and attention also contribute to resisting informational masking caused by babble noise (Schneider et al., 2007; Shinn-Cunningham and Best, 2008). Furthermore, other cognitive factors like Mild Cognitive Impairment (MCI) could occur in some older adults (Petersen et al., 1999)

that could influence neural processing of speech and speech perception (Bidelman et al., 2017). It is not clear how these factors may affect mediation of the phase-locked activities. Future studies should include screening of cognitive functions to better model how they and phase-locked activities together affect SiN perception following ageing. In addition, future work could use simultaneous neural and behavioural assessments (i.e., neural recordings conducted concurrently during the SiN perception tasks) which would better assess cortical phase-locked activities during processing at higher-level cognitive stages and inform how such higher-level processing may be related to SiN performances.

Second, the cortical phase-locked activity was measured with two electrode sites (C3 and C4; see 2.2.2). A whole-head configuration of EEG electrodes may help to better localize (e.g., via source localization with high-density EEGs) where this activity takes place besides the presumptive early-stage region of auditory cortex. It was suggested that adults with hearing loss tend to recruit larger or higher-level cortical networks in the brain even for early-stage auditory speech processing during passive listening (Campbell and Sharmer, 2013). The whole-head EEG may also help to reveal hemispheric laterality. One can hypothesize that dedifferentiation of laterality may occur following age and/or hearing loss due to the greater recruitment of larger cortical networks to aid speech processing. Future work on this in coordination with mediation analyses could further our understanding of how phase-locked activities in higher-level cortical regions (even during early processing stages) may be affected by age and hearing loss that contribute to SiN perception.

Finally, different types of hearing loss may have additional influences. Here, we tested hearing loss via air-conduction. Audiograms for majority of older participants showed hearing loss at high frequencies ( $\geq 2$  kHz), consistent with the pattern of presbycusis which is usually sensorineural hearing loss caused by age-related declines in the inner ear (Huang and Tang, 2010). Despite this, we cannot exclude that some participants may also suffer from conductive hearing loss (tested via bone-conduction). Furthermore, conductive hearing loss is often related to low-frequency hearing loss (Declau et al., 2007). As such, it is interesting to look into how conductive hearing loss may contribute to phase-locked encoding of speech (esp. FFR<sub>ENV\_F0</sub> which was shown to be differently affected by low- and high-frequency hearing loss in the present study). Therefore, future work could further include bone-conduction to better clarify the roles of different types of hearing loss.

#### 4.5. Summary and conclusion

The present study investigated early-stage neural phase-locked responses to speech acoustic properties (F<sub>0</sub> and slowly-fluctuating envelopes) in human adults across a wide age range. The aims were to address how these phase-locked activities are affected by ageing that may contribute to speech-in-noise (SiN) perception and whether mechanisms that underlie this process differ between neural processing of these two types of speech properties.

New approaches and insights were brought by this study. First, compared to previous studies, older adults recruited in this study had great inter-individual variability of audiometric hearing. Participants were thus more representative to reflect the hearing status of typical ageing populations in real life. Second, we applied PCA which largely separated age/high-frequency ( $\geq 2$  kHz) hearing loss from low-frequency ( $< 2$  kHz) hearing loss. We showed that the effects of these two components on phase-locked activities differed for the two types of phase-locked activities, the results of which have not been reported by previous studies. Specifically, FFR to F<sub>0</sub> cues decreased with age/high-frequency hearing loss but increased with low-frequency hearing loss, whilst

phase-locking to slowly-fluctuating envelopes increased with both age/high-frequency hearing loss and low-frequency hearing loss. Finally, we investigated how age and hearing loss affect SiN perception *through* the phase-locked activities via mediation analysis, an advanced approach that has not been employed so far for studying neural mechanisms underlying SiN perception during ageing. We showed that FFR had both facilitative and suppressive mediation whilst phase-locking to slowly-fluctuating envelopes had suppressive mediation for the effects of age and hearing loss on SiN perception. The suppressive mediation thus indicates that effects on neural activities following ageing may not necessarily lead to worse SiN perception as expected, but could also help alleviate SiN difficulties. Moreover, we showed different mediation effects under different types of background noise. Significant mediation of FFR occurred only under babble noise compared to mediation of slow envelope phase-locking under both SSN and babble noise, indicating the possible specific role of  $F_0$  encoding for resisting information masking in multi-talker babbles compared to the potential importance of encoding slowly-fluctuating envelopes for SiN perception regardless of whether information masking is present or not.

Taken together, our results demonstrate possible mechanisms through which age and hearing loss affect the early-stage phase-locked encoding of different speech acoustic properties and SiN perception. As the FFR has primary sources from the subcortex with neural phase-locking to slowly-fluctuating envelopes occurring at the cortex, the results could reflect distinct mechanisms between the subcortical and cortical levels, although cortical contribution of FFR needs to be clarified in the future. The present study should therefore further our understanding of the possible role of phase-locked neural activities for SiN perception during ageing. Future work may study how age-related declines in higher-level cognitive functions contribute to these mechanisms.

## Declaration of Competing Interest

The authors declare no competing conflicts of interest.

## CRediT authorship contribution statement

**Guangting Mai:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization, Funding acquisition. **Peter Howell:** Conceptualization, Methodology, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

## Data Availability

Data will be made available on request.

## Acknowledgements

The experiments were conducted with facilities support in the EEG laboratories based at UCL Department of Speech, Hearing and Phonetic Sciences. We thank Mr Andrew Clark for technical support. The first author was financially supported by the UCL Graduate Research Scholarship for Cross-Disciplinary Training program.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.heares.2022.108647](https://doi.org/10.1016/j.heares.2022.108647).

## References

Ahissar, E, Nagarajan, S, Ahissar, M, Protopapas, A, Mahncke, H, Merzenich, MM, 2001. Speech comprehension is correlated with temporal response patterns recorded from auditory cortex. *Proc. Natl Acad. Sci.* 98 (23), 13367–13372.

Aiken, SJ, Picton, TW, 2008. Envelope and spectral frequency-following responses to vowel sounds. *Hear. Res.* 245, 35–47.

Alain, C, Roye, A, Salloum, C, 2014. Effects of age-related hearing loss and background noise on neuromagnetic activity from auditory cortex. *Front. Syst. Neurosci.* 8 (8), 1–12.

Anderson, S, Parbery-Clark, A, Yi, HG, Kraus, N, 2011. A neural basis of speech-in-noise perception in older adults. *Ear Hear.* 32, 750.

Anderson, S, Parbery-Clark, A, White-Schwoch, T, Kraus, N, 2012. Aging affects neural precision of speech encoding. *J. Neurosci.* 32, 14156–14164.

Anderson, S, Parbery-Clark, A, White-Schwoch, T, Drehobl, S, Kraus, N, 2013. Effects of hearing loss on the subcortical representation of speech cues. *J. Acoust. Soc. Am.* 133 (5), 3030–3038.

Andrés, P, Parmentier, FB, 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, 2564–2568.

Barr, DJ, Levy, R, Scheepers, C, Tily, HJ, 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J. Memory Language* 68 (3), 255–278.

Bench, J, Kowal, Á, Bamford, J, 1979. The BKB (Bamford-Kowal-Bench) sentence lists for partially-hearing children. *Br. J. Audiol.* 13, 108–112.

Bidelman, GM, 2018. Subcortical sources dominate the neuroelectric auditory frequency-following response to speech. *Neuroimage* 175, 56–69.

Bidelman, GM, Lowther, JE, Tak, SH, Alain, C, 2017. Mild cognitive impairment is characterized by deficient brainstem and cortical representations of speech. *J. Neurosci.* 37 (13), 3610–3620.

Boersma P, Weenink D (2013) Praat: doing phonetics by computer [Computer program] Version 5.3.51.

Brown, RE, Basheer, R, McKenna, JT, Strecker, RE, McCarley, RW, 2012. Control of sleep and wakefulness. *Physiol. Rev.* 92, 1087–1187.

Coffey, EB, Herholz, SC, Chepeliuk, AM, Baillet, S, Zatorre, RJ, 2016. Cortical contributions to the auditory frequency-following response revealed by MEG. *Nat. Commun.* 7 (1), 1–11.

Coffey, EB, Nicol, T, White-Schwoch, T, Chandrasekaran, B, Krizman, J, Skoe, E, Zatorre, RJ, Kraus, N, 2019. Evolving perspectives on the sources of the frequency-following response. *Nat. Commun.* 10 (1), 1–10.

Campbell, J, Sharma, A, 2013. Compensatory changes in cortical resource allocation in adults with hearing loss. *Front. Syst. Neurosci.* 7, 71.

Carpenter, AL, Shahin, AJ, 2013. Development of the N1–P2 auditory evoked response to amplitude rise time and rate of formant transition of speech sounds. *Neurosci. Lett.* 544, 56–61.

Chandrasekaran, B, Kraus, N, 2010. The scalp-recorded brainstem response to speech: neural origins and plasticity. *Psychophysiology* 47, 236–246.

Cohen, J, 1992. Statistical power analysis. *Curr. Dir. Psychol. Sci.* 1 (3), 98–101.

Declau, F, Van Spaendonck, M, Timmermans, JP, Michaels, L, Liang, J, Qiu, JP, Van de Heyning, P, 2007. Prevalence of histological otosclerosis: an unbiased temporal bone study in Caucasians. In: *Otosclerosis and Stapes Surgery*. Karger Publishers, pp. 6–16 Vol. 65.

Decruy, L, Vanthornhout, J, Francart, T, 2020. Hearing impairment is associated with enhanced neural tracking of the speech envelope. *Hear. Res.* 393, 107961.

Doelling, KB, Arnal, LH, Ghita, O, Poeppel, D, 2014. Acoustic landmarks drive delta-theta oscillations to enable speech comprehension by facilitating perceptual parsing. *Neuroimage* 85, 761–768.

Efron, B, 1987. Better bootstrap confidence intervals. *J. Am. Statist. Assoc.* 82 (397), 171–185.

Efron, B, Tibshirani, RJ, 1994. *An Introduction to the Bootstrap*. CRC press.

Fujihira, H, Shiraishi, K, 2015. Correlations between word intelligibility under reverberation and speech auditory brainstem responses in elderly listeners. *Clin. Neurophysiol.* 126, 96–102.

Gransier, R, Hofmann, M, van Wieringen, A, Wouters, J, 2021. Stimulus-evoked phase-locked activity along the human auditory pathway strongly varies across individuals. *Sci. Rep.* 11 (1), 1–13.

Greenberg, S, Carvey, H, Hitchcock, L, Chang, S, 2003. Temporal properties of spontaneous speech—a syllable-centric perspective. *J. Phonetics* 31, 465–485.

Goossens, T, Vercammen, C, Wouters, J, Wieringen, AV, 2016. Aging affects neural synchronization to speech-related acoustic modulations. *Front. Aging Neurosci.* 8, 133.

Goossens, T, Vercammen, C, Wouters, J, van Wieringen, A, 2018. Neural envelope encoding predicts speech perception performance for normal-hearing and hearing-impaired adults. *Hear. Res.* 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. *Neurobiol. Aging* 74, 202–212.

Gopinath, B, Rochtchina, E, Wang, JJ, Schneider, J, Leeder, SR, Mitchell, P, 2009. Prevalence of age-related hearing loss in older adults: blue Mountains Study. *Arch. Intern. Med.* 169, 415–418.

Hartmann, T, Weisz, N, 2019. Auditory cortical generators of the Frequency Following Response are modulated by intermodal attention. *Neuroimage* 203, 116185.

Henry, KS, Heinz, MG, 2012. Diminished temporal coding with sensorineural hearing loss emerges in background noise. *Nat. Neurosci.* 15 (10), 1362–1364.

Henry, KS, Kale, S, Heinz, MG, 2014. Noise-induced hearing loss increases the temporal precision of complex envelope coding by auditory-nerve fibers. *Front. Syst. Neurosci.* 8, 20.

Herrmann, B, Henry, MJ, Scharinger, M, Obleser, J, 2013. Auditory filter width affects response magnitude but not frequency specificity in auditory cortex. *Hear. Res.* 304, 128–136.

Herrmann, B, Henry, MJ, Johnsrude, IS, Obleser, J, 2016. Altered temporal dynamics of neural adaptation in the aging human auditory cortex. *Neurobiol. Aging* 45, 10–22.

Hoentjen, M, Meunier, F, Grataloup, CL, Pellegrino, F, Grimault, N, Perrin, F, Perrot, X, Collet, L, 2007. Phonetic and lexical interferences in informational masking during speech-in-speech comprehension. *Speech Commun.* 49 (12), 905–916.

Howard, MF, Poeppel, D, 2010. Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. *J. Neurophysiol.* 104, 2500–2511.

Humes, LE, Kewley-Port, D, Fogerty, D, Kinney, D, 2010. Measures of hearing threshold and temporal processing across the adult lifespan. *Hear. Res.* 264, 30–40.

Huang, Q, Tang, J, 2010. Age-related hearing loss or presbycusis. *Eur. Arch. Otorhinolaryngol.* 267 (8), 1179–1191.

Kale, S, Heinz, MG, 2010. Envelope coding in auditory nerve fibers following noise-induced hearing loss. *J. Assoc. Res. Otolaryngol.* 11, 657–673.

Kim, A, Latchoumane, C, Lee, S, Kim, GB, Cheong, E, Augustine, GJ, Shin, HS, 2012. Optogenetically induced sleep spindle rhythms alter sleep architectures in mice. *Proc. Natl. Acad. Sci.* 109 (50), 20673–20678.

Krizman, J, Kraus, N, 2019. Analyzing the FFR: a tutorial for decoding the richness of auditory function. *Hear. Res.* 382, 107779.

Lakatos, P, Chen, CM, O'Connell, MN, Mills, A, Schroeder, CE, 2007. Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron* 53, 279–292.

Lin, FR, Yaffe, K, Xia, J, Xue, QL, Harris, TB, Purchase-Helzner, E, Satterfield, S, Ayonayon, HN, Ferrucci, L, Simonsick, EM, 2013. Health ABC Study Group F. Hearing loss and cognitive decline in older adults. *JAMA Intern Med* 173, 293–299.

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

MacKinnon, DP, Fairchild, AJ, Fritz, MS, 2007. Mediation analysis. *Annu. Rev. Psychol.* 58, 593–614.

MacKinnon, DP, Krull, JL, Lockwood, CM, 2000. Equivalence of the mediation, confounding and suppression effect. *Prev. Sci.* 1 (4), 173–181.

MacLeod, A, Summerfield, Q, 1990. A procedure for measuring auditory and audio-visual speech-reception thresholds for sentences in noise: Rationale, evaluation, and recommendations for use. *British Journal of Audiology* 24 (1), 29–43.

Mai, G, Howell, P, 2022. Causal relationship between the right auditory cortex and speech-evoked envelope-following response: evidence from combined transcranial stimulation and electroencephalography. *Cereb. Cortex* 298, 1–18.

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.

Mai, G, Tuomainen, J, Howell, P, 2018. Relationship between speech-evoked neural responses and perception of speech in noise in older adults. *J. Acoust. Soc. Am.* 143, 1333–1345.

McNeish, D, Bauer, DJ, 2020. Reducing incidence of nonpositive definite covariance matrices in mixed effect models. *Multivariate Behav. Res.* 57, 318–340.

Morillon, B, Liégeois-Chauvel, C, Arnal, LH, Bénar, CG, Giraud, AL, 2012. Asymmetric function of theta and gamma activity in syllable processing: an intra-cortical study. *Front. Psychol.* 3, 248.

Ng, BSW, Logothetis, NK, Kayser, C, 2013. EEG phase patterns reflect the selectivity of neural firing. *Cereb. Cortex* 23 (2), 389–398.

Noguchi, Y, Fujiwara, M, Hamano, S, 2015. Temporal evolution of neural activity underlying auditory discrimination of frequency increase and decrease. *Brain Topogr.* 28, 437–444.

Oya, H, Gander, PE, Petkov, CI, Adolphs, R, Nourski, KV, Kawasaki, H, Howard, MA, Griffiths, TD, 2018. Neural phase locking predicts BOLD response in human auditory cortex. *Neuroimage* 169, 286–301.

Peelle, JE, Gross, J, Davis, MH, 2013. Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cereb. Cortex* 23, 1378–1387.

Petersen, RC, Smith, GE, Waring, SC, Ivnik, RJ, Tangalos, EG, Kokmen, E, 1999. Mild cognitive impairment: clinical characterization and outcome. *Arch. Neurol.* 56 (3), 303–308.

Plomp, R, Mimpen, AM, 1979. Speech-reception threshold for sentences as a function of age and noise level. *J. Acoust. Soc. Am.* 66, 1333–1342.

Preacher, KJ, Kelley, K, 2011. Effect size measures for mediation models: quantitative strategies for communicating indirect effects. *Psychol. Methods* 16 (2), 93–115.

Presacco, A, Simon, JZ, Anderson, S, 2016. Evidence of degraded representation of speech in noise, in the aging midbrain and cortex. *J. Neurophysiol.* 116, 2346–2355.

Presacco, A, Simon, JZ, Anderson, S, 2019. Speech-in-noise representation in the aging midbrain and cortex: Effects of hearing loss. *PLoS One* 14 (3), e0213899.

Rimmele, JM, Golumbic, EZ, 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.

Rosen, S, Souza, P, Ekelund, C, Majeed, AA, 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, Tremblay, KL, Alain, C, 2020. Simultaneous EEG and MEG recordings reveal vocal pitch elicited cortical gamma oscillations in young and older adults. *Neuroimage* 204, 116253.

Schneider, BA, Li, L, Daneman, M, 2007. How competing speech interferes with speech comprehension in everyday listening situations. *J. Am. Acad. Audiol.* 18 (7), 559–572.

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

Schoof, T, Rosen, S, 2016. The role of age-related declines in subcortical auditory processing in speech perception in noise. *J. Assoc. Res. Otolaryngol.* 17, 441–460.

Shinn-Cunningham, BG, Best, V, 2008. Selective attention in normal and impaired hearing. *Trends Amplif.* 12 (4), 283–299.

Shrout, PE, Bolger, N, 2002. Mediation in experimental and nonexperimental studies: new procedures and recommendations. *Psychol. Methods* 7 (4), 422.

Skoe, E, Kraus, N, 2010. Auditory brainstem response to complex sounds: a tutorial. *Ear Hear.* 31, 302.

Slade, K, Plack, CJ, Nuttal, HE, 2020. The effects of age-related hearing loss on the brain and cognitive function. *Trends Neurosci.* 43 (10), 810–821.

Thatcher, RW, 2012. Coherence, phase differences, phase shift, and phase lock in EEG/ERP analyses. *Dev. Neuropsychol.* 37, 476–496.

Tlumak, AI, Durrant, JD, Delgado, RE, 2015. The effect of advancing age on auditory middle-and long-latency evoked potentials using a steady-state-response approach. *Am. J. Audiol.* 24, 494–507.

Waschke, L, Wöstmann, M, Obleser, J, 2017. States and traits of neural irregularity in the age-varying human brain. *Sci. Rep.* 7, 17381.

Yong, AG, Pearce, S, 2013. A beginner's guide to factor analysis: focusing on exploratory factor analysis. *Tutorials Quantitat. Methods Psychol.* 9 (2), 79–94.

Zhong, Z, Henry, KS, Heinz, MG, 2014. Sensorineural hearing loss amplifies neural coding of envelope information in the central auditory system of chinchillas. *Hear. Res.* 309, 55–62.