

Modulation of phase-locked neural responses to speech during different arousal states is age-dependent

Guangting Mai¹, Tim Schoof¹, Peter Howell^{1*}

¹Division of Psychology and Language Sciences, University College London, London UK

*Corresponding author (p.howell@ucl.ac.uk)

Running title: Arousal state and phase locked responses to speech

Number of pages: 29

Number of Figures: 8

Number of Tables: 2

Conflict of Interest: The authors declare no competing conflicts of interest.

Running head

Abstract

Phase-locked responses are vital for auditory perception and they may vary with participants' arousal state and age. Two phase-locked neural responses that reflect fine-grained acoustic properties of speech were examined in the current study: the frequency-following response (FFR) to the speech fundamental frequency (F_0), which originates primarily from the auditory brainstem, and the theta-band phase-locked response (θ -PLV) to the speech envelope that originates from the auditory cortices. The ways these responses were affected by arousal in adults across a wide age-range (19 ~ 75 years) were examined. Extracts from electroencephalographic (EEG) responses to repeated syllables were classified into either high or low arousal state based on the occurrence of sleep spindles. The magnitudes of both FFRs and θ -PLVs were statistically greater in the high, than in the low, arousal state. The difference in θ -PLV between the two arousal states was significantly associated with sleep spindle density in the young, but not the older, adults. The results show that (1) arousal affects phase-locked processing of speech at cortical/sub-cortical sensory levels; and that (2) there is an interplay between aging and arousal state which indicates that sleep spindles have an age-dependent neuro-regulatory role on cortical processes. The results lay the grounds for studying how cognitive states affect early-stage neural activity in the auditory system across the lifespan.

Keywords: Frequency following responses (FFR), theta-band phase-locked responses, electroencephalography (EEG), arousal states, sleep spindles, aging

1 Introduction

The frequency-following response (FFR) ([Worden and Marsh, 1968](#), [Moushegian et al., 1973](#)) and oscillatory activity in the theta-band (4 ~ 8 Hz) ([Luo and Poeppel, 2007](#); [Howard and Poeppel, 2010](#)) are two neural responses that can phase-lock to components of speech stimuli. FFRs to speech reflect phase-locked processing of fundamental frequency (F0) ([Skoe and Kraus, 2010](#)) and arise primarily at the level of the brainstem ([Chandrasekaran and Kraus, 2010](#); [Bidelman, 2018](#); but see [Coffey et al., 2016](#)). FFRs have important roles in pitch ([Wong et al., 2007](#); [Bidelman et al., 2011](#)), and speech-in-noise perception ([Song et al., 2010](#); [Coffey et al., 2017](#)). Theta-band phase-locked responses (θ -PLV) reflect processing of speech envelopes ([Luo and Poeppel, 2007](#); [Howard and Poeppel, 2010](#)). They originate in the auditory cortex ([Luo and Poeppel, 2007](#); [Peelle et al., 2013](#)) and are associated with speech intelligibility ([Peelle et al., 2013](#); [Mai et al., 2018](#)). These phase-locked auditory responses change during aging. As age increases, FFR magnitudes decline ([Anderson et al., 2012](#)) whilst θ -PLVs in response to amplitude-modulated tones increase ([Tlumak et al., 2015](#); [Goossens et al., 2016](#)).

Auditory signals are processed by the sleeping brain ([Issa and Wang, 2008](#); [Nir et al., 2015](#)). However, neural responses to speech in subcortical ([Portas et al., 2000](#)) and cortical ([Czisch et al., 2002, 2004](#); [Wilf et al., 2016](#)) auditory regions reduce during sleep compared to wakefulness. Phase-locked responses to complex auditory signals change in different arousal states. For example, [Makov et al. \(2017\)](#) examined relationships between episodes designated as wakefulness, non-Rapid Eye Movement (nREM) and REM and processing at different linguistic levels. EEG phase-locked responses at rates corresponding to higher-order linguistic units (words, phrases and sentences) were statistically greater in wakefulness than in sleep, but not at the rates corresponding to those of lower-order units (syllables). Currently evidence is lacking concerning whether phase-locked auditory responses to fine-grained speech acoustic properties (such as F0 and speech envelopes) are affected by arousal state at subcortical and cortical levels.

Sleep spindles can be used to locate episodes where arousal is low and to indicate whether and when arousal state changes within EEG sessions. Sleep spindles are bursts of oscillatory neural activity at frequencies of 12 ~ 16 Hz ([Warby et al., 2014](#)) that last for 0.5 s minimum ([De Gennaro and Ferrara, 2003](#)). Spindles are transmitted to the cortex over thalamo-thalamic and thalamo-cortical loops where they modulate neural sensitivity to external auditory stimuli

(Dang-Vu et al., 2011; Schabus et al., 2012). Aging affects the magnitude, duration and density of sleep spindles during nREM sleep (Martin et al., 2013; Mander et al., 2017). Consequently, spindles should neuromodulate auditory phase-locked responses at subcortical and cortical levels (FFR and θ -PLV) and the neuromodulation could differ across ages. Spindle activity within recording sessions, such as spindle density, has been used previously to indicate arousal state and sleep stability (Kim et al., 2012). However, such activity has not been used to explore the effects on phase-locked responses in speech perception at cortical and sub-cortical levels across age groups.

Here the links between arousal, sleep spindle density and speech-evoked phase-locked activity (FFRs and θ -PLV) in human adults across a wide age range (19 ~ 75 years old) were assessed. The predictions were: (1) State of arousal affects magnitudes of FFRs and θ -PLV; (2) FFR magnitude should decline with age whilst θ -PLV should increase; (3) Sleep spindle activity (density, magnitude and duration) should decline with age; (4) Arousal effects may covary with spindle density within the different age groups and for the different phase-locked measures.

2 Materials and Methods

2.1 Participants

Data from Schoof and Rosen (2016) and Mai et al. (2018) (Exp 1 and Exp 2) were used. Participants in both experiments were native English speakers who had no history of language or neurological disorders. In Exp 1 there were 20 young (19 ~ 29 yrs; Mean \pm SD = 23.7 \pm 2.9 yrs; 10 males) and 20 older adults (60 ~ 72 yrs; Mean \pm SD = 64.1 \pm 3.3 yrs; 3 males). They all had near-normal hearing defined as pure-tone thresholds \leq 25 dB between 0.25 and 4 kHz in both ears and at 6 kHz in at least one ear. In Exp 2 there were 23 young (19 ~ 42 yrs; Mean \pm SD = 26.3 \pm 5.5 yrs; 15 males) and 35 older adults¹ (53 ~ 75 yrs; Mean \pm SD = 67.6 \pm 5.1 yrs; 12 males). Pure-tone audiometric thresholds (PTA) were measured via a MAICO MA41 Audiometer (MAICO Diagnostics, Germany) in a sound-attenuating booth in both experiments. All young participants had normal hearing (thresholds \leq 25 dB) from 0.25 to 8 kHz in both ears except for one whose pure-tone thresholds on the left ear were 35 and 45 dB

¹ There were 47 older participants in total in Exp 2. This included 12 hearing aid users and 35 participants who did not use hearing aids (Mai et al., 2018). Hearing aid users were excluded from the present study.

at 6 and 8 kHz. For older adults, 27 out of 35 had normal hearing at low frequencies (≤ 1 kHz) but PTAs ranged from normal hearing to severe hearing loss at high frequencies (2 to 8 kHz). **Figure 1** gives individual PTAs separately for each experiment.

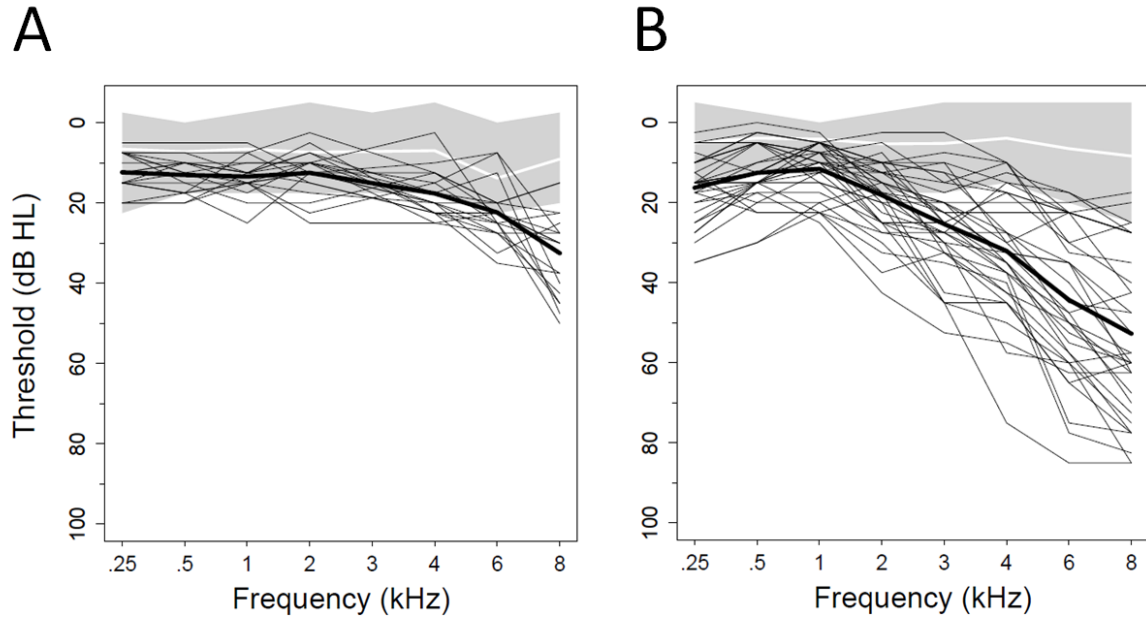


Figure 1. Individual audiograms for (A) Exp 1 and (B) Exp 2. Individual participant data were averaged across ears. The grey shaded areas and white lines represent the ranges and average thresholds for the young adults. The grey and bold lines represent the individual and average thresholds of the older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

2.2 Stimuli

In Exp 1, EEG was recorded in response to repeated presentations of a 100 ms /a/ vowel presented with an inter-stimulus interval (ISI) of 100 ms (5 syllables per second). The vowel had a flat fundamental frequency (F_0) at 160 Hz and F_1 , F_2 , F_3 and F_4 were at 710, 1200, 2900 and 3400 Hz, respectively (**Figure 2A**). The stimulus was ramped on/off for 6.25 ms with a cosine window. The syllables were presented at 80 dB SPL binaurally via electrically shielded ER-3 insert earphones (Intelligent Hearing Systems, Miami, FL).

In Exp 2, the stimuli were repeated presentations of a 120 ms /i/ and ISIs varied randomly between 60 and 120 ms (~ 4.8 syllables per second). The vowel had an F_0 contour that dropped from 160 to 110 Hz. F_1 , F_2 and F_3 were approximately 280, 2400 and 3100 Hz, respectively (**Figure 2B**). The vowel was ramped on and off with a 5 ms cosine window. The stimuli were presented over a Rogers LS3/5A loudspeaker (Falcon Acoustics, UK). The intensity at a distance of 1 metre from the loudspeaker at 0 degrees azimuth, which corresponded to where participants' heads were located during measurements, was 77 dB SPL.

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

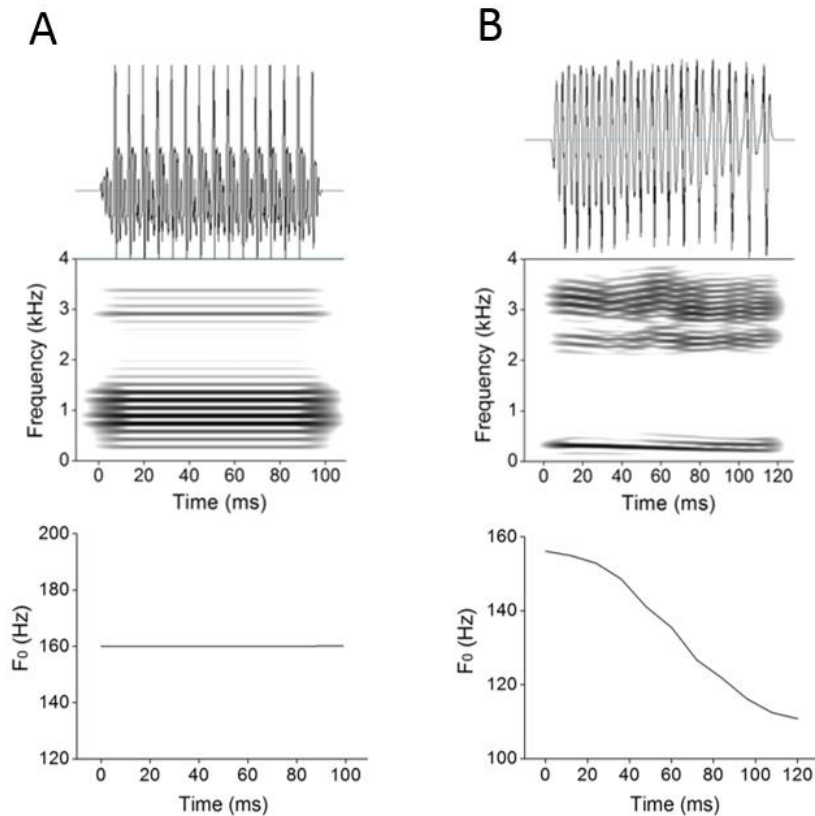


Figure 2. Vowel stimuli for (A) Exp 1 (/a/ with an F_0 at 160 Hz) and (B) Exp 2 (/i/ with an F_0 falling from 160 to 110 Hz). Top, middle and bottom panels show the acoustic waveforms, narrow-band spectrograms, and the F_0 contours respectively.

2.3 EEG recording procedure

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

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

2.4 Classification of arousal states

Subsequent analyses of EEG signals were conducted using Matlab R2014a (Mathworks, USA). Sleep spindles in the EEG sigma frequency band (12 ~ 16 Hz) were used as signatures of Stage 2 nREM sleep (Warby et al., 2014) using a method adapted from Martin et al. (2013). EEGs at Cz were filtered into alpha (8 ~ 11 Hz), sigma (12 ~ 16 Hz) and beta (17 ~ 20 Hz) bands using a 2nd-order zero-phase Butterworth filter. Then the filtered signals were divided into 250-ms-long successive segments. A spindle was labelled when the following criteria were met: (1) root-mean-square (RMS) voltage in the sigma band in a given segment exceeded the threshold of the 95th percentile of the sigma RMS of all segments; (2) RMS of the sigma band was higher than both alpha and beta RMS in the current segment; (3) two successive segments met both criteria (1) and (2). (1) and (2) were invoked because dominance of the sigma-band in the spectrum is the major characteristic of sleep spindles (Martin et al., 2013; Warby et al., 2014). The requirement to extend across two segments was included because sleep spindles usually last for at least 500 ms (De Gennaro and Ferrara, 2003).

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

The epochs were then classified into high arousal, low arousal, and transition between high and low arousal states. Low arousal epochs were those that contained at least one sleep spindle. High arousal epochs were those that contained no spindles and were not adjacent to an epoch with a sleep spindle. High arousal epochs approximate to wakefulness or nREM Stage 1, whilst low arousal epochs approximate to nREM Stage 2. Transition epochs were discarded.

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

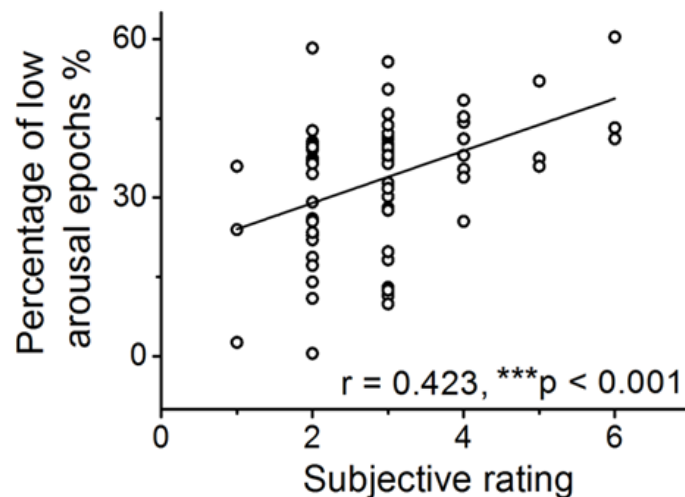


Figure 3. Correlations of the subjective rating of sleepiness with the percentage of epochs classified as low arousal in Exp 2. Higher ratings indicate higher levels of sleepiness.

2.5 Sleep spindle parameters

Density, magnitude and duration of sleep spindles were calculated. Spindle density was the number of spindles per minute averaged across the low arousal states². Magnitude of each spindle was quantified as the maximum power value in the Hilbert envelope during spindle activity. Spindle duration was the time between the start- and end-point values at half spindle magnitude in the amplitude envelope. Participants who had fewer than five epochs classified as the low arousal states (those that contained spindles) were excluded from these analyses. This left 91 participants (38 young and 53 older adults, i.e., 5 young and 2 older adults were removed).

2.6 Frequency-following responses (FFRs)

Baseline was adjusted using the 40-ms pre-stimulus period. EEGs were re-referenced to the average of bilateral earlobes and bandpass filtered between 70 and 2000 Hz using a zero-phase 2nd-order Butterworth filter. Sweeps exceeding $\pm 25 \mu\text{V}$ were rejected to reduce the incidence of movement artefacts (Schoof and Rosen, 2016; Mai et al., 2018). FFRs with positive (FFR_{pos}) and negative (FFR_{neg}) polarities were obtained by averaging across sweeps with their respective polarities. In Exp 1, FFR magnitudes were quantified as the magnitude along the F₀ trajectory of the /a/ vowel (160 Hz) based on either FFR_{pos} or FFR_{neg}. In Exp 2, FFR magnitudes were quantified as the magnitude along the F₀ trajectory of the /i/ vowel (160 ~ 110 Hz) using the waveform resulting from addition of FFR_{pos} and FFR_{neg} that was then divided by 2 (Aiken and Picton, 2008). The procedure that added alternate polarities minimized fine structure temporal information at the auditory periphery (i.e. the cochlear microphonic) and emphasized the processing of envelope cues at the brainstem (Aiken and Picton, 2008). The addition step was not conducted in Exp 1 because the sequential order of the polarities (positive polarity was followed by negative polarity) led to different temporal distributions of the two polarities. This raises the possibility that the magnitudes of FFR_{pos} and FFR_{neg} may differ because neural adaptation effects differ across the two polarities³. See *Minimizing adaptation effects and normalization of sweep numbers* for the procedures that checked for adaptation effects.

² Spindle density measures were based on the periods classified as low arousal states (when spindle occurred), not on the entire measurement period. This was because the purpose of using spindle density was to help characterize participants' status during the low arousal states, rather than their average status across the entire experiment.

³ As the addition step was not conducted in Exp 1, the cochlear microphonic would also exist at the frequency of the first harmonics. However, because Exp 1 used the /a/ vowel which has relatively low energy at the first harmonics compared to other frequencies (see **Figure 2(A)**), the cochlear microphonic would not significantly influence the FFR magnitude, as demonstrated in previous studies (c.f. Skoe and Kraus, 2010).

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

2.7 Theta-band phase-locked responses

Phase-locking values (PLV) at theta-band (θ -band) frequencies were measured (4 ~ 6 Hz, corresponding to stimulus repetition rates of ~ 5 syllables per second; see *Part 2.2*). 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](#)). A lower rejection threshold was used for θ -PLV ($\pm 15 \mu\text{V}$) compared to FFRs ($\pm 25 \mu\text{V}$) because the θ -band signal normally does not have excessively high amplitude since a relatively narrow frequency range (4 ~ 6 Hz) was used. More than 80% of the sweeps were retained in all participants after artefact rejection. PLV time series ($PLV(t)$) were calculated ([Morillon et al., 2012](#)) as follows:

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

where n denotes the total number of sweeps, $\phi_i(t)$ denotes the Hilbert phase series of the filtered EEG of the i th sweep time-locked to the syllable onset and j refers to $\sqrt{-1}$. Hilbert phase was used as it reflects phase-locking to stimuli even when EEG amplitude variation due to relaxation and eye closure occurred ([Thatcher, 2012](#)). The decision to measure θ -PLV was the desire to examine the degree of EEG phase coherence relative to syllable onset. The perfect scenario is that θ -band EEG will be reset to the same phase value at each onset. This requires similar lengths of one cycle of the θ -band EEG and the stimulus onset asynchrony (SOA) of the stimuli, in order that the same phase value of EEG can appear around each onset. Here, both one cycle of θ -band EEG (4 ~ 6 Hz) and SOA of the stimuli (~ 5 syllables per second) were at ~ 200 ms, which met this requirement.

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

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

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

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

2.8 Statistical analyses

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

2.8.1 Minimizing adaptation effects and normalization of sweep numbers

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

To tackle the first issue, the number of sweeps was normalized to around 1,500 for FFRs and around 500 for Logit- θ -PLV for both types of arousal period for each participant (c.f., Dajani et al., 2005; Wong et al., 2007). To ensure the data quality was adequate with respect to number

of syllable repetitions, participants whose artefact-free sweeps were fewer than 1,450 (for FFR) or 450 (for Logit- θ -PLV) in either low or high arousal states were not included in subsequent analyses. This gave 58 and 91 participants for FFRs and Logit- θ -PLV, respectively (see *Part 2.8.2 Combining data sets* for more details). Normalization was then conducted by randomly selecting epochs which contained the requisite numbers of artefact-free sweeps between 1,450 and 1,550 for FFRs, and between 450 and 550 for Logit- θ -PLV.

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

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

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

$$AI_{Within_Block} = P_{Low} - P_{High}$$

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

AI_{Across_Block} was defined as:

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

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

$$AI_{Across_Block} = P_{Low} - P_{High}$$

where P_i denotes the position of the i th high/low arousal epoch; N denotes the total number of high/low epochs across all blocks. P_{High} and P_{Low} thus represent the average across-block

positions of the high and low arousal epochs, respectively. Therefore, $AI_{Across_Block} > 0$ means that, on average, high arousal epochs were in earlier position than low arousal epochs across all blocks. Greater FFR/Logit- θ -PLV magnitudes in high arousal periods may be due to less neural adaptation, rather than being ascribed to the effect of arousal itself. To avoid this situation, AI_{Within_Block} and AI_{Across_Block} both needed to be ≤ 0 at the group level.

To combine approaches that normalize sweep numbers and minimize adaptation, the signals were processed for each participant as follows: (1) The normalization procedure using 1450-1550 (FFR) or 450-550 (Logit- θ -PLV) artefact-free sweeps was conducted 1000 times to generate 1000 sets of high and low arousal epochs. Within these 1000 sets, only those where $AI_{Within_Block} \leq 0$ were retained, unless AI_{Within_Block} was above 0 for all sets. (2) The set of epochs with the minimum absolute value of AI_{Within_Block} was chosen for FFR/Logit- θ -PLV measurements. (3) Steps (1) and (2) were repeated 500 times giving 500 estimates of AI values, density of sleep spindles in low arousal periods, and FFR/Logit- θ -PLV magnitudes in both arousal periods. Measures averaged over these 500 estimates were used in the final statistical analyses. The reason for refining AI_{Within_Block} rather than AI_{Across_Block} was because subsequent analyses found that both AI s were ≤ 0 at the group level when AI_{Within_Block} was ≤ 0 but not when the AI_{Across_Block} was ≤ 0 .

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

2.8.2 Combining data sets

Stimuli were presented at similar sound intensities in both experiments (80 and 77 dB, respectively, see *Stimuli*). [Gama et al., \(2017\)](#) showed that FFRs generated via free-field acoustic stimulation (loudspeaker) are comparable to those measured in close field (inserted earphones) with the same sound intensity. To further confirm the validity of combining the data from the two experiments, three-way mixed ANOVAs were conducted for magnitudes of

FFRs and Logit- θ -PLV with the within-subject factor of Arousal (high vs. low) and the between-subject factors of Age Group (young vs. older) and Data Set (Exp 1 vs. Exp 2; **Table 1**). Data from Exp 1 and Exp 2 were combined in subsequent analyses since there were no significant main effects or interactions involving Data Set. This resulted in data for 58 participants (25 young and 33 older) for FFRs and 91 participants (38 young and 53 older) for Logit- θ -PLV.

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

Table 1. Summary statistics for the three-way ANOVA with factors of Arousal, Age Group and Data Set. The top and bottom panels are for FFRs and Logit- θ -PLV respectively. *Df*, *F*, *p* and η_p^2 refer to degrees of freedom, *F*-values, *p*-values and partial eta-squared, respectively. Significant *p*-values are in bold. * = significance at $p < 0.05$; *** = significance at $p < 0.001$.

Dependent variables	Factors	<i>df1</i>	<i>df2</i>	<i>F</i>	<i>P</i>	η_p^2
FFRs	Arousal	1	54	6.357	0.015*	0.105
	Age Group	1	54	0.816	0.370	0.015
	Data Set	1	54	0.034	0.853	< 0.001
	Arousal \times Age Group \times Data Set	1	54	1.722	0.195	0.031
	Arousal \times Age Group	1	54	1.176	0.283	0.021
	Arousal \times Data Set	1	54	0.004	0.948	< 0.001
	Age Group \times Data Set	1	54	0.296	0.588	0.005

Logit- θ - PLV	Arousal	1	87	5.289	0.024*	0.057
	Age Group	1	87	22.217	< 0.001***	0.203
	Data Set	1	87	2.379	0.127	0.027
	Arousal \times Age Group \times Data Set	1	87	2.209	0.141	0.025
	Arousal \times Age Group	1	87	0.241	0.624	0.003
	Arousal \times Data Set	1	87	0.025	0.875	< 0.001
	Age Group \times Data Set	1	87	0.445	0.506	0.005

2.8.3 Effects of arousal, age and sleep spindle density

To address the question whether state of arousal affected the phase-locked responses cortically and subcortically, and whether such effects change with age and sleep spindle density, linear mixed-effect regressions were conducted for FFRs and Logit- θ -PLV. In these analyses arousal was the fixed-effect factor, Age and the spindle density in the low arousal states were fixed-effect covariates, and Participant was a random-effect factor. Both Age and Spindle Density were mean-centred. The covariance matrix that was chosen was the one that generated the smallest BIC value. Age (a continuous variable), rather than Age Group (with categorical levels), was used in the model here so that the effects of age itself and age-related variables could be compared. The extra age-related variables examined were pure-tone audiometric threshold (PTA) averaged across 0.25 and 4 kHz over both ears, and sleep spindle duration. The results of the analyses that used the age-related variables as fixed effect covariates instead of Age are given in the *Supplement*.

3 Results

3.1 Age effects on sleep spindles

Sleep spindle density, magnitude and duration in the low arousal epochs were compared between young and older adults using independent sample *t*-tests. Equal variances were not assumed during these *t*-tests, as Levene's test showed that variances differed significantly

between the two age groups (all $p < 0.02$). There were no significant differences between young and older adults for spindle density (**Figure 4A**, t-test: $t_{(63.772)} = 1.221$, $p = 0.227$) or spindle magnitude (**Figure 4B**, t-test: $t_{(54.723)} = 0.767$, $p = 0.447$). Spindle duration was significantly longer in young than in older adults (**Figure 4C**, t-test: $t_{(58.756)} = 3.006$, $p = 0.004$).

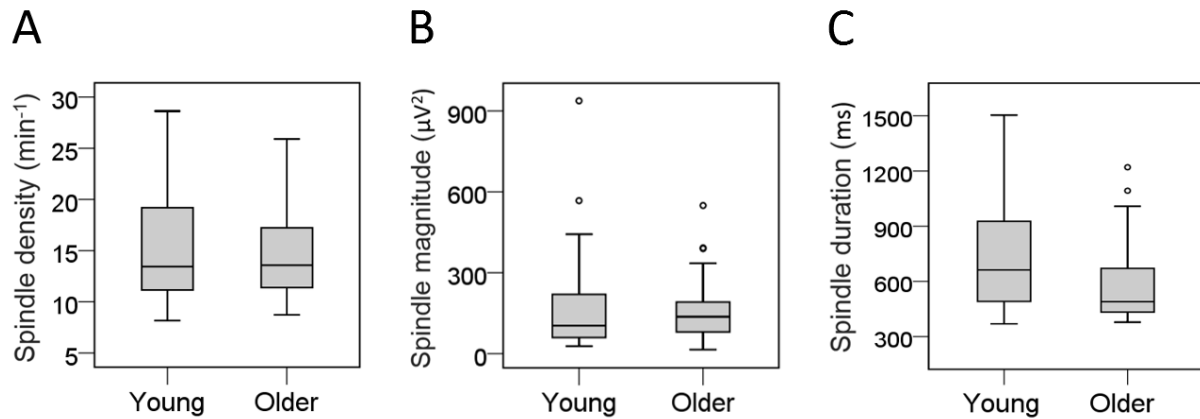


Figure 4. Boxplots of sleep spindle parameters for the two age groups. (A) Spindle density. (B) Spindle magnitude. (C) Spindle duration.

3.2 Effects of arousal, age and spindle density on FFRs and Logit- θ -PLV

Figure 5A shows waveforms and spectra of FFR responses for one participant and **Figure 5B** shows changes of Logit- θ -PLV across time for another participant. These participants were selected so that the differences in the respective measures between the two arousal states were closest to the group averages.

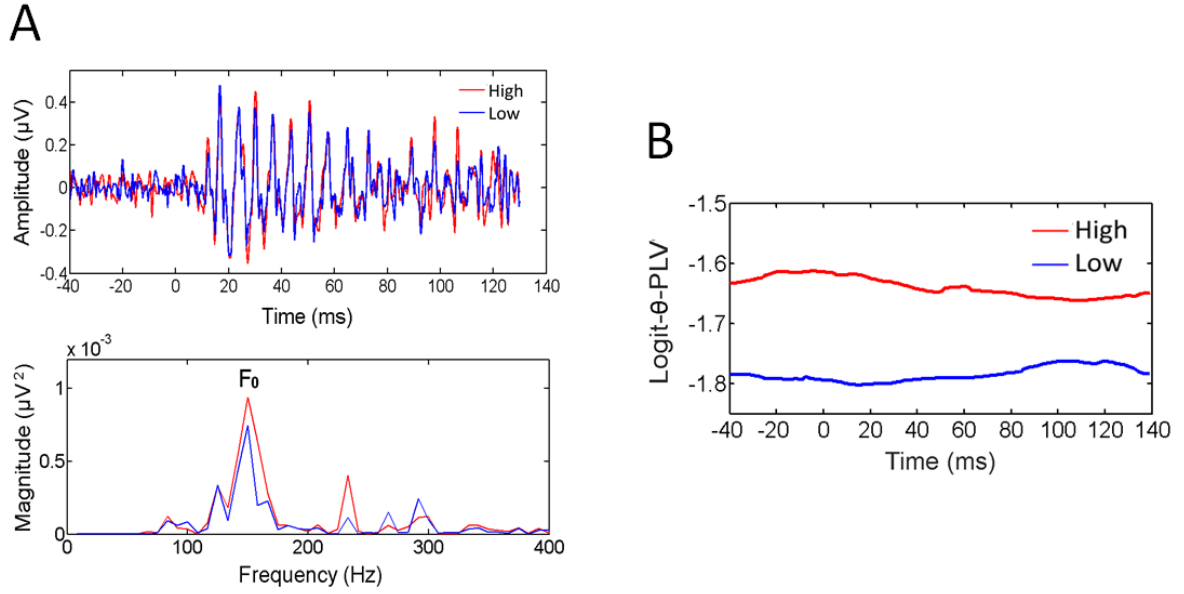


Figure 5. (A) FFRs and (B) Logit- θ -PLV for selected individual participants. Participants were chosen whose differences between the high and low arousal states were closest to the group averages for that measure. The red and blue lines indicate the high and the low arousal states, respectively. (A) Top: FFR waveforms across time; Bottom: the spectra for the sections between 0 ~ 120 ms. The spectra at the bottom of (A) peak at around F_0 frequency (labelled). (B) Changes of Logit- θ -PLV across time.

Distributions of FFR magnitudes and Logit- θ -PLV across the arousal states and age groups are shown as boxplots in **Figure 6** and **Figure 7**, respectively.

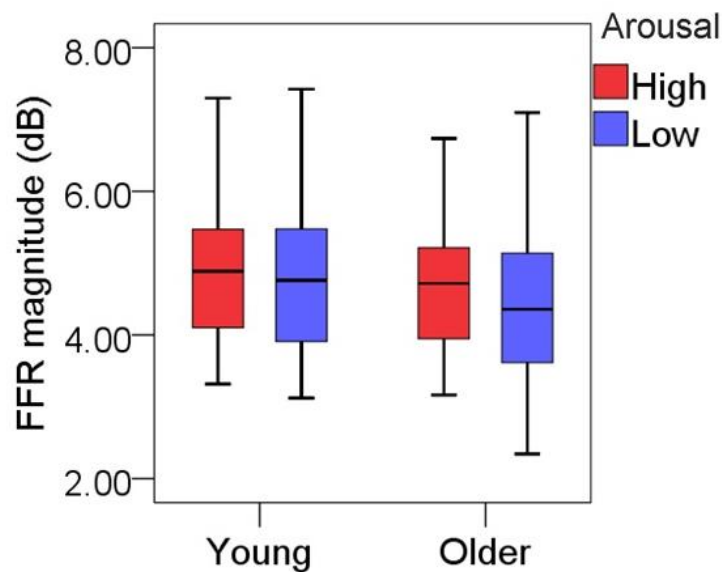


Figure 6. Boxplots of FFR magnitudes for the combined data sets. Magnitudes are shown for the high and low arousal states across the two age groups. Red and blue bars indicate the high and the low arousal states, respectively.

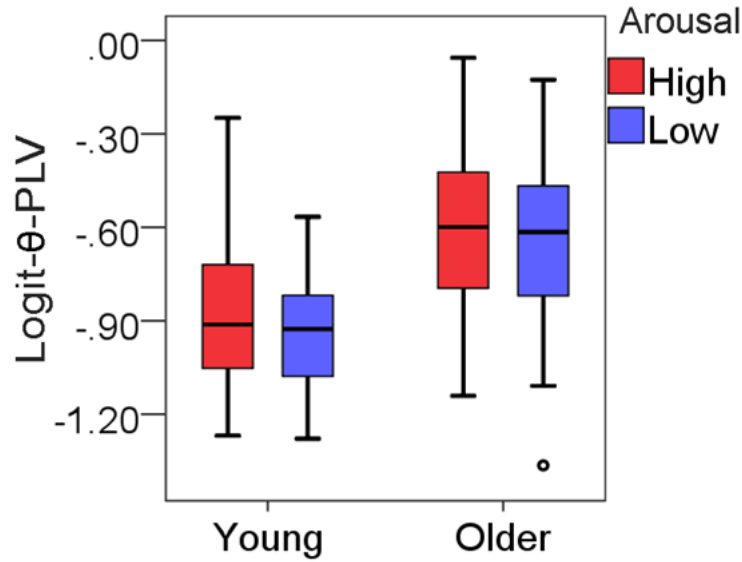


Figure 7. Boxplots of Logit-θ-PLV for the high and low arousal states across the two age groups. Red and blue bars indicate the high and the low arousal states, respectively.

Linear mixed-effect regressions were conducted using Arousal as the fixed-effect factor, Age and Spindle Density in the low arousal states as the fixed-effect covariates, and Participant as the random-effect factor. Statistics for the linear mixed-effect regressions are summarized in **Table 2**. Significant main effects of Arousal were found for both FFRs ($F_{(1,54)} = 6.263$, $p = 0.015$, $\eta_p^2 = 0.104$) and Logit-θ-PLV ($F_{(1,87)} = 5.520$, $p = 0.021$, $\eta_p^2 = 0.060$), with greater FFR magnitude and Logit-θ-PLV in the high than in the low arousal state. There was a significant main effect of Age for Logit-θ-PLV (Logit-θ-PLV increased with age; $F_{(1,87)} = 32.076$, $p < 0.001$, $\eta_p^2 = 0.199$) but not for FFRs.

The [Arousal \times Age] interactions were not significant either for FFRs or for Logit-θ-PLV. However, there was a significant three-way [Arousal \times Age \times Spindle Density] interaction for Logit-θ-PLV ($p = 0.010$; see **Table 2**) that suggests an interplay between age and the effect of arousal on the cortical phase-locked processing. To follow this up, a post-hoc analysis was conducted to examine how the [Arousal \times Spindle Density] interaction differed across ages.

Figure 8 shows the [Arousal \times Spindle Density] interaction (i.e., correlation between the effect of arousal on Logit- θ -PLV (Logit- θ -PLV_{High_vs_Low}) and spindle density) in the young and older adults. The interaction was significant for the young adults ($p = 0.012$, Logit- θ -PLV_{High_vs_Low} increased with spindle density; **Figure 8** left panel), but not for the older adults ($p = 0.661$; **Figure 8** right panel). Age-related variables (mean-centred PTA and spindle duration) were used respectively as covariates that replaced Age in the model to test whether the aging effect could result from age-related changes in peripheral hearing loss (PTA) or a spindle property (spindle duration). No significant three-way interactions relevant to PTA or spindle duration were found (see *Supplement*).

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

Table 2. Summary statistics for linear mixed-effect regression, using Arousal as the fixed-effect factor, Age and Spindle Density as fixed-effect covariates, and Participant as the random-effect factor. Df , F , p and η_p^2 refer to degrees of freedom, F -values, p -values and partial eta-squared, respectively. Significant p -values are in bold and * = significance at $p < 0.05$, and *** = significance at $p < 0.001$.

Dependent variables	Main effects/interactions	$df1$	$df2$	F	P	η_p^2
FFRs	Arousal	1	54	6.263	0.015*	0.104
	Age	1	54	1.401	0.242	0.025
	Spindle Density	1	54	0.004	0.947	< 0.001

	Arousal \times Age \times Spindle Density	1	54	0.242	0.625	0.004
	Arousal \times Age	1	54	0.352	0.555	0.006
	Arousal \times Spindle Density	1	54	0.207	0.651	0.004
	Age \times Spindle Density	1	54	1.101	0.299	0.020
Logit- θ -PLV	Arousal	1	87	5.520	0.021*	0.060
	Age	1	87	32.076	< 0.001***	0.199
	Spindle Density	1	87	0.976	0.326	0.009
	Arousal \times Age \times Spindle Density	1	87	6.848	0.010*	0.070
	Arousal \times Age	1	87	0.093	0.762	< 0.001
	Arousal \times Spindle Density	1	87	1.754	0.189	0.021
	Age \times Spindle Density	1	87	0.033	0.856	< 0.001

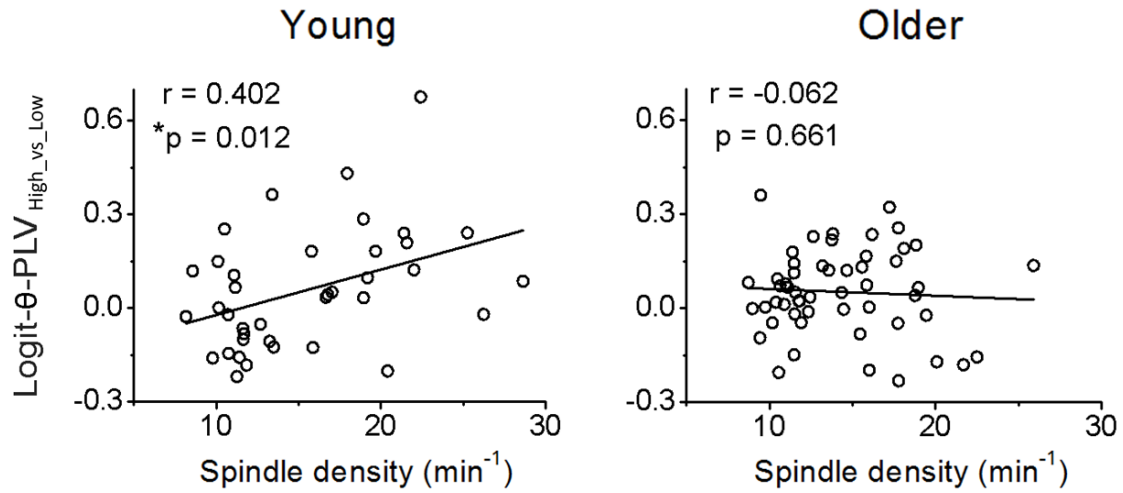


Figure 8. Interaction between Arousal and Spindle Density for Logit- θ -PLV (i.e. the correlation between the effect of arousal (Logit- θ -PLV_{High_vs_Low}) and spindle density for separate age groups. A significant [Arousal \times Spindle Density] correlation occurred for the young (left), but not for the older adults (right).

4 Discussion

The present study showed that two types of speech-evoked phase-locked responses, FFRs to the fundamental frequency (F_0) and θ -PLV to the acoustic envelopes, were affected by arousal levels in nREM sleep. Both types of response were statistically greater under high (no sleep spindles) than under low arousal states (with spindles) after potential neural adaptation had been ruled out. The effects were statistically significant and the effect sizes were in the medium range (Cohen, 1988; see **Table 2**). These results show that arousal affects the neuro-temporal precision of responses to speech at early sensory levels in the auditory system⁴. Age effects were found on θ -PLV and sleep spindle duration. As predicted, θ -PLV increased with age as spindle duration decreased. Furthermore, age interacted significantly with arousal and sleep spindle density in the low arousal states for θ -PLV. The arousal effect on θ -PLV increased significantly as spindle density increased in the young, but not the older, adults. Thus, incidence of sleep spindles during nREM sleep affects auditory processing differentially across ages.

4.1 Effect of arousal on phase-locked responses to speech

Previous studies have reported that magnitudes of auditory steady-state responses (ASSRs) which phase-lock to amplitude-modulated (AM) tones reduced significantly during sleep, but only at rates below around 70 Hz (Cohen et al., 1991; Lins et al., 1995; Picton et al., 2003). The speech-evoked FFRs here were at F_0 s above 100 Hz and were significantly affected by arousal. The ASSRs in previous studies involved responses to AM tones which were carried by a pure tone with a fixed frequency. In contrast, the FFRs here were responses to a complex harmonic carrier which excited formants that spanned a wide range of frequencies. The phase locking at frequencies higher than 70 Hz that occurred in the present study could have arisen because the stimuli had harmonic carriers. The properties of carriers (pure tone vs. harmonic complexes) that affect phase-locked responses in different arousal states merit further examination. A practical issue raised by the present results arises. Several previous studies have investigated the relationship between speech-evoked FFRs and behavioral functions (e.g., speech perception in reverberation, Fujihira and Shiraishi, 2015; pitch perception, Krishnan et al., 2005; Bidelman et al., 2011), where participants were allowed or even encouraged to fall

⁴ It is also noteworthy that FFR here was quantified as spectral magnitude. So mathematically it is not merely determined by temporal precision/synchrony of phase, but also by single-trial spectral magnitudes. Therefore, it is possible that arousal level affects FFRs by changing single-trial spectral magnitudes as well as temporal precision.

asleep during EEG recording without controlling for levels of arousal. In these studies, between-subject analyses were conducted to study the neural-behavioral relationship. The approach of allowing participants to sleep may need to be revised by including procedures that control for each individual's arousal state or by processing the data as here, since different levels of arousal across participants are potentially between-subject confounds that can bias the observed neural-behavioral relationship. Disregarding the influence of arousal can thus affect the validity of phase-locked activities as neuro-markers for behavioral performances.

Both experiments in the present study used a paradigm where participants listened passively to a repeated vowel that had no high-level linguistic features (it had no semantic or syntactic information). Hence, the present results showed that arousal affects early-stage neural processing of speech. It has been suggested that speech-evoked FFRs originate primarily in the auditory brainstem (Chandrasekaran and Kraus, 2010; Bidelman, 2018) and that theta phase-locked responses originate in the auditory cortex (Luo and Poeppel, 2007; Howard and Poeppel, 2010; Peelle et al., 2013). If this is the case, the findings that FFR magnitudes and θ -PLV changed across arousal states indicate arousal affects both subcortical and cortical levels. However, a recent study found no evidence that neural responses to speech signals were reduced at subcortical levels during nREM sleep (Wilf et al., 2016). Whilst FFRs were obtained in the present study by scalp-EEG, BOLD responses were obtained from fMRI in Wilf et al. (2016). Subcortical BOLD responses can be highly contaminated by systemic physiological signals (Brooks et al., 2013), making subcortical BOLD insensitive to auditory processing changes due to varying arousal levels arising in the auditory cortex (Czisch et al., 2002, 2004). Also, it is not clear how speech-evoked FFR is related to subcortical BOLD, and this needs to be investigated. An alternative account is that speech-evoked FFRs also originate in the auditory cortex (Coffey et al., 2016, 2017). It is possible that all the effects of arousal in the current study arose at the cortical level. Ascribing arousal effects to cortical processing would be compatible with the report that neural responses to speech signals in the auditory cortex were greater during wakefulness than nREM sleep (Czisch et al., 2002, 2004; Wilf et al., 2016). The cortical changes with arousal state could be explained by findings of sensory deafferentation as a result of reduced thalamocortical connectivity during the transition from wakefulness to nREM sleep (Spoormaker et al., 2010, 2011; Picchioni et al., 2014). This is consistent with the thalamic gating mechanism which maintains that most sensory information must pass through the thalamus before reaching the sensory cortex (Steriade et al., 1993; McCormick and Bal, 1994; 1997). One of the results such gating achieves is the selective

control of afferent sensory information flow at different levels of arousal. When the arousal level decreases, thalamic neurons fire in bursts, creating membrane hyperpolarization that raises the threshold potential for postsynaptic firing (Sherman, 2001). Incoming postsynaptic potentials are thus less likely to pass the threshold potential, leading to a partial blockage of sensory information ascending from thalamus to cortex (Coenen, 2010). Furthermore, sleep spindles that occur during nREM sleep are generated by the interplay between the thalamic reticular nucleus and thalamocortical neurons after hyperpolarization (De Gennaro and Ferrara, 2003). Spindles are transmitted to the cortex where they reflect the degree of sensory deafferentation that modulates neural sensitivity to external auditory stimuli (Dang-Vu et al., 2011; Schabus et al., 2012) (see *Part 4.2* for further discussion of the role of sleep spindles).

Both types of phase-locked responses that were investigated in the present study reflect neural precision for speech-related acoustic cues at millisecond levels. FFRs reflect the processing of speech F0s (Aiken and Picton, 2008; Skoe and Kraus, 2010) and they correlate significantly with speech-in-noise perception (Song et al., 2010; Coffey et al., 2017). Theta-band phase-locked responses reflect the neural sensitivity to speech acoustic envelopes (Luo and Poeppel, 2007; Howard and Poeppel, 2010) and are closely associated with speech intelligibility (Peelle et al., 2013; Mai et al., 2018). From this perspective, the current results indicate that arousal influences sensory levels of processing that aid perception and understanding of speech and such effects are not limited to higher-order linguistic processing (i.e., perception at the word, phrase or sentence levels; Makov et al., 2017).

4.2 The role of aging

Previous studies have reported that FFR magnitudes are smaller in older than young adults, arguing for an effect of aging on FFRs (Anderson et al., 2012; Presacco et al., 2016). However, the present results did not provide statistical evidence to support this claim. It is noteworthy that older participants in these studies (Anderson et al., 2012; Presacco et al., 2016) had normal peripheral hearing (PTAs ≤ 25 dB HL up to 4 kHz). On the other hand, other evidence has shown that hearing loss may lead to loss of neural inhibition, further leading to greater FFRs in response to envelope modulations at F₀ rates compared to normal hearing in both animals (Kale and Heinz, 2010; Henry et al., 2014; Zhong et al., 2014) and humans (Anderson et al., 2013; Goossens et al., 2018). As hearing loss and aging often co-occur (which is the case in the present study, especially in Exp 2), null effect of aging on FFRs here could be ascribed to the combined effects of aging itself and age-related hearing loss.

The present results showed significantly higher θ -PLV and shorter spindle duration in older adults compared to young adults. The finding of higher θ -PLV responses in older adults is consistent with previous studies that have shown age-related increases in theta-band ASSRs (Tlumak et al., 2015; Goossens et al., 2016). Such observations can be ascribed to hyperexcitability of the central auditory system as a result of aging (Casparly et al., 2008) since increased θ -PLV is associated with higher neural excitability in response to acoustic stimuli in the auditory cortex (Ng et al., 2013). This is also consistent with previous findings that older adults showed larger magnitudes of auditory-evoked responses (Hermann et al., 2013, 2016) and greater cortical tracking of speech envelopes (Presacco et al., 2016). It is worth mentioning that increased excitability may also be due to age-related changes such as hearing loss, which merits further investigation. The observation that spindle duration is shorter in older adults accords with findings that spindle properties change during aging where the changes include decreased density, magnitude and duration of spindles (Martin et al., 2013). Lack of statistical age-related changes in spindle density and magnitude in the present study may have occurred because measurements were made over different time scales across studies. While spindle properties are typically collected across whole nights of sleep (e.g., Martin et al., 2013), the present experiments lasted for much shorter times (~ 2 hrs for Exp 1 and ~ 70 min for Exp 2) and spindles were only measured during part of the experiments (periods when syllables were presented in the quiet background).

Although no significant interaction was found between Age and Arousal for magnitude of either FFRs or θ -PLVs, the current results indicated that aging could still interplay with the effect of arousal. A significant three-way [Arousal \times Age \times Spindle Density] interaction was found for Logit- θ -PLV. A subsequent post-hoc analysis showed a significant [Arousal \times Spindle Density] interaction where the effect of arousal on Logit- θ -PLV increased statistically with sleep spindle density for the young adults alone. Furthermore, it was shown that this discrepancy between the young and older adults was attributable to age itself, rather than age-related variables such as hearing loss (PTA) or shorter spindle duration (see *Supplement*). The lack of the two-way [Arousal \times Age] interaction may have arisen because Arousal can still be a crude proxy for characterizing the degree of arousal, as sleep spindle density in the low arousal states differed across participants. Hence, by adding Spindle Density as an additional covariate in the model, the interplay between age and arousal was assessed precisely.

The observed correlation between the effect of arousal and spindle density in young adults thus echoes previous studies that have shown that sleep spindles reduce neural sensitivity to external acoustic stimuli during nREM sleep (Dang-Vu et al., 2011; Schabus et al., 2012). Higher spindle density may reflect a greater degree of sensory deafferentation in sleep (Spoormaker et al., 2010, 2011; Picchioni et al., 2014) since spindle density reflects sleep stability (Kim et al., 2012) and tolerance to auditory disturbance during nREM sleep (Dang-Vu et al., 2010). On the other hand, the absence of this correlation in older adults could indicate that speech spindles play less of a role in regulating auditory processing as people age. While the present results found no evidence that the reduced regulatory role of sleep spindles was due to age-related changes in peripheral hearing or particular spindle properties (spindle duration here), the underlying neural mechanisms need further clarification. A possible explanation could be that the generation of sleep spindles does not solely involve changes in thalamocortical systems that impact on auditory processing, but they are also associated with hippocampal activity that is influential in memory consolidation (Schabus et al., 2007; Andrade et al., 2011; Bergmann et al., 2012). Consequently, age-related atrophy of the hippocampus could partly contribute to functional changes in sleep spindles during aging (Mander et al., 2017). In this way, it is possible that age-related changes in spindle properties could alter the regulatory role of spindles on auditory processing over ages.

4.3 Data in quiet vs. noisy backgrounds

The current study analysed the neural responses to speech in a quiet background. Data in noisy backgrounds were not investigated because different types of background noise and SNRs were used in the two data sets (see Schoof and Rosen, 2016 and Mai et al., 2018). FFR magnitudes and θ -PLV can change significantly across background types and SNRs (Mai et al., 2018). Here, the only background type that was shared across all participants was the quiet background. Hence it was not appropriate to use data in noisy backgrounds. It could be hypothesized that the results in noisy backgrounds would be similar to those in quiet background. Alternatively, as phase-locked responses would decrease as SNR decreases (Russo et al., 2004; Mai et al., 2018), it is possible that effects of aging or arousal alter at low SNRs. Since noisy environments provide a more ecological scenario for neural processing of auditory and speech signals than do quiet background, it is important to systematically investigate data in different types of noise backgrounds and at different SNRs in the future.

4.4 Summary

The present study was the first to investigate the effect of arousal on phase-locked neural responses to speech signals and to examine how aging interplays with these effects. The topic is important because the phase-locked responses investigated here reflect neuro-temporal processing of speech properties at a millisecond-scale precision that is essential for audition in general and speech perception in particular. The results highlight the significant role arousal plays in assisting processing of fine-grained acoustic properties of F0 and envelope modulations at the sensory level. A possible regulatory role of sleep spindles for phase-locked responses in the auditory cortex was revealed and it was further found that aging reduced the role of spindle regulation. The neuromodulation of speech-evoked responses and the role of aging that were revealed lays the groundwork for studying how cognitive states, such as arousal, anaesthesia and attention, affects early-stage neural activities in the auditory system across the lifespan.

Acknowledgements

Guangting Mai is supported by UCL Graduate Scholarship for Cross-Disciplinary Training. Tim Schoof is supported by Action on Hearing Loss, UK. We thank Mr. Andrew Clark for his technical support for the experiments.

References

- Aiken SJ, Picton TW (2008). Envelope and spectral frequency following responses to vowel sounds. *Hearing Research* 245:35–47.
- Anderson S, Parbery-Clark A, White-Schwoch T, Kraus N (2012) Aging affects neural precision of speech encoding. *Journal of Neuroscience* 32:14156-64.
- Anderson S, Parbery-Clark ., White-Schwoch T, Drehabl S, Kraus N (2013) Effects of hearing loss on the subcortical representation of speech cues. *J Acoust Soc Am* 133:3030-3038.

Andrade KC, Spoormaker VI, Dresler M, Wehrle R, Holsboer F, Sämann PG, Czeisler M (2011) Sleep spindles and hippocampal functional connectivity in human NREM sleep. *Journal of Neuroscience* 31:10331-9.

Aviyente S, Bernat EM, Evans WS, Sponheim SR (2011) A phase synchrony measure for quantifying dynamic functional integration in the brain. *Human Brain Mapping* 32:80-93.

Bergmann TO, Mölle M, Diedrichs J, Born J, Siebner HR (2012) Sleep spindle-related reactivation of category-specific cortical regions after learning face-scene associations. *NeuroImage* 59:2733-42.

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

Bidelman GM, Gandour JT, Krishnan A (2011). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *J Cogn Neurosci* 23:425-434

Brooks JC, Faull OK, Pattinson KT, Jenkinson M (2013) Physiological noise in brainstem fMRI. *Frontiers in Human Neuroscience* 7:623.

Brown RE, Basheer R, McKenna JT, Strecker RE, McCarle RW (2012) Control of sleep and wakefulness. *Physiological Reviews*, 92:1087-1187.

Caspary, DM, Ling L, Turner JG, Hughes LF (2008) Inhibitory neurotransmission, plasticity and aging in the mammalian central auditory system. *Journal of Experimental Biology* 211: 1781-1791.

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

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

Coffey EB, Herholz SC, Chepesiuk AM, Baillet S, Zatorre RJ (2016) Cortical contributions to the auditory frequency-following response revealed by MEG. *Nature Communications* 7:11070.

Coffey EB, Chepesiuk AM, Herholz SC, Baillet S, Zatorre RJ (2017) Neural correlates of early sound encoding and their relationship to speech-in-noise perception. *Front Neurosci* 11:479.

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

Cohen LT, Rickards FW, Clark GM (1991) A comparison of steady-state evoked potentials to modulated tones in awake and sleeping humans. *The Journal of the Acoustical Society of America* 90:2-4672479.

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

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

Dajani HR, Purcell D, Wong W (2005) Recording human evoked potentials that follow the pitch contour of a natural vowel. *IEEE Trans Biomed Eng* 52:1614–1618.

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

Dang-Vu TT, McKinney SM, Buxton OM, Solet JM, Ellenbogen JM (2010) Spontaneous brain rhythms predict sleep stability in the face of noise. *Current Biology* 20:R626-7.

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

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

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

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. *Hearing Research* 304:128-36.

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

Howard MF, Poeppel D (2010) Discrimination of speech stimuli based on neuronal response phase patterns depends on acoustics but not comprehension. *Journal of Neurophysiology* 104:2500-11.

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

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

Goossens T, Vercammen C, Wouters J, van Wieringen A (2018) The association between hearing impairment and neural envelope encoding at different ages. *Neurobiology of Aging* (in press).

Lakatos P, Chen CM, O'Connell MN, Mills A, Schroeder CE (2007) Neuronal oscillations and multisensory interaction in primary auditory cortex. *Neuron* 53:279-92.

Lins OG, Picton PE, Picton TW, Champagne SC, Durieux-Smith A (1995) Auditory steady-state responses to tones amplitude-modulated at 80-110 Hz. *The Journal of the Acoustical Society of America* 97:3051-63.

Luo H, Poeppel D (2007) Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron* 54:1001-10.

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. *Proceedings of the National Academy of Sciences* 109:20673-20678.
- Krishnan A, Xu Y, Gandour J (2005) Encoding of pitch in the human brainstem is sensitive to language experience. *Brain Res Cogn Brain Res* 25:161–168.
- Mai G, Tuomainen J, Howell P (2018) Relationship between speech-evoked neural responses and perception of speech in noise in older adults. *The Journal of the Acoustical Society of America* 143:1333-45.
- Makov S, Sharon O, Ding N, Ben-Shachar M, Nir Y, Golumbic EZ (2017) Sleep disrupts high-level speech parsing despite significant basic auditory processing. *Journal of Neuroscience* 37:7772-7781.
- Mander BA, Winer JR, Walker MP (2017) Sleep and human aging. *Neuron* 94:19-36.
- Martin N, Lafortune M, Godbout J, Barakat M, Robillard R, Poirier G, Carrier J (2013) Topography of age-related changes in sleep spindles. *Neurobiology of Aging*, 34:468-476.
- McCormick DA, Bal T (1994) Sensory gating mechanisms of the thalamus. *Current Opinion in Neurobiology* 4:550-556.
- McCormick DA, Bal T (1997) Sleep and arousal: thalamocortical mechanisms. *Annual Review of Neuroscience*. 20:185-215.
- 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. *Frontiers in Psychology* 3:248.
- Moushegian G, Rupert AL, Stillman RD (1973) Scalp-recorded early responses in man to frequencies in the speech range. *Electroencephalography and Clinical Neurophysiology* 35:665-7.
- Ng BSW, Logothetis NK, Kayser C (2013) EEG phase patterns reflect the selectivity of neural firing. *Cerebral Cortex* 23:389-398.

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

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

Peelle JE, Gross J, Davis MH (2012) Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex* 23:1378–1387.

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

Picchioni D, Pixa ML, Fukunaga M, Carr WS, Horovitz SG, Braun AR, Duyn, JH (2014) Decreased connectivity between the thalamus and the neocortex during human nonrapid eye movement sleep. *Sleep* 37:387-397.

Picton TW, John MS, Dimitrijevic A, Purcell D (2003) Human auditory steady-state responses. *International Journal of Audiology* 42:177-219.

Portas CM, Krakow K, Allen P, Josephs O, Armony JL, Frith CD (2000) Auditory processing across the sleep-wake cycle: simultaneous EEG and fMRI monitoring in humans. *Neuron* 28:991-9.

Presacco A, Simon JZ, Anderson S (2016) Evidence of degraded representation of speech in noise, in the aging midbrain and cortex. *Journal of Neurophysiology* 116:2346-55.

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

Schabus M, Dang-Vu TT, Albouy G, Balteau E, Boly M, Carrier J, Darsaud A, Degueldre C, Desseilles M, Gais S, Phillips C (2007) Hemodynamic cerebral correlates of sleep spindles during human non-rapid eye movement sleep. *Proceedings of the National Academy of Sciences* 104:13164-9.

Schabus MD, Dang-Vu TT, Heib DPJ, Boly M, Desseilles M, Vandewalle G, Degueldre C (2012) The fate of incoming stimuli during NREM sleep is determined by spindles and the phase of the slow oscillation. *Frontiers in Neurology* 3:40.

Schoof T, Rosen S (2016) The role of age-related declines in subcortical auditory processing in speech perception in noise. *Journal of the Association for Research in Otolaryngology* 17:441-60.

Sherman SM (2001) Tonic and burst firing: dual modes of thalamocortical relay. *Trends in Neurosciences* 24:122-6.

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

Song JH, Skoe E, Banai K, Kraus N (2010) Perception of speech in noise: Neural correlates. *J Cogn Neurosci* 23:2268–2279.

Spoormaker VI, Czeisler M, Maquet P, Jäncke L (2011) Large-scale functional brain networks in human non-rapid eye movement sleep: insights from combined electroencephalographic/functional magnetic resonance imaging studies. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 369:3708-29.

Spoormaker VI, Schröter MS, Gleiser PM, Andrade KC, Dresler M, Wehrle R, Czeisler M (2010). Development of a large-scale functional brain network during human non-rapid eye movement sleep. *Journal of Neuroscience* 30:11379-11387.

Steriade M, McCormick DA, Sejnowski TJ (1993) Thalamocortical oscillations in the sleeping and aroused brain. *Science* 262:679-85.

Thatcher RW (2012) Coherence, phase differences, phase shift, and phase lock in EEG/ERP analyses. *Developmental Neuropsychology* 37:476-96.

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.

Warby SC, Wendt SL, Welinder P, Munk EG, Carrillo O, Sorensen HB, Jennum P, Peppard PE, Perona P, Mignot E (2014) Sleep-spindle detection: crowdsourcing and evaluating performance of experts, non-experts and automated methods. *Nature Methods* 11:385.

Waschke L, Wöstmann M, Obleser J (2017) States and traits of neural irregularity in the age-varying human brain. *Scientific Reports* 7:17381.

Wilf M, Ramot M, Furman-Haran E, Arzi A, Levkovitz Y, Malach R (2016) Diminished auditory responses during NREM sleep correlate with the hierarchy of language processing. *PloS One* 11:e0157143.

Wong PC, Skoe E, Russo, NM, Dees T, Kraus N (2007) Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience* 10:420.

Worden FG, Marsh JT (1968) Frequency-following (microphonic-like) neural responses evoked by sound. *Clinical Neurophysiology* 25:42-52.

Zhong Z, Henry KS, Heinz MG (2014) Sensorineural hearing loss amplifies neural coding of envelope information in the central auditory system of chinchillas. *Hearing Research* 309:55-62.

Highlights

- Speech-evoked phase-locked activities were tested in adults across a wide age range
- State of arousal can affect both subcortical (FFR) and cortical (θ -PLV) activities
- Sleep spindles showed an age-dependent neuro-regulatory role on cortical activities

Supplement

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

The [Arousal \times PTA \times Spindle Density] (**Table S1**) and the [Arousal \times Spindle Duration \times Spindle Density] (**Table S2**) interactions were not significant. Hence there was no evidence that the interaction between Age, Arousal and Spindle Density resulted from age-related changes in PTA or Spindle Duration. This contrasted with the analysis which used age as covariate where there was a significant three-way [Arousal \times Age \times Spindle Density] interaction for Logit- θ -PLV (see *main text*).

Table S1 Linear mixed regression for Logit- θ -PLV, using Arousal as the fixed-effect factor, PTA and Spindle Density as fixed-effect covariates, and Participant as the random-effect factor. *Df*, *F*, *p* and η_p^2 refer to degrees of freedom, *F*-values, *p*-values and partial eta-squared, respectively. Significant *p*-values are in bold and * = significance at $p < 0.05$, and *** = significance at $p < 0.001$.

Dependent variable	Main effects/interactions	<i>df1</i>	<i>df2</i>	<i>F</i>	<i>P</i>	η_p^2
Logit- θ -PLV	Arousal	1	87	5.740	0.019*	0.062
	PTA	1	87	16.199	< 0.001***	0.157
	Spindle Density	1	87	0.126	0.723	0.001
	Arousal \times PTA \times Spindle Density	1	87	2.206	0.141	0.025

Arousal \times PTA	1	87	0.003	0.957	< 0.001
Arousal \times Spindle Density	1	87	1.647	0.203	0.019
PTA \times Spindle Density	1	87	0.862	0.356	0.010

Table S2 Linear mixed regression for Logit- θ -PLV, using Arousal as the fixed-effect factor, Spindle Duration and Spindle Density as fixed-effect covariates, and Participant as the random-effect factor. *Df*, *F*, *p* and η_p^2 refer to degrees of freedom, *F*-values, *p*-values and partial eta-squared, respectively. Significant *p*-values are in bold and * = significance at $p < 0.05$, and ** = significance at $p < 0.01$.

Dependent variable	Main effects/interactions	<i>df1</i>	<i>df2</i>	<i>F</i>	<i>P</i>	η_p^2
Logit- θ -PLV	Arousal	1	87	5.593	0.020*	0.060
	Spindle Duration	1	87	7.402	0.008**	0.078
	Spindle Density	1	87	3.209	0.077	0.036
	Arousal \times Spindle Duration \times Spindle Density	1	87	0.037	0.847	< 0.001
	Arousal \times Spindle Duration	1	87	2.118	0.149	0.024
	Arousal \times Spindle Density	1	87	6.209	0.015*	0.067
	Spindle Duration \times Spindle Density	1	87	0.810	0.370	0.009