Cerebral Cortex, 2022, 1–14

https://doi.org/10.1093/cercor/bhac302

Original Article

Speaker–listener neural coupling reveals a right-lateralized mechanism for non-native speech-in-noise comprehension

Zhuoran Li[1,2], Bo Hong[2,3], Daifa Wang[4], Guido Nolte[5], Andreas K. Engel[5] and Dan Zhang[1,2,*],

[1]Department of Psychology, School of Social Sciences, Tsinghua University, Beijing 100084, China,
[2]Tsinghua Laboratory of Brain and Intelligence, Tsinghua University, Beijing 100084, China,
[3]Department of Biomedical Engineering, School of Medicine, Tsinghua University, Beijing 100084, China,
[4]School of Biological Science and Medical Engineering, Beihang University, Beijing 100083, China,
[5]Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg Eppendorf, 20246 Hamburg, Germany
*Corresponding author: Department of Psychology, Tsinghua University, Room 334, Mingzhai Building, Beijing 100084, China. Email: dzhang@tsinghua.edu.cn

While the increasingly globalized world has brought more and more demands for non-native language communication, the prevalence of background noise in everyday life poses a great challenge to non-native speech comprehension. The present study employed an interbrain approach based on functional near-infrared spectroscopy (fNIRS) to explore how people adapt to comprehend non-native speech information in noise. A group of Korean participants who acquired Chinese as their non-native language was invited to listen to Chinese narratives at 4 noise levels (no noise, 2 dB, −6 dB, and − 9 dB). These narratives were real-life stories spoken by native Chinese speakers. Processing of the non-native speech was associated with significant fNIRS-based listener–speaker neural couplings mainly over the right hemisphere at both the listener's and the speaker's sides. More importantly, the neural couplings from the listener's right superior temporal gyrus, the right middle temporal gyrus, as well as the right postcentral gyrus were found to be positively correlated with their individual comprehension performance at the strongest noise level (−9 dB). These results provide interbrain evidence in support of the right-lateralized mechanism for non-native speech processing and suggest that both an auditory-based and a sensorimotor-based mechanism contributed to the non-native speech-in-noise comprehension.

Key words: speech-in-noise; non-native; right hemisphere; inter-brain; fNIRS.

Introduction

In an increasingly globalized world, more and more people are learning and acquiring languages that are not their native language for cross-cultural speech communication. However, the use of non-native languages often faces many challenges. In particular, the prevalence of background noise (e.g. other people's babblings at a cocktail party) in everyday life has a significant impact on speech communication in the form of a non-native language (reviewed in Lecumberri et al. 2010; Scharenborg and van Os 2019). Comprehending non-native speech information under noisy conditions often requires extra effort, even for a proficient user of that language (Tabri et al. 2011; Mendel and Widner 2016; Peng and Wang 2019; Regalado et al. 2019). However, the neural mechanism of how people adapt to comprehend non-native speech information in noise remains unclear.

Neuroimaging studies on native speech comprehension in noise suggest that both an auditory mechanism and a sensorimotor mechanism could facilitate the comprehension of noisy native speech (e.g. Du et al. 2014; Alain et al. 2018; Li et al. 2021). While the auditory mechanism suggested that efficient processing of a to-be-attended speech stream could be achieved by suppressing the encoding of the to-be-ignored noise streams based on the auditory features (Ding and Simon 2013; Zion Golumbic et al. 2013; Vander Ghinst et al. 2016), the sensorimotor mechanism is believed to compensate for noise-contaminated speech information by simulating the articulatory gestures associated with the perceived speech (Hickok et al. 2011; Parrell and Niziolek 2021; Raharjo et al. 2021). Although both mechanisms are well supported by the literature in the context of native speech-in-noise comprehension, applying them to the mechanistic interpretation of non-native speech-in-noise comprehension remains challenging. In terms of the auditory mechanism, while the native speech-in-noise processing relies on left-lateralized brain regions (Vigneau et al. 2006; Scott et al. 2009; Corballis 2015; Vander Ghinst et al. 2016), comprehending non-native speech in noise may employ a different auditory processing mechanism. Specifically, as the processing of non-native speech has mainly relied on right-lateralized brain regions, including the right superior temporal gyrus (STG), the right middle temporal gyrus (MTG), etc. (Archila-Suerte et al. 2013; Qi et al. 2019; Gao et al. 2020; Cotosck et al. 2021; Yi et al. 2021), it remains elusive whether the left-lateralized counterpart could contribute

Received: June 3, 2022. **Revised:** July 8, 2022. **Accepted:** July 9, 2022

to noise reduction as well. In terms of the sensorimotor mechanism, as a strong association between the auditory experience and the corresponding speech production experience is regarded as the prerequisite for a stable sensorimotor representation of the perceived speech information (Hickok and Poeppel 2007; Hickok et al. 2011; Liebenthal and Mottonen 2018; Ohashi and Ostry 2021), presumably over the brain regions such as the inferior frontal gyrus (IFG), the precentral gyrus (preCG), postcentral gyrus (postCG), inferior parietal lobe, etc. (Hickok and Poeppel 2007; Sehm et al. 2013; Du et al. 2014; Alain et al. 2018). Hereby, whether a non-native listener could benefit from such a mechanism is unclear (Willems and Hagoort 2007; Jones et al. 2013; Drijvers et al. 2019; Schmitz et al. 2019), especially for those who acquired the non-native languages after the critical period (Archila-Suerte et al. 2012, 2015). Moreover, it should be pointed out that the available studies to date have mainly focused on the neural responses to isolated and simplified speech materials (e.g. the last word in a fixed-length sentence) under noise (Coulter et al. 2021; Grant et al. 2022), while the understanding of the continuous naturalistic speech under noise is less explored.

The use of continuous naturalistic speech is expected to provide valuable insights into the neural mechanisms of non-native speech-in-noise comprehension. On one hand, continuous naturalistic speech materials, whether native or non-native, are expected to contain richer dynamic time–frequency information than isolated and simplified speech materials, thus would correspond to more reliable and more widely distributed neural activations (Huk et al. 2018; Sonkusare et al. 2019; Li et al. 2022). On the other hand, the continuous naturalistic speech could provide important context information for the listener to recover the missing information from the noise-contaminated speech, therefore facilitating speech-in-noise comprehension. In particular, context information such as acoustic and semantic cues, etc., has been shown to be of great help for non-native listeners (Bradlow and Alexander 2007; Song and Iverson 2018; Borghini and Hazan 2020). As both the auditory and the sensorimotor mechanisms could benefit from the rich dynamic time–frequency information and the context information, further studies are necessary to clarify how the non-native listener could make use of the continuous naturalistic speech materials for efficient speech-in-noise comprehension. However, the usage of naturalistic speech materials renders the conventional analytical methods (i.e. the event-related design with general linear modeling) ineffective. Specifically, the continuous and dynamic changing nature of the naturalistic speeches makes it difficult to define the required discrete and well-separated events.

The recently emerging interbrain approach could be a promising tool for studying the neural mechanisms of comprehending continuous naturalistic non-native speech in noise (Schoot et al. 2016; Zhang 2018; Redcay and Schilbach 2019; Kingsbury and Hong 2020; Farahzadi and Kekecs 2021). In contrast to the conventional single-brain approach that focuses on the coupling (response) of the listener's neural activities to external speech stimuli, the interbrain approach takes an integrated view and emphasizes the coupling between the listener's neural activities and the speaker's neural activities (Hasson et al. 2012; Czeszumski et al. 2020; Kelsen et al. 2022), thus evading the challenges of defining and extracting "event" information from naturalistic speech acoustics. By taking the speaker's neural activities as the reference to characterize the listener's neural activity patterns, the state-of-the-art interbrain studies have suggested that the listener–speaker neural coupling could reflect information beyond simple speech acoustics, such as a shared representation and an active interpretation from the listener's perspective on the delivered information from the speaker (Jiang et al. 2021; Yeshurun et al. 2021; Holroyd 2022). Therefore, the interbrain approach could provide added value to the study of non-native speech-in-noise comprehension: the listener–speaker neural coupling is expected to give a comprehensive overview of how the non-native listener understands the speaker, with the speaker's neural activities as a noise-free reference; the spatial pattern of the listener–speaker neural coupling can help reveal the possible contributions of the auditory and the sensorimotor mechanisms for the understanding of non-native speech in noise. However, the interbrain studies on speech comprehension have focused on native speeches, and no study to date has investigated non-native speech comprehension in noise.

The present study employed an interbrain approach with continuous naturalistic speech narratives to investigate the neural mechanisms underlying how the listener could comprehend non-native speech in noise. Fifteen Korean participants, who acquired Chinese after the critical period, were recruited as the listeners to listen to recorded Chinese narratives from 6 native Chinese speakers. Functional near-infrared spectroscopy (fNIRS) signals were collected from both the listeners and the speakers, with the same channel setups covering bilaterally distributed typical speech-related brain regions. The fNIRS was chosen for its suitability for speech-in-noise tasks: Compared with fMRI, the portability and the low operating noise of fNIRS is advantageous for auditory speech communication in naturalistic settings; compared with EEG, the relatively focused spatial sampling and tolerance to motion provide a better measurement of localized neural activities with less influence by speech production-related artifacts (Quaresima et al. 2012; Pinti et al. 2020). The present study manipulated the noise in a graded manner with 4 noise levels in order to reveal how the listener could adapt to different noise levels, as noise adaptation has been shown to be the key feature for speech-in-noise comprehension (Ding and Simon 2013; Du et al. 2014; Li et al. 2021). Following our previous speech-in-noise study

(Li et al. 2021), one brain region of the listener would be regarded as functionally relevant for the non-native speech-in-noise comprehension if the neural coupling between the corresponding fNIRS channel of the listener and the speaker's fNIRS signals were correlated with the listeners' individualized comprehension performance under at least one of the noise levels. A high correlation between the neural coupling and the comprehension performance would indicate that the specific brain region of the listener can respond flexibly to the processing of the non-native speech information in the presence of noise. The spatial patterns of the above-defined listener–speaker neural coupling on the listener's side would therefore inform us of the neural mechanisms for non-native speech-in-noise comprehension. The auditory mechanism would be supported if the listener's speech-auditory-related brain regions were involved, such as STG, MTG, etc. (Scott et al. 2009; Sulpizio et al. 2020). The sensorimotor mechanism would be supported if the listener's speech sensorimotor-integration-related brain regions were involved, such as left IFG, preCG, postCG, etc. (Hickok and Poeppel 2007; Schomers and Pulvermuller 2016; Alain et al. 2018). More importantly, the lateralization of the involved brain regions would further reveal the possible specificity of non-native speech processing as compared with native speech: a left-lateralized processing would imply the reliance on the listener's native speech processing modules for resolving the noise of the non-native speech, whereas a right-lateralized processing would suggest a distinct mechanism for non-native speech-in-noise processing.

Methods

Ethics statement

The study was conducted in accordance with the Declaration of Helsinki and was approved by the local Ethics Committee of Tsinghua University. Written informed consent was obtained from all participants.

Participants

Fifteen college students (6 males, 9 females; age ranged from 18 to 24 years old) from Tsinghua University participated in the study as the listeners. The sample size was determined to be sufficient by reference to existing studies, including our previous study with a similar design (Li et al. 2021) as well as other related fNIRS-based interbrain studies (Liu et al. 2017). These participants were all native Korean speakers from South Korea, which provides China's largest source of foreign students, with currently more than 500,000 South Korean students in China (the Ministry of Education, China, http://www.moe.gov.cn/jyb_xwfb/gzdt_gzdt/s5987/201904/t20190412_377692.html). All the participants started to learn Chinese after 12 years old (the first exposure time to Chinese ranged from 12 to 17 years old, and the length of learning ranged from 4 to 9 years), which is well above the critical period of language learning (Costa and Sebastian-Galles 2014; Sulpizio et al. 2020). Fourteen of them have moved to China for more than 2 years, and the other one has moved to China for half a year. They all passed the Chinese Proficiency Test Level VI (the official Chinese language test for non-native speakers), indicating that they were capable of fluent Chinese communication with native Chinese people. All participants are right-handed, with normal hearing and normal or corrected-to-normal vision by their self-report. The detailed demographic information and the language experience of the participants are listed in Table S1.

Stimuli

Thirty-two narrative audios from our previous study (Li et al. 2021) were used for the listener participants. All of them lasted for 85–90 s each and were about daily topics that were adapted from the National Mandarin Proficiency Test. These narratives were spoken by 6 college students (3 males, 3 females; age ranged from 21 to 25 years old) who were native Chinese speakers with professional training in broadcasting and hosting from Tsinghua University. All the narratives were about the speaker's personal experiences which were not familiar to the participants. These audios were recorded by a regular microphone at a sampling rate of 44,100 Hz in a sound-attenuated room. The fNIRS data from the speakers were obtained during their speech. More details of the speaker's data collection procedure can be found in our previous study (Li et al. 2021) and hereby were skipped in the present study.

Before being played to the listeners, these audios were processed into 4 versions at 4 different noise levels: a no noise (NN) level and three noise levels with the signal-to-noise ratio (SNR) equaling 2, −6, and − 9 dB. The noise level was manipulated by adding white noise to the original speech audios. Then, all the processed audios were matched in terms of their average root mean square sound pressure level. For each selected narrative, there were 4 four-choice questions concerning narrative details and themes. For example, one question following a narrative audio was, "What is the occupation of the person the speaker admires most? (说话人最欣赏的知名人物的职业是什么记者心理医生作家历史学家职业是什么)" and the four choices were 1) Journalist, 2) Psychologist, 3) Writer, and 4) Historian (1.记者, 2.心理医生, 3.作家, and 4.历史学家). These questions were prepared by the experimenters to assess the listener's comprehension performance. Each listener's comprehension performance per noise level is defined as the average accuracy across all the 8 narratives at the corresponding noise level.

Experimental procedure

The experimental procedure is illustrated in Fig. 1A and B. The participants listened to 32 Chinese narrative audios at different noise levels, organized as 32 trials. In each trial, the participants listened to one narrative audio at one of the four noise levels. They were then required to rate the clarity and the intelligibility of the audio with 7-point Likert scales and complete 4 four-choice questions about the content of the narrative audio. Afterward, the participants were asked to rest for at least 20 s and press the SPACE key on the computer keyboard to start the next trial. The order of the narrative audios and their assigned noise levels were randomized across the listeners. The listeners were informed that the narrative audios were pre-recorded.

Prior to the start of the experiment, the participants attended a resting-state session where they were required to keep relaxed with their eyes closed for 3 min. They were then given one practice trial, with an additional speech audio presented at −2 dB SNR, to get them familiar with the procedure.

The experimental procedure was programmed in MATLAB using the Psychophysics Toolbox 3.0 extensions (Brainard 1997).

Data acquisition

The neural signals of the listeners were recorded from 36 channels by an fNIRS system (NirScan Inc., HuiChuang, Beijing). As shown in Fig. 1c, three sets of optode probes were placed covering the prefrontal cortex and bilateral inferior frontal, pre- and post-central, inferior parietal, middle and STG, etc. The positions of CH21 and CH31 were placed at T3 and T4 according to the international 10–20 system. The center of the prefrontal probe set was placed at FPz position. The fNIRS signals were recorded at a sampling rate of 17 Hz, with the near-infrared light of 2 different wavelengths (740 and 850 nm). The concentration change of oxy-hemoglobin (HbO) and deoxy-hemoglobin (HbR) was obtained based on the modified Beer–Lambert law.

The neural signals of the speakers were recorded from the same 36 channels by another fNIRS system (NirScan Inc., HuiChuang, Beijing), with a sampling rate of 12 Hz and the near-infrared light of 3 different wavelengths (785, 808, and 850 nm). Similar to the listeners, both the neural signals in the task conditions (i.e. narrative speaking) and a 3-min resting-state condition were obtained.

To allow a probabilistic reference to cortical areas underlying each fNIRS channel, a procedure (Singh et al. 2005; Shattuck et al. 2008) which projects the topographic data based on skull landmarks into a 3D reference frame (MNI-space, Montreal Neurological Institute) was performed based on NIRS_SPM (Ye et al. 2009). The procedure is expected to provide a spatial registration of each channel with a standard deviation of 4.7∼7.0 mm (Singh et al. 2005). The anatomical labels with the percentage of overlap for each channel, as well as the corresponding MNI coordinates, are listed in Table S2.

Data analysis

Preprocessing

Two preprocessing steps were applied to remove possible motion artifacts by using HoMER2 software package (Huppert et al. 2009). First, the targeted principal component analysis (tPCA; function: hmrMotionCorrectPCArecurse; input parameters: tMotion = 0.5, tMask = 1, STDthresh = 30, AMPthresh = 0.5, nSV = 0.97, maxIter = 5) was applied to identify and correct the motion artifacts contained in raw data. The artifact-related principal components were removed, and the remaining principal components were back-projected to reconstruct the cleaned fNIRS signals (Yucel et al. 2014). Next, to further reduce possible artifacts, motion artifacts were identified (function hmrMotionArtifactByChannel; input parameters: tMotion = 0.5, tMask = 1, STDEVthresh = 30, AMPthresh = 0.5) and then corrected by a cubic spline interpolation method (function hmrMotionCorrectSpline; input parameters: $P = 0.99$) (Scholkmann et al. 2010). The parameters of the 2 algorithms were the same as our previous study (Li et al. 2021), allowing us to have a fair comparison of the results from the 2 studies. Then, the data of the listeners were downsampled to 12 Hz to match the sampling rate of the speakers' data.

Interbrain neural coupling analysis

The interbrain neural coupling between the speakers and the listeners in both the resting-state and the task sessions was analyzed. The MATLAB function "wcoherence" (Grinsted et al. 2004) was used to calculate the wavelet transform coherence (WTC), which assessed the cross-correlation between 2 series of physiological signals as a function of frequency and time (Cui et al. 2012; Gvirts and Perlmutter 2020; Hu et al. 2021). First, each trial in the task sessions was extended to 300 s, covering the 90-s trial duration and additional 105-s periods both preceding and after the trial. The extended periods were included for the WTC analysis to ensure a reliable calculation of the interbrain couplings over the frequency range of interest. The WTC was then calculated over the two 300-s fNIRS signal segments from the listener and the corresponding speaker and organized in a two-dimension matrix in time and frequency domains. In specific, the non-analytic Morlet wavelet ($\omega_0 = 6$, smallest scale $s_0 = 1/6$, spacing between scales $ds = 0.4875$) was used for the WTC calculation, resulting in the coherence values of the 3,600 time points over the whole 300-s duration at 121 frequency bins ranging from 0.0056 to 5.73 Hz. The above calculation was conducted for all listener–speaker channel combinations, forming 1,296 (36 channels from the speaker × 36 channels from the listener) cross-channel combinations in total. Second, the coherence values were time-averaged across the 90-s trial duration and then converted into Fisher-z values.

The follow-up analysis was focused on the frequency range of 0.01–0.032 Hz. This frequency band was decided based on our previous finding that neural activities of the speakers and the listeners during noisy speech communication were coupled in 0.01–0.032 Hz (Li et al. 2021). To verify the validity of the selection for the present dataset, we analyzed the neural coupling at each frequency bin.

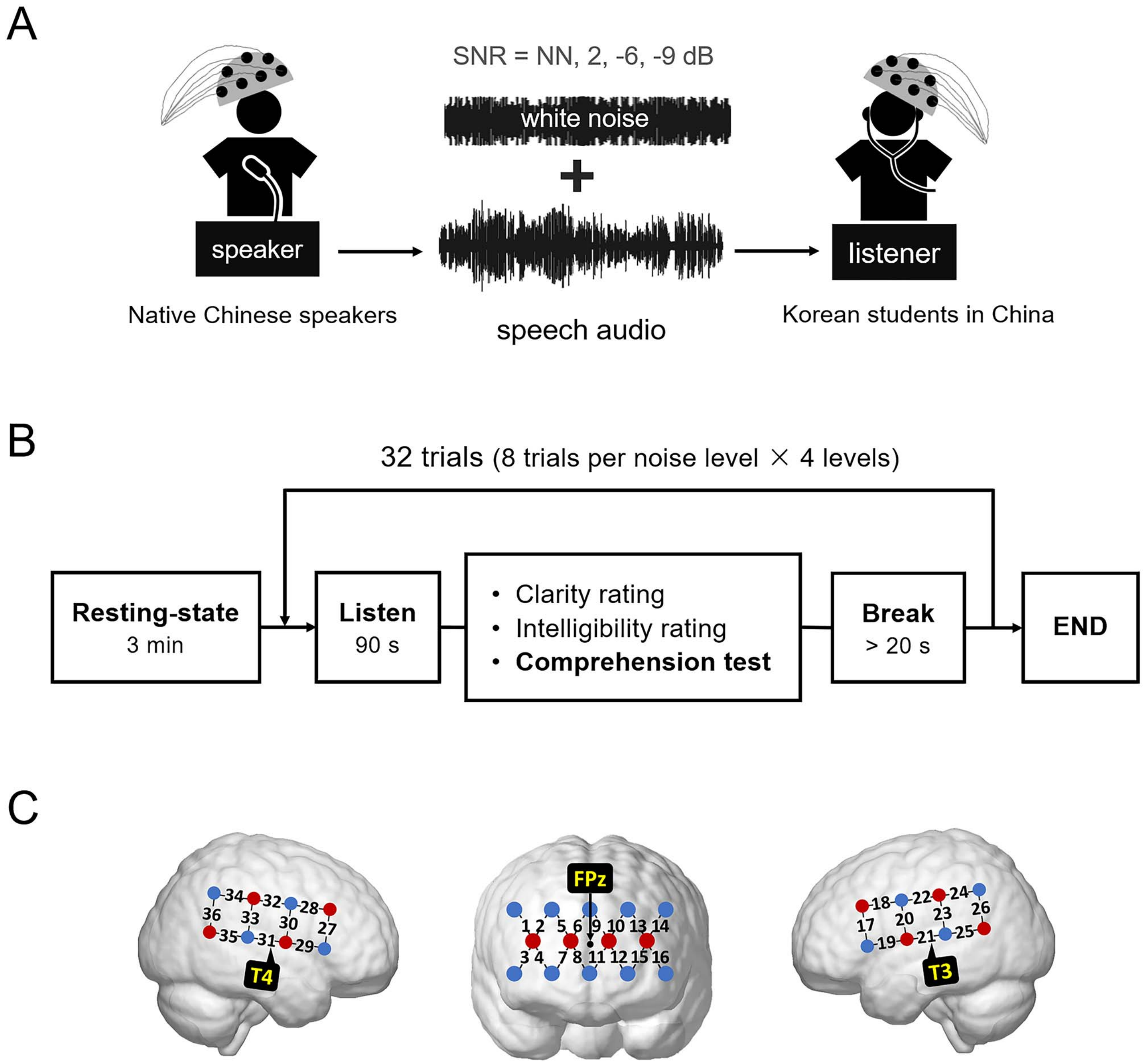

Fig. 1. Schematic of the experimental design. A) The experimental design. B) The experimental procedure for the non-native listener. C) The fNIRS optode probe set for both the listeners and the speakers. Channels of 21 and 31 were placed at T3 and T4, and the center of the prefrontal probe set was placed at FPz, in accordance with the international 10–20 system. The probe set covered typical speech-related brain regions, including the prefrontal cortex and bilateral IFG, pre- and post-central gyrus, MTG and STG, etc.

As shown in Fig. S1, the difference among neural coupling in various listening conditions was most significant in this frequency band.

Afterward, the coherence value in this frequency band was selected and averaged. The coherence values of the 8 trials with the same noise level were further averaged within each participant, resulting in averaged interbrain coherence values per participant per noise level for the statistical analysis.

The speaker–listener neural coupling in the resting-state condition was analyzed in a similar way. The 180-s resting-state data per listener were extended to a 300-s segment (both 60-s periods preceding and after the state were included), and the couplings were calculated between the listener and each of the 6 speakers. The coherence values of the middle 90 s were averaged, and then Fisher-z transformed. Next, the time-averaged coherence values were averaged in the same frequency band as above. Finally, for each listener, the coherence values to the speakers under the resting-state condition were averaged among all 6 speaker–listener pairs. The resting-state session served as the baseline for the tasking session. Based on the above calculation, each listener had the coupling values under 5 conditions, i.e. the 4 noise levels and the resting-state, over 1,296 channel combinations in the frequency band of interest (0.01–0.032 Hz).

At the group level, a repeated-measures ANOVA (rmANOVA) was first applied to identify the channel combinations that showed significant modulation of speaker–listener neural coupling by the within-participant factor of task condition. As no explicit hypothesis regarding the variations of neural couplings by the noise conditions could be formulated, the rmANOVA was conducted among all the 5 conditions, i.e. the resting-state condition and the 4 noise levels, to capture all possible changes of the neural couplings that could be related to speech-in-noise processing. The rmANOVA was performed separately for all the 1,296 channel combinations. To account for multiple comparisons, the false discovery rate (FDR) correction method was applied (Benjamini and Hochberg 1995). When significant, a post-hoc analysis was used to compare the pair-wise neural coupling levels between different conditions. One brain region on the listeners' side would be regarded to be relevant for the processing of the non-native speeches if the neural couplings of at least one of the noise levels were different from the resting-state condition.

To further explore the specificity of the identified channel combinations for non-native speech processing, the dataset from our previous study with the same paradigm but native listeners (Li et al. 2021) was used for comparison. Specifically, as the data processing procedure was kept consistent in both studies, the channel combinations of interest (i.e. with significant modulation by the within-participant factor of task condition) in the present study were applied to the fNIRS dataset from the previous study. The coherence at the resting-state condition was subtracted from those at the 4 noise levels to better compare the neural couplings from 2 different groups of listeners. A mixed-design ANOVA with the between-subject factor of listeners (native vs. non-native listeners) and the within-subject factor of noise level (4 noise levels) was then applied to the median fNIRS coherence over all the channel combinations of interest to get an overview of the difference between the 2 datasets. Any significant difference for the listener factor would inform us about the specificity of the present data for non-native speech processing.

To reveal the listeners' brain regions in support of an adaptive processing of noisy speeches, correlational analysis was applied between the neural couplings and the comprehension performance (accuracy) over the brain regions showing a significant task condition effect of the corresponding neural couplings (see above). Specifically, Spearman's correlation was computed between the coherence values of each listener and their speech comprehension performances at each of the 4 noise levels. A significant correlation would imply a behavioral relevance of the involved brain region of the listener at an individual listener's level that could support noise adaptation.

The above analyses were performed on both the HbO and the HbR signals. However, the HbR-based analyses revealed different speaker–listener neural coupling patterns from the results of HbO-based analyses: only one channel combination was reported to show a negative correlation between neural coupling and comprehension performance. As no positive behavioral correlation of HbR-based neural coupling was found, and the HbO signals have been suggested to own a higher SNR than the HbR signals (Mahmoudzadeh et al. 2013; Zhang et al. 2020), only the HbO-based results were reported in Results section. The HbR-based results are shown in Figs. S2 and S3.

Single-brain activation analysis

The listener's fNIRS channels that showed significant neural coupling were further selected for the following single-brain analyses of activations. The single-channel activations were calculated as the average of HbO values among trials in the same noise level. Before average, the HbO values were converter into the *z*-scores by the middle 90 s of the resting-state session. Similarly, for each channel, an rmANOVA was applied to analyze the activation patterns at the 4 noise levels. Also, behavioral correlations were calculated to examine the behavioral relevance of the neural activations at these noise levels.

Results

Behavioral performance

The speech comprehension performance was calculated by the average accuracy of the 4-choice questions over all trials within a noise level. The speech comprehension scores were 78.75 ± 3.29%, 68.13 ± 5.15%, 51.04 ± 4.22%, and 42.71 ± 3.74% (mean ± SE) at the noise levels of NN, 2 dB, −6 dB, and −9 dB, respectively, with lower performance score with lower SNR level (stronger noise). A significant effect of the noise level on the comprehension performance was observed (rmANOVA, $F(3, 42) = 54.224$, $P < 0.001$). Post-hoc analysis showed that all comparisons were significant (post-hoc *t*-test, $Ps < 0.05$, FDR corrected). Besides, even at −6 and −9 dB, the comprehension scores were significantly higher than random level 25% [$t(14) = 12.09$, 11.41, $Ps < 0.001$].

At the 4 noise levels, the clarity scores were 6.83 ± 0.05, 4.63 ± 0.21, 2.73 ± 0.24, and 2.12 ± 0.23 (mean ± SE), respectively, and the intelligibility scores were 6.00 ± 0.17, 4.75 ± 0.25, 2.84 ± 0.29, and 2.16 ± 0.30 (mean ± SE). The subjective ratings of clarity and intelligibility also showed significant effects of the noise level [rmANOVA, $F(3, 42) = 216.64$ and 94.50, $Ps < 0.001$]. Post-hoc *t*-tests revealed significant pairwise differences for all comparisons (post-hoc *t*-test $Ps < 0.05$, FDR corrected). The behavioral performances are summarized in Fig. 2. It is noteworthy to point out that while the non-native listeners self-reported a sharp decrease of the clarity and intelligibility ratings when the noise was stronger, they still achieved a moderate level of comprehension.

Neural couplings between speaker and listener

The rmANOVA of neural coupling revealed 10 channel combinations with a significant effect of task condition,

重复测量方差分析

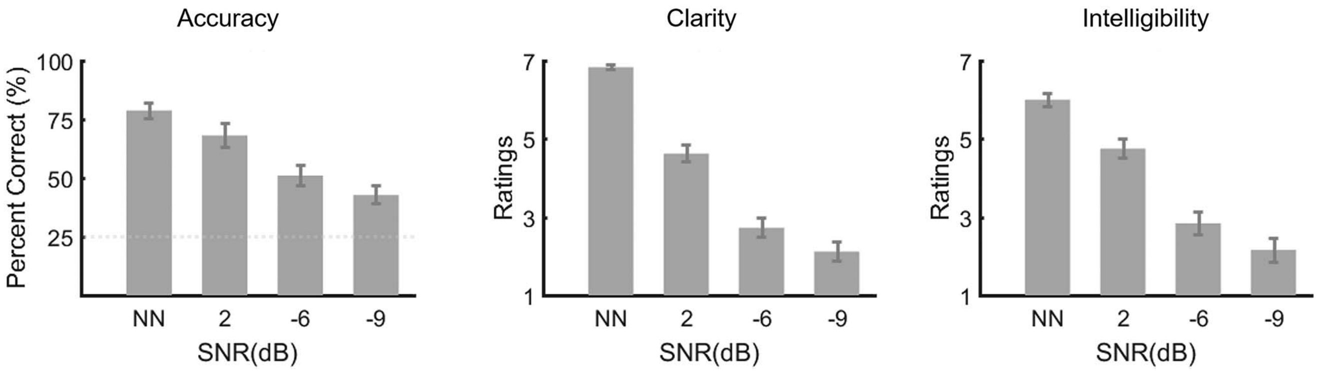

Fig. 2. The behavioral performance. Repeated measures ANOVA results for accuracy, clarity, and intelligibility were significant (Ps < 0.001). The post-hoc *t*-tests indicated that all comparisons were significant (post-hoc *t*-test, Ps < 0.05, FDR corrected). The error bar means the standard error.

as shown in Fig. 3. The channel combinations covered the listener's right middle frontal gyrus (right MFG, CH1), the left IFG (CH17), the right preCG (CH27), the right STG (CH29/31), the right postCG (CH30), and the right MTG (CH35). On the speaker's side, related brain areas were more restricted, centering over the right STG (CH29/31) and the right postCG (CH30).

For all these identified channel combinations, stronger speaker–listener neural couplings were observed at all the 4 noise levels than the resting-state condition (post-hoc *t*-test, Ps < 0.05, FDR corrected, Fig. 3B). For 6 out of these 10 channel combinations, no significant differences among the noise levels were found.

However, there are 4 channel combinations showing significant differences among the noise levels as well (post-hoc *t*-test, Ps < 0.05, FDR corrected). For CH30-CH27, the neural couplings at 2 dB were lower than those at −9 dB; for CH31-CH27, the neural couplings at 2 dB were lower than those at the other 3 levels; for CH30-CH29, the couplings showed an increasing trend, with lower couplings at NN and 2 dB than −9 dB; for CH31-CH30, the couplings were lower at 2 dB than −6 dB.

Comparison of non-native and native listeners' neural coupling to speakers

The mixed-design ANOVA revealed a marginally significant main effect of listeners [*F*(1, 28) = 3.574, *P* = 0.069] for the median speaker–listener neural coupling over all the 10 channel combinations showing a significant noise-level effect for non-native speech-in-noise processing. As shown in Fig. S4a, the neural coherence was generally higher for non-native speech-in-noise processing than the native counterpart. Post-hoc analysis revealed that the non-native vs. native difference was significant at the noise level of −9 dB (*P* = 0.037) and the differences were not significant in the other 3 noise levels (Ps > 0.05). The coherences at each individual channels are displayed in Fig. S4b, which showed similarly higher coherences over these channel combinations for non-native speech-in-noise processing.

Behavioral relevance of the speaker–listener neural couplings

The correlation analyses between the speech comprehension performance and the neural coupling revealed the brain regions functionally related to the comprehension of the noisy speech. Figure 4A shows the correlational *r*-values for the channel combinations with a significant task condition effect at all 4 noise levels. Only the neural couplings between the speakers' right postCG, the right STG and the listeners' right postCG (CH30-CH30, CH31-CH30), the speakers' right STG, and the listeners' right STG and the right MTG (CH31-CH31, CH31-CH35) were positively correlated with the comprehension performance at the noise level of −9 dB (*r* = 0.68, 0.54, 0.51, 0.54, uncorrected Ps = 0.005, 0.039, 0.051, 0.036). The other *r*-values were nonsignificant (Ps > 0.05).

Figure 4B demonstrates the scatter plots of the correlation results and the brain localization of the 4 channel combinations with significant behavioral relevance. We further compared the correlations among all 4 noise levels for each channel combination. For CH30-CH30, the correlation at −9 dB was significantly larger than that at 2 and −6 dB (Ps = 0.003, 0.065); for CH31-CH31, the correlation at −9 dB was larger than that at 2 dB (*P* = 0.056); and for CH31-CH35, the correlation at −9 dB was larger than that at −6 dB (*P* = 0.004). The other comparisons were nonsignificant (Ps > 0.05).

Single-brain activation of the listener

The single-brain activations over CH1/17/27/29/30/31/35 of the listeners' brain within the frequency range of 0.01–0.032 Hz were not significantly higher than the resting-state condition at the 4 noise levels, and the 4 noise levels did not differ from each other either (Ps > 0.05). Besides, no significant correlation was found between these single-brain activations and the speech comprehension performance (Ps > 0.05).

Discussion

The present study investigated the neural mechanism of comprehending non-native speech in a noisy

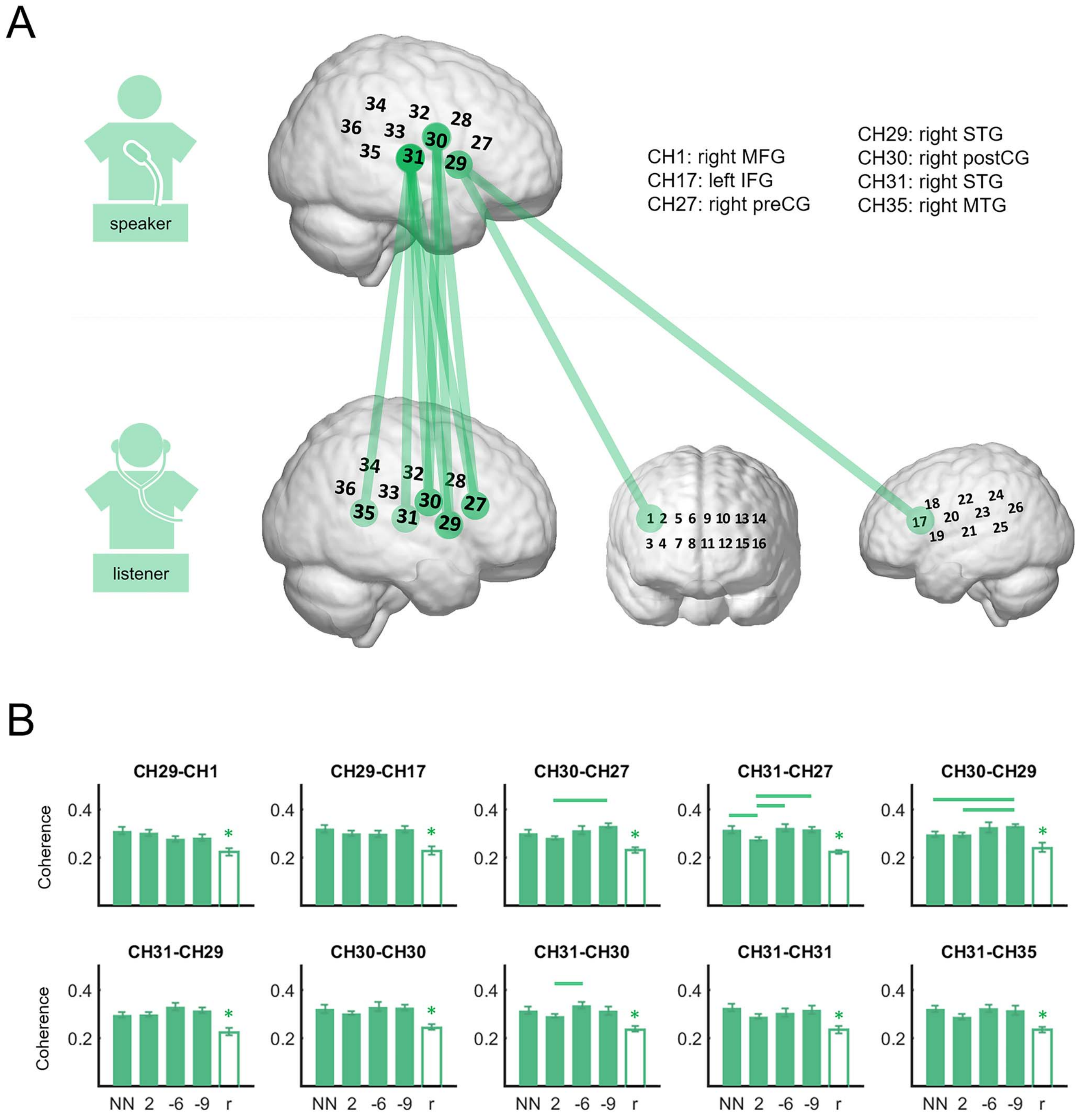

Fig. 3. The repeated measures ANOVA results of the speaker–listener neural coupling. A) The colored lines connect channel combinations between the speaker and the listener that show statistically significant differences across the 5 conditions ($Ps < 0.05$, FDR corrected). A right-lateralized neural pattern was observed for both the speaker and the listener. B) The neural coupling of these channel combinations in the 5 conditions. The error bar means the standard error. The asterisks over the resting-state condition bins indicate that the corresponding neural couplings were significantly lower than all the other 4 conditions (post-hoc *t*-test, $Ps < 0.05$, FDR corrected). The horizontal lines indicate significant pairwise differences (post-hoc *t*-test, $Ps < 0.05$, FDR corrected).NN means no noise; 2, −6, −9 mean noise levels whose SNR equaling to 2, −6 and − 9 dB; *r* means resting-state.

environment by an fNIRS-based interbrain approach. A group of native Korean participants who have studied Chinese as a non-native language after their critical language-learning period were invited to listen to Chinese narratives under different noise conditions. These narratives were spoken by another group of native Chinese speakers. The neural activities of both the listeners and the speakers were measured by fNIRS. The interbrain neural coupling analysis showed that the neural activities of the listeners' right STG, the right MTG, the right postCG, the right preCG, the right MFG, and the left IFG were coupled to the speakers' right postCG and the right STG. Furthermore, only the neural couplings of the listeners' right STG, the right MTG, and the right postCG were positively correlated with the comprehension performance at the strongest noise level

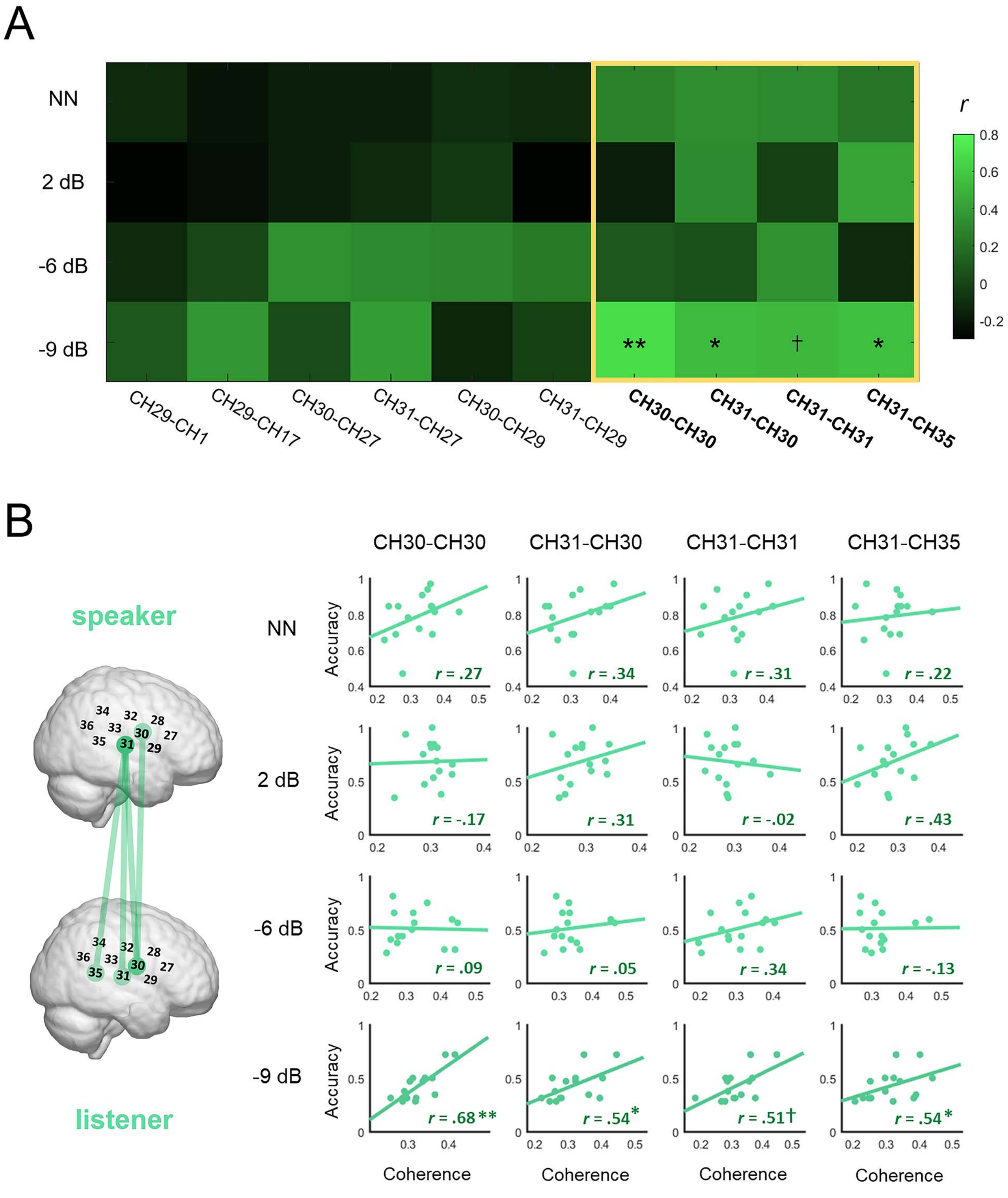

Fig. 4. The behavioral relevance of the speaker–listener neural couplings. A) The r-values of the correlations analysis between the couplings and the accuracy for all channel combinations with a significant task condition effect. B) For CH30-CH30, CH31-CH30, CH31-CH31, and CH31-CH35, the correlations at −9 dB were significant or marginally significant ($r = 0.68$, 0.54, 0.51, 54, uncorrected $Ps = 0.005$, 0.039, 0.051, 0.036). The other correlations were not significant ($Ps > 0.05$). b) The scatter plots show the correlations between the couplings of CH30-CH30, CH31-CH30, CH31-CH31, and CH31-CH35 and the accuracy score at 4 noise levels. ** means $P < 0.01$; * means $P < 0.05$; † means $P < 0.1$.

(−9 dB). These results revealed how the listener's brain was adaptively comprehending the non-native speeches in noise.

The right-lateralized distribution of the significant listener–speaker neural couplings in the speech conditions as compared with the resting-state condition provided new evidence in support of a right-lateralized mechanism to process non-native speech. Although the non-native speech processing has long been associated with right-lateralized brain regions (Friederici 2011; Archila-Suerte et al. 2013; Qi et al. 2019; Cotosck et al. 2021; Yi et al. 2021) of the listeners for key features of non-native language usage, such as language proficiency (Hull and Vaid 2007; Qi et al. 2015, 2019), learning outcome (Qi et al. 2019), etc., the investigations to date have mainly focused on the neural couplings of the listeners' brain to the speech information (e.g. the conventional event-related analysis). By taking the speakers' neural activities as references for an interbrain coupling analysis, our findings on the listeners' side further suggest the possible association between the listeners and the speakers. As the speakers' neural activities could reflect information from the simple speech acoustics level to a higher level about the intention of the speakers (Jiang et al. 2021; Yeshurun et al. 2021), our results are consistent and extending the present understanding of the possible functional roles of the listeners' right hemisphere for non-native speech processing. Specifically, the right hemisphere of the listeners could be responsible for the processing of the acoustic and phonological variation in the non-native speech stream (Lattner et al. 2005; Wolmetz et al. 2011; Di Liberto et al. 2018; Weed and Fusaroli 2020; Lakertz et al. 2021), as well as possibly the messages that the speakers intended to convey (Silbert et al. 2014; Dai et al. 2018; Pan et al. 2020). In this study, the coupled brain regions on the speakers' side were also right-lateralized over the right STG and the right postCG, which have been proposed to be mainly related to the representation and generation of auditory information during speech production (Tanaka and Kirino 2018; Yamamoto et al. 2019). These regions were coupled to wider areas on the listeners' side, including the auditory-related regions (the right STG, CH29/31; the right MTG, CH35), the sensorimotor-related regions (the left IFG, CH 17; the right preCG, CH27; the right postCG, CH30), and the high-level prefrontal region (the right MFG, CH1). To sum up, with the auditory representation from the speakers' brain as the reference, it is plausible to assume that the listeners employed both an auditory and a sensorimotor mechanism for the processing of the non-native speech toward a shared representation of the conveyed speech information.

It is interesting to note that the neural couplings from the sensorimotor-related regions on the listener's side were also right lateralized. While the right lateralization of the auditory processing for non-native language has been well documented (Weed and Fusaroli 2020; Lakertz et al. 2021), the possible neural mechanisms for right-lateralized sensorimotor processing for non-native speeches remain elusive. Nevertheless, emerging evidences have suggested the recruitment of bilateral sensorimotor system during speech production (Stephens et al. 2010; Silbert et al. 2014b) and speech-in-noise comprehension (Du et al. 2014, 2016). Recent studies have speculated that the left and right sensorimotor-related regions could have different functional roles for speech processing: while the phonemic processing is strongly left-lateralized, the sensorimotor processing of prosody has been suggested to be lateralized to the right hemisphere (Hickok and Poeppel 2007; Sammler et al. 2015; Tang et al. 2021). Hereby, it is plausible to assume that the right-lateralized neural couplings over the sensorimotor regions on the listener's side could reflect a shared representation of the prosodic information.

More importantly, the listener–speaker neural couplings from the speech-auditory-related regions of the listeners' brain suggest an adaptive mechanism for comprehending noisy non-native speeches. The couplings originated from the speech-auditory-related regions of the listeners, i.e. the right STG (CH31) and the right MTG (CH35), showed a significantly positive correlation with the individualized comprehension performance at the highest noise level (−9 dB): the higher the neural coupling between the listeners and the speakers over these regions, the better the comprehension performance. Hereby, the behavioral relevance would highlight the functional importance of these speech-auditory-related regions of the listeners for an adaptive processing of noisy speeches. It supported the hypothesis that non-native people relied on the auditory processing to comprehend noisy speeches (Qi et al. 2019; Borghini and Hazan 2020; Song et al. 2020).

Meanwhile, the behavioral relevance of listener–speaker neural coupling also reveals the involvement of the sensorimotor-related regions of the non-native listeners to comprehend the noisy speeches. As shown in Fig. 4B, the coupling from the right postCG (CH31) of the listeners was also positively correlated with the comprehension performance at the high noise level. It was consistent with previous studies on native speech that the postCG could adaptively maintain the speech representation under adverse conditions (Du et al. 2014, 2016; Du and Zatorre 2017). However, the recruitment of the sensorimotor-related regions for non-native listeners was only restricted to the right postCG, without an extension to more broader regions such as the IFG and the preCG, etc. (Sehm et al. 2013; Alain et al. 2018). This limited recruitment suggests that the non-native listeners might fail to establish a sufficiently strong sensorimotor-related representation of the perceived speech information in support of effective comprehension as the native listeners did (Jones et al. 2013; Archila-Suerte et al. 2015). In spite of this, the present study revealed right-lateralized and mixed mechanisms of both the auditory- and sensorimotor-based processing for non-native speech-in-noise comprehension.

Although the present study did not conduct the experiment with native listeners, the results can be compared with our previous study (Li et al. 2021), in which a group of native (Chinese) listeners attended an experiment paradigm with the same speech materials from the same speakers. For both studies, the listener–speaker neural couplings were found to be within the same frequency range, suggesting a common spectral mechanism for listener–speaker communication. However, the spatial patterns of the listener–speaker neural couplings were substantially different. First, while the coupled brain regions on the native listeners' side were localized to the left IFG and the right MTG/AG, additional right-lateralized brain regions extending to the anterior part were included for the non-native listeners. The specificity of the right-lateralized regions for non-native speech-in-noise processing was further supported by the comparison analysis using the dataset from the present study and the dataset from our previous study (Li et al. 2021): As shown in Fig. S4, the neural couplings to the speakers were generally higher for the non-native listeners than the native listeners, suggesting that these regions could be more related for non-native rather than speech-in-noise processing in general. Second, the noise adaptation mechanism as reflected by the behavioral relevance analysis highlighted the left IFG for native listeners but the right speech-auditory-related regions, i.e. the right STG and the right MTG, and one right sensorimotor-related region, i.e. the right postCG, for non-native listeners. It not only provides important support for the notion of a distinct auditory-based mechanism for the non-native listeners to adaptively process the perceived speeches but also reveals the different recruitments of the sensorimotor-based mechanism for native and non-native listeners. While both the right postCG and the left IFG belonged to the sensorimotor-integration-related regions (Hickok et al. 2011; Du et al. 2014), they were supposed to support speech-in-noise comprehension in distinct ways: while the postCG has been proposed to compensate for the noise-masked phonological information by articulatory simulation (Du et al. 2014; Du and Zatorre 2017), the left IFG has been shown to be more activated when the speech was more intelligible or predictable (Davis and Johnsrude 2003; Okada et al. 2010; Abrams et al. 2013) and thus considered to promote speech-in-noise comprehension by high-level linguistic processing, such as the semantic prediction in a top-down manner (Sehm et al. 2013; Alain et al. 2018). Following this line, it could be inferred that the speech information processed by the sensorimotor integration has reached to a high linguistic level for the native listeners, but it remained at the phonological level for the non-native listeners (Bidelman and Dexter 2015; Drijvers et al. 2019). Last but not least, the coupled brain regions on the speakers' side were more widely distributed for the native listeners (frontal and bilateral temporal regions) than the non-native listeners (only right-lateralized temporal regions). From the speaker's side, it further supported the above discussion that non-native listeners were mainly dealing with the auditory information during speech-in-noise comprehension. Hereby, the non-native listeners might have a limited capability to understand the speaker, resulting in a generally worse performance, especially in noisy conditions.

It should also be mentioned that the observed neural couplings were mainly focused in an ultra-low frequency band of 0.01–0.032 Hz. Interbrain coupling at a similar frequency range has been previously reported inZhuoran Li *et al.* studies about interpersonal verbal communication (Zheng et al. 2018; Liu et al. 2019). The frequency range could correspond to the theme or scene-level processing of the speech narratives as suggested by a recent fMRI study (Baldassano et al. 2017). Due to the limited temporal resolution of the fNIRS technique, the neural activities corresponding to the word or syllable level might not be effectively captured. This might explain the absence of the left-hemisphere brain regions that have been frequently reported in previous EEG or MEG single-brain studies for the processing of fast-changing speech dynamics (Giraud and Poeppel 2012; Ding et al. 2016; Teng et al. 2020). Alternatively, it could be possible that the right hemisphere was dominant for non-native speech processing, as a left-lateralized neural coupling pattern was observed in our previous study on native speeches with a similar paradigm and data analysis procedure (Li et al. 2021). Further studies with simultaneously EEG and fNIRS/fMRI recordings could help elucidate this issue.

This study has some limitations that should be noted. First, the present study only recruited the Korean participants as non-native listeners. Although the results were compared with our previous study (Li et al. 2021) with native listeners, the possible difference in the demographic information of the 2 group of listeners such as their cultural background, language experience, etc., could complicate the interpretation. It would be ideal to have a group of listeners to comprehend speeches in both their native and second (non-native) languages, in order to have a within-participant design to further evaluate the specificity of the right-lateralized mechanism for non-native speech-in-noise comprehension. Second, the present study adopted a sequential interbrain approach rather than having real-time speech communications. While the sequential design is advantageous for its high flexibility in controlling and manipulating the noise level of speeches, its ecological validity might be limited for its lack of real-time bidirectional interaction between speakers and listeners. Real-time communications could further facilitate speech-in-noise comprehension by having an enhanced shared representation between the speaker and the listener and more active prediction of the speaker, leading to broader and stronger neural couplings (Jiang et al. 2021; Kelsen et al. 2022). Third, given the spatial resolution of the fNIRS technique and the precision of the localization method, the results were interpreted at a relatively coarse anatomical level (at the scale of the cerebral sulci and gyri). While the distinction between the sensorimotor-related and the auditory-related regions was not likely to be affected, a finer spatial localization would require imaging techniques such as fMRI and ECoG (e.g. Stephens et al. 2010; Yi et al. 2021; Zhang et al. 2021). Lastly, the direction of the neural coupling was not analyzed in the present study. While the exploration of the directionality would provide more information about the temporal dynamics of the speaker–listener neural coupling

(Stephens et al. 2010; Dai et al. 2018), additional hypotheses need to be formulated in order to further promote our understanding of the non-native speech-in-noise processing.

Acknowledgments

We thank Jiawei Li and Xinmiao Zhang for critical comments.

Supplementary material

Supplementary material can be found at *Cerebral Cortex* online.

Funding

National Natural Science Foundation of China (NSFC) and the German Research Foundation (DFG) in project Crossmodal Learning (NSFC 62061136001/DFG TRR-169/C1, B1); National Natural Science Foundation of China (61977041).

Conflict of interest statement: All the co-authors declare that they have no conflict of interest.

References

Abrams DA, Ryali S, Chen TW, Balaban E, Levitin DJ, Menon V. Multivariate activation and connectivity patterns discriminate speech intelligibility in Wernicke's, Broca's, and Geschwind's areas. *Cereb Cortex*. 2013:23(7):1703–1714.

Alain C, Du Y, Bernstein LJ, Barten T, Banai K. Listening under difficult conditions: An activation likelihood estimation meta-analysis. *Hum Brain Mapp*. 2018:39(7):2695–2709.

Archila-Suerte P, Zevin J, Bunta F, Hernandez AE. Age of acquisition and proficiency in a second language independently influence the perception of non-native speech. *Bilingualism*. 2012: 15(1):190–201.

Archila-Suerte P, Zevin J, Ramos AI, Hernandez AE. The neural basis of non-native speech perception in bilingual children. *NeuroImage*. 2013:67:51–63.

Archila-Suerte P, Zevin J, Hernandez AE. The effect of age of acquisition, socioeducational status, and proficiency on the neural processing of second language speech sounds. *Brain Lang*. 2015:141:35–49.

Baldassano C, Chen J, Zadbood A, Pillow JW, Hasson U, Norman KA. Discovering event structure in continuous narrative perception and memory. *Neuron*. 2017:95(3):709–721.e5.

Benjamini Y, Hochberg Y. Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodology*. 1995:57(1):289–300.

Bidelman GM, Dexter L. Bilinguals at the "cocktail party": Dissociable neural activity in auditory-linguistic brain regions reveals neurobiological basis for nonnative listeners' speech-in-noise recognition deficits. *Brain Lang*. 2015:143:32–41.

Borghini G, Hazan V. Effects of acoustic and semantic cues on listening effort during native and non-native speech perception. *J Acoust Soc Am*. 2020:147(6):3783–3794.

Bradlow AR, Alexander JA. Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *J Acoust Soc Am*. 2007:121(4):2339–2349.

Brainard DH. The psychophysics toolbox. *Spat Vis*. 1997:10(4): 433–436.

Corballis MC. What's left in language? Beyond the classical model. *Ann N Y Acad Sci*. 2015:1359(1):14–29.

Costa A, Sebastian-Galles N. How does the bilingual experience sculpt the brain? *Nat Rev Neurosci*. 2014:15(5):336–345.

Cotosck KR, Meltzer JA, Nucci MP, Lukasova K, Mansur LL, Amaro E. Engagement of language and domain general networks during word monitoring in a native and unknown language. *Brain Sci*. 2021:11(8):1063.

Coulter K, Gilbert AC, Kousaie S, Baum S, Gracco VL, Klein D, Titone D, Phillips NA. Bilinguals benefit from semantic context while perceiving speech in noise in both of their languages: Electrophysiological evidence from the N400 ERP. *Biling-Lang Cogn*. 2021:24(2):344–357.

Cui X, Bryant DM, Reiss AL. NIRS-based hyperscanning reveals increased interpersonal coherence in superior frontal cortex during cooperation. *NeuroImage*. 2012:59(3):2430–2437.

Czeszumski A, Eustergerling S, Lang A, Menrath D, Gerstenberger M, Schuberth S, Schreiber F, Rendon ZZ, Konig P. Hyperscanning: A valid method to study neural inter-brain underpinnings of social interaction. *Front Hum Neurosci*. 2020: 14:39.

Dai B, Chen C, Long Y, Zheng L, Zhao H, Bai X, Liu W, Zhang Y, Liu L, Guo T, et al. Neural mechanisms for selectively tuning in to the target speaker in a naturalistic noisy situation. *Nat Commun*. 2018:9(1):2405.

Davis MH, Johnsrude IS. Hierarchical processing in spoken language comprehension. *J Neurosci*. 2003:23(8):3423–3431.

Di Liberto GM, Peter V, Kalashnikova M, Goswami U, Burnham D, Lalor EC. Atypical cortical entrainment to speech in the right hemisphere underpins phonemic deficits in dyslexia. *NeuroImage*. 2018:175:70–79.

Ding N, Simon JZ. Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. *J Neurosci*. 2013:33(13):5728–5735.

Ding N, Melloni L, Zhang H, Tian X, Poeppel D. Cortical tracking of hierarchical linguistic structures in connected speech. *Nat Neurosci*. 2016:19(1):158–164.

Drijvers L, van der Plas M, Ozyurek A, Jensen O. Native and non-native listeners show similar yet distinct oscillatory dynamics when using gestures to access speech in noise. *NeuroImage*. 2019:194:55–67.

Du Y, Zatorre RJ. Musical training sharpens and bonds ears and tongue to hear speech better. *Proc Natl Acad Sci U S A*. 2017:114(51):13579–13584.

Du Y, Buchsbaum BR, Grady CL, Alain C. Noise differentially impacts phoneme representations in the auditory and speech motor systems. *PNAS Nexus*. 2014:111(19):7126–7131.

Du Y, Buchsbaum BR, Grady CL, Alain C. Increased activity in frontal motor cortex compensates impaired speech perception in older adults. *Nat Commun*. 2016:7(1):12241.

Farahzadi Y, Kekecs Z. Towards a multi-brain framework for hypnosis: a review of quantitative methods. *Am J Clin Hypn*. 2021: 63(4):389–403.

Friederici AD. The brain basis of language processing: From structure to function. *Physiol Rev*. 2011:91(4):1357–1392.

Gao Z, Guo X, Liu CR, Mo Y, Wang JJ. Right inferior frontal gyrus: An integrative hub in tonal bilinguals. *Hum Brain Mapp*. 2020:41(8):2152–2159.

Giraud AL, Poeppel D. Cortical oscillations and speech processing: emerging computational principles and operations. *Nat Neurosci*. 2012:15(4):511–517.

Grant AM, Kousaie S, Coulter K, Gilbert AC, Baum SR, Gracco V, Titone D, Klein D, Phillips NA. Age of acquisition modulates alpha power during bilingual speech comprehension in noise. *Front Psychol*. 2022:13:865857.

Grinsted A, Moore JC, Jevrejeva S. Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Process Geophys*. 2004:11(5/6):561–566.

Gvirts HZ, Perlmutter R. What guides us to neurally and behaviorally align with anyone specific? A neurobiological model based on fNIRS hyperscanning studies. *Neuroscientist*. 2020:26(2):108–116.

Hasson U, Ghazanfar AA, Galantucci B, Garrod S, Keysers C. Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends Cogn Sci*. 2012:16(2):114–121.

Hickok G, Poeppel D. The cortical organization of speech process. *Nat Rev Neurosci*. 2007:8(5):393–402.

Hickok G, Houde J, Rong F. Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron*. 2011:69(3):407–422.

Holroyd CB. Interbrain synchrony: on wavy ground. *Trends Neurosci*. 2022:45(5):346–357.

Hu YY, Wang ZX, Song B, Pan YF, Cheng XJ, Zhu Y, Hu Y. How to calculate and validate inter-brain synchronization in a fNIRS hyperscanning study. *J Vis Exp*. 2021:175.

Huk A, Bonnen K, He BYJ. Beyond trial-based paradigms: Continuous behavior, ongoing neural activity, and natural stimuli. *J Neurosci*. 2018:38(35):7551–7558.

Hull R, Vaid J. Bilingual language lateralization: A meta-analytic tale of two hemispheres. *Neuropsychologia*. 2007:45(9):1987–2008.

Huppert TJ, Diamond SG, Franceschini MA, Boas DA. HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain. *Appl Opt*. 2009:48(10):D280–D298.

Jiang J, Zheng LF, Lu CM. A hierarchical model for interpersonal verbal communication. *Soc Cogn Affect Neurosci*. 2021: 16(1–2):246–255.

Jones OP, Seghier ML, Duncan KJK, Leff AP, Green DW, Price CJ. Auditory-motor interactions for the production of native and non-native speech. *J Neurosci*. 2013:33(6):2376–2387.

Kelsen BA, Sumich A, Kasabov N, Liang SHY, Wang GY. What has social neuroscience learned from hyperscanning studies of spoken communication? *Biobehav Rev*. 2022:132:1249–1262.

Kingsbury L, Hong WZ. A multi-brain framework for social interaction. *Trends Neurosci*. 2020:43(9):651–666.

Lakertz Y, Ossmy O, Friedmann N, Mukamel R, Fried I. Single-cell activity in human STG during perception of phonemes is organized according to manner of articulation. *NeuroImage*. 2021:226:117499.

Lattner S, Meyer ME, Friederici AD. Voice perception: Sex, pitch, and the right hemisphere. *Hum Brain Mapp*. 2005:24(1):11–20.

Lecumberri MLG, Cooke M, Cutler A. Non-native speech perception in adverse conditions: A review. *Speech Comm*. 2010: 52(11–12):864–886.

Li ZR, Li JW, Hong B, Nolte G, Engel AK, Zhang D. Speaker-Listener neural coupling reveals an adaptive mechanism for speech comprehension in a noisy environment. *Cereb Cortex*. 2021: 31(10):4719–4729.

Li JW, Hong B, Nolte G, Engel AK, Zhang D. Preparatory delta phase response is correlated with naturalistic speech comprehension performance. *Cogn Neurodyn*. 2022:16(2):337–352.

Liebenthal E, Mottonen R. An interactive model of auditory-motor speech perception. *Brain Lang*. 2018:187:33–40.

Liu Y, Piazza EA, Simony E, Shewokis PA, Onaral B, Hasson U, Ayaz H. Measuring speaker-listener neural coupling with functional near infrared spectroscopy. *Sci Rep*. 2017:7(1):43293.

Liu W, Branigan HP, Zheng L, Long Y, Bai X, Li K, Zhao H, Zhou S, Pickering MJ, Lu C. Shared neural representations of syntax during online dyadic communication. *NeuroImage*. 2019:198: 63–72.

Mahmoudzadeh M, Dehaene-Lambertz G, Fournier M, Kongolo G, Goudjil S, Dubois J, Grebe R, Wallois F. Syllabic discrimination in premature human infants prior to complete formation of cortical layers. *Proc Natl Acad Sci U S A*. 2013:110(12):4846–4851.

Mendel LL, Widner H. Speech perception in noise for bilingual listeners with normal hearing. *Int J Audiol*. 2016:55(2):126–134.

Ohashi H, Ostry DJ. Neural development of speech sensorimotor learning. *J Neurosci*. 2021:41(18):4023–4035.

Okada K, Rong F, Venezia J, Matchin W, Hsieh IH, Saberi K, Serences JT, Hickok G. Hierarchical organization of human auditory cortex: Evidence from acoustic invariance in the response to intelligible speech. *Cereb Cortex*. 2010:20(10):2486–2495.

Pan YF, Dikker S, Goldstein P, Zhu Y, Yang CR, Hu Y. Instructor-learner brain coupling discriminates between instructional approaches and predicts learning. *NeuroImage*. 2020:211:116657.

Parrell B, Niziolek CA. Increased speech contrast induced by sensorimotor adaptation to a nonuniform auditory perturbation. *J Neurophysiol*. 2021:125(2):638–647.

Peng ZE, Wang LM. Listening effort by native and nonnative listeners due to noise, reverberation, and talker foreign accent during English speech perception. *J Speech Hear Res*. 2019:62(4): 1068–1081.

Pinti P, Tachtsidis I, Hamilton A, Hirsch J, Aichelburg C, Gilbert S, Burgess PW. The present and future use of functional near-infrared spectroscopy (fNIRS) for cognitive neuroscience. *Ann N Y Acad Sci*. 2020:1464(1):5–29.

Qi ZH, Han M, Garel K, Chen ES, Gabrieli JDE. White-matter structure in the right hemisphere predicts Mandarin Chinese learning success. *J Neurolinguistics*. 2015:33:14–28.

Qi ZH, Han M, Wang YX, de los Angeles C, Liu Q, Garel K, San Chen E, Whitfield-Gabrieli S, G, Perrachione TK. Speech processing and plasticity in the right hemisphere predict variation in adult foreign language learning. *NeuroImage*. 2019:192:76–87.

Quaresima V, Bisconti S, Ferrari M. A brief review on the use of functional near-infrared spectroscopy (fNIRS) for language imaging studies in human newborns and adults. *Brain Lang*. 2012:121(2):79–89.

Raharjo I, Kothare H, Nagarajan SS, Houde JF. Speech compensation responses and sensorimotor adaptation to formant feedback perturbations. *J Acoust Soc Am*. 2021:149(2):1147–1161.

Redcay E, Schilbach L. Using second-person neuroscience to elucidate the mechanisms of social interaction. *Nat Rev Neurosci*. 2019:20(8):495–505.

Regalado D, Kong J, Buss E, Calandruccio L. Effects of language history on sentence recognition in noise or two-talker speech: Monolingual, early bilingual, and late bilingual speakers of English. *Am J Audiol*. 2019:28(4):935–946.

Sammler D, Grosbras MH, Anwander A, Bestelmeyer PE, Belin P. Dorsal and ventral pathways for prosody. *Curr Biol*. 2015:25(23): 3079–3085.

Scharenborg O, van Os M. Why listening in background noise is harder in a non-native language than in a native language: A review. *Speech Comm*. 2019:108:53–64.

Schmitz J, Bartoli E, Maffongelli L, Fadiga L, Sebastian-Galles N, D'Ausilio A. Motor cortex compensates for lack of sensory and motor experience during auditory speech perception. *Neuropsychologia*. 2019:128:290–296.

Scholkmann F, Spichtig S, Muehlemann T, Wolf M. How to detect and reduce movement artifacts in near-infrared imaging using moving standard deviation and spline interpolation. *Physiol Meas*. 2010:31(5):649–662.

Schomers MR, Pulvermuller F. Is the sensorimotor cortex relevant for speech perception and understanding? *Front Hum Neurosci*. 2016:10:435.

Schoot L, Hagoort P, Segaert K. What can we learn from a two-brain approach to verbal interaction? *Neurosci Biobehav Rev*. 2016: 68:454–459.

Scott SK, Rosen S, Beaman CP, Davis JP, Wise RJS. The neural processing of masked speech: Evidence for different mechanisms in the left and right temporal lobes. *J Acoust Soc Am*. 2009: 125(3):1737–1743.

Sehm B, Schnitzler T, Obleser J, Groba A, Ragert P, Villringer A, Obrig H. Facilitation of inferior frontal cortex by transcranial direct current stimulation induces perceptual learning of severely degraded speech. *J Neurosci*. 2013:33(40):15868–15878.

Shattuck DW, Mirza M, Adisetiyo V, Hojatkashani C, Salamon G, Narr KL, Poldrack RA, Bilder RM, Toga AW. Construction of a 3D probabilistic atlas of human cortical structures. *NeuroImage*. 2008:39(3):1064–1080.

Silbert LJ, Honey CJ, Simony E, Poeppel D, Hasson U. Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proc Natl Acad Sci U S A*. 2014: 111(43):4687–4696.

Singh AK, Okamoto M, Dan H, Jurcak V, Dan I. Spatial registration of multichannel multi-subject fNIRS data to MNI space without MRI. *NeuroImage*. 2005:27(4):842–851.

Song J, Iverson P. Listening effort during speech perception enhances auditory and lexical processing for non-native listeners and accents. *Cognition*. 2018:179:163–170.

Song J, Martin L, Iverson P. Auditory neural tracking and lexical processing of speech in noise: Masker type, spatial location, and language experience. *J Acoust Soc Am*. 2020:148(1):253–264.

Sonkusare S, Breakspear M, Guo C. Naturalistic stimuli in neuroscience: Critically acclaimed. *Trends Cogn Sci*. 2019:23(8):699–714.

Stephens GJ, Silbert LJ, Hasson U. Speaker-listener neural coupling underlies successful communication. *Proc Natl Acad Sci U S A*. 2010:107(32):14425–14430.

Sulpizio S, Del Maschio N, Fedeli D, Abutalebi J. Bilingual language processing: a meta-analysis of functional neuroimaging studies. *Neurosci Biobehav Rev*. 2020:108:834–853.

Tabri D, Abou Chacra KMS, Pring T. Speech perception in noise by monolingual, bilingual and trilingual listeners. *Int J Lang Commun Disord*. 2011:46(4):411–422.

Tanaka S, Kirino E. The parietal opercular auditory-sensorimotor network in musicians: a resting-state fMRI study. *Brain Cogn*. 2018:120:43–47.

Tang DL, Möttönen R, Asaridou SS, Watkins KE. Asymmetry of Auditory-Motor Speech Processing is Determined by Language Experience. *J Neurosci*. 2021:41(5):1059–1067.

Teng X, Ma M, Yang J, Blohm S, Cai Q, Tian X. Constrained structure of ancient chinese poetry facilitates speech content grouping. *Curr Biol*. 2020:30(7):1299–1305.e7.

Vander Ghinst M, Bourguignon M, Op de Beeck M, Wens V, Marty B, Hassid S, Choufani G, Jousmaki V, Hari R, Van Bogaert P, et al. Left superior temporal gyrus is coupled to attended speech in a cocktail-party auditory scene. *J Neurosci*. 2016:36(5): 1596–1606.

Vigneau M, Beaucousin V, Herve PY, Duffau H, Crivello F, Houde O, Mazoyer B, Tzourio-Mazoyer N. Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *NeuroImage*. 2006:30(4):1414–1432.

Weed E, Fusaroli R. Acoustic measures of prosody in right-hemisphere damage: A systematic review and meta-Analysis. *J Speech Lang Hear R*. 2020:63(6):1762–1775.

Willems RM, Hagoort P. Neural evidence for the interplay between language, gesture, and action: A review. *Brain Lang*. 2007: 101(3):278–289.

Wolmetz M, Poeppel D, Rapp B. What does the right hemisphere know about phoneme categories? *J Cogn Neurosci*. 2011: 23(3):552–569.

Yamamoto AK, Jones OP, Hope TMH, Prejawa S, Oberhuber M, Ludersdorfer P, Yousry TA, Green DW, Price CJ. A special role for the right posterior superior temporal sulcus during speech production. *NeuroImage*. 2019:203:116184.

Ye JC, Tak S, Jang KE, Jung J, Jang J. NIRS-SPM: statistical parametric mapping for near-infrared spectroscopy. *NeuroImage*. 2009:44(2):428–447.

Yeshurun Y, Nguyen M, Hasson U. The default mode network: where the idiosyncratic self meets the shared social world. *Nat Rev Neurosci*. 2021:22(3):181–192.

Yi HG, Chandrasekaran B, Nourski KV, Rhone AE, Schuerman WL, Howard MA, Chang EF, Leonard MK. Learning nonnative speech sounds changes local encoding in the adult human cortex. *Proc Natl Acad Sci U S A*. 2021:118(36):e2101777118.

Yucel MA, Selb J, Cooper RJ, Boas DA. Targeted principle component analysis: A new motion artifact correction approach for near-infrared spectroscopy. *J Innov Opt Heal Sci*. 2014:7(02): 1350066.

Zhang D. Computational EEG analysis for hyperscanning and social neuroscience. *In Computational EEG Analysis*. 2018:215–228.

Zhang X, Noah JA, Dravida S, Hirsch J. Optimization of wavelet coherence analysis as a measure of neural synchrony during hyperscanning using functional near-infrared spectroscopy. *Neurophotonics*. 2020:7(1):015010.

Zhang Y, Ding Y, Huang J, Zhou W, Ling Z, Hong B, Wang X. Hierarchical cortical networks of "voice patches" for processing voices in human brain. *Proc Natl Acad Sci U S A*. 2021:118(52):e2113887118.

Zheng L, Chen C, Liu W, Long Y, Zhao H, Bai X, Zhang Z, Han Z, Liu L, Guo T, et al. Enhancement of teaching outcome through neural prediction of the students' knowledge state. *Hum Brain Mapp*. 2018:39(7):3046–3057.

Zion Golumbic EM, Ding N, Bickel S, Lakatos P, Schevon CA, McKhann GM, Goodman RR, Emerson R, Mehta AD, Simon JZ, et al. Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party". *Neuron*. 2013:77(5):980–991.