Cerebral Cortex, 2021;00: 1–11

doi: 10.1093/cercor/bhab118
Original Article

ORIGINAL ARTICLE

Speaker–Listener Neural Coupling Reveals an Adaptive Mechanism for Speech Comprehension in a Noisy Environment

Zhuoran Li[1,2,†], Jiawei Li[1,2,†], Bo Hong[2,3], Guido Nolte[4], Andreas K. Engel[4] 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 and [4]Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg Eppendorf, Hamburg 20246, Germany

Address correspondence to Dan Zhang, Department of Psychology, Tsinghua University, Room 334, Mingzhai Building, Beijing 100084, China. Email: dzhang@tsinghua.edu.cn.

[†]Zhuoran Li and Jiawei Li contributed equally to this work.

Abstract

Comprehending speech in noise is an essential cognitive skill for verbal communication. However, it remains unclear how our brain adapts to the noisy environment to achieve comprehension. The present study investigated the neural mechanisms of speech comprehension in noise using an functional near-infrared spectroscopy-based inter-brain approach. A group of speakers was invited to tell real-life stories. The recorded speech audios were added with meaningless white noise at four signal-to-noise levels and then played to listeners. Results showed that speaker–listener neural couplings of listener's left inferior frontal gyri (IFG), that is, sensorimotor system, and right middle temporal gyri (MTG), angular gyri (AG), that is, auditory system, were significantly higher in listening conditions than in the baseline. More importantly, the correlation between neural coupling of listener's left IFG and the comprehension performance gradually became more positive with increasing noise level, indicating an adaptive role of sensorimotor system in noisy speech comprehension; however, the top behavioral correlations for the coupling of listener's right MTG and AG were only obtained in mild noise conditions, indicating a different and less robust mechanism. To sum up, speaker–listener coupling analysis provides added value and new sight to understand the neural mechanism of speech-in-noise comprehension.

Key words: fNIRS, inter-brain, neural coupling, speech comprehension, speech-in-noise

Introduction

Real-life verbal communication often takes place in a noisy environment, such as the whispers of goodbye in a crowded train station or the orders of operations besides the machine that fires on all cylinders. Speech-in-noise comprehension has long been known to be closely associated with the auditory system. These regions (e.g., the auditory cortices, superior temporal gyrus [STG], superior temporal sulcus, and middle temporal gyrus [MTG]) could selectively process the to-be-attended speech stream, while effectively suppressing the environmental noise or the to-be-ignored speech streams (Mesgarani and Chang 2012; Vander Ghinst et al. 2016; Etard and Reichenbach 2019; Vander Ghinst et al. 2019), possibly by encoding the

regularity of the acoustic features of the speech. In parallel, it has been suggested that the sensorimotor system (e.g., the posterior part of the inferior frontal gyrus [IFG], the premotor cortex [PM], etc.) is recruited to facilitate speech perception and comprehension as well, by identifying articulatory gestures associated with the perceived speech sound (Liberman and Mattingly 1985; Hickok and Poeppel 2007; Pulvermuller and Fadiga 2010; Guediche et al. 2014; Si et al. 2017; but see Cheung et al. 2016 for a different opinion). While both the auditory system and the sensorimotor system have been shown to be able to adapt to noisy speech situations, the sensorimotor system appears to be more robust against noise. For instance, the neural representation specificity for different phonemes was better reserved with increasing noise level in sensorimotor-related brain regions such as IFG and PM, but not in the auditory regions such as STG (Du et al. 2014). It suggests a compensatory recruitment of the sensorimotor system in the comprehension of the degraded speech, which is proposed to be achieved by forwarding prediction in a top-down way (Pulvermuller and Fadiga 2010; Hickok et al. 2011; Pickering and Garrod 2013; Schomers and Pulvermuller 2016).

Although the investigation of neural mechanisms for speech-in-noise comprehension is regaining popularity in recent years (Ding and Simon 2013; Du et al. 2016; Khalighinejad et al. 2019), it remains to be elucidated how our brain adapts to the noisy environment for effective speech comprehension and how the auditory system and the sensorimotor system differently contribute to the adaptation. While classical studies have focused on the neural responses to short-duration speech units such as phonemes (Assmann and Summerfield 2004; Song et al. 2011; Du et al. 2014), emerging studies have adopted naturalistic speech materials (Ding and Simon 2013; Zou et al. 2019). Specifically, the continuity of naturalistic speech might provide a better basis for top-down prediction and thus facilitates the adaptation of the sensorimotor system to the comprehension of the noisy speech (Alday 2018; Sonkusare et al. 2019). However, analyzing the neural responses to the naturalistic speech stimuli is still a challenging task: On the one hand, it is difficult to code the speech stimuli in the event-related fashion due to the complexity and dynamicity of the naturalistic speech stimuli (Nastase et al. 2019; Sonkusare et al. 2019); on the other hand, the to-be-communicated speech information could go beyond the acoustics of the speech (Hasson et al. 2018; Hagoort 2019; Holler and Levinson 2019), making it even more difficult to establish effective stimulus–response models.

The inter-brain approach could be a promising candidate to explore the mechanisms of natural speech-in-noise comprehension (Redcay and Schilbach 2019; Czeszumski et al. 2020). While previous literature has been dominated by single-brain studies that focus on how the listener's neural activities are related or coupled to the speech stimuli, emerging research in recent years has demonstrated the feasibility of using speaker–listener neural coupling to explore the neural mechanisms of human speech processing. As speech is a tool for interpersonal communication, the coupling of the neural activities between the speakers and the listeners could offer added value by having a more integrated view of the speech process from an interactive perspective (Stephens et al. 2010; Silbert et al. 2014; Liu et al. 2017, 2020; Perez et al. 2017). Accordingly, recent inter-brain studies have found significant speaker–listener coupling during speech communication, mainly over brain regions within both the auditory system and the sensorimotor system of both the speaker and the listener. The couplings were shown to be generally reduced when the speech communications were less effective, such as during a back-to-back dialogue (Jiang et al. 2012), with low predictability language materials (Dikker et al. 2014), unattended during a cocktail party scenario (Kuhlen et al. 2012; Dai et al. 2018), etc. However, these studies employed a fixed level of noise and did not manipulate the speech adversity in a graded way. Nonetheless, a direct, graded manipulation of the speech adversity is necessary to investigate the neural mechanism of adaptation and to reveal how the inter-brain neural coupling changes as a function of the degree of speech adversity.

The present study aimed to investigate how our brain adapted to the noisy environment and comprehended speech from an inter-brain perspective. A group of speakers were invited to tell real-life stories with their functional near-infrared spectroscopy (fNIRS) signals recorded, covering bilaterally typical speech-related brain regions in both the auditory system and the sensorimotor system. The speakers' speech audios were then presented to a group of listeners with fNIRS recordings of the same channels as well. In this study, fNIRS was applied for its suitability for speech tasks: The low operating noise of fNIRS is important for auditory speech comprehension tasks as compared to fMRI; the optical measurement method makes fNIRS relatively tolerant to motion artifacts generated in speech production tasks as compared to EEG; in addition, fNIRS is better in spatially localizing brain activity than does EEG. Four noise levels were introduced during the presentation of the speech audios to the listeners. The speaker–listener neural couplings were analyzed, and its behavioral relevance (as reflected by the correlations with individual listener's comprehension performance) under different noise levels was considered as the indicator for adaptation (Stephens et al. 2010; Dai et al. 2018). A high behavioral relevance of neural coupling would highlight the importance of the listener's specific brain region for speech comprehension. Following previous single-brain studies on speech-in-noise perception or comprehension (Du et al. 2014; Zou et al. 2019), we expected the speaker–listener neural couplings originated from the auditory and the sensorimotor systems of the listeners to be differently affected by noise, with the sensorimotor regions being more behaviorally relevant in high noise conditions. In addition, the present study also highlighted the brain regions of the speakers that were coupled to the listeners' brain, which could shed light on the type of information shared between the speaker and the listener.

Materials and 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

Six college students (3 females; age ranged 21–25 years old) from Tsinghua University participated in the study as speakers. All of them had professional training in broadcasting and hosting. Sixteen college students (8 females; age ranged 18–24 years old) from Tsinghua University were recruited as listeners. All 22 participants were native Chinese speakers, reported right-handed, and having normal hearing and normal or corrected-to-normal vision. Prior to the accomplishment of the experiment, both speakers and listeners were blind to the research question.

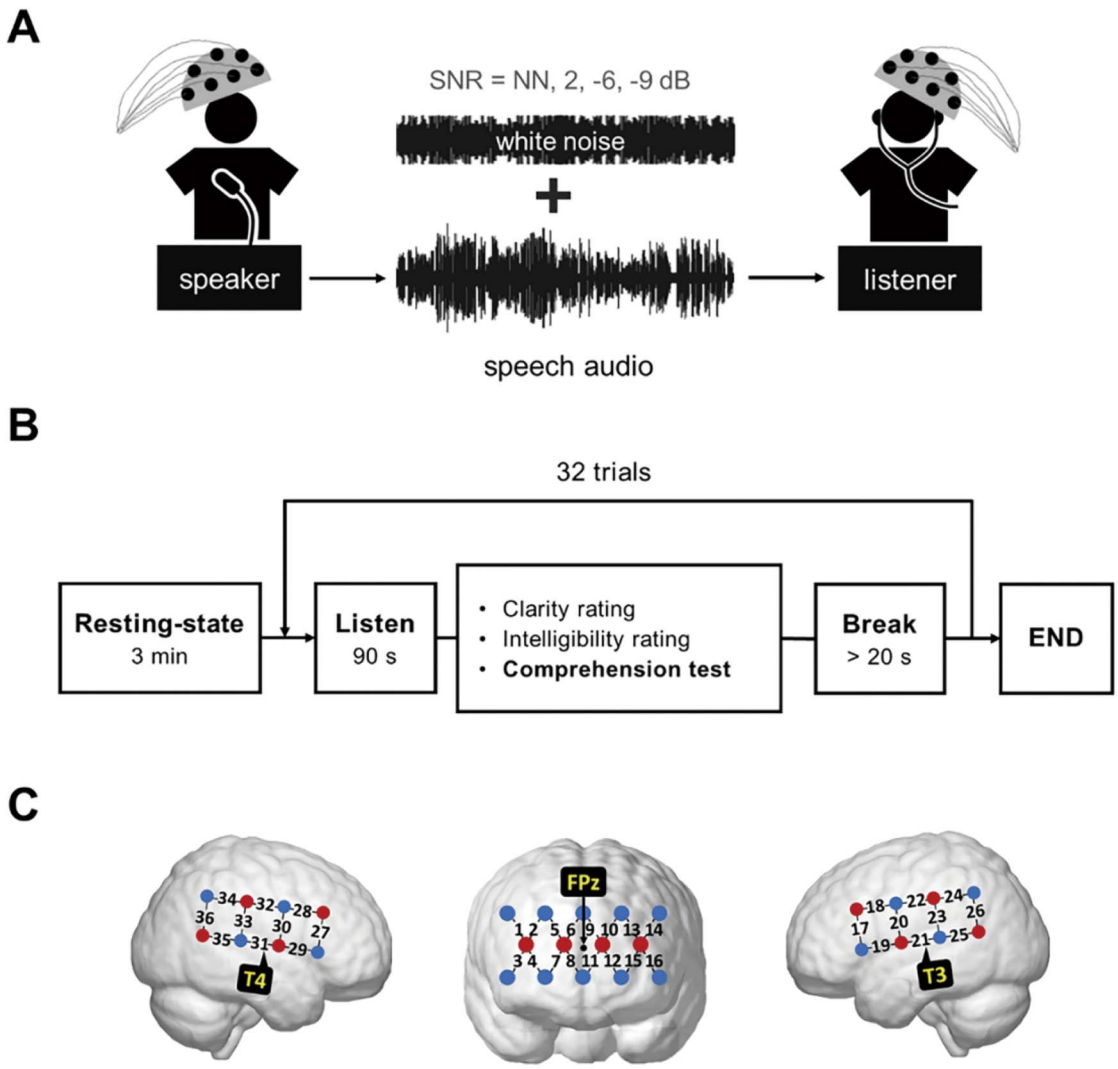

Figure 1. Schematic of the experimental design. (A) experimental design. (B) Experimental procedure of listeners. For the comprehension test, four choice questions based on the narrative audio were used. (C) fNIRS optode probe set. Channels of 21 and 21 were placed at T3 and T4, and the center of the prefrontal probe set was placed at FPz, in accordance with the international 10-20 system.

Speakers were told to give narratives on the given topics, and listeners were told to comprehend the speech in noise.

Data Acquisition

The fNIRS signals of both the speakers and the listeners were recorded from the same 36 channels using the same fNIRS system (NirScan Inc., HuiChuang, Beijing) at a sampling rate of 12 Hz. Near-infrared light of three different wavelengths (785, 808, and 850 nm) was used to detect the concentration change of oxy-hemoglobin (HbO) and deoxy-hemoglobin (HbR). As shown in Figure 1C, three sets of optode probes are placed covering the prefrontal cortex and bilateral inferior frontal, pre- and postcentral, inferior parietal, superior temporal cortices, etc. The positions of CH21 and CH31 are placed at T3 and T4 according to the international 10-20 system. The center of the prefrontal probe set is placed at FPz position.

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 anatomical labels and percentage of overlap for each channel were shown in Supplementary Table S1, which covered bilaterally typical speech-related brain regions in the auditory and the sensorimotor systems.

Experimental Procedure for the Speakers

The present study adopted a sequential inter-brain approach (Redcay and Schilbach 2019), in which the neural activities of the speakers were recorded prior to the listeners. The sequential design was more suitable for the present study as compared to a real-time interactive design, for its convenience of manipulation of the speaker's audios (Stephens et al. 2010; Leong et al. 2017; Liu et al. 2017).

The speaker's experiment was carried out in a sound-attenuated room to ensure the quality of the recorded speech

stimuli. Prior to the start of the experiment, the speakers participated in a resting-state session, during which they were instructed to relax with their eyes closed for 3 min. The resting-state served as the baseline for other conditions.

During the experiment, each speaker participated in 12 story-telling trials. The speaker was required to tell one narrative story in standard Mandarin Chinese while sitting in front of a computer with his/her 36-channel fNIRS signals recorded. The topic of the story was adapted from the National Mandarin Proficiency Test and was closely related to the daily life of college students. One example topic was "'说一说你最喜欢的一本书吧？那本书都讲了什么?" (presented in Chinese, the English translation of this topic is: What is your favorite book? And what is it about?). The speaker was presented with the topic in text on the computer screen and could prepare freely before telling the story (usually 1–2 min). Once ready, the speaker pressed the SPACE key of the computer keyboard to send a start signal to the recording system and then started to speak. The speaker was required to speak for about 90 s. To help the speaker keep track of the time, a number was displayed on the computer screen that counted down from 90 to 1. If the audio did not meet the requirements of either the experimenter (e.g., speech duration) or the speakers themselves (e.g., speech coherence), speaker needed to speak again. The order of 12 topics for each speaker was randomized. Before the experiment, there was one practice trial for the speaker to get familiar with the procedure. A complete list of all the topics is provided in Supplementary Table S2. The story or narrative audios were recorded by a regular microphone at a sampling rate of 44 100 Hz.

Stimuli for the Listeners

A total number of 72 trials with both fNIRS signals and narrative audios were obtained from the six speakers. Thirty-two narrative audios were selected as the stimuli for the listener's experiment. Two speakers (one male, one female) contributed eight stories, and the other four speakers (two male, two female) contributed four stories each. The actual speaking time ranged from 85 to 90 s per trial.

The selected speech stimuli were further processed into four versions at four different noise levels: no noise (NN), 2 dB, −6 dB, and −9 dB. The selection of these signal-to-noise ratio (SNR) levels was decided following previous studies (Ding and Simon 2013; Du et al. 2014). 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 (SPL).

In addition, four four-choice questions per the audio were prepared by the experimenters. These questions and the corresponding choices concerned narrative details and themes that required significant attentional efforts. For instance, one question following a narrative audio about one's major was, "What is the speaker's major as graduate student?(说话人的研究生专业是什么?)," and the four choices were 1) Social science, 2) International politics, 3) Pedagogy, and 4) Psychology (1.社会科学, 2.国际政治, 3.教育学, and 4.心理学). These questions were used to assess the speech comprehension performance of individual listeners.

Experimental Procedure for the Listeners

Each listener participated in 32 speech comprehension trials. In each trial, the listener was required to listen to a narrative audio in one of the four noise levels. The order of the narrative audios and their assigned noise levels were designed following a complete Latin square design for the 16 listeners for the purpose of reducing stimulus-specific confounding (see Supplementary Fig. S1 for the design matrix). After each trial, the listener rated the clarity, intelligibility of the audio with 7-point Likert scales, and completed four four-choice questions about the content of the narrative audio. After that, he/she took at least 20 s to rest and pressed the SPACE key on the computer keyboard to start another trial.

Before the start of the experiment, the listener had a 3-min resting-state session with eyes closed and one practice trial to get familiar with the task, using an additional speech audio stimulus at −2 dB SNR. The experimental procedure was programmed in MATLAB using the Psychophysics Toolbox 3.0 extensions (Brainard 1997). The overall design of the experiment and the structure of one listener's trial are shown in Figure 1A and 1B, respectively.

Data Analysis

Preprocessing

Two preprocessing steps were applied to remove possible motion artifacts by using HoMER2 software package (Huppert et al. 2009). First, motion artifacts were identified and corrected using 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). Artifact-related principal components were removed and the remaining principal components were back-projected to reconstruct the cleaned fNIRS signals (Yücel 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 corrected by a cubic spline interpolation method (function hmrMotionArtifactByChannel; input parameters: $P = 0.99$) (Scholkmann et al. 2010).

Inter-brain Analysis

The fNIRS data collected during the resting-state and task sessions were analyzed. Wavelet transform coherence (WTC) was used to assess the cross-correlation between two fNIRS time series generated by pairs of participants as a function of frequency and time. In the present study, the analysis was used to estimate brain-to-brain coupling between speakers and listeners. For each speech trial, WTC was applied to calculate the neural coupling between listener and the corresponding speaker. There were 36 channels measurement for each participant, leading to 1296 (36 channels from the speaker × 36 channels from the listener) channel combinations for speaker–listener pairs. The frequency range for the WTC analysis was 0.01–0.7 Hz, divided into 112 frequency bins. Following previous studies, data above 0.7 Hz were not included to avoid aliasing of higher-frequency physiological noise, such as cardiac activity (~0.8 to 2.5 Hz), and data below 0.01 Hz were not used to remove very-low-frequency fluctuations (Guijt et al. 2007; Tong et al. 2011; Liu et al. 2019).

The trial-based WTC analysis for each channel combination was performed by using two 300-s fNIRS signal segments from both the listener and the speaker. The 300-s segments covered the 90-s trial duration but were extended to cover an additional 105-s period preceding the trial and an additional 105-s period after the trial. The extended periods were included

for the WTC analysis to ensure a reliable calculation of the inter-brain couplings over the frequency range of interest. The calculated coupling values, organized as a 2D matrix with the coherence value in both time and frequency domains per trial, were then averaged across the 90-s trial duration and converted into Fisher z values after time average. Next, for each listener, the coupling values of trials under the same speech-noise-level condition (eight trials per condition) were averaged, resulting in averaged inter-brain coupling values per listener–speaker pair per condition at the 112 frequency bins for the follow-up statistical analyses. The resting-state condition was analyzed similarly but with a different organization of the trial data: The 180-s resting-state data per listener were extended to a 300-s segment and the couplings were calculated between the listener and each of the six speakers (with their resting-state data), resulting in six trials per listener for the resting-state condition. Then, the coupling values corresponding to the middle 90 s were extracted and averaged. WTC was calculated using the MATLAB function "wcoherence" (Grinsted et al. 2004; Zheng et al. 2020). In sum, each listener had the coupling values under five conditions, that is, four noise levels and the resting state, over 1296 channel combinations and 112 frequency bins (0.01–0.7 Hz). One listener was excluded because of the data acquisition interruption during the experiment. Then, 15 listeners were included in the follow-up analysis.

The channel combinations that showed significant modulation of speaker–listener coupling by speech condition were first identified by a repeated measures ANOVA (rmANOVA). The rmANOVA was performed for all 1296 channel combinations and 112 frequency bins over the range of 0.01–0.7 Hz with the factor of speech conditions. Note that the rmANOVA method was performed on the coupling values averaged over all trials within each speech condition, without considering the specificity of the speakers or the narrative audios. The rationale of making this choice was based on the following considerations: 1) the complete Latin square design of the order of the narrative audios and their assigned noise levels for the 16 listeners could effectively reduce the possible influence by the specificity of the speakers or the narrative audios; 2) such a design could also make it difficult to include the speaker or narrative audio information into analysis, due to the limited amount of inter-brain data of a specific listener–speaker pair at one given speech condition; 3) in addition, ignoring the speaker or narrative audio information would be beneficial to increase the reliability of the inter-brain coupling values, by supporting the practice of averaging the coupling values over all trials within each speech condition.

To account for multiple comparisons, a nonparametric cluster-based permutation method was applied (Maris and Oostenveld 2007). In this procedure, neighboring frequency bins with an uncorrected P-value below 0.05 were combined into clusters, for which the sum of the F-statistics was obtained. The cluster was formed only for the frequency bins but not the channel combinations to avoid potential confounding of global physiological noise (Zheng et al. 2020). A null distribution was obtained by shuffling the labels of five conditions ($n = 1000$ permutations), which defined the maximum cluster-level test statistics and corrected P-values for each real cluster. A cluster would be considered as significant only if its cluster-level statistics was significantly larger than the cluster-level statistics calculated from the permutated data at the 0.05 level. A significant cluster would imply a significant difference in speaker–listener coupling in different conditions within the defined cluster that expand over several frequency bins of one specific channel combination.

The significant clusters were then subject to a correlational analysis for their possible behavioral relevance at different noise levels. Specifically, the neural coupling values within each cluster were averaged in each noise level. Then, Spearman's correlations were computed between the neural coupling values of individual participants and the corresponding speech comprehension performances as reflected by the average accuracy of the four-choice questions of the trials in each noise level.

All the above inter-brain analyses were performed on both the HbO and the HbR signals. However, the HbR-based analyses revealed no significant behavioral correlations. Therefore, only the HbO-based results were reported in the Results section and the HbR-based results are presented in Supplementary Figure S2.

Single-Brain Control Analysis

The listener's fNIRS channels that showed significant inter-brain coupling were further subject to the following single-brain analyses for their single-channel activations and intra-brain couplings among these channels.

The single-channel activations were defined as the standardized HbO values, which were converted into the z-scores by the middle 90 s of resting state. Then, the activations were averaged among trials in the same noise level. A rmANOVA was applied to analyze the activation patterns in four noise levels. Also, a correlation analysis was applied to examine the behavioral relevance of neural activation.

The intra-brain couplings were operationalized as the coherence values based on WTC calculation. The coherence was time-averaged for each trial and then averaged among the corresponding frequency range. Next, after the Fisher z transformation, the coherence values were averaged among trials in the same noise level or resting state. A rmANOVA was also applied to analyze the intra-brain functional connectivity in five conditions. Both these results were then summarized and compared to the inter-brain results to explore the possible uniqueness of the inter-brain approach for understanding speech-in-noise comprehension.

Results

Behavioral Performance

The speech comprehension performance as reflected by the average accuracy of the four-choice questions over all trials within one noise level was found to be significantly different in different noise levels (rmANOVA, $F(3, 42) = 219.81$, $P < 0.001$). The speech comprehension performance reached $91.2 \pm 4.8\%$ (SE) in the NN level and decreased to $71.3 \pm 11.2\%$ in the -9 dB level. Specifically, while the NN level did not significantly differ from the 2 dB level (post hoc t-test, $P = 0.621$), all the other comparisons were significant ($Ps < 0.05$, corrected by false discovery rate [FDR]).

The subjective ratings of clarity and intelligibility showed a similar pattern, with significant effects of the noise level (rmANOVA, $F(3, 42) = 90.86$ and 27.77, $Ps < 0.001$). Post hoc t-tests revealed significant pairwise differences for all possible comparisons ($Ps < 0.05$, FDR corrected). The behavioral performances are summarized in Figure 2. These results suggested that the effect of noise on speech comprehension and perception was effectively manipulated.

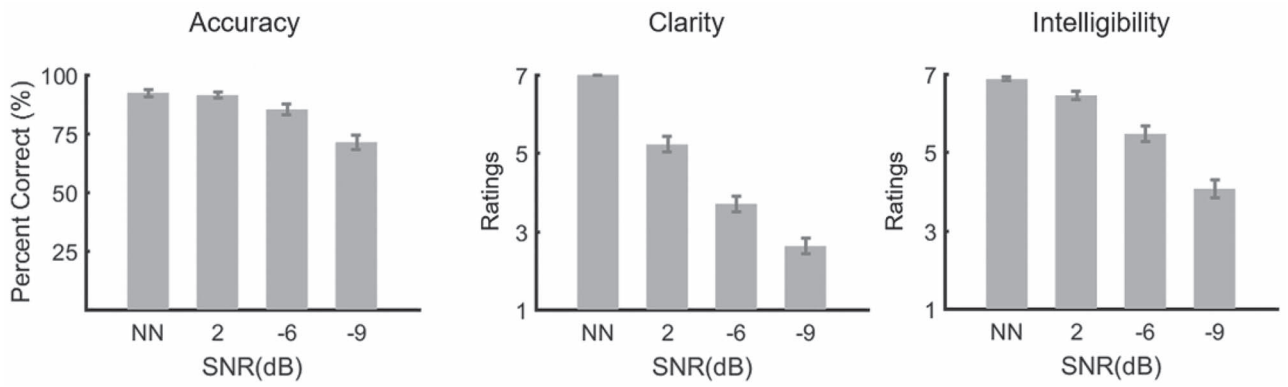

Figure 2. Behavioral performance. The rmANOVA results for accuracy, clarity, and intelligibility were significant (Ps < 0.001). The post hoc *t*-tests indicated that all comparisons were significant (FDR corrected Ps < 0.05) except the accuracy in the NN and SNR = 2 dB condition ($P = 0.621$).

Modulation of Speaker–Listener Neural Coupling by the Noise Level

The nonparametric cluster-based permutation analysis revealed a significant difference in neural coupling for four noise levels and the resting state. For all significant clusters, the frequency bands were around 0.01–0.03 Hz, as shown in Figure 3A. The channel combinations were mainly centered over the left inferior frontal and precentral gyri (left IFG, CH17) and the right middle temporal and angular gyri (right MTG and AG, CH35-36) for listeners. More distributed brain areas were involved on the speaker's side: Manipulation of the noise levels affected the couplings between the listener's left IFG and the speaker's left superior frontal gyrus (SFG), left middle frontal gyrus (MFG), left supramarginal gyrus (SMG), left angular gyrus (AG), right MFG, right postcentral gyrus (postCG), right SMG, and right AG; the noise-dependent couplings of the listener's right MTG and AG were linked to the speaker's left SFG, left MFG, right MFG, right postCG, and right SMG. All significant channel combinations were shown in Figure 3B,C. The specific information of all significant clusters was shown in Supplementary Table S3, including the channels of the speaker and the listener, frequency ranges, and necessary statistics.

Post hoc analysis of the neural couplings within these significant clusters showed two types of pairwise difference patterns. As shown in Figure 3D, the neural couplings of the listener's left-lateralized clusters (CH17) illustrate a significantly higher coupling value at the NN condition as compared to the three noisy conditions (Ps < 0.05, FDR corrected), whereas the neural couplings of the listener's right-lateralized clusters (CH35-36) show comparable strengths at all the listening conditions (Ps > 0.05, FDR corrected). Nevertheless, the neural couplings of the resting-state condition were always significantly lower than the listening conditions for all clusters (Ps < 0.05, FDR corrected).

Behavioral Relevance of Speaker–Listener Neural Coupling

The correlation analyses between the speech comprehension performance and the neural coupling further suggested distinct behavioral relevance patterns of the left- and right-lateralized clusters. Figure 4A summarizes the correlational *r*-values as a function of noise level from all the channel combinations of listener's left IFG (CH17). The *r*-values for each channel combination increased significantly with increasing noise levels, with average *r*-values of -0.11 ± 0.18 (SE), -0.18 ± 0.27, -0.16 ± 0.23, and 0.37 ± 0.30 for the NN, 2 dB, −6 dB, and −9 dB conditions, respectively. A rmANOVA revealed a significant effect of noise level ($F(3, 42) = 25.44$, $P < 0.001$). Post hoc *t*-tests revealed a significantly higher *r*-values at −9 dB than other three noise levels (Ps < 0.05, FDR corrected). The scatter plots in Figure 4B show a typical correlational result by the neural coupling between the listener's left IFG and the speaker's left SMG (CH23): There was a gradual increase of the correlation values from 0.13 at the NN condition to 0.86 at the −9 dB condition.

The correlational results by the neural couplings originated from listener's right MTG (CH35) and AG (CH36) are shown in Figure 4C,D. Similarly, a rmANOVA revealed a significant effect of noise level ($F(3, 24) = 7.23$, $P = 0.001$), with average *r*-values of -0.16 ± 0.28 (SE), 0.03 ± 0.19, 0.15 ± 0.19, and -0.21 ± 0.15 for the NN, 2 dB, −6 dB, and −9 dB conditions, respectively. Post hoc *t*-tests revealed significantly higher *r*-values at −6 dB than NN and −9 dB condition, and higher *r*-values at 2 dB than −9 dB (Ps < 0.05, FDR corrected). Figure 4D gives an exemplar result of the neural coupling between the listener's right MTG and the speaker's right postCG (CH30).

Single-Brain Activation of the Listener

The activation of CH17 and CH35-36 with frequency of 0.01–0.03 Hz was analyzed. The corresponding single-brain activation was not significantly different (rmANOVA for CH17, 35, 36, $F(3, 42) = 1.153, 0.617, 0.390$, $Ps = 0.339, 0.608, 0.761$). No significant correlation was found between speech comprehension performance and the fNIRS activations at four noise levels (Ps > 0.05). Figure 5A illustrates the activation of CH17, 35, 36 at the four noise levels. Figure 5B further illustrates the correlations between the activation and the comprehension performance. There was no significant result.

Intra-brain Coupling of the Listener

The CH17 and CH35-36 were selected as the seed channels to analyze their functional connectivity to the other channels under four noise levels and the resting state. The frequency of 0.01–0.03 Hz was selected. A rmANOVA revealed connectivity of seven channel combinations that were significantly affected by conditions: CH35-CH34, CH36-CH28/30 ($F(4, 56) = 9.288, 14.313, 7.097$, Bonferroni corrected $Ps < 0.05$). The main results of

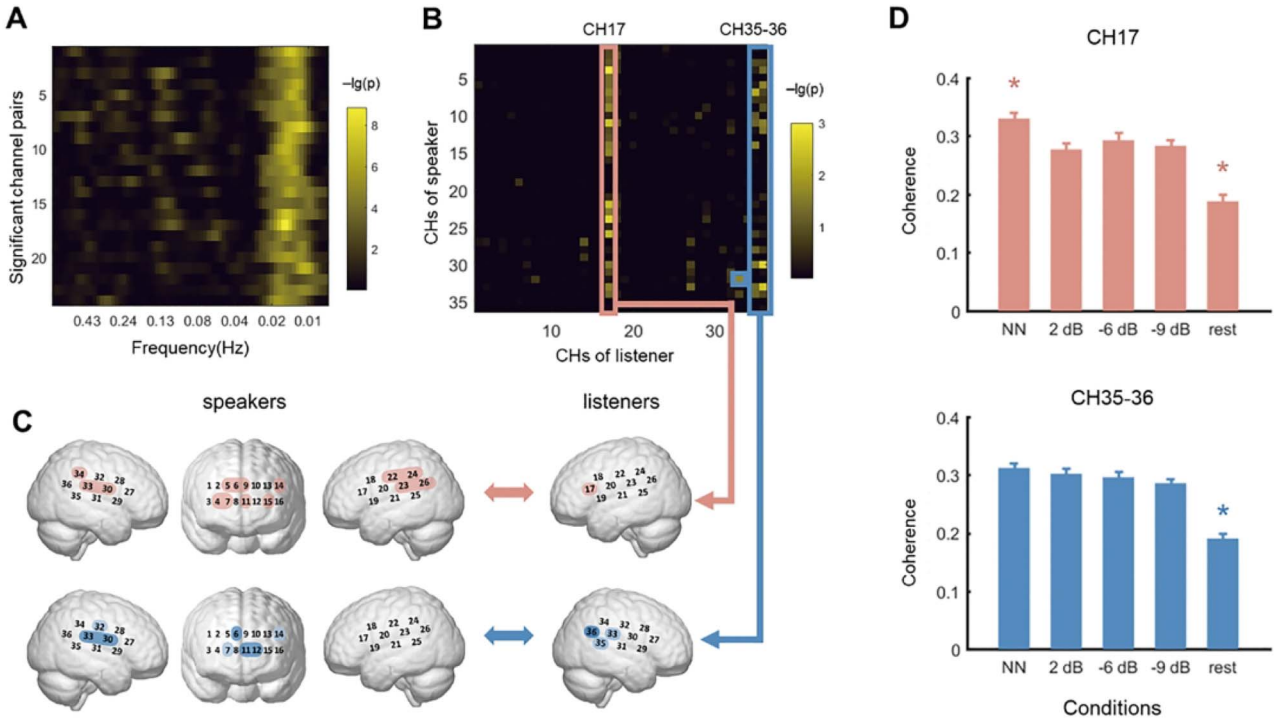

Figure 3. The rmANOVA results. (*A*) Significant frequency bands. To highlight the frequency range of each significant cluster, each row draws the −log10(*p*) of the rmANOVA result of each frequency bin of significant channel combinations. The significant frequency bands range from 0.01 to 0.03 Hz. (*B*) The −log10(*p*) of the maximum cluster for each channel combinations. The significant channel combinations were mainly centered on listener's CH17 and CH35-36. (*C*) Significant channel combinations. The top row shows combinations with listener's CH17, corresponding to left IFG. The red indicates channel combinations with listener's CH17 and speaker's CH4/5/6/7/9/11/14/15/22/23/24/26/30/33/34. The bottom row shows combinations with listener's CH35 and CH36, corresponding to right MTG and right AG. The light blue illustrates three channel combinations with listener's CH35 and speaker's CH7/14/30, and the blue illustrates the channel combinations with listener's CH36 and speaker's CH6/11/12/30/33. There is only one channel combination with listener's CH33 (corresponding to right SMG) and speaker's CH32. (*D*) Mean coupling of all channel combinations with listener's CH17 and CH35-36 (including CH33) in four noise levels and resting state. For CH17, coupling at NN condition is significantly higher than other three noisy conditions (FDR corrected $Ps < 0.05$). For both channel combinations with listener's CH17 and CH35-36, the coupling between speaker and listener at resting state was lower than the four speech conditions (FDR corrected $Ps < 0.001$).

post hoc analysis showed a lower connectivity in resting-state than task conditions. However, the average connectivity between CH17-CH35/36 was not affected by task conditions ($F(4,56) = 0.478$, $P = 0.752$), indicating that the direct connectivity between these two regions (CH17 and CH35-36) was not sensitive to noise.

Discussion

The present study investigated the neural mechanisms of speech comprehension in noise using an fNIRS-based inter-brain approach. Comparison of the speaker–listener neural couplings in different noise conditions revealed two clusters of couplings, one of the listener's left IFG and the other of the listener's right MTG and AG. Both clusters were linked to distributed brain regions on the speaker's side. More importantly, they were associated with different behavioral relevance patterns. Specifically, within the cluster of the listener's left IFG, the correlation between one's neural coupling and comprehension performance was more positive at high noise level; however, top behavioral correlations for the coupling within the cluster based on the listener's right MTG and AG were only obtained in mild noise conditions.

The brain regions showing significant inter-brain coupling on the listener's side were in accordance with previous single-brain and inter-brain studies. The left IFG regions have been frequently reported in classical single-brain studies. They are activated in both speech perception and production and regarded as representative brain areas of sensorimotor system (Hickok et al. 2011; Schomers and Pulvermuller 2016; Glanz Iljina et al. 2018). Meanwhile, the right MTG and AG regions have been found to be typically associated with lexical or semantic processing (Hickok and Poeppel 2007; Friederici 2012; Price 2012). Besides, these brain regions have been found to be related to speech processing in recent inter-brain studies as well (Jiang et al. 2012; Silbert et al. 2014).

Most importantly, our study extended previous understanding on the functional role of left frontal regions in speech processing, as reflected by the quantitative exploration of the behavioral relevance of the inter-brain couplings as a function of noise level. Specifically, the significant increase of the behavioral relevance with increasing noise level suggested an adaptive mechanism for the sensorimotor system for speech processing. As the sensorimotor system on the listener's side has been proposed to be more involved in generating top-down predictions about the perceived speech information during high noise condition to compensate for the degraded speech information (Pickering and Garrod 2013; Schomers and Pulvermuller 2016), the relatively high behavioral relevance of the inter-brain coupling under the high noise condition

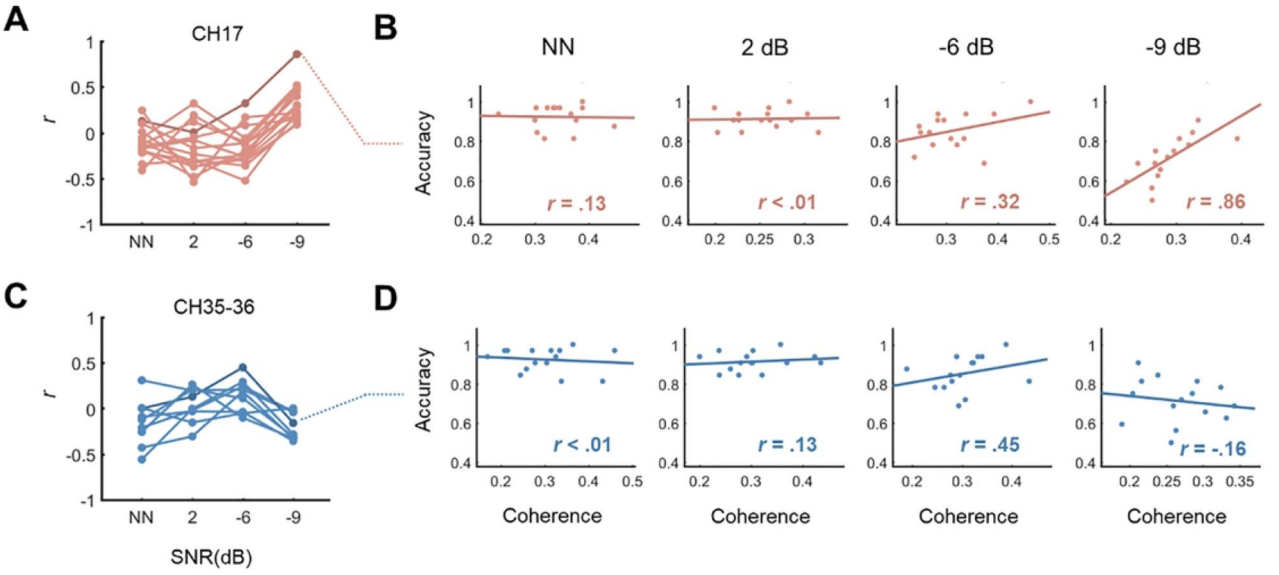

Figure 4. Behavioral relevance of speaker–listener neural coupling. (*A*, *C*) Correlations between coupling and accuracy for all channel combinations of listener's CH17 and CH35-36 in four noise levels. For CH17, the correlation at −9 dB was significantly higher than the other three levels (*P*s < 0.05, FDR corrected). For CH35-36, the correlation at −6 dB was significantly higher than NN and −9 dB condition, and the correlation at 2 dB was significantly higher than −9 dB condition (*P*s < 0.05, FDR corrected). (*B*) Correlations between coupling of listener's left IFG (CH17) to speaker's left SMG (CH23) and accuracy at four noise levels. Only at −9 dB, the correlation was significant (uncorrected *P* < 0.001). (*D*) Correlations between coupling of listener's right MTG (CH35) to speaker's right postCG (CH30) and accuracy at four noise levels. No correlation was significant (*P*s > 0.05).

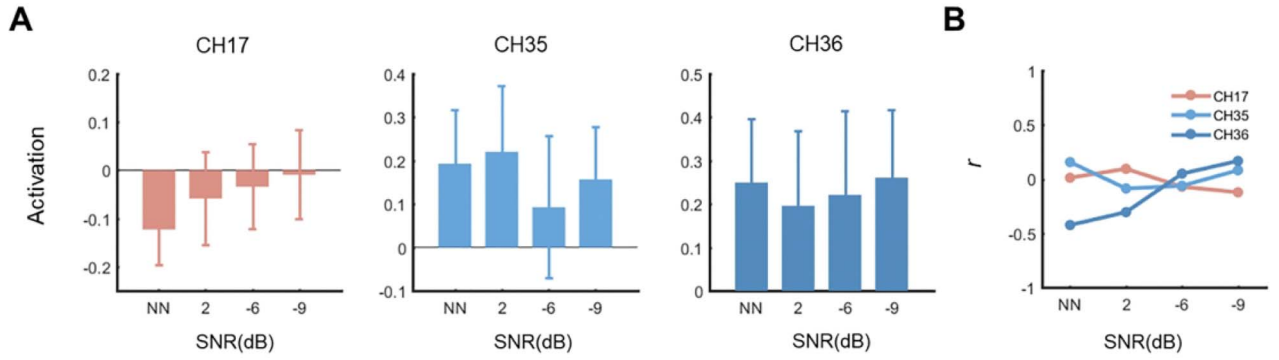

Figure 5. Results of listener's activation and its behavioral relevance. (*A*) Listener's activation of left IFG (CH17), right MTG (CH35), and right AG (CH36). No difference was significant. (*B*) The correlations between activation and accuracy for CH17, 35, and 36 in four noise levels. No correlation was significant (*P*s > 0.05).

implied that the listeners achieved a better understanding of the noisy speech information when they could be more coupled to the speaker's (no-noise) neural activities over the sensorimotor regions. The relatively low behavioral relevance under the low noise conditions, on the other hand, suggests that the sensorimotor regions might not play a key role when the speech information was sufficiently clear. As the sensorimotor regions were proposed to be responsible for articulation-oriented representation of the perceive speech information (Hickok et al. 2011; Schomers and Pulvermuller 2016), our results therefore provide further evidence for the proposal of the active perception of speech by articulation-based internal predictions for noise adaptation (Pulvermuller and Fadiga 2010; Alain et al. 2018).

In contrast, although significant speaker–listener couplings were also observed during speech comprehension as compared to the resting-state baseline for channel combinations originated from the listener's right MTG and AG, these couplings were associated with a distinct behavioral relevance pattern. Specifically, the correlations between the couplings and comprehension performances were more positive for the mild noise condition (−6 dB) as compared to the clear condition (NN) and the high noise condition (−9 dB). While the MTG and AG regions have been traditionally suggested to participate in lexical and semantic processing (reviewed in Hickok and Poeppel 2007; Friederici 2012), they were also found to be responsible for the processing of lower-level speech acoustic features, such as prosody (Kreitewolf et al. 2014), lexical tone, and pitch information (Si et al. 2017). The inter-brain coupling, therefore, can reflect a shared representation of acoustic or lexical-level information by the auditory system between the listener and the speaker, which is not necessarily correlated to high-level speech comprehension performance. This explains why even correlations at −6 dB were not as high as that of left

IFG. The relatively more positive behavioral correlations in the mild noisy conditions could demonstrate the effort to separate speech from noise by the auditory system (Mesgarani and Chang 2012; Vander Ghinst et al. 2016; Etard and Reichenbach 2019). It should also be noted that our results seemed to be inconsistent with a recent inter-brain fNIRS study, in which the neural coupling between the listener and the attended speaker in a multi-talker paradigm was found to be positively correlated with the level of mutual understanding over the left temporal–parietal junction (including AG and SMG) but not over the sensorimotor-related regions (Dai et al. 2018). The inconsistency could be attributed to the difference in the experimental design of the two studies: The noise level in their multi-talker paradigm could be not sufficiently high for the engagement of the sensorimotor system, and the auditory system might be underestimated in the present study due to the lack of real-time bidirectional interaction. Nonetheless, the contrast of the results originated from the listeners' two systems could suggest a different mechanism for speech-in-noise comprehension of the listener's auditory system, which might be less adaptive than the sensorimotor system.

The speaker–listener coupling pattern would shed light on the type of information shared between the two sides. The relatively broader coverage of speech-related brain regions in the present study as compared to previous fNIRS studies (Jiang et al. 2012; Liu et al. 2017; Dai et al. 2018) enabled us to have a more comprehensive overview of the speaker–listener couplings. While the couplings were found to be associated with highly restricted brain regions on the listener's side, widely distributed brain regions from the speaker's side were involved. As shown in Figure 3C, the neural activities of a distributed and bilateral set of brain areas of the speakers were shown to be coupled to the listeners, including right postCG, left SFG, bilateral SMG, bilateral MFG, and bilateral AG. The extensive brain regions of the speakers might be related to a unified language production network that included multiple cognitive systems, from memory retrieval, linguistic organization, to motor planning and execution (Hickok 2012; Walker and Hickok 2016). Therefore, the coupling could reflect speech-related information beyond speech acoustics, with possible involvement of shared sensorimotor representation, interpretive framework, etc. Admittedly, it was possible that the relatively extensive coverage of brain regions on the speaker's side was affected by the global physiological noise, for example, due to motion artifacts by speech production that were not completely resolved by the present preprocessing procedure. The relatively sparse distribution of these regions could argue against such possibility, since global noise would most likely result in continuously distributed brain regions. Nevertheless, further studies with quantitative analyses on speech production-related artifacts are necessary to clarify this issue.

It is also worth noting that the single-brain activation and intra-brain coupling analyses on the listener's side revealed no significant results by the manipulation of the noise level. The activation of left IFG, right MTG, and AG was not significantly affected by noise level, and the behavioral relevance was neither significant at all four noise levels. Besides, noise did not modulate the functional connectivity between sensorimotor system and auditory system, indicating that the function of these two systems operated in a relatively independent way or was managed by some other brain areas or networks. Compared to single-brain analysis, inter-brain coupling could be more sensitive when tracking ongoing social interactions because it considers the neural dynamics from all interacting agents simultaneously (Pan et al. 2018). Besides, inter-brain measures could have a higher signal-to-noise ratio than single-brain measures, as suggested by several recent studies (Parkinson et al. 2018; Nastase et al. 2019). Thus, inter-brain perspective could provide added value and new sight to understand the neural mechanism of natural speech communication. However, previous studies, for example, Du et al. 2014 revealed that noise modulated the activation of speech-related brain regions. One possible reason is different language materials (e.g., simple stimuli, such as phoneme) used in previous studies compared to present study. Alternatively, the current null results could be explained by the relatively low spatial and temporal resolutions of fNIRS, which could be improved in future studies by employing neuroimaging technologies with higher spatial or temporal resolutions, such as fMRI, ECoG etc.

Moreover, the present findings could be extended in the following aspects. Firstly, the fNIRS optode probes only covered language-related brain regions, while some regions related to social interaction, such as theory-of-mind and mentalization networks (Yeshurun et al. 2017; Finn et al. 2020), were not covered. Further studies need to consider wider coverage. Secondly, the present study utilized a sequential inter-brain approach, which owned the advantage of flexible stimuli manipulation and control. However, compared to real-time online communication, it lacks of bidirectional interaction between speaker and listener. To obtain a complete landscape of how human processes noisy speech in real-life condition, a more natural and interactive experimental setting is needed. Thirdly, to obtain a reliable estimation of speech comprehension performance, the comprehension questionnaire accuracies of trials were averaged within each participant. Thus, our results do not explain the behavioral variability within each individual. In future studies, a more efficient and dynamic estimation of speech processing state, even within a trial, could be developed. Lastly, although the relatively small sample size of the present study was decided in reference to previous studies with either a similar noisy speech comprehension paradigm (e.g., Du et al. 2014) or similar inter-brain fNIRS methods (e.g., Liu et al. 2017; Hou et al. 2020), future studies with a larger sample size would help understand the generalizability of the present findings.

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, B4); National Natural Science Foundation of China (61977041).

Notes

The authors wish to thank Dr Yi Du for critical comments. *Conflict of Interest*: All the co-authors declare that they have no conflict of interest.

References

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

Alday PM. 2018. M/EEG analysis of naturalistic stories: a review from speech to language processing. *Language, Cognition and Neuroscience*. 34:457–473.

Assmann P, Summerfield Q. 2004. The perception of speech under adverse conditions. In: Greenberg S, Ainsworth WA, Arthur NP, Fay RR, editors, *Speech processing in the auditory system*. New York: Springer, p. 231–308.

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

Cheung C, Hamiton LS, Johnson K, Chang EF. 2016. The auditory representation of speech sounds in human motor cortex. *Elife*. 5:e12577.

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

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

Dikker S, Silbert LJ, Hasson U, Zevin JD. 2014. On the same wavelength: predictable language enhances speaker-listener brain-to-brain synchrony in posterior superior temporal gyrus. *J Neurosci*. 34:6267–6272.

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

Du Y, Buchsbaum BR, Grady CL, Alain C. 2014. Noise differentially impacts phoneme representations in the auditory and speech motor systems. *Proc Natl Acad Sci U S A*. 111:7126–7131.

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

Etard O, Reichenbach T. 2019. Neural speech tracking in the theta and in the delta frequency band differentially encode clarity and comprehension of speech in noise. *J Neurosci*. 39:5750–5759.

Finn ES, Glerean E, Khojandi AY, Nielson D, Molfese PJ, Handwerker DA, Bandettini PA. 2020. Idiosynchrony: from shared responses to individual differences during naturalistic neuroimaging. *Neuroimage*. 215:116828.

Friederici AD. 2012. The cortical language circuit: from auditory perception to sentence comprehension. *Trends Cogn Sci*. 16:262–268.

Glanz Iljina O, Derix J, Kaur R, Schulze-Bonhage A, Auer P, Aertsen A, Ball T. 2018. Real-life speech production and perception have a shared premotor-cortical substrate. *Sci Rep*. 8:8898.

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

Guediche S, Blumstein SE, Fiez JA, Holt LL. 2014. Speech perception under adverse conditions: insights from behavioral, computational, and neuroscience research. *Front Syst Neurosci*. 7:126.

Guijt AM, Sluiter JK, Frings-Dresen MH. 2007. Test-retest reliability of heart rate variability and respiration rate at rest and during light physical activity in normal subjects. *Arch Med Res*. 38:113–120.

Hagoort P. 2019. The neurobiology of language beyond single-word processing. *Science*. 366:55–58.

Hasson U, Egidi G, Marelli M, Willems RM. 2018. Grounding the neurobiology of language in first principles: the necessity of non-language-centric explanations for language comprehension. *Cognition*. 180:135–157.

Hickok G. 2012. Computational neuroanatomy of speech production. *Nat Rev Neurosci*. 13:135–145.

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

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

Holler J, Levinson SC. 2019. Multimodal language processing in human communication. *Trends Cogn Sci*. 23:639–652.

Hou Y, Song B, Hu Y, Pan Y, Hu Y. 2020. The averaged inter-brain coherence between the audience and a violinist predicts the popularity of violin performance. *Neuroimage*. 211:116655.

Huppert TJ, Diamond SG, Franceschini MA, Boas DA. 2009. HomER: a review of time-series analysis methods for near-infrared spectroscopy of the brain. *Appl Optics*. 48:280–298.

Jiang J, Dai B, Peng D, Zhu C, Liu L, Lu C. 2012. Neural synchronization during face-to-face communication. *J Neurosci*. 32:16064–16069.

Khalighinejad B, Herrero JL, Mehta AD, Mesgarani N. 2019. Adaptation of the human auditory cortex to changing background noise. *Nat Commun*. 10:2509.

Kreitewolf J, Friederici AD, von Kriegstein K. 2014. Hemispheric lateralization of linguistic prosody recognition in comparison to speech and speaker recognition. *Neuroimage*. 102(Pt 2):332–344.

Kuhlen AK, Allefeld C, Haynes JD. 2012. Content-specific coordination of listeners' to speakers' EEG during communication. *Front Hum Neurosci*. 6:266.

Leong V, Byrne E, Clackson K, Georgieva S, Lam S, Wass S. 2017. Speaker gaze increases information coupling between infant and adult brains. *Proc Natl Acad Sci U S A*. 114:13290–13295.

Liberman AM, Mattingly IG. 1985. The motor theory of speech perception revised. *Cognition*. 21(1):1–36.

Liu L, Zhang Y, Zhou Q, Garrett DD, Lu C, Chen A, Qiu J, Ding G. 2020. Auditory-articulatory neural alignment between listener and speaker during verbal communication. *Cereb Cortex*. 30:942–951.

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

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

Maris E, Oostenveld R. 2007. Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods*. 164:177–190.

Mesgarani N, Chang EF. 2012. Selective cortical representation of attended speaker in multi-talker speech perception. *Nature*. 485:233–236.

Nastase SA, Gazzola V, Hasson U, Keysers C. 2019. Measuring shared responses across subjects using intersubject correlation. *Soc Cogn Affect Neurosci*. 14:667–685.

Pan Y, Novembre G, Song B, Li X, Hu Y. 2018. Interpersonal synchronization of inferior frontal cortices tracks social interactive learning of a song. *Neuroimage*. 183:280–290.

Parkinson C, Kleinbaum AM, Wheatley T. 2018. Similar neural responses predict friendship. *Nat Commun*. 9:332.

Perez A, Carreiras M, Dunabeitia JA. 2017. Brain-to-brain entrainment: EEG interbrain synchronization while speaking and listening. *Sci Rep*. 7:4190.

Pickering MJ, Garrod S. 2013. An integrated theory of language production and comprehension. *Behav Brain Sci*. 36: 329–347.

Price CJ. 2012. A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *Neuroimage*. 62:816–847.

Pulvermuller F, Fadiga L. 2010. Active perception: sensorimotor circuits as a cortical basis for language. *Nat Rev Neurosci*. 11:351–360.

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

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

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

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

Si X, Zhou W, Hong B. 2017. Cooperative cortical network for categorical processing of Chinese lexical tone. *Proc Natl Acad Sci U S A*. 114:12303–12308.

Silbert LJ, Honey CJ, Simony E, Poeppel D, Hasson U. 2014. Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proc Natl Acad Sci U S A*. 111:E4687–E4696.

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

Song JH, Skoe E, Banai K, Kraus N. 2011. Perception of speech in noise: neural correlates. *J Cogn Neurosci*. 23:2268–2279.

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

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

Tong Y, Lindsey KP, de BFB. 2011. Partitioning of physiological noise signals in the brain with concurrent near-infrared spectroscopy and fMRI. *J Cereb Blood Flow Metab*. 31:2352–2362.

Vander Ghinst M, Bourguignon M, Niesen M, Wens V, Hassid S, Choufani G, Jousmaki V, Hari R, Goldman S, De Tiege X. 2019. Cortical tracking of speech-in-noise develops from childhood to adulthood. *J Neurosci*. 39:2938–2950.

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. 2016. Left superior temporal gyrus is coupled to attended speech in a cocktail-party auditory scene. *J Neurosci*. 36:1596–1606.

Walker GM, Hickok G. 2016. Bridging computational approaches to speech production: the semantic-lexical-auditory-motor model (SLAM). *Psychon Bull Rev*. 23:339–352.

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

Yeshurun Y, Swanson S, Simony E, Chen J, Lazaridi C, Honey CJ, Hasson U. 2017. Same story, different story. *Psychol Sci*. 28:307–319.

Yücel MA, Selb J, Cooper RJ, Boas DA. 2014. Targeted principle component analysis: a new motion artifact correction approach for near-infrared spectroscopy. *J Innov Opt Health Sci*. 7:1350066.

Zheng L, Liu W, Long Y, Zhai Y, Zhao H, Bai X, Zhou S, Li K, Zhang H, Liu L, et al. 2020. Affiliative bonding between teachers and students through interpersonal synchronisation in brain activity. *Soc Cogn Affect Neurosci*. 15:97–109.

Zou J, Feng J, Xu T, Jin P, Luo C, Zhang J, Pan X, Chen F, Zheng J, Ding N. 2019. Auditory and language contributions to neural encoding of speech features in noisy environments. *Neuroimage*. 192:66–75.