Cognitive Neurodynamics (2024) 18:371–382
https://doi.org/10.1007/s11571-022-09924-w

REVIEW PAPER

How does the human brain process noisy speech in real life? Insights from the second-person neuroscience perspective

Zhuoran Li[1,2] · Dan Zhang[1,2]

Received: 10 October 2022 / Revised: 20 November 2022 / Accepted: 19 December 2022 / Published online: 5 January 2023

Abstract

Comprehending speech with the existence of background noise is of great importance for human life. In the past decades, a large number of psychological, cognitive and neuroscientific research has explored the neurocognitive mechanisms of speech-in-noise comprehension. However, as limited by the low ecological validity of the speech stimuli and the experimental paradigm, as well as the inadequate attention on the high-order linguistic and extralinguistic processes, there remains much unknown about how the brain processes noisy speech in real-life scenarios. A recently emerging approach, i.e., the second-person neuroscience approach, provides a novel conceptual framework. It measures both of the speaker's and the listener's neural activities, and estimates the speaker-listener neural coupling with regarding of the speaker's production-related neural activity as a standardized reference. The second-person approach not only promotes the use of naturalistic speech but also allows for free communication between speaker and listener as in a close-to-life context. In this review, we first briefly review the previous discoveries about how the brain processes speech in noise; then, we introduce the principles and advantages of the second-person neuroscience approach and discuss its implications to unravel the linguistic and extralinguistic processes during speech-in-noise comprehension; finally, we conclude by proposing some critical issues and calls for more research interests in the second-person approach, which would further extend the present knowledge about how people comprehend speech in noise.

Keywords Speech-in-noise · Speech comprehension · Second-person neuroscience · Speaker-listener neural coupling · Naturalistic stimuli.

Introduction

Noise is unavoidable during our daily speech comprehension, such as another speaker at a cocktail party, the sound of traffic horns on the street, and the echoes in an empty hall. Maintaining a relatively robust speech comprehension in a noisy environment is of great importance to human life. Large quantities of psychological, cognitive and neuroscientific research have investigated how people comprehend noisy speech and achieved a wealth of discoveries (Coffey et al. 2017; Dryden et al. 2017; Alain et al. 2018). For example, multiple mechanisms, e.g., auditory mechanism and sensorimotor mechanism, etc., have been found to support speech-in-noise comprehension in distinct ways and correspond to different neural basis (Du et al. 2014; Guediche et al. 2014; Etard and Reichenbach 2019). However, most previous studies have long investigated this issue by using highly-controlled and short-duration artificial stimuli, such as phonemes and words (Hamilton and Huth 2020), which failed to resemble the naturalistic speech used in real-life scenarios. Moreover, traditional neuroscience routinely adopted a single-brain or third-person neuroscience approach. Participants were often isolated from the natural environment of interpersonal communication and required to accomplish a series of simple tasks with the only instruction of a computerized program (Hasson et al. 2012). Such a single-brain approach instructed participants to solely and passively perceive a

✉ Dan Zhang
dzhang@tsinghua.edu.cn

[1] Department of Psychology, School of Social Sciences, Tsinghua University, Room 334, Mingzhai Building, Beijing 100084, China

[2] Tsinghua Laboratory of Brain and Intelligence, Tsinghua University, Beijing 100084, China

non-interactive stimulus (Redcay and Schilbach 2019), neglecting the nature of interpersonal communication through language (Jiang et al. 2021). These experimental settings are quite different from naturalistic speech situations. Consequently, the neurocognitive mechanisms for speech-in-noise comprehension remain much unclear.

In recent years, modern advances in the simultaneous dual- or multiple-brain measurement technique (also known as 'hyperscanning', Montague et al. 2002) have given rise to a new approach to neuroscience: the inter-brain or second-person neuroscience approach (Schilbach et al. 2013; Hasson and Frith 2016; Redcay and Schilbach 2019). In contrast to the traditional single-brain or third-person approach that focuses on estimating each individual's neural responses to the highly-controlled and simplified stimuli, e.g., phonemes and words, the second-person neuroscience approach measures the neural activities of the socially interactive agents (i.e., speaker and listener) during interaction and analyzes how the coherence or coupling of their neural activities varies among different conditions or correlates to the interactive behavior (Czeszumski et al. 2020; Kingsbury and Hong 2020; Holroyd 2022). It gives a novel perspective for investigating the neural basis of speech-in-noise comprehension from an integrative view. In this review, we first briefly review the previous findings about how the brain processed speech in noise and discuss their limitations; next, the second-person neuroscience approach and its advantages over the classical third-person approach are introduced; then, how the second-person neuroscience approach could help to reveal the linguistic and extralinguistic processes during speech-in-noise comprehension are discussed respectively; finally, we conclude by proposing some critical issues and calls for more research interests on the second-person approach for studying the neural mechanisms of speech-in-noise comprehension.

How does the brain process speech in noise?

Dual mechanisms, i.e., auditory mechanism and sensorimotor mechanism, have long been reported to support speech-in-noise comprehension (Du et al. 2014; Alain et al. 2018; Etard and Reichenbach 2019). The auditory mechanism refers to the faithful processing of multi-level linguistic information in a bottom-up way. It is associated with the brain regions responsible for acoustic, phonological, syntactic and lexical-semantic processing, which is mainly located in the temporal lobe (Hickok and Poeppel 2007; Price 2012) and with an extension to frontal regions for complex linguistic computation (Friederici 2012; Fedorenko and Blank 2020). The auditory mechanism could filter out the noise by selectively processing the target speech while suppressing the encoding of noise based on their various acoustic statistics (Guediche et al. 2014; Herrmann et al. 2014; Etard and Reichenbach 2019; Vander Ghinst et al. 2019; Marrufo-Perez et al. 2020), or resolve the noise-induced ambiguity of speech information by integrating it to the linguistic context (Zekveld et al. 2011; Golestani et al. 2013; Shi and Koenig 2016; Rysop et al. 2021).

In contrast to the auditory mechanism, the sensorimotor mechanism refers to the generation of linguistic information and the subsequent integration to the actual sensory input. It was associated with production-related regions covering the left posterior frontal lobe and the sensorimotor interface located at the posterior dorsal-most aspect of the left temporal lobe, etc. (Hickok and Poeppel 2007; Pulvermuller and Fadiga 2010; Sehm et al. 2013; Du et al. 2014; Alain et al. 2018). It supports speech-in-noise comprehension by compensating for the noise-masked linguistic information through motor simulation (Liberman et al. 1967; Hickok and Poeppel 2007; Pulvermuller and Fadiga 2010) or the content-based prediction (Hickok et al. 2011; Pickering and Garrod 2013; Schomers and Pulvermuller 2016). The sensorimotor-related regions in the frontal and parietal regions were broadly reported to be activated in noisy conditions (Du et al. 2014; Alain et al. 2018).

Whereas both of the dual mechanisms play supportive roles during speech-in-noise comprehension, the sensorimotor mechanism seems to be more robust against the noise than the auditory mechanism. It is because the sensorimotor mechanism could benefit from the linguistic information generated from the internal model, while the auditory mechanism relies on the relative completeness of the external auditory input. In this way, when an increasing intensity of background noise has interrupted many acoustic and linguistic details of the speech, the sensorimotor mechanism might be more adaptive and supportive for speech-in-noise processing as the auditory mechanism might have failed to function to support speech comprehension. Du et al. 2014 adopted fMRI to measure the neural activities from auditory-related and sensorimotor-related regions when people listened to phoneme tokens under various signal-to-noise ratio (SNR) levels, i.e., no noise, 8, -2, -6, -9 and − 12 dB. While the activation of anterior regions of superior temporal gyrus (STG), the anterior and posterior regions of middle temporal gyrus (MTG), etc., were decreased by increasing background noise, the neural activities of the sensorimotor regions, e.g., anterior insular and adjacent Broca's area, the ventral premotor cortex, etc., were enhanced. Furthermore, the multivoxel patterns of neural activities in the sensorimotor regions exhibited effective phoneme representation even when the intensity of noise became stronger than the

original speech. Meanwhile, the phoneme representation of the neural activities in the auditory regions was disrupted by even very mild background noise (Du et al. 2014, 2016; Du and Zatorre 2017). These findings suggested that the sensorimotor mechanism was more adaptive and might play a more fundamental role when the environmental noise became strong.

In addition to the above dual mechanisms, several studies have also highlighted the importance of extralinguistic processing for speech-in-noise comprehension (McGowan 2015; Hernandez et al. 2020). The extralinguistic processing is an indispensable part of speech comprehension (Hasson et al. 2018; Hagoort 2019). On the one hand, as speech is intended for interpersonal communication, speech comprehension is not limited to the linguistic processes but related to a broad range of extralinguistic processes to cope with the situational and (both individually and socially) personal content contained in the speech (Hasson et al. 2018; Redcay and Moraczewski 2020; Yuan 2020). They refer to a series of non-linguistic-specific but domain-general cognitive processes, including mentalizing, perspective taking, personal memory and knowledge, self- and social-cognition, social emotion, and etc. (Redcay and Moraczewski 2020; Yeshurun et al. 2021). On the other hand, the extralinguistic processes could reversely modulate the hierarchical encoding and the content-driven prediction of the linguistic information in a top-down fashion. For example, manipulating the listeners' beliefs about the age, gender, race, etc., of a speaker could influence the processing of speech signals (Hanulikova 2021; Kutlu et al. 2022; Yu 2022). Therefore, the extralinguistic processing might also help resolve the interference of the noise. For example, when presented with a congruent cue about the speaker's social identity, i.e., race, people could better comprehend the speech in noise (McGowan 2015).

Although lots of efforts have been devoted to exploring the neural mechanisms of speech-in-noise comprehension, it remains to be elucidated on how people comprehend speech in noise in real-life scenarios. This is because that traditional neuroscience typically adopts a reductionist and deductive approach to investigate the neural response to a particular stimulus (Hasson et al. 2012; Sonkusare et al. 2019; Kingsbury and Hong 2020; Holroyd 2022). Researchers often used highly-controlled and short-duration speech stimuli, such as phonemes, words, and single sentences (Anderson and Kraus 2010; Scharenborg and van Os 2019; Hennessy et al. 2022), to measure the behavioral or neural response to a particular speech stimulus in noise. However, these isolated materials didn't resemble the continuity, complexity and dynamics of naturalistic speech. Moreover, they lacked of continuous contexts, which formed the essential basis for recovering the missing part from the noise-contaminated speech (Golestani et al. 2013; Hennessy et al. 2022). In this way, the neural mechanism of naturalistic speech comprehension in noise is still little understood. Besides, except for the over-simplified stimuli, the corresponding simplification of the speech-related task, such as the Quick Speech-in-Noise Test (QuickSIN), Hearing in Noise Test (HINT), or words in noise (WIN), etc. (Wilson et al. 2007, 2012; Holder et al. 2018), encouraged participants to simply perceive or comprehend speech in a decontextualized and non-social way (Guediche et al. 2014; Sonkusare et al. 2019; Hitczenko et al. 2020; Jaaskelainena et al. 2021). Such neglect of the interpersonal nature of speech led to the underestimate of the extralinguistic processing, i.e., mentalizing, perspective taking, etc., (Redcay and Moraczewski 2020). As discussed above, the extralinguistic processing not only influenced the linguistic processing of speech but also helped to resolve the interference of noise. Thus, to obtain a complete vision of the neurocognitive mechanisms for speech-in-noise comprehension, both the naturalistic speech stimuli and paradigm encouraging people to comprehend speech as naturally as in real-life scenarios are needed.

From third-person neuroscience to second-person neuroscience

The naturalistic stimuli paradigm has been gaining popularity recently (Sonkusare et al. 2019). It refers to the employment of naturalistic stimuli, such as natural speech, videos, and music, that people typically encounter in everyday life. While the naturalistic stimuli paradigms can be employed in laboratory settings, they are expected to give an ecologically reasonable approximation of real-life situations by resembling the complexity, diversity and dynamics of everyday stimuli (Sonkusare et al. 2019). The use of natural speech not only improves the ecological validity of neuroscientific research but also extends the previous knowledge about the neural mechanism of speech processing (for a review, Hamilton and Huth 2020). For example, more widespread brain regions beyond the classical language-specific areas, i.e., the Wernicke's and the Broca's areas, were found to be activated when comprehending natural speech as compared to isolated and simple language materials (Huth et al. 2016; de Heer et al. 2017). Besides, a less left-lateralized response was observed when people listened to natural speech than simple language stimuli (Hamilton and Huth 2020). While some researchers proposed that the rich meaning and long duration of the naturalistic speech contributed to a more extensive activation of bilaterally higher-order cortical areas (Price 2012), some other researchers explained it as increased involvement of the right hemisphere for the processing of

prosody (Si et al. 2017; Weed and Fusaroli 2020), emotion (Schirmer and Kotz 2006), social information (Alexandrou et al. 2017), etc., which were fully activated by the natural speech.

However, the use of natural speech poses a great challenge for the classical single-brain or third-person neuroscience approach, which routinely calculates the brain-to-stimulus contingency with the measurement of individual's brain response to a particular stimulus. For one thing, the multiple high-level linguistic information is difficult to be coded quantitatively and objectively, let alone the more implicit extralinguistic processes (Armeni et al. 2017). While the recent advance in natural language processing algorithms seemed to give quantitative and human-like descriptions to speech at linguistic levels, such as syntax (Nelson et al. 2017) and semantics (Broderick et al. 2018; Grand et al. 2022), and even at the extralinguistic levels, such as sentiment or emotion (Tanana et al. 2021), social state (Badal et al. 2021), etc. These labels still require validation and verification by human behavioral and neural data (Kingsbury and Hong 2020). Also, the multi-level linguistic and extralinguistic information were often interwoven with each other. It's hard to neatly extract one particular feature from naturalistic speech. For another, even with the quantitative labels of natural speech from human coding or the computational language models, the continuous, time-varying and multivariate properties of natural speech would still render the conventional analytical method, i.e., the event-related design with general linear modelling, ineffective. To address this issue, some powerful mathematical models, e.g., temporal response function (Ding and Simon 2012; Mesgarani and Chang 2012; Golumbic et al. 2013; Broderick et al. 2018; Li et al. 2022a), are developed or introduced to the neuroscience. In line with the traditional event-related modelling, they typically model one or several features to the measured neural data to estimate how the listener's brain processes particular linguistic information. However, these models often pre-assume some hypotheses about the brain and its correspondence to the stimulus, which are sometimes too abstract and over-simplified (Sonkusare et al. 2019). For instance, the temporal response function approximates the brain as a linear time-invariant system, while the brain is neither linear nor time-invariant (Crosse et al. 2016). These assumptions will somewhat limit the validity of the explanation for the brain.

The recent advent of inter-brain or second-person neuroscience (Hasson et al. 2012; Schilbach et al. 2013; Redcay and Schilbach 2019; Kingsbury and Hong 2020) provides a novel solution for investigating natural speech comprehension in no-noise or noisy conditions. As shown in Fig. 1, in contrast to the single-brain approach relying on the modelling of people's neural response to a particular stimulus or feature, the inter-brain approach collects data from both speaker's and listener's brains, and estimates how the time series of their neural signals were synchronized or coupled to each other (Czeszumski et al. 2020; Kelsen et al. 2022). Actually, the synchronization or alignment between the listener and the speaker is the basis for successful comprehension (Garrod and Pickering 2004; Hasson and Frith 2016). It entails the shared processes of the multi-level linguistic information, i.e., acoustic, phonology, syntax and semantics, and the extralinguistic information, such as situational model, etc. (Garrod and Pickering 2004; Hasson and Frith 2016). Following this line, the listener's neural activities underlying these multiple processes would also be synchronized or coupled to the speaker. Emerging studies have demonstrated that the neural activities of the speaker and the listener were significantly coupled to each other (e.g., Stephens et al. 2010; Jiang et al. 2012; Kuhlen et al. 2012; Dikker et al. 2014).

Speaker-listener neural coupling is achieved by the transfer of speech from the speaker to the listener (Hasson et al. 2012; Schoot et al. 2016; Kelsen et al. 2022). In essence, the speaker's and the listener's brains together could be analogous to a coupled two-source system that communicates via the wireless transmission of sound-based physical signals, i.e., speech (Hasson et al. 2012; Schoot et al. 2016; Kelsen et al. 2022). The emergence of the brain-to-brain coupling relies on the brain-to-stimulus coupling at both the speaker's and the listener's sides. Thus, the inter-brain neural coupling would disappear when no verbal communication took place between the speaker and the listener (Stephens et al. 2010). Moreover, as the speech was originally organized and generated by the speaker, the production-related neural activity inside the speaker's brain could be regarded as a standardized reference to estimate how the listener's brain processes the speech. With this logic, the more coupled the listener's neural activity is to the speaker, the better the listener would comprehend the speaker. Numerous studies gave supportive evidence that the speaker-listener neural coupling level was positively correlated to the listener's comprehension (e.g., Stephens et al. 2010; Dai et al. 2018; Liu et al. 2020). Thus, the strength of speaker-listener neural coupling could reflect whether or to how much degree the listener's brain was (correctly) dealing with the speech.

As compared to the single-brain approach, the inter-brain approach owns some advantages. Firstly, the inter-brain neural coupling analysis provides a model-free and data-driven method by modelling the neural activities of listener's brain to the speaker (Sonkusare et al. 2019). It doesn't propose any assumption of the explicit model of the complex contents of dynamic stimuli and the brain-to-stimulus correspondence (Redcay and Schilbach 2019;

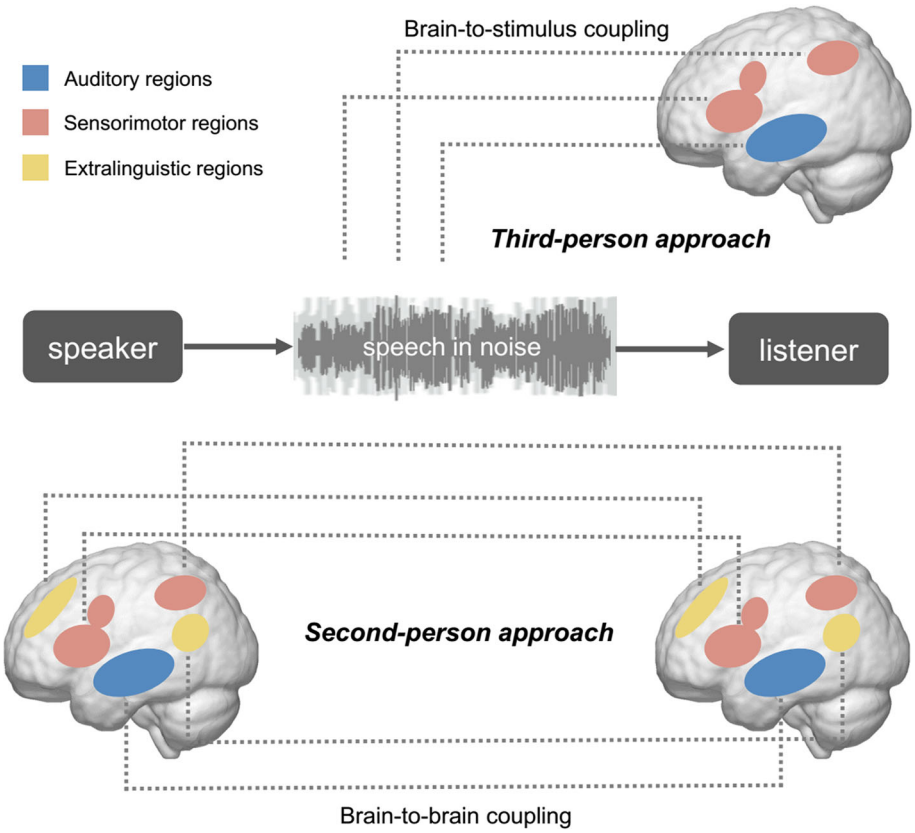

Fig. 1 The third-person and second-person neuroscience approaches for studying speech-in-noise comprehension. The third-person neuroscience approach has typically investigated the brain-to-stimulus contingencies and revealed both involvements of the auditory and sensorimotor regions for linguistic processing in noisy conditions. The auditory-related regions (marked in blue) are mainly located in the temporal lobes; the sensorimotor-related regions (marked in red) are mainly covering the left posterior frontal lobe, ventral premotor cortex, and the posterior dorsal-most aspect of the left temporal lobe. As previous single-brain studies often neglected the interpersonal nature of speech, the extralinguistic processes, e.g., mentalizing, lack of adequate concern. In contrast, the second-brain neuroscience approach collects data from both speaker's and listener's brain from an integrative view and calculates the coupling between their neural activities. The speaker-listener neural coupling not only originates from the auditory- and sensorimotor-related regions but also from the extralinguistic-related regions (marked in orange), such as the middle frontal gyrus, temporal-parietal junction, etc.

Sonkusare et al. 2019), and thus gets rid of the pre-assumed bias about the brain, e.g., a linearized approximation of the neural system. Secondly, the inter-brain approach gives a powerful tool to fill the gap on the high-level linguistic and extralinguistic processes of natural speech, which are often underestimated by existing studies (Hamilton and Huth 2020). As the interpersonal alignment comprehensively covers the multiple linguistic and extralinguistic processes of speech, the corresponding speaker-listener neural coupling also emerges from a wide range of speech-related regions. It was found to cover the linguistic-related regions over the fronto-temporo-parietal cortex and further extend to extralinguistic regions associated with the processing of semantic and social aspects of the story, such as the precuneus, striatum, dorsolateral prefrontal cortex, orbitofrontal cortex and medial prefrontal cortex (Stephens et al. 2010; Silbert et al. 2014). Thirdly, the inter-brain approach restores the nature of interpersonal communication through speech. In contrast to the single-brain approach that solely analyzes the listener's neural response to the speech stimuli, the inter-brain approach simultaneously includes the listener and the producer of the speech, i.e., the speaker. It encourages the listener to comprehend the speech in an interpersonal and social way, and even allows the listener to communicate with the speaker in a real-time fashion (Jiang et al. 2021).

In sum, the recently well-developed inter-brain approach proposes a novel perspective to examine the neural mechanisms of speech-in-noise comprehension by modelling the speech-evoked neural response of the listener to the production-related neural response of the speaker. It not only promotes the use of naturalistic speech but also gives a new conceptual framework, i.e., speaker-listener neural coupling, to measure the listener's neural processing from an integrative view. Thus, the inter-brain approach owns the potential to deepen the understanding of the neural mechanisms of speech-in-noise comprehension.

In the next parts, its implications for unfolding the linguistic and extralinguistic processing during speech-in-noise comprehension are to be discussed.

Inter-brain neural coupling underlies the linguistic processing in noise

The linguistic processes are the basic parts of speech processing. As both the auditory and the sensorimotor mechanisms are suggested to support speech-in-noise comprehension (e.g., Ding and Simon 2013; Du et al. 2014), the speaker-listener neural coupling from the corresponding brain regions could underlie the alignment of the listener's auditory or sensorimotor processing of linguistic information to the speaker. By examining how these inter-brain neural couplings from the auditory- and sensorimotor-related regions occur and further correlate to the comprehension in noisy conditions, researchers could understand how these various linguistic processes are involved and contribute to the speech-in-noise comprehension, respectively.

One recent study has used the inter-brain approach to explore how the auditory and sensorimotor mechanisms of linguistic processing supported natural speech-in-noise comprehension (Li et al. 2021). In this study, both Chinese speaker and listener participants were recruited. The speakers were invited to give unrehearsed narratives based on given topics. Their speeches were recorded and added with different intensities of meaningless white noise, which were manipulated into four conditions with the SNR equaling to no noise, 2, -6 and − 9 dB. The listeners then listened to these narratives in noisy conditions and finished comprehension tests about the content of the narratives. Both of the speakers' and the listeners' neural activities were measured by functional near-infrared spectroscopy (fNIRS). Results showed that the neural activity from the listener's auditory-related regions, i.e., right MTG and angular gyrus (AG), and sensorimotor-related regions, i.e., left IFG, were coupled to the speaker in both clear and noisy conditions. However, only the neural coupling from the left IFG was correlated to the listener's comprehension performance at the strong noise level. These results suggested that while both the auditory and sensorimotor processes were activated in noisy conditions, the sensorimotor processes played a more supportive role in comprehension when noise became strong (Li et al. 2021).

This study validated the feasibility of the inter-brain approach for revealing the neural mechanism of speech-in-noise comprehension. To further investigate how people comprehend non-native speech in noise and explain the non-native disadvantage in noisy conditions, Li et al. 2022b recruited another group of Korean listeners who had learnt Chinese for years to listen to Chinese narratives at different noise levels. Their neural coupling to the Chinese speakers was calculated. They found that the non-native listener relied on a right-lateralized mechanism for linguistic processing. In specific, the neural activities from the non-native listener's right dorsolateral prefrontal cortex, pre- and post-central gyrus (preCG/postCG), MTG and STG, as well as the left IFG, were coupled to the speaker. Among these regions, the neural coupling from right postCG, MTG and STG was positively correlated to the comprehension at the strong noise level. As the right postCG was responsible for sensorimotor processing and the right MTG/STG for auditory processing, it suggested that non-native listeners recruited a mixed and right-lateralized mechanism of auditory and sensorimotor processing to support speech-in-noise comprehension.

Moreover, the speaker-listener coupling pattern at the speaker's side can bring extra insights for explaining the specific linguistic processing inside the listener's brain. In specific, as speaker's neural activity is regarded as a standardized reference for listener, the brain region at the speaker's side could tell what type of linguistic information is processed by the listener in noisy conditions. In Li et al. 2021, the inter-brain neural coupling for native listeners covered a distributed and bilateral set of brain areas at the speaker's side, including the right postCG, left superior frontal gyrus, bilateral supramarginal gyrus, bilateral middle frontal gyrus, and bilateral AG, which might represent a unified language production network for the semantic-level linguistic generation. Meanwhile, in Li et al. 2022b>, for non-native listeners, the inter-brain coupling pattern at the speaker's side was restricted to the right postCG and STG, which were responsible for the generation of phonological-level linguistic information. Taken together, with the regard of the same group of speakers' neural activities during narrative speaking as a reference, the neural coupling at the speaker's side further highlighted that people relied on various linguistic information for native and non-native speech comprehension in noise.

Except for this, the temporal dynamic of the inter-brain neural coupling could help to further distinguish various processing modalities of listeners, such as the follow-up auditory encoding of the speech vs. the forward prediction of the upcoming information, during speech-in-noise comprehension. The temporal dynamic here means the neural activities of the speaker and the listener are not necessarily synchronized to each other, but coupled with a time lag. The neural coupling with the speaker's precedence to the listener might underlie the listener's delayed linguistic processing, while the neural coupling with the listener's precedence to the speaker underlies the listener's predictive coding of the speech or the speaker (Jiang et al. 2021). Although no study has been done in noisy

conditions yet, some studies in no-noise conditions have given supportive evidence for its potential. For example, Liu et al. 2020 found that the listener's neural activity lagged behind the speaker in order along the temporal progressing of speech processing. In specific, the listener's neural activities in the primary auditory cortex synchronized to the speaker's articulatory-related neural activity without delay, but lagged by 2 and 4 s in the STS/STG and MTG, respectively. This temporal sequence underly the bottom-up information flow from lower-level acoustic-processing areas to higher-level semantic processing areas. More generally, Stephens et al. 2010 showed that the listener's neural activities lagged behind the speaker's activities in most areas, but the striking listener's precedence to the speaker was observed in the striatum and anterior frontal areas. This listener's precedence might indicate an anticipatory neural response to predict the upcoming words of the speaker in a top-down fashion. What's more, the listener-preceded neural coupling was highly correlated to the listener's comprehension, suggesting that this prediction-based process was essential for speech comprehension and might play a more supportive role in noisy conditions. Following this logic, more futural efforts can also be paid to investigate how noise modulated these temporal dynamics of the speaker-listener neural coupling. It would deepen the understanding of how these follow-up and top-down linguistic processes differently activate and support speech-in-noise comprehension.

Inter-brain neural coupling underlies extralinguistic processing in noise

Except for the linguistic processing, the extralinguistic processing contributes to speech-in-noise comprehension as well. The neural alignment between the speaker and the listener could also take place at the extralinguistic level, such as emotion (Smirnov et al. 2019). Many studies have shown that the speaker-listener neural coupling (in no-noise condition) not only originated from the linguistic-related regions but also emerged from those extralinguistic-related regions, such as medial and dorsolateral prefrontal cortex, temporal-parietal junction (TPJ), precuneus, etc., during natural speech communication (for a review, Jiang et al. 2021). Also, these extralinguistic-related speaker-listener neural coupling was sensitive to interpersonal interaction, e.g., visual gaze (Jiang et al. 2012; Leong et al. 2017), interactive style (Pan et al. 2018; Zheng et al. 2018), etc., instead of (solely) the linguistic content. Thus, the speaker-listener neural coupling could also help to examine the extralinguistic processes during speech-in-noise comprehension, which are often underestimated by previous studies.

Dai et al. 2018 have adopted the inter-brain approach to reveal the functional role of extralinguistic processes in speech-in-noise comprehension. They recruited three-person groups, i.e., one listener and two speakers, to the laboratory. Two speakers were simultaneously speaking to the listener, while the listener was required to attend to one of them and ignore the other. Results showed that the listener's neural activity from the left TPJ was more coupled to the attended speaker than the unattended speaker, with the listener's neural activity preceding the attended speaker for several seconds. Moreover, the strength of the speaker-listener neural coupling from TPJ was positively correlated to their speech-in-noise communication. As the left TPJ was a critical region for mentalizing the other's mind or concept, these results might suggest that people selectively focused on the target speaker by predicting what the speaker intended to express. It was in favor of the previous hypothesis that the prediction promoted the selective focus and comprehension of the to-be-attended speech by gaining more weights for the processes of the relevant information (Schwartz et al. 2012).

Although previous studies have highlighted the importance of extralinguistic processing for speech comprehension in no-noise conditions (Jiang et al. 2021), there is no more inter-brain study examining its functional involvement in noisy conditions except for the one study above. Noteworthy, some researchers have suggested that the extralinguistic-related inter-brain neural coupling may serve as the neural base for successful speech comprehension and mutual understanding (Schoot et al. 2016). According to the mutual prediction theory, the integration of predicting others' actions and enacting one's own action by each individual led to the dynamic neural similarity among them, which formed the basis for successful reciprocal social interaction (Kingsbury et al. 2019), including speech interaction. In this line, the extralinguistic-related neural coupling serves an important purpose for the listener's interpretation of the speaker. A recent study showed that the speaker-listener neural coupling from the emotion-related regions, such as the middle frontal gyrus, superior parietal lobule, precuneus, amygdala, etc., modulated the emotional feelings shared between them: the more the listener's neural activities synchronized to the speaker, the more similar the listener's emotional feelings were to the speaker (Smirnov et al. 2019). Such an extralinguistic-related speaker-listener neural coupling might be more important with the existence of background noise, as the listener could refer to the overall representation of both themselves and the speaker (Sebanz et al. 2006; Yeshurun et al. 2021) to resolve the interference of noise. Therefore, more futural studies are needed to employ the inter-brain approach to examine the extralinguistic processing during speech-in-noise comprehension.

Discussion

As compared to the classical single-brain approach, the inter-brain approach has provided a novel methodology to investigate the linguistic and extralinguistic processes during speech-in-noise comprehension by analyzing the relationship between the speaker's and the listener's neural activities. It is suitable for naturalistic settings by promoting the use of naturalistic speech and even allowing for real-time communication between speaker and listener. Some recent studies have respectively validated its potential by highlighting the essential roles of linguistic (Li et al. 2021, 2022b) and extralinguistic processes (Dai et al. 2018) in speech-in-noise comprehension. However, the number of existing inter-brain studies on speech-in-noise comprehension is still quite limited. There remains much unknown and calls for more research interests in the future.

Firstly, how the speaker-listener neural coupling varies with various types of background noise remains unclear. Previous behavioral evidence has suggested that meaningless noise (e.g., white noise), meaningless speech (e.g., speech in an unknown language) and competing meaningful speech interfered people's speech-in-noise comprehension in different ways (Oswald et al. 2000; Wong et al. 2012). In particular, while both meaningless noise and meaningful speech cause acoustic masking to the original speech, the latter often brings additional interference to the high-level linguistic and extralinguistic processes of speech (Scharenborg and van Os 2019). However, existing inter-brain studies only focused on how either white noise (Li et al. 2021, 2022b) or competing speech (Dai et al. 2018) affected the linguistic or extralinguistic processes, respectively. It remains to be elucidated how various types of noise differently modulate the speaker-listener neural coupling from the speech-related regions, i.e., auditory, sensorimotor and extralinguistic-related regions. Future inter-brain studies with a direct comparison of different types of noises are needed to clarify this question, which are expected to bring more insights on the noise effect from an inter-brain perspective.

Another important issue is the causality between the speaker-listener neural coupling and speech comprehension. While large quantities of studies have revealed significant speaker-listener neural coupling during successful speech comprehension in no-noise (e.g., Stephens et al. 2010; Liu et al. 2020) and noisy (e.g., Dai et al. 2018; Li et al. 2021) conditions, it remains controversial whether it causally determines the comprehensive outcome of speech processing, or is just an epiphenomenal consequence for sharing the same environment or performing the same task (Hamilton 2021; Novembre and Iannetti 2021). In order to resolve this controversy, causal protocols, such as multi-brain stimulation (MBS), would give a solution (Novembre and Iannetti 2021). MBS refers to the simultaneous stimulation of multiple brains engaged in social interaction (Novembre and Iannetti 2021; Pan et al. 2021). The investigation of whether the direct manipulation of the speaker-listener neural coupling influences speech comprehension would clarify its causal role. Moreover, if the causality were proved, MBS could be further used to help people to listen better, especially for those populations with difficulty in speech-in-noise comprehension, such as the elders (Panouilleres and Mottonen 2018) or people with hearing loss (Healy and Yoho 2016).

Besides, the relevant inter-brain studies on speech-in-noise comprehension are all based on fNIRS measurement (Dai et al. 2018; Li et al. 2021, 2022b). The fNIRS was often chosen for its high tolerance to motion and the little operating noise (e.g., Li et al. 2021), making it broadly used in close-to-life (e.g., Dai et al. 2018) and even real-life communication scenarios (for a review, Kelsen et al. 2022). However, the fNIRS owns some disadvantages. For instance, both of its spatial and temporal resolutions are not high. The temporal resolution of fNIRS is around 1 s, and the spatial resolution is up to 1 cm (Dieler et al. 2012). To allow for a more precise description of the temporal dynamics or spatial localization of the speaker-listener neural coupling during speech-in-noise comprehension, other neuroimaging technologies with a higher spatial or temporal resolution, such as EEG, MEG, fMRI, ECoG, etc., could be further implemented.

Last but not least, as speech communication is a dynamic process with continuous mutual adaptation and coordination between the speaker and the listener (Hasson and Frith 2016), advanced mathematical and computational methods are necessary to further estimate how background noise influences the emergence, direction and dynamics of the speaker-listener neural coupling. For instance, by taking the communicators' brains together as an integrated neuronal network, computational neuronal models, such as the Rulkov map, could offer a promising tool to investigate the phenomenon of speech-in-noise comprehension by modeling how noise modulates the coherence and stochastic resonance over the network (Wang et al. 2008, 2009).

Conclusion

Comprehending speech with the interruption of background noise is of great importance for human life. In the past decades, a large number of psychological, cognitive and neuroscientific research has explored the neurocognitive mechanisms of speech-in-noise comprehension. However, as limited by the low ecological validity of the

speech stimuli and the experimental paradigm, as well as the inadequate attention on the high-order linguistic and extralinguistic processes, there remains much unknown about how people comprehend noisy speech in real-life scenarios. A recently emerging approach, i.e., the second-person or inter-brain neuroscience approach, provides a novel conceptual framework to address these issues by measuring the neural activities of both the speaker and the listener and calculating their inter-brain neural coupling from an integrative view. It promotes the use of naturalistic speech and allows for real-time communication between speaker and listener as in real-life scenarios. Several studies have validated its potential to investigate the linguistic and extralinguistic processes during speech-in-noise processing. More research interests in the inter-brain approach would further extend the present knowledge about the neural mechanism of speech-in-noise comprehension.

Funding This work was supported by the National Natural Science Foundation of China (NSFC) under grant (61977041), the Tsinghua University Spring Breeze Fund (2021Z99CFY037), and the National Natural Science Foundation of China (NSFC) and the German Research Foundation (DFG) in project Crossmodal Learning (NSFC 62061136001/DFG TRR-169/C1, B1).

Declarations

Conflict of interest The authors declare no competing interests.

References

Alain C, Du Y, Bernstein LJ et al (2018) Listening under difficult conditions: an activation likelihood estimation meta-analysis. Hum Brain Mapp 39(7):2695–2709. https://doi.org/10.1002/hbm.24031

Alexandrou AM, Saarinen T, Makela S et al (2017) The right hemisphere is highlighted in connected natural speech production and perception. NeuroImage 152:628–638. https://doi.org/10.1016/j.neuroimage.2017.03.006

Anderson S, Kraus N (2010) Sensory-cognitive interaction in the neural encoding of speech in noise: a review. J Am Acad Audiol 21(9):575–585. https://doi.org/10.3766/jaaa.21.9.3

Armeni K, Willems RM, Frank SL (2017) Probabilistic language models in cognitive neuroscience: promises and pitfalls. Neurosci Biobehav Rev 83:579–588. https://doi.org/10.1016/j.neubiorev.2017.09.001

Badal VD, Nebeker C, Shinkawa K et al (2021) Do words Matter? Detecting social isolation and loneliness in older adults using Natural Language Processing. Front Psychiatry 12:728732. https://doi.org/10.3389/fpsyt.2021.728732

Broderick MP, Anderson AJ, Di Liberto GM et al (2018) Electrophysiological Correlates of Semantic Dissimilarity reflect the comprehension of Natural, Narrative Speech. Curr Biol 28(5):803–809e803. https://doi.org/10.1016/j.cub.2018.01.080

Coffey EBJ, Mogilever NB, Zatorre RJ (2017) Speech-in-noise perception in musicians: a review. Hear Res 352:49–69. https://doi.org/10.1016/j.heares.2017.02.006

Crosse MJ, Di Liberto GM, Bednar A et al (2016) The multivariate temporal response function (mTRF) toolbox: a MATLAB Toolbox for relating neural signals to continuous stimuli. Front Hum Neurosci 10:604. https://doi.org/10.3389/fnhum.2016.00604

Czeszumski A, Eustergerling S, Lang A et al (2020) Hyperscanning: a valid method to study neural inter-brain underpinnings of Social Interaction. Front Hum Neurosci 14:39. https://doi.org/10.3389/fnhum.2020.00039

Dai B, Chen C, Long Y et al (2018) Neural mechanisms for selectively tuning in to the target speaker in a naturalistic noisy situation. Nat Commun 9(1):2405. https://doi.org/10.1038/s41467-018-04819-z

de Heer WA, Huth AG, Griffiths TL et al (2017) The hierarchical cortical organization of human speech processing. J Neurosci 37(27):6539–6557. https://doi.org/10.1523/Jneurosci.3267-16.2017

Dieler AC, Tupak SV, Fallgatter AJ (2012) Functional near-infrared spectroscopy for the assessment of speech related tasks. Brain Lang 121(2):90–109. https://doi.org/10.1016/j.bandl.2011.03.005

Dikker S, Silbert LJ, Hasson U et al (2014) On the same wavelength: predictable language enhances speaker-listener brain-to-brain synchrony in posterior superior temporal gyrus. J Neurosci 34(18):6267–6272. https://doi.org/10.1523/JNEUROSCI.3796-13.2014

Ding N, Simon JZ (2012) Emergence of neural encoding of auditory objects while listening to competing speakers. Proc Natl Acad Sci U S A 109(29):11854–11859. https://doi.org/10.1073/pnas.1205381109

Ding N, Simon JZ (2013) Adaptive temporal encoding leads to a background-insensitive cortical representation of speech. J Neurosci 33(13):5728–5735. https://doi.org/10.1523/JNEUROSCI.5297-12.2013

Dryden A, Allen HA, Henshaw H et al (2017) The association between cognitive performance and speech-in-noise perception for adult listeners: a systematic literature review and meta-analysis. Trends Hear. https://doi.org/10.1177/2331216517744675

Du Y, Buchsbaum BR, Grady CL et al (2014) Noise differentially impacts phoneme representations in the auditory and speech motor systems. Proc Natl Acad Sci U S A 111(19):7126–7131. https://doi.org/10.1073/pnas.1318738111

Du Y, Buchsbaum BR, Grady CL et al (2016) Increased activity in frontal motor cortex compensates impaired speech perception in older adults. Nat Commun 7:12241. https://doi.org/10.1038/ncomms12241

Du Y, Zatorre RJ (2017) Musical training sharpens and bonds ears and tongue to hear speech better. Proc Natl Acad Sci U S A 114(51):13579–13584. https://doi.org/10.1073/pnas.1712223114

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(29):5750–5759. https://doi.org/10.1523/JNEUROSCI.1828-18.2019

Fedorenko E, Blank IA (2020) Broca's area is not a Natural Kind. Trends Cogn Sci 24(4):270–284. https://doi.org/10.1016/j.tics.2020.01.001

Friederici AD (2012) The cortical language circuit: from auditory perception to sentence comprehension. Trends Cogn Sci 16(5):262–268. https://doi.org/10.1016/j.tics.2012.04.001

Garrod S, Pickering MJ (2004) Why is conversation so easy? Trends Cogn Sci 8(1):8–11. https://doi.org/10.1016/j.tics.2003.10.016

Golestani N, Hervais-Adelman A, Obleser J et al (2013) Semantic versus perceptual interactions in neural processing of speech-in-noise. NeuroImage 79:52–61. https://doi.org/10.1016/j.neuroimage.2013.04.049

Golumbic EMZ, Ding N, Bickel S et al (2013) Mechanisms underlying selective neuronal Tracking of attended Speech at a "Cocktail Party". Neuron 77(5):980–991. https://doi.org/10.1016/j.neuron.2012.12.037

Grand G, Blank IA, Pereira F et al (2022) Semantic projection recovers rich human knowledge of multiple object features from word embeddings. Nat Hum Behav. https://doi.org/10.1038/s41562-022-01316-8

Guediche S, Blumstein SE, Fiez JA et al (2014) Speech perception under adverse conditions: insights from behavioral, computational, and neuroscience research. Front Syst Neurosci 7:126. https://doi.org/10.3389/fnsys.2013.00126

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

Hamilton LS, Huth AG (2020) The revolution will not be controlled: natural stimuli in speech neuroscience. Lang Cogn Neurosci 35(5):573–582. https://doi.org/10.1080/23273798.2018.1499946

Hamilton AFC (2021) Hyperscanning: beyond the hype. Neuron 109(3):404–407. https://doi.org/10.1016/j.neuron.2020.11.008

Hanulikova A (2021) Do faces speak volumes? Social expectations in speech comprehension and evaluation across three age groups. PLoS ONE 16(10):e0259230. https://doi.org/10.1371/journal.pone.0259230

Hasson U, Ghazanfar AA, Galantucci B et al (2012) Brain-to-brain coupling: a mechanism for creating and sharing a social world. Trends Cogn Sci 16(2):114–121. https://doi.org/10.1016/j.tics.2011.12.007

Hasson U, Frith CD (2016) Mirroring and beyond: coupled dynamics as a generalized framework for modelling social interactions. Philos Trans R Soc Lond B Biol Sci 371(1693). https://doi.org/10.1098/rstb.2015.0366

Hasson U, Egidi G, Marelli M et al (2018) Grounding the neurobiology of language in first principles: the necessity of non-language-centric explanations for language comprehension. Cognition 180:135–157. https://doi.org/10.1016/j.cognition.2018.06.018

Healy EW, Yoho SE (2016) Difficulty understanding speech in noise by the hearing impaired: underlying causes and technological solutions. In: Annual international conference IEEE engineering in medicine and biology society 2016, pp 89–92. https://doi.org/10.1109/EMBC.2016.7590647

Hennessy S, Mack WJ, Habibi A (2022) Speech-in-noise perception in musicians and non-musicians: a multi-level meta-analysis. Hear Res 416:108442. https://doi.org/10.1016/j.heares.2022.108442

Hernandez LM, Green SA, Lawrence KE et al (2020) Social attention in Autism: neural sensitivity to Speech over background noise predicts encoding of Social Information. Front Psychiatry 11:343. https://doi.org/10.3389/fpsyt.2020.00343

Herrmann B, Schlichting N, Obleser J (2014) Dynamic range adaptation to spectral stimulus statistics in human auditory cortex. J Neurosci 34(1):327–331. https://doi.org/10.1523/Jneurosci.3974-13.2014

Hickok G, Poeppel D (2007) The cortical organization of speech processing. Nat Rev Neurosci 8(5):393–402. https://doi.org/10.1038/nrn2113

Hickok G, Houde J, Rong F (2011) Sensorimotor integration in speech processing: computational basis and neural organization. Neuron 69(3):407–422. https://doi.org/10.1016/j.neuron.2011.01.019

Hitczenko K, Mazuka R, Elsner M et al (2020) When context is and isn't helpful: a corpus study of naturalistic speech. Psychon Bull Rev 27(4):640–676. https://doi.org/10.3758/s13423-019-01687-6

Holder JT, Levin LM, Gifford RH (2018) Speech Recognition in noise for adults with normal hearing: age-normative performance for AzBio, BKB-SIN, and QuickSIN. Otol Neurotol 39(10):e972–e978. https://doi.org/10.1097/MAO.0000000000002003

Holroyd CB (2022) Interbrain synchrony: on wavy ground. Trends Neurosci 45(5):346–357. https://doi.org/10.1016/j.tins.2022.02.002

Huth AG, de Heer WA, Griffiths TL et al (2016) Natural speech reveals the semantic maps that tile human cerebral cortex. Nature 532(7600):453–458. https://doi.org/10.1038/nature17637

Jaaskelainena IP, Sams M, Glerean E et al (2021) Movies and narratives as naturalistic stimuli in neuroimaging. NeuroImage 224:117445. https://doi.org/10.1016/j.neuroimage.2020.117445

Jiang J, Dai B, Peng D et al (2012) Neural synchronization during face-to-face communication. J Neurosci 32(45):16064–16069. https://doi.org/10.1523/JNEUROSCI.2926-12.2012

Jiang J, Zheng LF, Lu CM (2021) A hierarchical model for interpersonal verbal communication. Soc Cogn Affect Neurosci 16(1–2):246–255. https://doi.org/10.1093/scan/nsaa151

Kelsen BA, Sumich A, Kasabov N et al (2022) What has social neuroscience learned from hyperscanning studies of spoken communication? A systematic review. Neurosci Biobehav Rev 132:1249–1262. https://doi.org/10.1016/j.neubiorev.2020.09.008

Kingsbury L, Huang S, Wang J et al (2019) Correlated neural activity and encoding of Behavior across brains of socially interacting animals. Cell 178(2):429–446e416. https://doi.org/10.1016/j.cell.2019.05.022

Kingsbury L, Hong WZ (2020) A multi-brain framework for social interaction. Trends Neurosci 43(9):651–666. https://doi.org/10.1016/j.tins.2020.06.008

Kuhlen AK, Allefeld C, Haynes JD (2012) Content-specific coordination of listeners' to speakers' EEG during communication. Front Hum Neurosci 6:266. https://doi.org/10.3389/fnhum.2012.00266

Kutlu E, Tiv M, Wulff S et al (2022) Does race impact speech perception? An account of accented speech in two different multilingual locales. Cogn Res Princ Implic 7(1):7. https://doi.org/10.1186/s41235-022-00354-0

Leong V, Byrne E, Clackson K et al (2017) Speaker gaze increases information coupling between infant and adult brains. Proc Natl Acad Sci U S A 114(50):13290–13295. https://doi.org/10.1073/pnas.1702493114

Li ZR, Li JW, Hong B et al (2021) Speaker-Listener neural coupling reveals an adaptive mechanism for Speech Comprehension in a noisy environment. Cereb Cortex 31(10):4719–4729. https://doi.org/10.1093/cercor/bhab118

Li JW, Hong B, Nolte G et al (2022a) Preparatory delta phase response is correlated with naturalistic speech comprehension performance. Cogn Neurodyn 16(2):337–352. https://doi.org/10.1007/s11571-021-09711-z

Li ZR, Hong B, Wang D et al (2022b) Speaker-listener neural coupling reveals a right-lateralized mechanism for non-native speech-in-noise comprehension. Cereb Cortex. https://doi.org/10.1093/cercor/bhac302

Liberman AM, Cooper FS, Shankweiler DP et al (1967) Perception of the speech code. Psychol Rev 74(6):431–461. https://doi.org/10.1037/h0020279

Liu L, Zhang Y, Zhou Q et al (2020) Auditory-articulatory neural alignment between Listener and Speaker during Verbal Communication. Cereb Cortex 30(3):942–951. https://doi.org/10.1093/cercor/bhz138

Marrufo-Perez MI, Sturla-Carreto DDP, Eustaquio-Martin A et al (2020) Adaptation to noise in Human Speech Recognition

depends on noise-level statistics and fast dynamic-range Compression. J Neurosci 40(34):6613–6623. https://doi.org/10.1523/JNEUROSCI.0469-20.2020
McGowan KB (2015) Social expectation improves speech perception in noise. Lang Speech 58(Pt 4):502–521. https://doi.org/10.1177/0023830914565191
Mesgarani N, Chang EF (2012) Selective cortical representation of attended speaker in multi-talker speech perception. Nature 485(7397):233–236. https://doi.org/10.1038/nature11020
Montague PR, Berns GS, Cohen JD et al (2002) Hyperscanning: simultaneous fMRI during linked social interactions. NeuroImage 16(4):1159–1164. https://doi.org/10.1006/nimg.2002.1150
Nelson MJ, El Karoui I, Giber K et al (2017) Neurophysiological dynamics of phrase-structure building during sentence processing. Proc Natl Acad Sci U S A 114(18):E3669–E3678. https://doi.org/10.1073/pnas.1701590114
Novembre G, Iannetti GD (2021) Hyperscanning alone cannot prove causality. Multibrain Stimulation can. Trends Cogn Sci 25(2):96–99. https://doi.org/10.1016/j.tics.2020.11.003
Oswald CJ, Tremblay S, Jones DM (2000) Disruption of comprehension by the meaning of irrelevant sound. Memory 8(5):345–350. https://doi.org/10.1080/09658210050117762
Pan Y, Novembre G, Song B et al (2018) Interpersonal synchronization of inferior frontal cortices tracks social interactive learning of a song. NeuroImage 183:280–290. https://doi.org/10.1016/j.neuroimage.2018.08.005
Pan Y, Novembre G, Song B et al (2021) Dual brain stimulation enhances interpersonal learning through spontaneous movement synchrony. Soc Cogn Affect Neurosci 16(1–2):210–221. https://doi.org/10.1093/scan/nsaa080
Panouilleres MTN, Mottonen R (2018) Decline of auditory-motor speech processing in older adults with hearing loss. Neurobiol Aging 72:89–97. https://doi.org/10.1016/j.neurobiolaging.2018.07.013
Pickering MJ, Garrod S (2013) An integrated theory of language production and comprehension. Behav Brain Sci 36(4):329–347. https://doi.org/10.1017/S0140525X12001495
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(2):816–847. https://doi.org/10.1016/j.neuroimage.2012.04.062
Pulvermuller F, Fadiga L (2010) Active perception: sensorimotor circuits as a cortical basis for language. Nat Rev Neurosci 11(5):351–360. https://doi.org/10.1038/nrn2811
Redcay E, Moraczewski D (2020) Social cognition in context: a naturalistic imaging approach. NeuroImage 216:116392. https://doi.org/10.1016/j.neuroimage.2019.116392
Redcay E, Schilbach L (2019) Using second-person neuroscience to elucidate the mechanisms of social interaction. Nat Rev Neurosci 20(8):495–505. https://doi.org/10.1038/s41583-019-0179-4
Rysop AU, Schmitt LM, Obleser J et al (2021) Neural modelling of the semantic predictability gain under challenging listening conditions. Hum Brain Mapp 42(1):110–127
Scharenborg O, van Os M (2019) Why listening in background noise is harder in a non-native language than in a native language: a review. Speech Commun 108:53–64. https://doi.org/10.1016/j.specom.2019.03.001
Schilbach L, Timmermans B, Reddy V et al (2013) Toward a second-person neuroscience. Behav Brain Sci 36(4):393–414. https://doi.org/10.1017/S0140525x12000660
Schirmer A, Kotz SA (2006) Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. Trends Cogn Sci 10(1):24–30. https://doi.org/10.1016/j.tics.2005.11.009
Schomers MR, Pulvermuller F (2016) Is the sensorimotor cortex relevant for speech perception and understanding? An integrative review. Front Hum Neurosci 10:435. https://doi.org/10.3389/fnhum.2016.00435
Schoot L, Hagoort P, Segaert K (2016) What can we learn from a two-brain approach to verbal interaction? Neurosci Biobehav Rev 68:454–459. https://doi.org/10.1016/j.neubiorev.2016.06.009
Schwartz JL, Basirat A, Menard L et al (2012) The perception-for-action-control theory (PACT): a perceptuo-motor theory of speech perception. J Neurolinguist 25(5):336–354
Sebanz N, Bekkering H, Knoblich G (2006) Joint action: bodies and minds moving together. Trends Cogn Sci 10(2):70–76. https://doi.org/10.1016/j.tics.2005.12.009
Sehm B, Schnitzler T, Obleser J et al (2013) Facilitation of Inferior Frontal Cortex by Transcranial Direct Current Stimulation induces perceptual learning of severely degraded Speech. J Neurosci 33(40):15868–15878. https://doi.org/10.1523/Jneurosci.5466-12.2013
Shi LF, Koenig LL (2016) Relative weighting of semantic and syntactic cues in native and non-native listeners' recognition of english sentences. Ear Hear 37(4):424–433
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(46):12303–12308. https://doi.org/10.1073/pnas.1710752114
Silbert LJ, Honey CJ, Simony E et al (2014) Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. Proc Natl Acad Sci U S A 111(43):E4687–4696. https://doi.org/10.1073/pnas.1323812111
Smirnov D, Saarimaki H, Glerean E et al (2019) Emotions amplify speaker-listener neural alignment. Hum Brain Mapp 40(16):4777–4788. https://doi.org/10.1002/hbm.24736
Sonkusare S, Breakspear M, Guo C (2019) Naturalistic stimuli in neuroscience: critically acclaimed. Trends Cogn Sci 23(8):699–714. https://doi.org/10.1016/j.tics.2019.05.004
Stephens GJ, Silbert LJ, Hasson U (2010) Speaker-listener neural coupling underlies successful communication. Proc Natl Acad Sci U S A 107(32):14425–14430. https://doi.org/10.1073/pnas.1008662107
Tanana MJ, Soma CS, Kuo PB et al (2021) How do you feel? Using natural language processing to automatically rate emotion in psychotherapy. Behav Res Methods 53(5):2069–2082. https://doi.org/10.3758/s13428-020-01531-z
Vander Ghinst M, Bourguignon M, Niesen M et al (2019) Cortical tracking of speech-in-noise develops from childhood to adulthood. J Neurosci 39(15):2938–2950. https://doi.org/10.1523/Jneurosci.1732-18.2019
Wang Q, Duan Z, Perc M et al (2008) Synchronization transitions on small-world neuronal networks: effects of information transmission delay and rewiring probability. Europhys Lett 83(5):50008. https://doi.org/10.1209/0295-5075/83/50008
Wang Q, Perc M, Duan Z et al (2009) Delay-induced multiple stochastic resonances on scale-free neuronal networks. Chaos 19(2):023112. https://doi.org/10.1063/1.3133126
Weed E, Fusaroli R (2020) Acoustic measures of Prosody in Right-Hemisphere damage: a systematic review and Meta-analysis. J Speech Lang Hear Res 63(6):1762–1775. https://doi.org/10.1044/2020_Jslhr-19-00241
Wilson RH, McArdle RA, Smith SL (2007) An evaluation of the BKB-SIN, HINT, QuickSIN, and WIN materials on listeners with normal hearing and listeners with hearing loss. J Speech Lang Hear Res 50(4):844–856. https://doi.org/10.1044/1092-4388(2007/059)
Wilson RH, Trivette CP, Williams DA et al (2012) The effects of energetic and informational masking on the words-in-noise test (WIN). J Am Acad Audiol 23(7):522–533. https://doi.org/10.3766/jaaa.23.7.4

Wong LL, Ng EH, Soli SD (2012) Characterization of speech understanding in various types of noise. J Acoust Soc Am 132(4):2642–2651. https://doi.org/10.1121/1.4751538

Yeshurun Y, Nguyen M, Hasson U (2021) The default mode network: where the idiosyncratic self meets the shared social world. Nat Rev Neurosci 22(3):181–192. https://doi.org/10.1038/s41583-020-00420-w

Zheng L, Chen C, Liu W et al (2018) Enhancement of teaching outcome through neural prediction of the students' knowledge state. Hum Brain Mapp 39(7):3046–3057. https://doi.org/10.1002/hbm.24059

Yu ACL (2022) Perceptual cue weighting is influenced by the Listener's gender and subjective evaluations of the speaker: the case of English Stop Voicing. Front Psychol 13:840291. https://doi.org/10.3389/fpsyg.2022.840291

Yuan JJ (2020) Cognitive neuroscience of emotional susceptibility (in Chinese). Science Press, Beijing

Zekveld AA, Rudner M, Johnsrude IS et al (2011) The influence of semantically related and unrelated text cues on the intelligibility of sentences in noise. Ear Hear 32(6):E16–E25