Neurosci Bull June 1, 2014, 30(3): 490–496. http://www.neurosci.cn
DOI: 10.1007/s12264-013-1428-6

·Review·

Interaction between auditory and motor systems in speech perception

Zhe-Meng Wu, Ming-Li Chen, Xi-Hong Wu, Liang Li
Department of Psychology, Speech and Hearing Research Center, Key Laboratory of Machine Perception (Ministry of Education), PKU-IDG/McGovern Institute for Brain Research, Peking University, Beijing 100871, China
Corresponding author: Liang Li. E-mail: liangli@pku.edu.cn

Based on the Motor Theory of speech perception, the interaction between the auditory and motor systems plays an essential role in speech perception. Since the Motor Theory was proposed, it has received remarkable attention in the field. However, each of the three hypotheses of the theory still needs further verification. In this review, we focus on how the auditory-motor anatomical and functional associations play a role in speech perception and discuss why previous studies could not reach an agreement and particularly whether the motor system involvement in speech perception is task-load dependent. Finally, we suggest that the function of the auditory-motor link is particularly useful for speech perception under adverse listening conditions and the further revised Motor Theory is a potential solution to the "cocktail-party" problem.

Keywords: auditory-motor interaction; Motor Theory of speech perception; motor cortex; "cocktail-party" problem.

Introduction

How listeners process the acoustic signals of speech is a hot question. Traditionally, studies of this question have mainly focused on the functions of the auditory system. However, speech processing is not the pure and simple analysis of speech sound signals, but a quite complicated integrated process involving multisensory modalities and even the motor system. In this review, we focus on the interaction between the auditory system and the motor system in speech perception and emphasize that the motor processing component plays an essential role: activation of the perceptual-motor loop enables listeners to both track the speaker over time and form the intention to speak, especially under adverse listening conditions, such as a noisy and reverberating environment.

The Motor Theory of Speech Perception

The Motor Theory of speech perception was first proposed by Liberman and colleagues[1,2] after an unexpected failure in a reading-machine study. In their experiments, although blind people could recognize independent linguistic units, they could not perceive alphabetic sequences and understand synthesized speech, because the linguistic units they perceived tended to merge into a blur[3]. This problem related to speech perception is called coarticulation, in which speech acoustic signals are highly context-sensitive; a single phoneme can be influenced by its surrounding phonemes[2]. However, normal listeners are able to conquer coarticulation and perceive the original phonemes well[4]. Based on the results, Liberman and colleagues assumed that what we really perceive when hearing speech signals is not only sound waves, but also body "gestures" that reflect the speaker's intention. Liberman proposed three hypotheses in both a weak[1] and a strong version of the Motor Theory[5]. (1) The object of speech perception is the "gesture"; (2) speech processing is special and requires a specific phonetic module; and (3) activation of the motor cortex is involved in speech perception.

When Liberman advanced the Motor Theory, he asked

a critical question, “when articulation and sound wave go their separate ways, which way does perception go?”. The answer he provided was that perception goes with articulation[1]. In more detail, the theory suggests that the speech sound wave that we perceive carries the speaker’s information indirectly, and there is a direct way to transmit information, that is through the “gesture” bearing the speaker’s intention. In other words, perceiving the “gesture” is just perceiving the actual movement of the speaker’s vocal tract, including motion of the larynx, tongue, and lips[1]. Our understanding of this theory is that although the listener may not be aware of tracking movement of the vocal tract, he/she automatically uses this motor cue to recognize the speaker’s intention, just as in imitative behavior. When a speaker talks, the listener tries to follow his speech style in mind and make a prediction before the speaker says the next word. Thus, the speaker and the listener must converge on the same “linguistic currency”, the “gesture”, to communicate.

It is well known that “gesture” information can affect speech perception in different ways, such as the McGurk effect. When the listener sees the speaker producing the syllable (/ga/) while listening to another syllable (/ba/), the mouth movement may mislead the listener into hearing a different syllable (/da/)[6]. This visuomotor cue strongly influences what we actually hear. Also, another study focused on the role of articulatory organ movement in a noisy environment[7]. Listeners perceive speech more accurately when they can see the speaker’s articulatory organ movement than when they cannot. Also, under adverse listening conditions, lip-reading associated with the target sentence can act as a cue to improve listener recognition of speech[8,9]. Thus, perceiving “gesture” signals provides visuomotor cues, which help listeners take advantage of the speaker’s motor actions during speech in an adverse noisy environment and facilitate the perceptual performance. In other words, listeners actively, rather than passively, receive speech information. Supporting this view, Alho *et al.* reported that stronger activation in the left premotor cortex is associated with better identification of syllables that are embedded in noise, and the cortical activation is quite different between active and passive listening[10]. Also, Calla *et al.* reported stronger activity in correct trials over incorrect trials within both the ventral premotor cortex and Broca’s area[11]. However, it is still not clear whether the enhanced activity of the cortical areas is specific to speech perceptual performance or just reflects an increase in general processing load.

Anatomical and Functional Associations between the Auditory and Motor Systems

To confirm the involvement of the motor system in speech perception, evidence of both anatomical and functional links between the motor and auditory systems is needed. Indeed, some models emphasize the auditory-motor link in speech perception. For example, the dual-stream processing model suggests that there are two pathways in audition: one is the ventral pathway down to the temporal lobe regulating “what” in acoustic information, and the other is the dorsal pathway from primary sensory areas up to the posterior cortex regulating “how” speech production takes place[12,13]. It is also known that the ventral pathway is involved in analyzing phonetic characters, acoustic features, and speech intelligibility[14,15], and the dorsal pathway is associated with sensorimotor mapping between auditory and motor representations[16,17], speech production[14,18], and silent articulatory organ movement[19]. Although this dual-stream model proposes that each of the pathways plays a specific role in speech perception, how the streams interact with each other is still not clear.

The other model, the forward-inverse model, proposes that the motor cortical regions predict the consequences of motor commands and revise the signals with the changing environment[20,21]. In more detail, before motor commands reach the effectors, the forward-inverse model produces predicted sensory consequences of the motor commands, and then compares the predicted results with the real sensory information. This comparison provides more information for the central system to produce a more appropriate performance. With time delays and interruptions from the surroundings, the motor commands need to be up-dated from time to time in order to produce the desired outcome. Thus, when a speech signal is distorted by environmental noise and/or time delays, the motor representation of the previous signals modifies the current auditory representation through inverse mapping. Due to the role of motor representation in revising distorted signals, listeners can recognize the speaker’s intention and predict the outcome of motor commands before making

responses. In other words, anticipation of a motor signal can be combined with both signal characteristics and the speaker's articulatory information, producing a more desirable response. Based on the basic principles of the dual-stream processing and forward-inverse models, we propose that further-developed models should emphasize how the auditory-motor interaction is modulated by both processing load (due to complex inputs) and prediction/estimation (due to task goals and feedback) (Fig. 1).

So far, co-activations between auditory and motor regions in speech perception have been clearly demonstrated. When exposed to novel speech distortions, such as time-compressed sentences, listeners can rapidly distinguish distorted sentences from normal-speed sentences, with increased activation associations between the auditory cortices and the left ventral premotor cortex[22]. Moreover, compared to listening to pseudo-words and reversed-words, listening to normal words induces broad activation connectivity in the auditory-motor network, which may be useful for facilitating semantic processing[23]. Further investigation is needed to verify whether this enhanced dynamic auditory-motor network promotes the transition from a sound stream into a series of meaningful motor-based units and results in speech comprehension.

In addition to the well-known fact that speech production is tightly related to the motor cortex, some studies have shown that the motor cortex is activated in speech perception tasks[24-30]. For example, when listeners

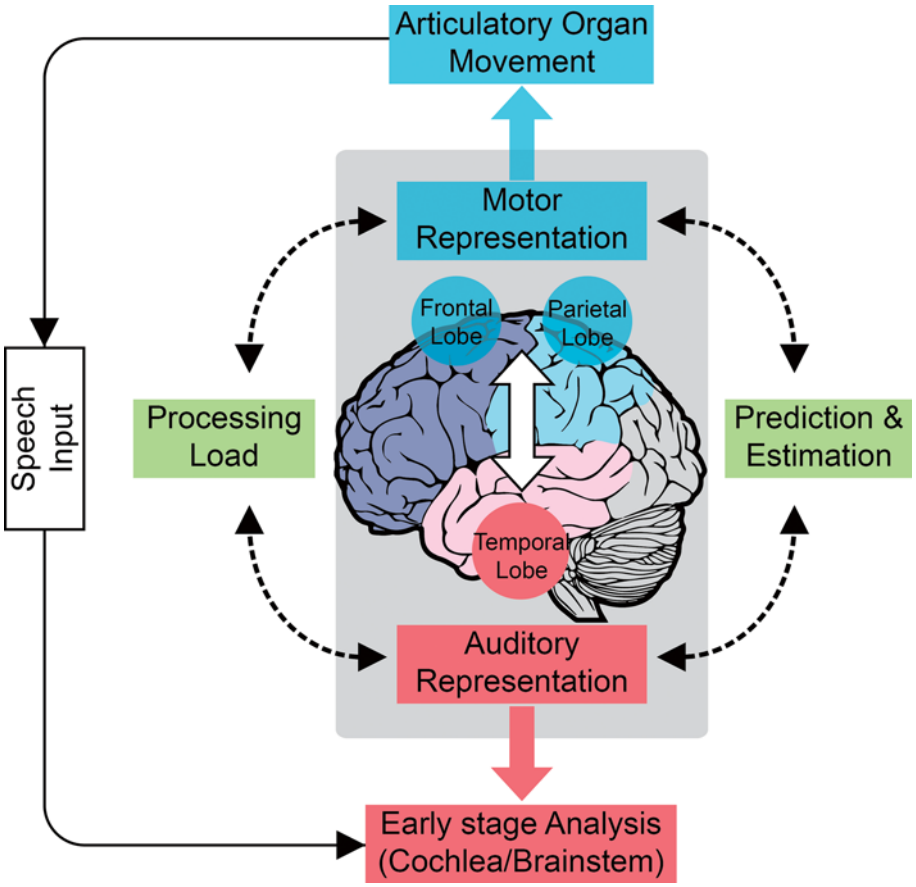

Fig. 1. Modified Auditory-Motor Interaction model based on the Dual-stream Processing model in combination with the Forward-inverse model. This model emphasizes that both processing load and prediction/estimation affect the auditory-motor interaction (white two-headed arrow). Red: the auditory processing system including the ventral stream pathway; blue: the motor processing system including the dorsal stream pathway; green: the systems that mediate the processing load, prediction, and estimation that modulate the auditory-motor interaction (arrows with dashed lines).

hear a lip-related phoneme [p] or a tongue-related phoneme [t], motor regions are activated differentially[26,27], suggesting that different speech stimuli activate motor cortical regions with different patterns. In other words, listening to various verbal stimuli may cause differential automatic activations of cortical regions involved in speech production. Also, it has been suggested that the activation of the motor cortex may reflect a mediating role of the motor cortex in speech perception[28-30].

Moreover, studies using either functional magnetic resonance imaging (fMRI) or transcranial magnetic stimulation (TMS) have confirmed the role of the motor cortex in speech perception. For example, in a speech perception task, strong activation in the motor cortex can be induced only when participants perceive the target speech[24,31]. Some studies using TMS of the motor cortex have demonstrated that stimulation of speech-related regions affects speech perception[32-34]. For example, using TMS to suppress the left premotor cortex, which is activated both during speech production and speech perception (Fig. 2), Meister *et al.* found that participants with a suppressed premotor cortex were impaired in discriminating voiceless stop consonants under white-noise masking conditions. Thus, they suggested that the premotor cortex is essentially involved in speech perception[33].

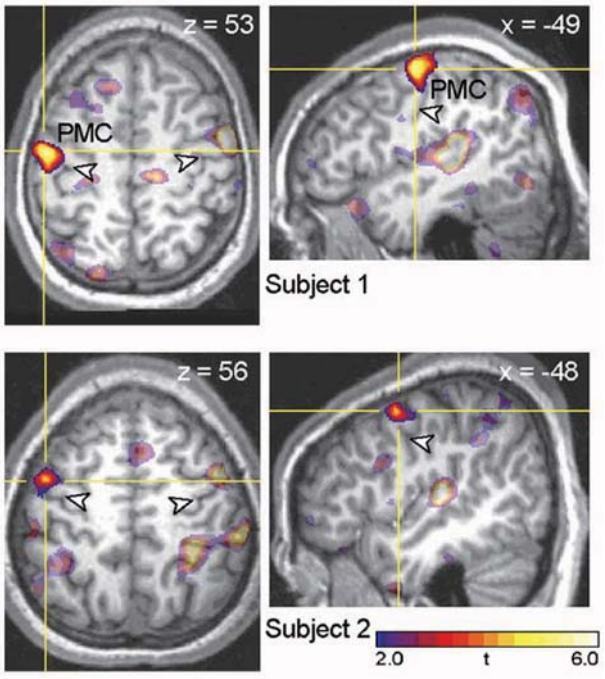

Fig. 2. Representative fMRI activation in the premotor cortex (PMC) associated with discriminating voiceless stop consonants in single syllables masked by white noise in two representative participants. Regions selected for stimulation are shown in bright colors. Arrowheads indicate the location of the central sulcus (adapted from Meister *et al.*[33] with permission).

However, the results of some clinical studies appear not to support the view that there is an association between impairment of speech perception and impairment of speech production. For example, patients with expressive aphasia exhibit impairment of speech production but not speech perception or comprehension[35]. Also, although patients with receptive aphasia exhibit impairment of speech perception and comprehension, they can speak fluently[36,37]. The dissociations in expressive and receptive aphasia support another view that speech perception and production are two distinct processes. Moreover, patients with lesions in Broca's area perform well in both word-comprehension and syllable-identification tests, but patients with temporal lobule damage perform poorly in these tests[38]. These studies also negate the role of motor regions in speech perception, but support the view that temporal regions rather than motor areas are important in speech perception[16,17].

Research in child development has also shown dissociations between speech perception and speech production. Children born with hearing loss can learn to speak if they gain enough positive somatosensory feedback, even though the learning process is much harder than in healthy children[39]. Also, children with severe dysarthria are unable to produce meaningful sentences, but they can accurately understand spoken content[40]. Furthermore, infants usually learn to understand speech first and then begin to learn how to produce their own words[41,42]. These studies indicate that dissociations between speech perception and speech production occur during development. As speech perception and speech production do not appear at the same time during development, motor cortical areas may not be as important for speech perception as the Motor Theory proposes.

As reviewed above, some fMRI and TMS studies support the view that motor cortical areas are important in speech perception, while clinical and developmental studies have shown significant dissociations between

speech perception and speech production. Also note that some fMRI studies did not reveal an increased activation in motor cortical regions during speech perception and comprehension[19,43,44].

The different results in the studies described above may be due to different task demands. The Hickok and Poeppel study showed that when the task load is high, requiring both speech identification and speech categorization, the activated frontal region extends to the premotor cortex[17]. Also, when speech signals are distorted, the motor cortex is markedly activated[45,46]. Interestingly, non-verbal signals can activate motor areas[31,47], and there is no difference in activation magnitude in the motor cortex between perceiving speech and perceiving non-verbal sounds[31]. Some studies have further shown that blurred speech causes even stronger activation in the bilateral premotor cortex, compared to clear speech[45] and perceiving a foreign language causes larger activation in the motor cortex than the native language[25]. Also, low-frequency words induce higher activation in the motor cortex than high-frequency words[48]. Thus, facing unfamiliar stimuli (such as distorted speech, a foreign language, or low-frequency words), the motor cortex may play a role in facilitating the association with the auditory system to improve speech perception. More interestingly, in a mixed visual and auditory task, weaker visual stimuli evoke stronger activation in the motor cortex than clear pictures of speakers[49], suggesting that a heavy-load task, such as analyzing distorted auditory or visual signals, requires involvement of the motor cortex. These studies suggest that whether the motor system is involved in speech perception is task-load dependent. In future, this assumption will be tested to confirm whether the dissociations between speech perception and speech production under either clinical or developmental conditions are task-load related.

Speech Perception under "Cocktail Party" Conditions

Speech perception is not just for hearing speech sounds, but more essentially, for recognizing and understanding speech signals, requiring that multisensory modalities interact. In fact, speech understanding and speech hearing do not share the same brain network, including the motor areas[40].

In a noisy environment (like a cocktail party), although there are many acoustic sources from various directions, listeners are still able to identify and follow target speech sounds in this high perceptual-load situation. How can listeners separate various speakers' signals and understand target sentences? Although this "cocktail party" problem advanced by Cherry[50] has not been fully solved, several lines of evidence suggest that the motor system plays a role in solving this problem when the perceptual load is high.

First, observing a speaker's articulator movements can induce better understanding of speech. Listeners perceive speech in noise-masking or speech-masking environments more accurately when they can see speaker's articulatory organ movement than when they cannot[7,9]. In addition, signals from the motor system help a listener to track a speaker talking over time[51,52]. It has been suggested that one of the functions of motor activation is tracking the talker's speed and rhythm over time, and provides the timing signals to the auditory cortex. Particularly in a conversation, the monitoring role of the motor system in interacting with the auditory system over time can induce fluent conversation[53].

Under "cocktail-party" conditions, listeners are able to take advantage of certain perceptual cues to facilitate their selective attention to target speech. Selective attention allocates more cognitive resources to the motor representation of speech so that a listener can capture a speaker's intention and improve speech recognition. Under noise-masking conditions, selective attention affects both active and passive listening. As Alho *et al.* have reported, attention modulates the magnitude of activation of the left premotor cortex, which influences the performance of phonetic categorization[10].

In patients with schizophrenia, both speech-perception deficits and increased vulnerability to masking stimuli generally occur. More specifically, speech recognition in both first-episode and chronic patients with schizophrenia is more vulnerable to masking stimuli, particularly speech-masking stimuli, than in healthy people[54]. Thus, whether functional impairments of motor cortical regions contribute to the enhanced vulnerability to speech-masking stimuli in patients with schizophrenia will be an important research issue in the future.

Conclusion

This review summarizes the studies showing that interactions between the auditory system and the motor system are related to speech perception. The anatomical and functional connections between the auditory and motor systems are important for improving speech recognition, particularly under difficult listening conditions (such as the cocktail-party environment). With the involvement of the motor system, the listener can better identify the speaker's intention and follow the target stream. Thus, investigation of the auditory-motor association in speech perception is important for solving the "cocktail party" problem.

ACKNOWLEDGEMENTS

This review was supported by the National Basic Research Development Program of China (2009CB320901, 2011CB707805, 2013CB329304), the National Natural Science Foundation of China (31170985, 91120001, 61121002), and "985" project grants from Peking University.

Received date: 2013-05-06; Accepted date: 2013-08-20

REFERENCES

[1] Liberman AM, Delattre P, Cooper FS. The role of selected stimulus-variables in the perception of the unvoiced stop consonants. Am J Psychol 1952, 65: 497–516.
[2] Liberman AM, Cooper FS, Shankweiler DP, Studdert KM. Perception of the speech code. Psychol Rev 1967, 74: 431–461.
[3] Liberman AM. Speech: A Special Code. Cambridge, MA: The MIT Press 1996.
[4] Kent RD, Minifie FD. Coarticulation in recent speech production models. J Phon 1977, 5: 115–133.
[5] Liberman AM, Mattingly IG. The motor theory of speech perception revised. Cognition 1985, 21: 1–36.
[6] McGurk H, MacDonald J. Hearing lips and seeing voices. Nature 1976, 264: 746–748.
[7] Sumby WH, Pollack I. Visual contribution to speech intelligibility in noise. J Acoust Soc Am 1954, 26: 212.
[8] Rudmann DS, McCarley JS, Kramer AF. Bimodal displays improve speech comprehension in environments with multiple speakers. Hum Fac Erg Soc P 2003, 45: 329–336.
[9] Wu C, Cao S, Wu X, Li L. Temporally pre-presented lipreading cues release speech from informational masking. J Acoust Soc Am 2013, 133: 281–285.
[10] Alho J, Sato M, Sams M, Schwartz JL, Tiitinen H, Jääskeläinen IP. Enhanced early-latency electromagnetic activity in the left premotor cortex is associated with successful phonetic categorization. Neuroimage 2012, 60: 1937–1946.
[11] Callan D, Callan A, Gamez M, Sato MA, Kawato M. Premotor cortex mediates perceptual performance. Neuroimage 2010, 51: 844–858.
[12] Schreiner CE, Winer JA. Auditory cortex mapmaking: principles, projections, and plasticity. Neuron 2007, 56: 356–365.
[13] Recanzone GH, Sutter ML. The biological basis of audition. Annu Rev Psychol 2008, 59: 119–142.
[14] Belin P, Zatorre RJ. Adaptation to speaker's voice in right anterior temporal lobe. Neuroreport 2003, 14: 2105–2109.
[15] Scott SK, Blank CC, Rosen S, Wise RJ. Identification of a pathway for intelligible speech in the left temporal lobe. Brain 2000, 123: 2400–2406.
[16] Hickok G, Poeppel D. Towards a functional neuroanatomy of speech perception. Trends Cogn Sci 2000, 4: 131–138.
[17] Hickok G, Poeppel D. The cortical organization of speech processing. Nat Neurosci 2007, 8: 393–402.
[18] Baddeley A, Lewis V, Vallar G. Exploring the articulatory loop. Q J Exp Psychol 1984, 36: 233–252.
[19] Wise RJ, Scott SK, Blank SC, Mummery CJ, Murphy K, Warburton EA. Separate neural subsystems within Wernicke's area. Brain 2001, 124: 83–95.
[20] Andersen RA, Buneo CA. Intentional maps in posterior parietal cortex. Annu Rev Neurosci 2002, 25: 189–220.
[21] Wolpert DM, Doya K, Kawato M. A unifying computational framework for motor control and social interaction. Philos Trans R Soc Lond B Biol Sci 2003, 358: 593–602.
[22] Adank P, Devlin JT. On-line plasticity in spoken sentence comprehension: Adapting to time-compressed speech. Neuroimage 2010, 49: 1124.
[23] Londei A, D'Ausilio A, Basso D, Sestieri C, Gratta CD, Romani GL, *et al.* Sensory-motor brain network connectivity for speech comprehension. Hum Brain Mapp 2010, 31: 567–580.
[24] Wilson SM. Listening to speech activates motor areas involved in speech production. Nat Neurosci 2004, 7: 701–702.
[25] Wilson SM, Iacoboni M. Neural responses to non-native phonemes varying in producibility: evidence for the sensorimotor nature of speech perception. Neuroimage 2006, 33: 316–325.
[26] Fadiga L, Craighero L, Buccino G, Rizzolatti G. Speech listening specifically modulates the excitability of tongue muscles: a TMS study. Eur J Neurosci 2002, 15: 399–402.
[27] Pulvermüller F, Huss M, Kherif F, del Prado Martin FM, Hauk

O, Shtyrov Y. Motor cortex maps articulatory features of speech sounds. Proc Natl Acad Sci U S A 2006, 103: 7865–7870.

[28] Bever TG, Poeppel D. Analysis by synthesis: a (re-) emerging program of research for language and vision. Biolinguistics 2010, 4: 174–200.

[29] Callan DE, Jones JA, Callan AM, Akahane-Yamada R. Phonetic perceptual identification by native-and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. Neuroimage 2004, 22: 1182–1194.

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

[31] Watkins KE, Strafella AP, Paus T. Seeing and hearing speech excites the motor system involved in speech production. Neuropsychologia 2003, 41: 989–994.

[32] D'Ausilio A, Pulvermüller F, Salmas P, Bufalari I, Begliomini C, Fadiga L. The motor somatotopy of speech perception. Curr Biol 2009, 19: 381–385.

[33] Meister IG, Wilson SM, Deblieck C, Wu AD, Iacoboni M. The essential role of premotor cortex in speech perception. Curr Biol 2007, 17: 1692–1696.

[34] Watkins K, Paus T. Modulation of motor excitability during speech perception: the role of Broca's area. J Cogn Neurosci 2004, 16: 978–987.

[35] Mohr JP, Pessin MS, Finkelstein S, Funkenstein HH, Duncan GW, Davis KR. Broca aphasia Pathologic and clinical. Neurology 1978, 28: 311–311.

[36] Crinion JT, Warburton EA, Lambon-Ralph MA, Howard D, Wise RJ. Listening to narrative speech after aphasic stroke: the role of the left anterior temporal lobe. Cereb Cortex 2006, 16: 1116–1125.

[37] Bogen JE, Bogen GM. Wernick's region-Where is it? Ann NY Acad Sci, 1976, 280: 834–843.

[38] Baker E, Blumstein SE, Goodglass H. Interaction between phonological and semantic factors in auditory comprehension. Neuropsychologia 1981, 19: 1–15.

[39] Bishop D, Mogford-Bevan K. Language Development in Exceptional Circumstances. Psychology Press 1993.

[40] Bishop CW, Miller LM. A multisensory cortical network for understanding speech in noise. J Cogn Neurosci 2009, 21: 1790–1804.

[41] Werker JF, Yeung HH. Infant speech perception bootstraps word learning. Trends Cogn Sci 2005, 9: 519–527.

[42] Tsao FM, Liu HM, Kuhl PK. Speech perception in infancy predicts language development in the second year of life: a longitudinal study. Child Dev 2004, 75: 1067–1084.

[43] Basso A, Casati G, Vignolo LA. Phonemic identification defect in aphasia. Cortex 1977, 13: 85.

[44] Scott SK, Rosen S, Lang H, Wise RJ. Neural correlates of intelligibility in speech investigated with noise vocoded speech-a positron emission tomography study. J Acoust Soc Am 2006, 120: 1075.

[45] Davis MH, Johnsrude IS. Hierarchical processing in spoken language comprehension. J Neurosci 2003, 23: 3423–3431.

[46] Uppenkamp S, Johnsrude IS, Norris D, Marslen-Wilson W, Patterson RD. Locating the initial stages of speech? Sound processing in human temporal cortex. Neuroimage 2006, 31: 1284–1296.

[47] Warren JE, Sauter DA, Eisner F, Wiland J, Dresner M, Wise RJ, *et al.* Positive emotions preferentially engage an auditory–motor "mirror" system. J Neurosci 2006, 26: 13067–13075.

[48] Roy AC, Craighero L, Fabbri-Destro M, Fadiga L. Phonological and lexical motor facilitation during speech listening: A transcranial magnetic stimulation study. J Physiol Paris 2008, 102: 101–105.

[49] Fridriksson J, Moss J, Davis B, Baylis GC, Bonilha L, Rorden C. Motor speech perception modulates the cortical language areas. Neuroimage 2008, 41: 605–613.

[50] Cherry EC. Some experiments on the recognition of speech, with one and with two ears. J Acoust Soc Am 1953, 25: 975.

[51] McFarland DH. Respiratory markers of conversational interaction. J Speech Lang Hear Res 2001, 44: 128.

[52] Pickering MJ, Garrod S. Do people use language production to make predictions during comprehension? Trends Cogn Sci 2007, 11: 105–110.

[53] Scott SK, McGettigan C, Eisner F. A little more conversation, a little less action-candidate roles for the motor cortex in speech perception. Nat Rev Neurosci 2009, 10: 295–302.

[54] Wu C, Cao S, Zhou F, Wang C, Wu X, Li L. Masking of speech in people with first-episode schizophrenia and people with chronic schizophrenia. Schizophr Res 2012a, 134: 33–41.