frontiers in Human Neuroscience

ORIGINAL RESEARCH
published: 10 February 2017
doi: 10.3389/fnhum.2017.00034

Selective Attention Enhances Beta-Band Cortical Oscillation to Speech under "Cocktail-Party" Listening Conditions

Yayue Gao[1†], Qian Wang[1†], Yu Ding[1], Changming Wang[2,3], Haifeng Li[4], Xihong Wu[5,6], Tianshu Qu[5,6] and Liang Li[1,3,5]**

[illegible]

Human listeners are able to selectively attend to target speech in a noisy environment with multiple-people talking. Using recordings of scalp electroencephalogram (EEG), this study investigated how selective attention facilitates the cortical representation of target speech under a simulated "cocktail-party" listening condition with speech-on-speech masking. The result shows that the cortical representation of target-speech signals under the multiple-people talking condition was specifically improved by selective attention relative to the non-selective-attention listening condition, and the beta-band activity was most strongly modulated by selective attention. Moreover, measured with the Granger Causality value, selective attention to the single target speech in the mixed-speech complex enhanced the following four causal connectivities for the beta-band oscillation: the ones (1) from site FT7 to the right motor area, (2) from the left frontal area to the right motor area, (3) from the central frontal area to the right motor area, and (4) from the central frontal area to the right frontal area. However, the selective-attention-induced change in beta-band causal connectivity from the central frontal area to the right motor area, but not other beta-band causal connectivities, was significantly correlated with the selective-attention-induced change in the cortical beta-band representation of target speech. These findings suggest that under the "cocktail-party" listening condition, the beta-band oscillation in EEGs to target speech is specifically facilitated by selective attention to the target speech that is embedded in the mixed-speech complex. The selective attention-induced unmasking of target speech may be associated with the improved beta-band functional connectivity from the central frontal area to the right motor area, suggesting a top-down attentional modulation of the speech-motor process.

Keywords: selective attention, speech unmasking, long-term neural activities, neural network, motor theory, informational masking

Edited by:
[illegible]

Reviewed by:
[illegible]

****Correspondence:***
[illegible]

Received: 09 [illegible] 2016
Accepted: 16 [illegible] 2017
Published: 10 [illegible] 2017

Citation:
[illegible] (2017)
[illegible]. 11:34.
: 10.3389/[illegible].2017.00034

INTRODUCTION

The "cocktail-party" problem (Cherry, 1953) indicates the astonishing ability of human listeners to recognize target speech in noisy environments with multiple-people talking. It has been confirmed that selective attention plays a critical role in this perceptual/cognitive capacity (e.g., Brungart, 2001; Freyman et al., 2001, 2004; Roman et al., 2003; Li et al., 2004; Bidet-Caulet et al., 2007; Ezzatian et al., 2011; Golumbic et al., 2012; Mesgarani and Chang, 2012). On the other hand, non-selective attention provides more generalized and sustain alertness for preparing the emergence of high-priority signals (Posner and Petersen, 1990, 2012). The relationship between selective attention and non-selective attention has been an attractive issue in the visual research field (e.g., Coull et al., 1998; Matthias et al., 2010), but has not been systematically investigated in the auditory research field.

Recently, a few studies on how selective attention affects the cortical representation of target speech have been reported (e.g., Lalor and Foxe, 2010; Ding and Simon, 2012, 2013; Golumbic et al., 2013; Kong et al., 2014). Particularly, under "cocktail-party" listening conditions, selective attention modulates low-frequency oscillations of cortical responses to speech stimuli, exhibiting both enhanced tracking of target-speech signals and enhanced suppression of masker-speech signals (Kerlin et al., 2010; Lalor and Foxe, 2010; Power et al., 2010, 2012; Mesgarani and Chang, 2012; O'Sullivan et al., 2014). It is of interest to know how the neural representation of speech signals under "cocktail-party" conditions is affected by shifting non-selective attention to selective attention.

It has been proposed that low-frequency (alpha and beta bands) oscillations of cortical activation mainly carries top-down modulation information, while high-frequency (gamma) oscillations mainly carries bottom-up information (Wang, 2010; Bastos et al., 2012; Weiss and Mueller, 2012; Bressler and Richter, 2015; Friston et al., 2015; Lewis and Bastiaansen, 2015). Particularly, top-down signals that come to lower-level brain structure underlies the attentional processing that is associated with the synchrony in the beta frequency band (Hanslmayr et al., 2007; Womelsdorf and Fries, 2007; Donner and Siegel, 2011; Bressler and Richter, 2015; Saarinen et al., 2015; Todorovic et al., 2015). More specifically, for example, beta-band activity is related to various top-down cognitive/perceptual processes (review in Engel and Fries, 2010), including prediction (Engel et al., 2001; Ahveninen et al., 2013; Todorovic et al., 2015; Lewis et al., 2016) and motor control (Brittain and Brown, 2014; Piai et al., 2015). Also, the beta-band oscillation represents functional connectivity between the frontal cortex and motor cortex in attention tasks (Thorpe et al., 2012; Piai et al., 2015). It is of interest to know whether neural oscillations in the beta band are involved in speech unmasking based on selective attention.

The present study investigated whether neural oscillations of scalp-recoded electroencephalogram (EEGs) to multiple-talker (voice) speech are modulated by selective attention and what are the potential underlying mechanisms. EEG signals were recorded from participants who either selectively attended to one of the talker's voice or non-selectively attended to the whole mixed-speech complex. Four frequency bands (theta: 4–8 Hz; alpha: 8–12 Hz; beta: 13–30 Hz; gamma: 30–48 Hz) of recorded EEGs were analyzed to reveal both the cortical representation of speech signals and the differences in cortical causal connections between the selective attention condition and the non-selective attention condition. Across EEG correlations were used to estimate whether the cortical speech representation becomes more correlated to the attended target speech under the selective attention condition than the non-selective attention condition.

MATERIALS AND METHODS

Participants

Twelve younger adults (five males and seven females) with the mean age of 23.6 years old (from 19 to 25 years old) were recruited from Peking University as the participants in this study. They provided informed consent to participate in this study and were paid a modest stipend for their participation. All the participants were right-handed native Mandarin Chinese speakers with normal and balanced (no more than 15 dB difference between the two ears) pure-tone hearing thresholds between 125 and 8000 Hz. The participants gave their written informed consent for participation in this study. The experimental procedures used in this study were approved by the Committee for Protecting Human and Animal Subjects of the Department of Psychology at Peking University.

Speech Stimuli

The speech stimuli used in this study were Chinese "nonsense" sentences. "Nonsense" sentences are syntactically correct but not semantically meaningful (e.g., Freyman et al., 1999; Li et al., 2004; Yang et al., 2007; Gao et al., 2014). Direct English translations of these Chinese sentences are similar but not identical to the English "nonsense" sentences used in previous studies (Helfer, 1997; Freyman et al., 1999, 2004; Li et al., 2004). For example, the English translation of one Chinese nonsense sentence is "That corona removes the crest-span bag". The development of the Chinese "nonsense" sentences has been described elsewhere (Yang et al., 2007).

In this study, three different younger-adult female talkers recited the speech stimuli with different sentences. In a typical recording trial, during the mixed-speech presentation when EEGs were recorded (Phase III in **Figure 1**), the three voices reciting differences sentences were presented at the same time, simulating a "cocktail-party" listening condition. Before the 3-voice mixed-speech presentation, one of the speech stimuli was presented alone (Phase I in **Figure 1**) to indicate that either the repeatedly presented speech in the mixed-speech presentation was the target speech when the pre-presented speech was recited by Voice 1 or 2, or there was no particular (single) target speech in the mixed-speech presentation when the pre-presented speech was recited by Voice 3. Consequently, the target speech was determined (when recited by Voice 1 or 2), and the other two speech stimuli formed the masker. In other words, the target

FIGURE 1 | Illustration of the six phases within each trial of EEG recordings. Phase I: a trial was started with the presentation of a single-voiced speech (Voice 1, 2, or 3) to indicate which stimulation condition the present trial belonged to [Voice 1, selective attention to Voice 1 (top panel); Voice 2, selective attention to Voice 2 (middle panel); Voice 3, non-selective attention to the whole mixed-speech complex (bottom panel)]. Phase II: a period of silence lasting 1 s. Phase III: the presentation of the mixed three-voiced speech. Phase IV: a period of silence lasting 1.2 s. Phase V and Phase VI: the repetition of Phase III and Phase IV, respectively. Under the selective-attention condition (with Voice 1 or 2), participants were instructed to press a button if they had heard a wrong words probe (yellow waves); under the non-selective-attention condition, participants were instructed to press a button if they heard a click probe (yellow waves). The blue, green, and red waves indicate the single speech of Voice 1, Voice 2, and Voice 3, respectively.

speech was presented against a two-talker-speech background. Note that two-talker speech maskers were the most effective in inducing informational masking (Freyman et al., 2004). Each of the three voices recited different sentences and the sound pressure level of the three voices were the same. The mean duration of the sentences was 3.26 s (ranged from 3.1 to 3.5 s).

All speech signals were digitized at a sampling rate of 22.05 kHz using a 24-bit Creative Sound Blaster PCI128 with a built-in anti-aliasing filter (Creative Technology, Ltd., Singapore). All the stimuli, including the single-voice speech, mixed-voice speech, and click sounds were transferred using a Creative Extigy sound blaster and presented to participants at the two ears without any interaural time disparities using two tube-ear inserts (Neuroscan, El Paso, TX, USA). The sound pressure level of a single voice was set at 56 dB SPL, calibrated by a Larson Davis Audiometer Calibration and Electroacoustic Testing System (Audit and System 824, Larson Davis, USA). Since the sound pressure level of the three voices were the same, the signal-to-masker ratio (SMR) was −3 dB when a target speech was determined in the mixed-speech presentation.

Electrophysiological Recordings

Scalp EEG recordings (with the reference electrode located on the nose) were conducted in a dim double-walled sound-attenuating booth (EMI Shielded Audiometric Examination Acoustic Suite) that was equipped with a 64-channel NeuroScan SynAmps System (Compumedics Limited, Abbotsford, VIC, Australia). EEG signals were processed with a sample rate of 1000 Hz, on-line amplified 500 times, and low-pass filtered below 200 Hz. Eye movements and eye blinks were recorded from electrodes superior and inferior to the left eye and also at the outer canthi of the two eyes. The impedances of all the recording electrodes were kept below 5 kΩ.

Procedures

The effect of selective attention was estimated by examining the differences in EEGs between the selective attention condition and the non-selective attention condition. Voice 1 and Voice 2 were used as either the target voice or the masking voice, and Voice 3 was used only as the masking voice. There were three stimulation conditions for the mixed-speech presentation: (1) Condition 1: selective attention only to Voice 1, (2) Condition 2: selective attention only to Voice 2, and (3) Condition 3: non-selective attention to the whole speech complex (**Figure 1**).

In addition to the 3-voice mixed-speech presentation, each of the speech stimuli was presented alone to obtain EEGs to the single-speech presentation.

In this study, five "nonsense" sentences from a pool with totally 360 sentences were randomly assigned to a participants (two sentences for Voice 1; other two different sentences for Voice 2; one sentence for Voice 3) and different participants listened to difference sentences. For each participants, there were four different mixed-speech presentations.

As shown in **Figure 1**, each trial contained six phases: In Phase I, a trial was started with the presentation of a single-voiced speech (Voice 1, 2, or 3) with the duration about 3.2 s as the cue to indicate which stimulation condition the present trial belonged to (Voice 1, Condition 1; Voice 2, Condition 2; Voice 3, Condition 3). Phase I was followed by Phase II, which was a period of silence lasting 1 s.

In Phase III, the mixed three-voiced speech (about 3.2 s) was presented (the same stimuli under different conditions for a participant). Phase IV was also a period of silence lasting 1.2 s. The Phase V and Phase VI were the repetition of Phase III and Phase IV, respectively. In other words, the mixed speech presentation occurred twice in a trial.

Under a selective attention condition (Condition 1 or 2), participants were instructed to pay attention to the target voice and press a button if they had heard a novel "predicate-object" structure presented with the same voice as the to-be-attended talker (as the false-word probe, with four syllables and the possibility of 14.2%, **Figure 1**). Under non-selective attention condition (Condition 3), the participants were instructed to pay attention to the whole speech complex and press a button if they heard a "click" (as the probe with the possibility of 14.2%) at a random time position (**Figure 1**). To ensure that the participants could understand and follow the instructions, a training session was conducted before EEG recordings. The percent correct in detecting the probe in each of the participants were required to be no less than 85%.

In total, there were 96 stimulation presentations for EEG recordings (after the removal of the presentations with probes) for each of the three conditions, and these 96 presentations were randomly assigned into four blocks. Each block contained 24 stimulation presentations for each of the three conditions whose presenting order was arranged randomly for a participant. It took about 10 mins to complete one block. To limit eye movements, participants were also asked to stare a cross in the front in a trial.

Data Analyses

Using the EEGLAB toolbox (Delorme and Makeig, 2004) in MATLAB, raw EEG data were filtered by three different band-pass filters (alpha: 8–12 Hz; beta: 12–30 Hz; gamma: 30–48 Hz), and then segmented into epochs from −300 to 3500 ms relative to the onset of a mixed-speech presentation. The baseline correction was conducted in the period of −300 to 0 ms before the presentation onset. The epochs that contained more than ±30 μV potential were rejected as artifacts. The rest of epochs were averaged for each condition to analyze the grange causality and across EEG correlations.

To avoid the onset and offset (above 3000-ms) effect (Pasley et al., 2012]TJ 30t4.2%, the rest of period 300 (30046304(ms4630(a)8(fter4630[(t)-4(he)6304(mixed-spee)-6(c)6(h)6306(pre)2(sent)-6(ation)6306(o

Thechg(he)-43indplexwas φav1of

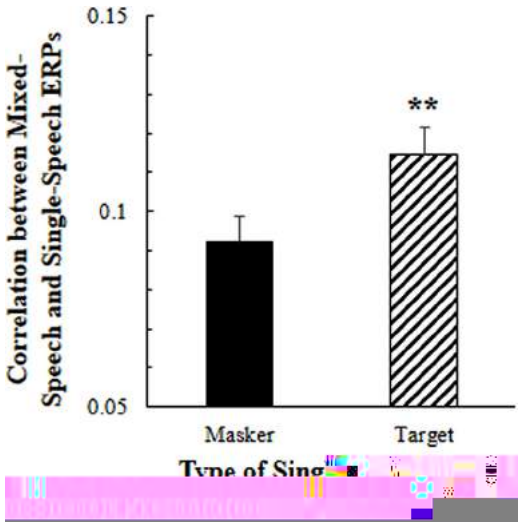

FIGURE 2 | Under the selective attention condition, the correlation between the all-site-averaged EEGs to the mixed-speech complex and the all-site-averaged EEGs to the single speech that was either the target or the masker speech in the mixed-speech complex. ** < 0.01, paired -test. The error bar indicates the standard errors of the mean.

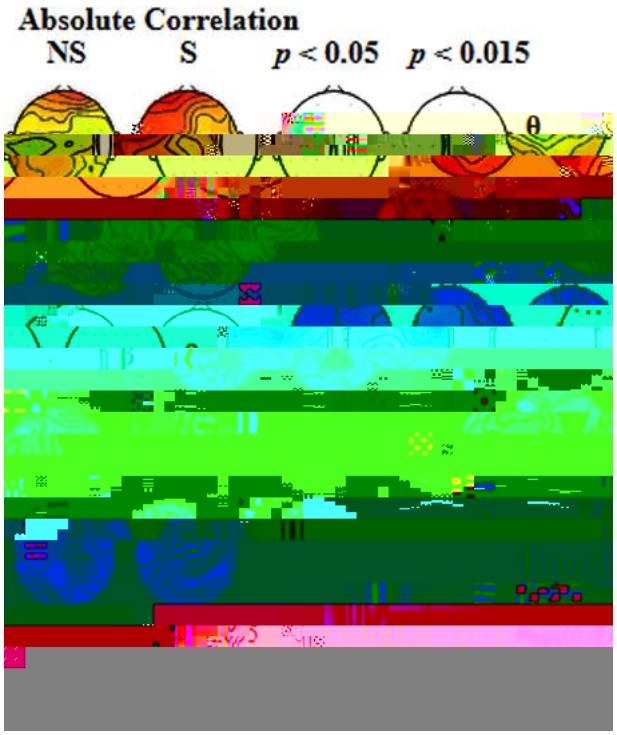

FIGURE 3 | The two left columns: for each of the five types of frequency bands [theta (θ), alpha (α), beta (β), gamma (γ), broad], the scalp topographical maps showing location distributions of absolute correlations between the EEGs to the mixed-speech complex and EEGs to a single speech under either the non-selective attention (NS) condition or the selective attention (S) condition. The two right columns: for each of the frequency bands, the recordings sites at which the correlation difference between the two attention conditions was significant when the level was either 0.05 and or 0.015.

attention condition (NS) and the EEGs to a single speech for all the recording sites; the second left column shows the absolute correlation coefficients between the EEGs to the mixed speech under the selective attention condition (S) and the EEGs to the target single speech for all the recording sites.

To reveal the frequency band that was the most vulnerable to selective attention, **Figure 3** also shows the statistically thresholded topographical map (the two right columns) indicating the electrode sites exhibiting significant differences in absolute correlation coefficient between the selective attention condition (S) and the non-selective attention condition (NS). When the p level was 0.05 (the second right column), both beta- and gamma-band components of EEGs recorded from a few electrode sites exhibited significant differences between the two attention conditions. **Table 1** shows the p-values for these electrode sites. Also shown in **Table 1**, only the beta-band component of EEGs recorded from the site Cz exhibited a significant difference between the two attentional conditions when the p was as low as 0.011. In other words, the beta-band obtained at the site Cz was the only component exhibiting a significant difference between the two attention conditions when the p value was less than 0.020. The right column in **Figure 3** presents the results indicating that the beta-band component of EEGs at the site Cz was the only one exhibiting a significant difference between the two attention conditions when the p value was 0.015 (which was just larger than 0.011 but smaller than 0.020). More in detail, at the p level of 0.015, the mixed-speech-evoked EEGs at site Cz were significantly more correlated with the single-speech-evoked EEGs under the selective attention condition than under the non-selective attention condition for beta band [$t(11) = 3.029$, $p = 0.011$, paired t-test], but not for other bands (both $p > 0.05$, paired t-test), indicating that the EEG beta-band component at the site Cz was the most vulnerable to selective attention (Supplementary Figure S2).

TABLE 1 | Electrode sites at which beta and gamma bands were significantly different between the two attention conditions.

Band	Sites	*df*	*t*	*p*
Beta	CZ	11	3.029	0.011
Beta	F7	11	2.642	0.023
Beta	F1	11	2.494	0.030
Beta	FT7	11	2.408	0.035
Beta	F3	11	2.384	0.036
Beta	FP1	11	2.325	0.040
Beta	FPZ	11	2.258	0.045
Beta	F5	11	2.203	0.050
Gamma	PO7	11	2.593	0.025
Gamma	PO5	11	2.554	0.027
Gamma	P5	11	2.372	0.037
Gamma	TP7	11	2.311	0.041
Gamma	PO3	11	2.283	0.043

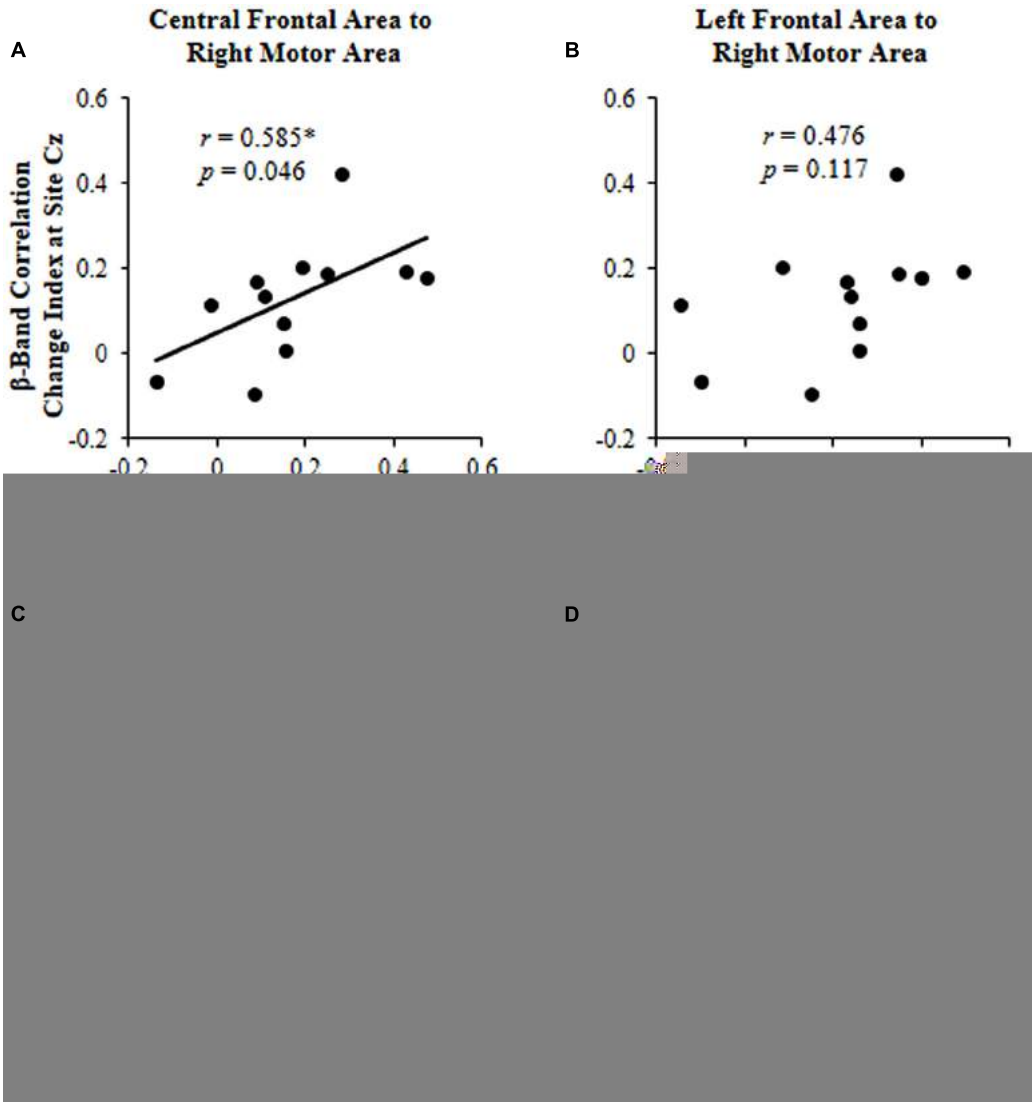

FIGURE 5 | For each of the four significant GCs shown in Figure 4, the correlation between the beta (β)-band correlation change index induced by selective attention and the beta (β)-band GC change index induced by selective attention. (A) Causal connectivity from the central frontal area to the right motor area; **(B)** Causal connectivity from the left frontal area to the right motor area; **(C)** Causal connectivity from the central frontal area to the right frontal area; **(D)** Causal connectivity from site TP7 to the right motor area. * < 0.05.

alpha-band component or the gamma-band component, in the mixed-speech-evoked EEGs, was significantly more correlated with the single-speech-evoked EEGs under the selective attention condition (where the target single-voice speech was attended) than under the non-selective attention condition. Thus, the EEG beta-band component was the most vulnerable to selective attention.

Beta oscillations are associated with attention and predictions (Engel and Fries, 2010; Donner and Siegel, 2011; Weiss and Mueller, 2012; Todorovic et al., 2015), which are critical to speech cognition. Particularly, the top-down propagation of predictions reflected by beta oscillations (Engel et al., 2001; Bastos et al., 2012; Ahveninen et al., 2013; Lewis and Bastiaansen, 2015; Todorovic et al., 2015; Lewis et al., 2016) may be more critical for selective-attention-induced unmasking of speech, probably through enhancing the mechanism underlying binding distributed sets of neurons into a coherent representation of speech contents (Weiss and Mueller, 2012).

Selective-Attention Facilitated Beta-Band Causal Connectivity from the Central Frontal Area to the Right Motor Area

The results of this study also showed that in total four beta-band causal connectivities (measured as GCs) were enhanced by selective attention, including the ones (1) from site FT7 to the right motor area, (2) from the left frontal area to the

right motor area, (3) from the central frontal area to the right motor area, and (4) from the central frontal area to the right frontal area. However, only the selective-attention-induced enhancement of beta-band GC from the central frontal area to the right motor area was significantly correlated to the selective-attention-induced enhancement of the correlation between beta-band oscillations to the mixed speech complex and beta-band oscillations to the single speech. The results suggest that the selective-attention-induced improvement of beta-band representation of target speech signals is associated with the enhanced top-down modulation of the motor areas in the right hemisphere by the central frontal cortical areas. In other words, selective attention improves speech-related motor processes. However, due to the low spatial resolution of EEGs, whether the beta activities over central areas are based on the auditory or motor activity need further investigation in the future.

The *Motor Theory* of speech perception proposes that the interaction between the auditory and motor systems plays an essential role in speech perception (Liberman et al., 1952, 1967; Liberman and Mattingly, 1985; for review see Wu et al., 2014). It has been evident that speech perception activates the motor cortex (Fadiga et al., 2002; Callan et al., 2004; Wilson et al., 2004; Pulvermüller et al., 2006; Wilson and Iacoboni, 2006; Meister et al., 2007; Bever and Poeppel, 2010; Hickok et al., 2011; Elemans et al., 2015). Thus, under adverse listening conditions (such as the cocktail-party environment) where the perceptual load is high (Hickok and Poeppel, 2007; Fridriksson et al., 2008; Bishop and Miller, 2009), with the involvement of the motor system the listener can better identify the speaker's intention and follow the target stream (Wu et al., 2014).

CONCLUSION

(1) The cortical representation of target-speech signals under the multiple-people talking condition is specifically improved by selective attention, and the beta-band EEG component is the most vulnerable to selective attention.
(2) The selective-attention-induced enhancement of beta-band causal connectivity from the central frontal area to the right motor area is correlated with the selective-attention-induced enhancement of the cortical beta-band representation of target speech.
(3) Selective attention to a single-voiced target speech, which is embedded in a mixed-speech complex (with speech-on-speech masking), improves the cortical representation of the target speech by facilitating the top-down frontal modulation of the motor cortical areas.
(4) The unmasking of target speech based on selective attention may be caused by top-down attentional modulation of the speech-motor interactions.

AUTHOR CONTRIBUTIONS

YG, QW, and YD: Experimental design, experiment set up, experiment conduction, data analyses, figure/table construction, and paper writing. CW and HL: Experimental design, data analyses, and paper writing. XW: Experimental design and paper writing. LL: Experimental design, figure/table construction, and paper writing. TQ: Experimental design, experiment set up, and paper writing.

ACKNOWLEDGMENTS

This work was supported by supported by the '973' National Basic Research Program of China (2015CB351800), the National High Technology Research and Development Program of China (863 Program: 2015AA016306), the Beijing Municipal Science and Tech Commission (Z161100002616017), and the National Natural Science Foundation of China (81501155, 61171186, 61671187).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fnhum.2017.00034/full#supplementary-material

REFERENCES

Ahveninen, J., Huang, S., Belliveau, J. W., Chang, W. T., and Hämäläinen, M. (2013). Dynamic oscillatory processes governing cued orienting and allocation of auditory attention. *J. Cogn. Neurosci.* 25, 1926–1943. doi: 10.1162/jocn_a_00452

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., and Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron* 76, 695–711. doi: 10.1016/j.neuron.2012.10.038

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

Bidet-Caulet, A., Fischer, C., Besle, J., Aguera, P. E., Giard, M. H., and Bertrand, O. (2007). Effects of selective attention on the electrophysiological representation of concurrent sounds in the human auditory cortex. *J. Neurosci.* 27, 9252–9261. doi: 10.1523/JNEUROSCI.1402-07.2007

Bishop, C. W., and Miller, L. M. (2009). A multisensory cortical network for understanding speech in noise. *J. Cogn. Neurosci.* 21, 1790–1804. doi: 10.1162/jocn.2009.21118

Bressler, S. L., and Richter, C. G. (2015). Interareal oscillatory synchronization in top-down neocortical processing. *Curr. Opin. Neurobiol.* 31, 62–66. doi: 10.1016/j.conb.2014.08.010

Brittain, J. S., and Brown, P. (2014). Oscillations and the basal ganglia: motor control and beyond. *Neuroimage* 85, 637–647. doi: 10.1016/j.neuroimage.2013.05.084

Brungart, D. S. (2001). Informational and energetic masking effects in the perception of two simultaneous talkers. *J. Acoust. Soc. Am.* 109, 1101–1109. doi: 10.1121/1.1345696

Callan, D. E., Jones, J. A., Callan, A. M., and Akahane-Yamada, R. (2004). 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* 22, 1182–1194. doi: 10.1016/j.neuroimage.2004.03.006

Cherry, E. C. (1953). Some experiments on the recognition of speech, with one and with two ears. *J. Acoust. Soc. Am.* 25, 975–979. doi: 10.1121/1.1907229

Coull, J. T., Frackowiak, R. S. J., and Frith, C. D. (1998). Monitoring for target objects: activation of right frontal and parietal cortices with increasing

time on task. *Neuropsychologia* 36, 1325–1334. doi: 10.1016/S0028-3932(98)00035-9

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

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

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

Donner, T. H., and Siegel, M. (2011). A framework for local cortical oscillation patterns. *Trends Cogn. Sci.* 15, 191–199. doi: 10.1016/j.tics.2011.03.007

Elemans, C. P. H., Rasmussen, J. H., Herbst, C. T., Düring, D. N., Zollinger, S. A., Brumm, H., et al. (2015). Universal mechanisms of sound production and control in birds and mammals. *Nat. Commun.* 6:8978. doi: 10.1038/ncomms9978

Engel, A. K., and Fries, P. (2010). Beta-band oscillations—signalling the status quo? *Curr. Opin. Neurobiol.* 20, 156–165. doi: 10.1016/j.conb.2010.02.015

Engel, A. K., Fries, P., and Singer, W. (2001). Dynamic predictions: oscillations and synchrony in top–down processing. *Nat. Rev. Neurosci.* 2, 704–716. doi: 10.1038/35094565

Ezzatian, P., Li, L. A., Pichora-Fuller, K., and Schneider, B. (2011). The effect of priming on release from informational masking is equivalent for younger and older adults. *Ear Hear.* 32, 84–96. doi: 10.1097/AUD.0b013e3181ee6b8a

Fadiga, L., Craighero, L., Buccino, G., and Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: a TMS study. *Eur. J. Neurosci.* 15, 399–402. doi: 10.1046/j.0953-816x.2001.01874.x

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (2001). Spatial release from informational masking in speech recognition. *J. Acoust. Soc. Am.* 109, 2112–2122. doi: 10.1121/1.1354984

Freyman, R. L., Balakrishnan, U., and Helfer, K. S. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *J. Acoust. Soc. Am.* 115, 2246–2256. doi: 10.1121/1.1689343

Freyman, R. L., Helfer, K. S., McCall, D. D., and Clifton, R. K. (1999). The role of perceived spatial separation in the unmasking of speech. *J. Acoust. Soc. Am.* 106, 3578–3588. doi: 10.1121/1.428211

Fridriksson, J., Moss, J., Davis, B., Baylis, G. C., Bonilha, L., and Rorden, C. (2008). Motor speech perception modulates the cortical language areas. *Neuroimage* 41, 605–613. doi: 10.1016/j.neuroimage.2008.02.046

Friston, K. J., Bastos, A. M., Pinotsis, D., and Litvak, V. (2015). LFP and oscillations—what do they tell us? *Curr. Opin. Neurobiol.* 31, 1–6. doi: 10.1016/j.conb.2014.05.004

Gao, Y.-Y., Cao, S.-Y., Qu, T.-S., Wu, X.-H., Li, H.-F., Zhang, J.-S., et al. (2014). Voice-associated static face image releases speech from informational masking. *Psych. J.* 3, 113–120. doi: 10.1002/pchj.45

Golumbic, E. M. Z., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., et al. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a "Cocktail Party". *Neuron* 77, 980–991. doi: 10.1016/j.neuron.2012.12.037

Golumbic, E. M. Z., Poeppel, D., and Schroeder, C. E. (2012). Temporal context in speech processing and attentional stream selection: a behavioral and neural perspective. *Brain Lang.* 122, 151–161. doi: 10.1016/j.bandl.2011.12.010

Hanslmayr, S., Aslan, A., Staudigl, T., Klimesch, W., Herrmann, C. S., and Bäuml, K. H. (2007). Prestimulus oscillations predict visual perception performance between and within subjects. *Neuroimage* 37, 1465–1473. doi: 10.1016/j.neuroimage.2007.07.011

Helfer, K. S. (1997). Auditory and auditory-visual perception of clear and conversational speech. *J. Speech Lang. Hear. Res.* 40, 432–443. doi: 10.1044/jslhr.4002.432

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

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

Kerlin, J. R., Shahin, A. J., and Miller, L. M. (2010). Attentional gain control of ongoing cortical speech representations in a "cocktail party". *J. Neurosci.* 30, 620–628. doi: 10.1523/JNEUROSCI.3631-09.2010

Kong, Y. Y., Mullangi, A., and Ding, N. (2014). Differential modulation of auditory responses to attended and unattended speech in different listening conditions. *Hear. Res.* 316, 73–81. doi: 10.1016/j.heares.2014.07.009

Lalor, E. C., and Foxe, J. J. (2010). Neural responses to uninterrupted natural speech can be extracted with precise temporal resolution. *Eur. J. Neurosci.* 31, 189–193. doi: 10.1111/j.1460-9568.2009.07055.x

Lewis, A. G., and Bastiaansen, M. (2015). A predictive coding framework for rapid neural dynamics during sentence-level language comprehension. *Cortex* 68, 155–168. doi: 10.1016/j.cortex.2015.02.014

Lewis, A. G., Schoffelen, J. M., Schriefers, H., and Bastiaansen, M. (2016). A predictive coding perspective on beta oscillations during sentence-level language comprehension. *Front. Hum. Neurosci.* 10:85. doi: 10.3389/fnhum.2016.00085

Li, L., Daneman, M., Qi, J. G., and Schneider, B. A. (2004). Does the information content of an irrelevant source differentially affect spoken word recognition in younger and older adults? *J. Exp. Psychol. Hum. Percept. Perform.* 30, 1077–1091.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychol. Rev.* 74, 431–461. doi: 10.1037/h0020279

Liberman, A. M., Delattre, P., and Cooper, F. S. (1952). The role of selected stimulus-variables in the perception of the unvoiced stop consonants. *Am. J. Psychol.* 65, 497–516. doi: 10.2307/1418032

Liberman, A. M., and Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition* 21, 1–36. doi: 10.1016/0010-0277(85)90021-6

Matthias, E., Bublak, P., Müller, H. J., Schneider, W. X., Krummenacher, J., and Finke, K. (2010). The influence of alertness on spatial and nonspatial components of visual attention. *J. Exp. Psychol. Hum. Percept. Perform.* 36, 38–56. doi: 10.1037/a0017602

Meister, I. G., Wilson, S. M., Deblieck, C., Wu, A. D., and Iacoboni, M. (2007). The essential role of premotor cortex in speech perception. *Curr. Biol.* 17, 1692–1696. doi: 10.1016/j.cub.2007.08.064

Mesgarani, N., and Chang, E. F. (2012). Selective cortical representation of attended speaker in multi-talker speech perception. *Nature* 485, 233–236. doi: 10.1038/nature11020

O'Sullivan, J. A., Power, A. J., Mesgarani, N., Rajaram, S., Foxe, J. J., Shinn-Cunningham, B. G., et al. (2014). Attentional selection in a cocktail party environment can be decoded from single-trial EEG. *Cereb. Cortex* 25, 1697–1706. doi: 10.1093/cercor/bht355

Pasley, B. N., David, S. V., Mesgarani, N., Flinker, A., Shamma, S. A., Crone, N. E., et al. (2012). Reconstructing speech from human auditory cortex. *PLoS Biol.* 10:e1001251. doi: 10.1371/journal.pbio.1001251

Piai, V., Roelofs, A., Rommers, J., Dahlslätt, K., and Maris, E. (2015). Withholding planned speech is reflected in synchronized beta-band oscillations. *Front. Hum. Neurosci.* 9:549. doi: 10.3389/fnhum.2015.00549

Posner, M. I., and Petersen, S. E. (1990). The attention system of the human brain. *Annu. Rev. Neurosci.* 13, 25–42. doi: 10.1146/annurev.neuro.13.1.25

Posner, M. I., and Petersen, S. E. (2012). The attention system of the human brain: 20 years after. *Annu. Rev. Neurosci.* 35, 73–89. doi: 10.1146/annurev-neuro-062111-150525

Power, A. J., Foxe, J. J., Forde, E. J., Reilly, R. B., and Lalor, E. C. (2012). At what time is the cocktail party? A late locus of selective attention to natural speech. *Eur. J. Neurosci.* 35, 1497–1503. doi: 10.1111/j.1460-9568.2012.08060.x

Power, A. J., Lalor, E. C., and Reilly, R. B. (2010). Endogenous auditory spatial attention modulates obligatory sensory activity in auditory cortex. *Cereb. Cortex* 21, 1223–1230. doi: 10.1093/cercor/bhq233

Pulvermüller, F., Huss, M., Kherif, F., del Prado Martin, F. M., Hauk, O., and Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proc. Natl. Acad. Sci. U.S.A.* 103, 7865–7870. doi: 10.1073/pnas.0509989103

Roman, N., Wang, D., and Brown, G. J. (2003). Speech segregation based on sound localization. *J. Acoust. Soc. Am.* 114, 2236–2252. doi: 10.1121/1.1610463

Saarinen, T., Jalava, A., Kujala, J., Stevenson, C., and Salmelin, R. (2015). Task-sensitive reconfiguration of corticocortical 6–20 Hz oscillatory coherence in naturalistic human performance. *Hum. Brain Mapp.* 36, 2455–2469. doi: 10.1002/hbm.22784

Tadel, F., Baillet, S., Mosher, J. C., Pantazis, D., and Leahy, R. M. (2011). Brainstorm: a user-friendly application for MEG/EEG analysis. *Comput. Intell. Neurosci.* 2011:879716. doi: 10.1155/2011/879716

Thorpe, S., D'Zmura, M., and Srinivasan, R. (2012). Lateralization of frequency-specific networks for covert spatial attention to auditory stimuli. *Brain Topogr.* 25, 39–54. doi: 10.1007/s10548-011-0186-x

Todorovic, A., Schoffelen, J. M., van Ede, F., Maris, E., and de Lange, F. P. (2015). Temporal expectation and attention jointly modulate auditory oscillatory activity in the beta band. *PLoS ONE* 10:e0120288. doi: 10.1371/journal.pone.0120288

Wang, X. J. (2010). Neurophysiological and computational principles of cortical rhythms in cognition. *Physiol. Rev.* 90, 1195–1268. doi: 10.1152/physrev.00035.2008

Weiss, S., and Mueller, H. M. (2012). "Too many betas do not spoil the broth": the role of beta brain oscillations in language processing. *Front. Psychol.* 3:201. doi: 10.3389/fpsyg.2012.00201

Wilson, S. M., and Iacoboni, M. (2006). Neural responses to non-native phonemes varying in producibility: evidence for the sensorimotor nature of speech perception. *Neuroimage* 33, 316–325. doi: 10.1016/j.neuroimage.2006.05.032

Wilson, S. M., Saygin, A. P., Sereno, M. I., and Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nat. Neurosci.* 7, 701–702. doi: 10.1038/nn1263

Womelsdorf, T., and Fries, P. (2007). The role of neuronal synchronization in selective attention. *Curr. Opin. Neurobiol.* 17, 154–160. doi: 10.1016/j.conb.2007.02.002

Wu, Z.-M., Chen, M.-L., Wu, X.-H., and Li, L. (2014). Interaction between auditory system and motor system in speech perception. *Neurosci. Bull.* 30, 490–496. doi: 10.1007/s12264-013-1428-6

Yang, Z., Chen, J., Huang, Q., Wu, X., Wu, Y., Schneider, B. A., et al. (2007). The effect of voice cuing on releasing Chinese speech from informational masking. *Speech Commun.* 49, 892–904. doi: 10.1097/AUD.0b013e3181db6dc2

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.