Perceptual Fusion Tendency of Speech Sounds

Ying Huang, Jingyu Li, Xuefei Zou, Tianshu Qu, Xihong Wu, Lihua Mao, Yanhong Wu, and Liang Li

Abstract

■ To discriminate and to recognize sound sources in a noisy, reverberant environment, listeners need to perceptually integrate the direct wave with the reflections of each sound source. It has been confirmed that perceptual fusion between direct and reflected waves of a speech sound helps listeners recognize this speech sound in a simulated reverberant environment with disrupting sound sources. When the delay between a direct sound wave and its reflected wave is sufficiently short, the two waves are perceptually fused into a single sound image as coming from the source location. Interestingly, compared with nonspeech sounds such as clicks and noise bursts, speech sounds have a much larger perceptual fusion tendency. This study investigated why the fusion tendency for speech sounds is so large. Here we show that when the temporal amplitude fluctuation of speech was artificially time reversed, a large perceptual fusion tendency of speech sounds disappeared, regardless of whether the speech acoustic carrier was in normal or reversed temporal order. Moreover, perceptual fusion of normal-order speech, but not that of time-reversed speech, was accompanied by increased coactivation of the attention-control-related, spatial-processing-related, and speech-processing-related cortical areas. Thus, speech-like acoustic carriers modulated by speech amplitude fluctuation selectively activate a cortical network for top–down modulations of speech processing, leading to an enhancement of perceptual fusion of speech sounds. This mechanism represents a perceptual-grouping strategy for unmasking speech under adverse conditions. ■

INTRODUCTION

Investigation of how various sound sources (including speech sounds) are discriminated and a target source is correctly recognized in a noisy, reverberant environment is critical for reaching the understanding of why the human brain is able to extract target information under conditions with sensory-input "flooding." It has been confirmed that to distinguish signals from various speech sources and to correctly recognize the target speech source in a simulated reverberant environment, listeners need to not only perceptually integrate the direct wave with the reflections of the target speech source (Huang, Huang, Chen, Wu, & Li, 2009; Huang, Huang, et al., 2008) but also perceptually integrate the direct wave with the reflections of the masking speech source (Rakerd, Aaronson, & Hartmann, 2006; Brungart, Simpson, & Freyman, 2005).

When the delay between a leading sound (such as the direct wave from a sound source) and a correlated lagging sound (such as a reflection of the direct wave) is sufficiently short, attributes of the lagging sound are perceptually captured by the leading sound (Li, Qi, He, Alain, & Schneider, 2005), causing a single fused sound image from a location near the leading source (the precedence effect, see Litovsky, Colburn, Yost, & Guzman, 1999; Freyman, Clifton, & Litovsky, 1991; Zurek, 1980; Wallach, Newman, & Rosenzweig, 1949). This perceptual fusion is able to produce *perceived* spatial separation between uncorrelated sound sources, and the *perceived* spatial separation plays a role in reducing masking for speech recognition (Huang, Huang, et al., 2008, 2009; Rakerd et al., 2006; Wu et al., 2005; Li, Daneman, Qi, & Schneider, 2004; Freyman, Helfer, McCall, & Clifton, 1999). For example, when both the target and the masker are presented by a loudspeaker to the listener's left and by another loudspeaker to the listener's right, the perceived location of the target and that of the masker can be manipulated by changing the interloudspeaker interval for the target and that for the masker (Li et al., 2004). More specifically, for both the target and the masker, when the sound onset of the right loudspeaker leads that of the left loudspeaker by a short time (e.g., 3 msec), both a single target image and a single masker image are perceived by the human listener as coming from the right loudspeaker. However, if the onset delay between the two loudspeakers is reversed only for the masker, the target is still perceived as coming from the right loudspeaker, but the masker is perceived as coming from the left loudspeaker. The perceived colocation and the perceived separation are based on perceptual integration of correlated sound waves delivered from each of the two energy at each ear nor the stimulus-image compactness/diffusiveness is substantially changed (e.g., Li et al., 2004).

The minimum delay allowing a listener to perceive the lagging sound as a discrete echo (when the perceptual fusion is just broken) is called the echo threshold (Litovsky et al., 1999; Freyman et al., 1991), which indicates the perceptual fusion tendency (i.e., a large echo threshold suggests a large perceptual fusion tendency). Interestingly, speech sounds, which are the most important acoustic stimuli for human communication and contain distinct patterns of periodicities and transients, have much larger echo thresholds than other types of sounds such as clicks and noise bursts (Rakerd, Hartmann, & Hsu, 2000; Litovsky et al., 1999; Lochner & Burger, 1958; Cherry & Taylor, 1954; Wallach et al., 1949).

This study investigated why the tendency of the perceptual fusion of speech sounds is so large. Because the temporal properties of speech sounds are critical for achieving communication (Rosen, 1992), the focus of this study was particularly on the role played by the overall temporal amplitude fluctuation (TAF) of speech in maintaining the perceptual fusion tendency and in eliciting perceptual-fusion-related cortical activations.

The TAF of Chinese speech was extracted using the Hilbert transform as previously used by other investigators (Zeng et al., 2005; Smith, Delgutte, & Oxenham, 2002). When the speech TAF is artificially removed from speech, the remaining acoustic component is called the speech acoustic carrier (AC), which still contains some speech acoustic features including harmonic structures, frequency modulation, and periodic occurrences of noise-like consonants. When the TAF and the AC are artificially separated, they can be time reversed either independently or at the same time. Any of the temporal reversals does not change the long-term spectrum, the fundamental frequency, and other spectrotemporal characters, including the overall spectral modulation density, the temporal modulation frequency, and the modulation amplitude. The modified speech sound is still "speech-like" but has reduced or no semantic content.

EXPERIMENT 1

Methods

Participants

Eighteen university students (12 women and 6 men, 19–28 years old, mean age = 23 years) participated in this experiment. In this and the following three experiments (Experiments 2–4), each participant participated in only one experiment of this study, and all the participants had normal (no more than 25 dB) and balanced (no more than 15 dB difference between the two ears) pure-tone hearing thresholds at frequencies from 0.125 to 8 kHz. They all gave their written informed consent to participate in the experiment and were paid a modest stipend for their participation.

Apparatus

During a testing session, the participant was seated at the center of an anechoic chamber (Beijing CA Acoustics, Beijing, China), which was 560 cm in length, 400 cm in width, and 193 cm in height. Acoustic signals were digitized using the 24-bit Creative Sound Blaster PCI128 (which had a built-in anti-aliasing filter) (Creative Technology, Ltd., Singapore) and audio editing software [Cooledit Pro 2.0 (Syntrillium Software Corp., Phoenix, AZ)]. The analog outputs were delivered to two loudspeakers (Dynaudio Acoustics, BM6A; Risskov, Denmark) in the frontal azimuthal plane at previous investigators (Zeng et al., 2005; Smith et al., 2002): spectrum-matched steady-state noise, normal-order speech, and time-reversed speech. To obtain the normal-order speech AC, the normal-order speech was multiplied by the ratio of the TAF of the spectrum-matched steady-state noise to the TAF of the normal-order speech. Thus, the normal-order speech AC had the fine structures of the normal-order speech and the TAF of the spectrum-matched steady-state noise (Drullman, Festen, & Plomp, 1994). Similarly, to obtain the time-reversed speech AC, the time-reversed speech was multiplied by the ratio of the TAF of the spectrum-matched noise to the TAF of the time-reversed speech.

Procedure

Participants pressed a button of the response box to initiate a test session. For each of the three sound types, sounds delivered from the two loudspeakers were identical in a trial. The right loudspeaker always led the left loudspeaker. Participants were instructed to indicate whether they perceived a discrete sound from the location around the left loudspeaker by pressing the left button of the response box or nothing from the location around the left loudspeaker by pressing the right button. The time lag between the two loudspeakers (called the lead/lag delay) was decreased following the participant's three successive responses, indicating that a sound around the left loudspeaker location was perceived, and increased following one response, indicating that no sound was perceived from the left loudspeaker, using a three-down-one-up procedure (Levitt, 1971). No feedback was given to participants. The testing order for the three types of sounds was counterbalanced among the 18 participants according to the Latin square design.

Each session was started with the lead/lag delay of 72 msec. The initial step size of changing the lead/lag delay was 16 msec, and the step size was altered by a factor of .5 with each reversal of direction until the minimum size of 1 msec was reached. A test session was terminated following 10 reversals in direction, and the echo threshold for that session was defined as the averaged lead/lag delay for the last six reversals. For each participant under each condition, there were four test sessions. The averaged echo threshold of the four sessions was used as the echo threshold for the participant under the condition. During testing, the participant was instructed to keep his or her head still and face the midline in the frontal area, but the head was not physically fixed.

In this and in the following psychophysical experiments (Experiments 1–3), sounds from the loudspeakers were calibrated using a B&K sound level meter (Type 2230) whose microphone was placed at the position of the head when the listener was absent, using a "slow"/"RMS" meter response. All sounds were presented at a level such that each loudspeaker, playing alone, would produce a comfortable sound pressure of 56 dBA SPL.

Results

In Experiment 1, both TAF information of normal-order speech and TAF information of time-reversed speech were removed by replacement of these speech TAFs with the TAF of spectrum-matched steady-state noise. The echo threshold of the spectrum-matched steady-state noise was significantly lower than both the echo threshold of the normal-order speech AC (p = .011, two-tailed paired samples t test with the significant level of .05/3 = .0167) and the echo threshold of the time-reversed speech AC (p = .006), but there was no significant difference between the two types of speech ACs ($p > .05$) (Figure 1).

EXPERIMENT 2

Methods

Participants

Eighteen university students (13 women and 5 men, 17–27 years old, mean age = 22 years) participated in this experiment.

Apparatus

The apparatus was the same as used in Experiment 1.

Stimuli

Three types of sounds were used: (a) spectrum-matched steady-state noises, (b) spectrum-matched noises modulated by the TAF of normal-order speech, and (c) spectrum-matched noises modulated by the TAF of time-reversed

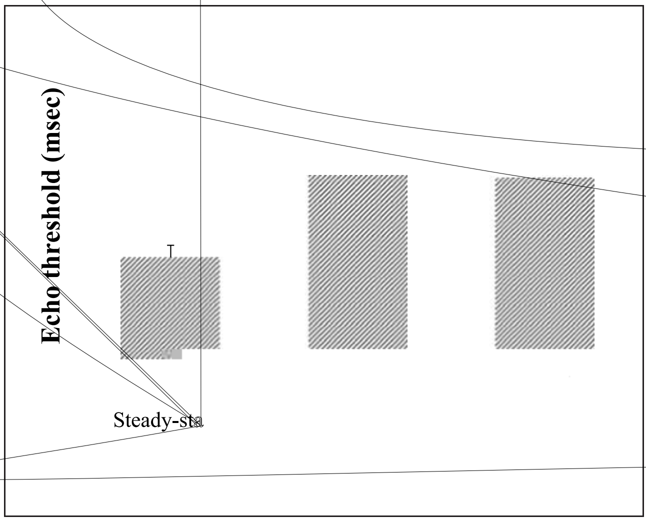


Figure 1

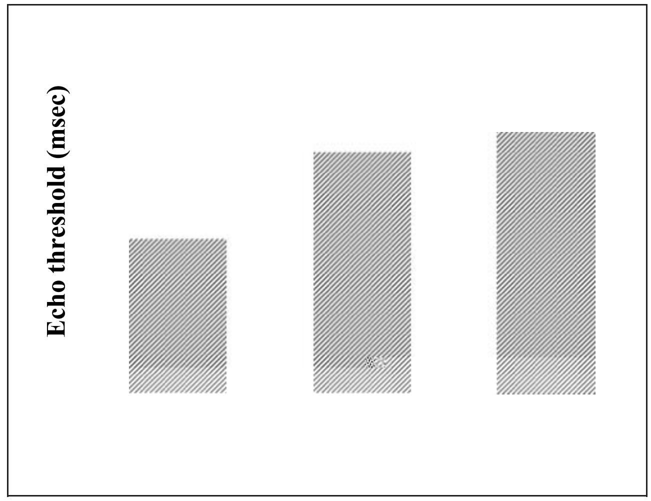
speech. To obtain the spectrum-matched noise modulated by the TAF of normal-order speech (the modulated noise had the same AC as the spectrum-matched steady-state noise and the TAF of the normal-order speech), the spectrum-matched steady-state noise was multiplied by the ratio of the TAF of normal-order speech to the TAF of the spectrum-matched steady-state noise (Drullman et al., 1994). Similarly, to obtain the spectrum-matched noise modulated by the TAF of time-reversed speech, the spectrum-matched steady-state noise was multiplied by the ratio of the TAF of time-reversed speech to the TAF of the steady-state noise.

Procedure

The procedure was the same as used in Experiment 1. The testing order for these three types of sounds was counterbalanced among the 18 participants according to the Latin square design.

Results

In Experiment 2, both the TAF of normal-order speech and the TAF of time-reversed speech were extracted, and each of the two types of TAFs was used to modulate spectrum-matched steady-state noise. The echo threshold of spectrum-matched steady-state noise significantly increased after the noise was amplitude modulated by either the TAF of normal-order speech ($p < .001$, two-tailed paired samples t test with the significant level of .05/3 = .0167) or the TAF of time-reversed speech ($p = .001$), and the two types of amplitude-modulated noises did not differ significantly ($p > .05$) (Figure 2).



EXPERIMENT 3

Methods

Participants

Twenty university students (8 women and 12 men, 19–25 years old, mean age = 22 years) participated in this experiment.

Apparatus

The apparatus was the same as used in Experiment 1.

Stimuli

Four types of Chinese speech or speech-like sounds were used in this experiment, which had different TAF and AC combinations: (a) normal-order TAF and normal-order AC, (b) time-reversed TAF and time-reversed AC, (c) normal-order TAF and time-reversed AC, and (d) time-reversed TAF and normal-order AC. To obtain speech with the TAF of normal-order speech and the AC of time-reversed speech, the time-reversed speech was multiplied by the ratio of the TAF of normal-order speech to the TAF of time-reversed speech (Drullman et al., 1994). Similarly, to obtain speech with the TAF of time-reversed speech and the AC of normal-order speech, the normal-order speech was multiplied by the ratio of the TAF of time-reversed speech to the TAF of normal-order speech.

Procedure

The procedure was the same as used in Experiment 1. The testing order for the four types of sounds was counterbalanced among the 20 participants according to the Latin square design.

Results

Experiment 3 investigated whether the TAF of normal-order speech and the TAF of time-reversed speech are different in affecting the perceptual fusion tendency when the AC is of speech or speech-like. Echo-threshold comparisons were made across the following four types of Chinese speech or speech-like sounds with different TAF and AC combinations: (a) normal-order TAF and normal-order AC, (b) time-reversed TAF and time-reversed AC, (c) normal-order TAF and time-reversed AC, and (d) time-reversed TAF and normal-order AC. As shown in Figure 3, unlike the results of Experiment 2, speech sounds with the normal-order TAF had markedly larger echo thresholds than those with the time-reversed TAF.

An ANOVA confirms that the interaction between the temporal order of TAF and the temporal order of AC was not significant ($p > .05$), the main effect of temporal order of AC was not significant ($p > .05$), but the main effect of temporal order of TAF was significant, $F(1, 19) = 51.875$, $p < .001$.

EXPERIMENT 4

Methods

Participants

Eight university students (two women and six men, 19–27 years old, mean age = 24 years) participated in this experiment.

Stimuli and Apparatus

A Knowles Electronic Manikin for Acoustic Research (KEMAR) was located at the center of the anechoic chamber that is described for Experiment 1. Two types of sounds were used in this experiment: normal-order speech (which were created and reproduced as in Experiment 1) and time-reversed speech, and they were the same as used in Experiment 3. Similar to Experiment 3, speech analog signals were delivered to two loudspeakers (Dynaudio Acoustics, BM6A) in the frontal azimuthal plane at the left and right 45° positions with respect to the median plane. For a recording trial, the two loudspeakers presented identical speech (either normal order or time reversed), with the left loudspeaker lagging behind the right loudspeaker by either 2 or 40 msec.

The loudspeaker height was 140 cm, which was at the ear level of KEMAR, and the distance between the loudspeaker and the center of the KEMAR was 200 cm. Sound waves were recorded using the KEMAR, which was equipped with the ear simulators (RA0045, G.R.A.S, Sound and Vibration, Holte, Denmark), a programmable front-end (BEQ II.0, Head-acoustic, Herzogenrath, Germany), and a sound-progressed software (ArtemiS 6lefveve(o)1519.4(t)07.8(K)05.2(c7-187(c7 functional image acquisition for coregistration and normalization of functional images.

fMRI Data Analyses

Under each stimulus condition, the first image in response to speech stimuli and both the first and the last images for the silence baseline were omitted from analyses. The remaining 228 scans for each participant were preprocessed using SPM2 software (http://www.fil.ion.ucl.ac.uk). The preprocessing of the functional images included realignment (rigid-body transformation),u-1.2019rmtoransti04.1(n)19.8gd
he198(a)10.6d8 m(o)189(va)10.6(e)0mees on ics9ai04.1sciti04.1oin dy,erlan (c)10.6eoi04.1sc-
rcw abe-n an-adcTe m.eser-e tt-
m(al)189(izded)-450.7(t)0.1od therecuthe
using trns(fomzatioes)-604.8(aed)-564.9esooct40.1(hdr)-62053(w-3.6ith.)-382.2(a)]TJ 0 -1.1
stical an eiltn0e4f.1(m)17.3puleenthneer
(e).8(re)-41976(m).6oel ttiat
tf f24.1oi76.6upr.m47.2(u)(lui76.6se)-254.4(c31.7(on76.6die54.1tie54.1on-)]TJ)220-3880 TD 0.0331i
baem47.2eli31.7ne-1143,.e47.2v(e47.2r)50.6ys11.3ede47.2cdlyeli31.7n e47.2,.
evai54.1(se)107(d/)40o-mnse(c)-011.2d-01(e)1476lny vhes)-2918.4(w)11.1(e)0(r)58.1(e)-045.9cta1976lclt
e r a - 9 0 . 7 (t -) 1 1 .
c01on-t-c01atrstnatie04.1(m-)1.1a1.201tce
ib15.5gh-passe (f24.5(i52.5al3024ted60.1rhdr)-335.1at40.1ta)-35(2d60.156.)-32846(se60.1(c)-360
r c 1 3 8 8 6
ehresr-283(holdl)-


15 20 25 30

from statistics if their absolute values were less than 0.0014. The results of the statistics show that the ratio of the mean from those associated with perceptual fusion of time-reversed speech. The results of the fMRI experiment (Experiment 4) of this study for the first time show that compared with normal-order speech without perceptual fusion (when the lead/lag delay was 40 msec), perceptually fused normal-order speech (when the lead/lag delay was 2 msec) was associated with increased BOLD contrast activations in both the right ACC, which is involved in top–down attentional modulation of auditory processing (Crottaz-Herbette & Menon, 2006), and the left middle temporal gyrus, which is involved in both phonemic perception (Liebenthal, Binder, Spitzer, Possing, & Medler, 2005) and semantic processing (Friederici, 2002). However, compared with time-reversed speech without perceptual fusion, perceptually fused time-reversed speech was associated with increased BOLD contrast activations only in the left superior frontal gyrus.

Compared with perceptually fused time-reversed speech, perceptually fused normal-order speech was associated with more BOLD activations in the attention-control-related right ACC (BA 32), the right inferior parietal lobule (BA 40), which is a cortical area in the "where" pathway processing auditory spatial information (Wang, Wu, & Li, 2008; Arnott, Binns, Grady, & Alain, 2004; Alain, Arnot, Hevenor, Graham, & Grady, 2001; Weeks et al., 1999), and the left anterior part of the superior temporal gyrus (BA 38), which is involved in processing speech signals for comprehension (Hickok & Poeppel, 2007; Binder et al., 1997, 2000; Scott, Blank, Rosen, & Wise, 2000). These patterns of cortical activations one of the loudspeakers. When the absolute value of intermasker interval was 32 msec or shorter, there was consistent evidence of release from speech masking for target speech recognition. If the masker became speech-spectrum noise, significant release of target speech occurred only at a few short intermasker intervals less than 4 msec. Thus, the release of target speech from speech masking over a range of intermasker intervals between 4 and 32 msec cannot be explained by a reduction in energetic masking, and perceptual integration of the leading and lagging speech maskers must play a role in reducing informational masking of target speech. Although in the Rakerd et al. study the time-reversed speech was not used, on the basis of both the results of the present study and the results of the Freyman et al. (2001) study, it can be expected that if the masker becomes time-reversed speech under the experimental conditions used in the Rakerd et al. study, the maximum intermasker interval for significantly releasing target speech must be shorter than 32 msec because time-reversed speech has both smaller perceptual fusion tendency and smaller informational-masking effectiveness than normal-order speech.

This fusion-enhancing mechanism underlying the improvement of speech recognition in noisy, reverberant environments may reflect a more general cross-time perceptual "grouping" strategy of the brain for unmasking speech signals. As summarized by Moore (2003), temporal integration of acoustic signals over time, which improves detection and discrimination under masking conditions, does not involve a simple summation or accumulation process but is based on the formation of the internal representation of acoustic inputs. Because temporal integration is also based on temporal storage of acoustic details (Huang, Huang, et al., 2009; Huang, Wu, & Li, 2009; Li, Huang, Wu, Qi, & Schneider, 2009; Huang, Kong, Fan, Wu, & Li, 2008), the fusion-tendency enhancement specifically induced by the normal-order speech TAF may be due to an expansion of the temporal storage of the internal representation of speech signals when top–down modulations of the speech-processing cortical areas are elicited.

Acknowledgment

This study was supported by the National Natural Science Foundation of China (30670704; 30711120563; 60535030), the "973" National Basic Research Program of China (2009CB320901), and "985" grants from Peking University.

Reprint requests should be sent to Liang Li, Department of Psychology, Peking University, Beijing, 100871, China, or via e-mail: liangli@pku.edu.cn.

REFERENCES

Alain, C., Arnot, S. R., Hevenor, S., Graham, S., & Grady, C. (2001). "What" and "where" in the human auditory system. *Proceedings of the National Academy of Sciences, U.S.A., 98,* 12301–12306.

Arbogast, T. L., Mason, C. R., & Kidd, G. J. R. (2002). The effect of spatial separation on informational and energetic masking of speech. *Journal of the Acoustical Society of America, 112,* 2086–2098.

Arbogast, T. L., Mason, C. R., & Kidd, G. J. R. (2005). The effect of spatial separation on informational masking of speech in normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America, 117,* 2169–2180.

Arnott, S. R., Binns, M. A., Grady, C. L., & Alain, C. (2004). Assessing the auditory dual-pathway model in humans. *Neuroimage, 22,* 401–408.

Binder, J. R., Frost, J. A., Hammeke, T. A., Bellgowan, P. S. F., Springer, J. A., Kaufman, J. N., et al. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex, 10,* 512–528.

Binder, J. R., Frost, J. A., Hammeke, T. A., Cox, R. W., Rao, S. M., & Prieto, T. (1997). Human brain language areas identified by functional magnetic resonance imaging. *Journal of Neuroscience, 17,* 353–362.

Brett, M., Johnsrude, I. S., & Owen, A. M. (2002). The problem of functional localization in the human brain. *Nature Reviews Neuroscience, 3,* 243–250.

Brungart, D. S., Simpson, B. D., & Freyman, R. L. (2005). Precedence-based speech segregation in a virtual auditory environment. *Journal of the Acoustical Society of America, 118,* 3241–3251.

Cherry, E. C., & Taylor, W. K. (1954). Some further experiments upon the recognition of speech with one and with two ears. *Journal of the Acoustical Society of America, 26,* 554–559.

Crottaz-Herbette, S., & Menon, V. (2006). Where and when the anterior cingulate cortex modulates attentional response: Combined fMRI and ERP evidence. *Journal of Cognitive Neuroscience, 18,* 766–780.

Drullman, R., Festen, J. M., & Plomp, R. (1994). Effect of temporal envelope smearing on speech reception. *Journal of the Acoustical Society of America, 95,* 1053–1064.

Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2001). Spatial release from informational masking in speech recognition. *Journal of the Acoustical Society of America, 109,* 2112–2122.

Freyman, R. L., Balakrishnan, U., & Helfer, K. S. (2004). Effect of number of masking talkers and auditory priming on informational masking in speech recognition. *Journal of the Acoustical Society of America, 115,* 2246–2256.

Freyman, R. L., Clifton, R. K., & Litovsky, R. Y. (1991). Dynamic processes in the precedence effect. *Journal of the Acoustical Society of America, 90,* 874–884.

Freyman, R. L., Helfer, K. S., McCall, D. D., & Clifton, R. K. (1999). The role of perceived spatial separation in the unmasking of speech. *Journal of the Acoustical Society of America, 106,* 3578–3588.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6,* 78–84.

Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Huang, Y., Huang, Q., Chen, X., Qu, T. S., Wu, X. H., & Li, L. (2008). Perceptual integration between target speech and target-speech reflection reduces masking for target-speech recognition in younger adults and older adults. *Hearing Research, 244,* 51–65.

Huang, Y., Huang, Q., Chen, X., Wu, X. H., & Li, L. (2009). Transient auditory storage of acoustic details is associated with release of speech from informational masking in reverberant conditions. *Journal of Experimental Psychology: Human Perception and Performance, 35,* 1618–1628.

Huang, Y., Kong, L.-Z., Fan, S.-L., Wu, X.-H., & Li, L. (2008). Both frequency and interaural delay affect ERP responses to binaural gap. *NeuroReport, 19,* 1673–1678.

Huang, Y., Wu, X. H., & Li, L. (2009). Detection of the break in interaural correlation is affected by interaural delay, aging, and center frequency. *Journal of the Acoustical Society of America, 126,* 300–309.

Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, C. S., Rainey, L., et al. (2000). Automated Talairach atlas labels for functional brain mapping. *Human Brain Mapping, 10,* 120–131.

Leff, A. P., Schofield, T. M., Stephan, K. E., Crinion, J. T., Friston, K. J., & Price, C. J. (2008). The cortical dynamics of intelligible speech. *Journal of Neuroscience, 28,* 13209–13215.

Levitt, H. (1971). Transformed up–down methods in psychoacoustics. *Journal of the Acoustical Society of America, 49,* 467–477.

Li, L., Daneman, M., Qi, J. G., & Schneider, B. A. (2004). Does the information content of an irrelevant source differentially affect speech recognition in younger and older adults? *Journal of Experimental Psychology: Human Perception and Performance, 30,* 1077–1091.

Li, L., Huang, J., Wu, X. H., Qi, J. G., & Schneider, B. (2009). The effects of aging and interaural delay on the detection of a break in the interaural correlation between two sounds. *Ear and Hearing, 30,* 273–286.

Li, L., Qi, J. G., He, Y., Alain, C., & Schneider, B. (2005). Attribute capture in the precedence effect for long-duration noise sounds. *Hearing Research, 202,* 235–247.

Liebenthal, E., Binder, J. R., Spitzer, S. M., Possing, E. T., & Medler, D. A. (2005). Neural substrates of phonemic perception. *Cerebral Cortex, 15,* 1621–1631.

Litovsky, R. Y., Colburn, H. S., Yost, W. A., & Guzman, S. J. (1999). The precedence effect. *Journal of the Acoustical Society of America, 106,* 1633–1654.

Lochner, J. P. A., & Burger, J. F. (1958). The subjective masking of short time delayed echoes, their primary sounds, and their contribution to the intelligibility of speech. *Acustica, 8,* 1–10.

Moore, B. C. J. (2003). Temporal integration and context effects in hearing. *Journal of Phonetics, 31,* 563–574.

Oppenheim, A. V., Schafer, R. W., & Buck, J. R. (1999). *Discrete-time signal processing*. Upper Saddle River, NJ: Prentice-Hall Press.

Rakerd, B., Aaronson, N. L., & Hartmann, W. M. (2006). Release from speech-on-speech masking by adding a delayed masker at a different location. *Journal of the Acoustical Society of America, 119,* 1597–1605.

Rakerd, B., & Hartmann, W. M. (1986). Localization of sound in rooms: III. Onset and duration effects. *Journal of the Acoustical Society of America, 80,* 1695–1706.

Rakerd, B., Hartmann, W. M., & Hsu, J. (2000). Echo suppression in the horizontal and median sagittal planes. *Journal of the Acoustical Society of America, 107,* 1061–1064.

Rosen, S. (1992). Temporal information in speech—Acoustic, auditory and linguistic aspects. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 336,* 367–373.

Schubert, E. D., & Wernick, J. (1969). Envelope versus microstructure in the fusion of dichotic signals. *Journal of the Acoustical Society of America, 45,* 1525–1531.

Scott, S. K., Blank, C. C., Rosen, S., & Wise, R. J. S. (2000). Identification of a pathway for intelligible speech in the left temporal lobe. *Brain, 123,* 2400–2406.

Smith, Z. M., D0Teutt, n&-315(DOxenham)-331.13(A)-3128J. S(20002). MChima