



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at ScienceDirect

## Neuroscience and Biobehavioral Reviews

journal homepage: [www.elsevier.com/locate/neubiorev](http://www.elsevier.com/locate/neubiorev)

## Review

## Auditory frequency-following response: A neurophysiological measure for studying the “cocktail-party problem”

Yi Du, Lingzhi Kong, Qian Wang, Xihong Wu, Liang Li\*

Department of Psychology, Speech and Hearing Research Center, Key Laboratory on Machine Perception (Ministry of Education), Peking University, Beijing 100871, China

## ARTICLE INFO

## Article history:

Received 1 September 2010

Received in revised form 12 May 2011

Accepted 19 May 2011

## Keywords:

Binaural interaction

Cocktail-party problem

Frequency-following responses

Inferior colliculus

Lateral nucleus of the amygdala

Perceptual cues

Selective attention

Unmasking of speech

Phase locking

Auditory aging

## ABSTRACT

How do we recognize what one person is saying when others are speaking at the same time? The “cocktail-party problem” proposed by Cherry (1953) has puzzled scientific societies for half a century. This puzzle will not be solved without using appropriate neurophysiological investigation that should satisfy the following four essential requirements: (1) certain critical speech characteristics related to speech intelligibility are recorded; (2) neural responses to different speech sources are differentiated; (3) neural correlates of bottom-up binaural unmasking of responses to target speech are measurable; (4) neural correlates of attentional top-down unmasking of target speech are measurable. Before speech signals reach the cerebral cortex, some critical acoustic features are represented in subcortical structures by the frequency-following responses (FFRs), which are sustained evoked potentials based on precisely phase-locked responses of neuron populations to low-to-middle-frequency periodical acoustical stimuli. This review summarizes previous studies on FFRs associated with each of the four requirements and suggests that FFRs are useful for studying the “cocktail-party problem”.

© 2011 Elsevier Ltd. All rights reserved.

## Contents

|   |      |
|---|------|
| 1. Introduction.....  | 2047 |
| 1.1. What is the “cocktail-party problem”? .....  | 2047 |
| 1.2. The four requirements for electrophysiological investigation of the “cocktail-party problem” ..... | 2047 |
| 1.3. The frequency-following responses.....   | 2047 |
| 2. Basic characteristics of FFRs .....  | 2048 |
| 2.1. Subcortical origins of FFRs .....  | 2048 |
| 2.2. Neural phase locking induces FFRs .....  | 2048 |
| 3. Representation of critical speech characteristics in FFRs .....                                      | 2048 |
| 4. FFRs are useful for studying the “cocktail-party problem”.....                                       | 2049 |
| 4.1. Stimulus selectivity of FFRs under multiple-source conditions .....                                | 2049 |
| 4.2. Binaural unmasking of FFRs .....   | 2049 |
| 4.3. Mechanisms underlying bottom-up binaural unmasking of IC FFRs .....                                | 2051 |
| 4.4. Attentional top-down modulation of FFRs .....  | 2052 |
| 4.5. Experience-dependent plasticity of FFRs .....  | 2053 |
| 4.6. Aging effects on speech recognition in “cocktail-party” situations and FFRs .....                  | 2053 |
| 5. Summary and future studies.....  | 2054 |
| Acknowledgments .....   | 2054 |
| References .....  | 2054 |

*Abbreviations:* A1, primary auditory cortex; ABR, auditory brainstem response; BMLD, binaural masking level difference; CN, cochlear nucleus; DNLL, dorsal nucleus of the lateral lemniscus; FFRs, frequency-following responses; F0, fundamental frequency; fMRI, functional magnetic resonance imaging; IC, inferior colliculus; ITD, interaural time difference; LA, lateral nucleus of the amygdala; LL, lateral lemniscus; SMR, signal-to-masker ratio.

\* Corresponding author. Tel.: +86 10 6278 5419; fax: +86 10 6276 1081.

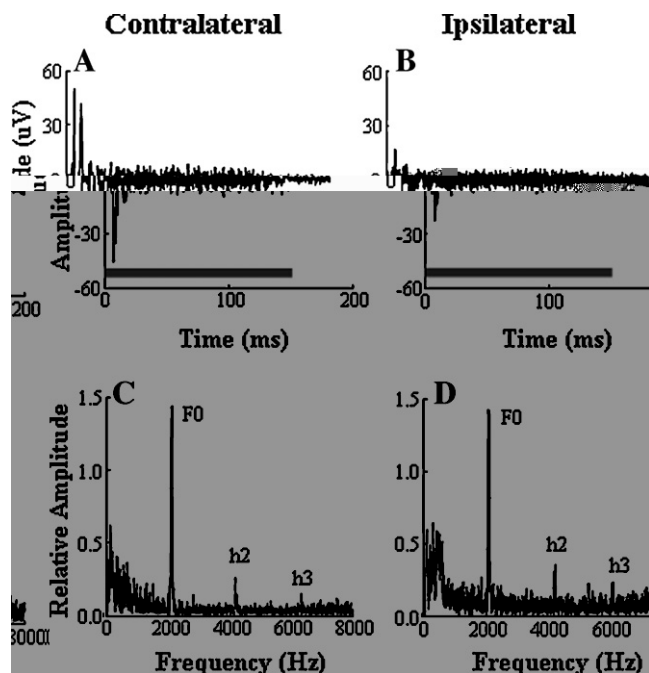
E-mail address: [liangli@pku.edu.cn](mailto:liangli@pku.edu.cn) (L. Li).

## 1. Introduction

### 1.1. What is the “cocktail-party problem”?

In a noisy, multiple-people-talking condition, listeners with normal hearing can still recognize and understand the attended speech and simultaneously ignore background noise and irrelevant speech stimuli. How do we recognize what one person is saying when others are speaking at the same time? This *cocktail-party problem*, first proposed by Cherry (1953), has puzzled the societies of psychology, neurophysiology, signal processing, and computer engineering for half a century. It reflects human's remarkable ability to selectively detect, locate, discriminate, and identify individual speech sources in noisy, multiple-people-talking conditions. More specifically, listeners can use various cues available to facilitate their attention to target speech and follow the target stream against irrelevant-speech influences. These cues include precedence-effect-induced spatial separation between the target image and the masker image (e.g., Freyman et al., 1999; Huang et al., 2008, 2009a; Li et al., 2004; Rakerd et al., 2006; Wu et al., 2005), prior knowledge about where and/or when target speech will occur (Best et al., 2008; Kidd et al., 2005a), knowledge/familiarity of the target-talker's voice (Brungart et al., 2001; Helfer and Freyman, 2009; Huang et al., 2010; Newman and Evers, 2007; Yang et al., 2007), prior knowledge about the topic of the target sentence (Helfer and Freyman, 2008), and viewing a speaker's movements of the speech articulators (Grant and Seitz, 2000; Helfer and Freyman, 2005; Rosenblum et al., 1996; Rudmann et al., 2003; Sumbly and Pollack, 1954; Summerfield, 1979, 1992). It appears that any perceptual or cognitive cue that facilitates listeners' selective attention to target speech can improve recognition of target speech against competing speech. Among these cues, both the effect of voice

recent progresses in investigating the neural mechanisms underlying



**Fig. 1.** Typical response waveforms to the chatter presented at the contralateral ear (panel A) or the ipsilateral ear (panel B) and the correspondent fast-Fourier spectral analyses (panels C and D) of FFRs recorded in the inferior colliculus (IC). Note that the recording site contralateral to the stimulated ear (panel A) exhibits a much larger onset evoked potential than the site ipsilateral to the stimulated ear (panel B), but contralateral FFRs and ipsilateral FFRs exhibit similar F0 and h2 amplitudes. The horizontal bar in panels A and B represents the duration of the chatter stimulus. (From Du et al., 2009b).

In particular, FFRs preserve spectral peaks corresponding to a few formants of steady-state vowel-like sounds (Krishnan, 1999, 2002; Russo et al., 2004), time-varying consonant-like sounds and the formant transition (Krishnan and Parkinson, 2000; Plyler and Ananthanarayan, 2001; Song et al., in press). Pitch-relevant information is also preserved in the phase-locked neural activity that generates FFRs not only for steady-state complex tones (Greenberg et al., 1987) but also for lexical tones such as Mandarin syllables with time-varying pitch contours (Krishnan et al., 2004, 2005, 2009; Xu et al., 2006). Moreover, FFRs can track time-varying pitch prosody (Russo et al., 2008) and convey emotional status of complex speech sounds (Strait et al., 2009). Using the 40-ms /da/ syllable to elicit brainstem responses, Kraus and co-workers in a series of studies have demonstrated how transiently responding components and sustained FFRs separately encode sources and filter characteristics of speech signals in representing paralinguistic and linguistic information (for reviews see Johnson et al., 2005; Kraus and Nicol, 2005).

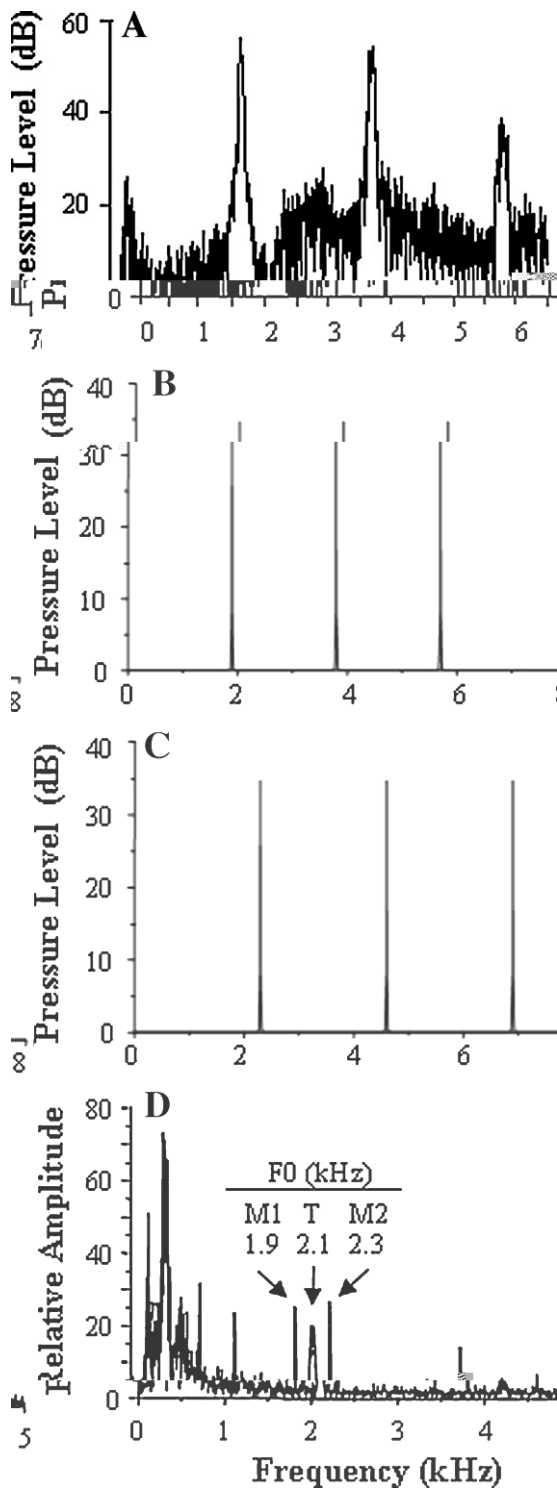
FFRs to speech-like stimuli were also investigated in rats. Du et al. (2009a,b) have found that the F0 component (2.1 kHz) of vowel-like rat tail-pain chatter elicits FFRs in all recorded sites in the IC (Fig. 1) and the LA, the h2 component (4.2 kHz) elicits FFRs in all recorded sites in the IC but 22 out of the 51 recorded sites in the LA, and the h3 component (6.3 kHz) barely elicits FFRs in the two structures.

#### 4. FFRs are useful for studying the “cocktail-party problem”

##### 4.1. Stimulus selectivity of FFRs under multiple-source conditions

The second critical requirement for FFRs to be useful for studying the “cocktail-party problem” is that when a target speech and a masker are presented at the same time with a considerably low signal-to-masker ratio (SMR), FFRs to the target speech should be

clearly differentiated from those to the masker. Russo et al. (2004) recorded brainstem responses to the syllable /da/ and found that both the transient component and the sustained component (FFRs) of the brainstem responses to the speech syllable can be reliably obtained with high test-retest stability and low variability across listeners. More importantly, FFRs to the harmonics of the syllable, particularly F0 and F1, are much more resistant to the deleterious effects of background noise than the transient responses to the syllable. Since encoding of the F0 and F1 is important for both recognizing the speech content and identifying the speaker and voice emotion, the robustness of the neural representation of the F0 and F1 components in FFRs allows FFRs to be useful for investigating the neural mechanisms underlying how speech recognition is achieved under masking conditions. Li and Jeng (2011) recently reported that the frequency error, slope error, and tracking accuracy of FFRs to the Mandarin syllable

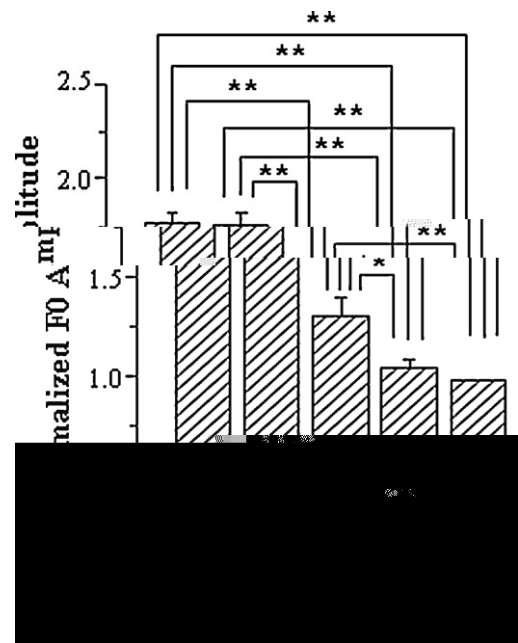


**Fig. 2.** Panels A, B, and C show the spectra of the rat's pain call ( $F_0=2.1$  kHz,  $h_2=4.2$  kHz,  $h_3=6.3$  kHz), tone-complex masker 1 (M1: 1.9, 3.8, and 5.7 kHz), and tone-complex masker 2 (M2: 2.3, 4.6, and 6.9 kHz), respectively. Panel D shows FFRs recorded in a rat's IC to the diotically presented mixture of the three stimuli with the signal-to-masker ratio (SMR) at each ear being 0 dB. Obviously, FFR components to the  $F_0$ s of the three stimuli can be distinguished and some low-frequency missing fundamentals occur in the FFRs.

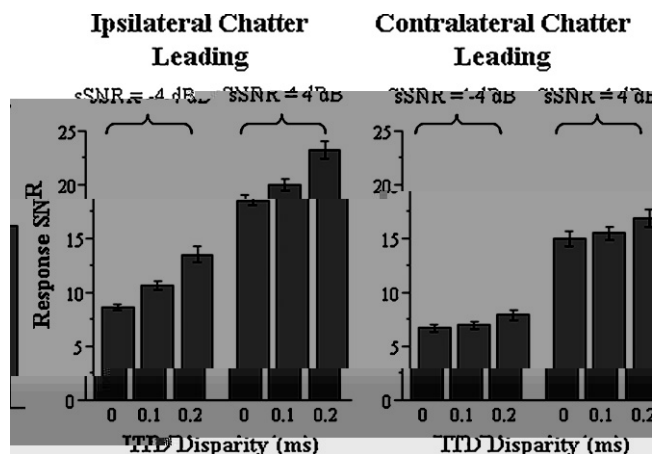
cessing (Wilson and Krishnan, 2005). The binaural masking level difference (BMLD) is a well-studied psychophysical phenomenon showing that the signal, which is presented at both ears and masked by a noise masker presented at both ears, becomes more detectable when either the interaural phase of the signal or that of the masker is reversed (Hirsh, 1948). Thus, the BMLD measures the ability of listeners to use a difference between signal and masker in binaural attributes to improve their detection of the signal against the masking noise. In the Wilson and Krishnan study (2005), the FFR amplitudes to the noise-masked 500-Hz tone bursts under antiphasic conditions ( $S\pi$ No or  $SoN\pi$ , with a  $180^\circ$  interaural phase delay between the tone signal and noise masker) were substantially larger than those under homophasic conditions (SoNo).

One of the advantages of intracranially recorded FFRs is that FFRs of a particular brain structure can be recorded and differentiated from those recorded from other structures. This structural resolution cannot be achieved by human scalp-recorded FFRs, especially those recorded by the electrode in the vertex. Binaural properties of FFRs in the rat IC were investigated by Du et al. (2009b). The results of the Du et al. study have shown that although the rat's pain call (the chatter) presented at the contralateral ear evokes much larger transient onset responses than the chatter presented at the ipsilateral ear (Fig. 1A and B), the spectral amplitude of FFRs to the contralateral chatter is similar to that to the ipsilateral chatter (Fig. 1C and D). Moreover, IC FFRs to binaural chatter stimulation exhibit a feature of ipsilateral predominance: FFRs are markedly stronger when the ipsilateral chatter either leads or starts simultaneously with the contralateral chatter than when the ipsilateral chatter lags behind the contralateral chatter (Fig. 3).

More importantly, under noise masking conditions, FFRs to the chatter signal are markedly improved by introducing an ITD disparity between the signal and the white-noise masker when FFRs are recorded in either the rat IC (Du et al., 2009b) (Fig. 4) or the



**Fig. 3.** Mean normalized  $F_0$  spectral amplitudes in IC FFRs under various monaural and binaural stimulation conditions.  $F_0$  amplitude evoked by contralateral stimulation only (C) served as the baseline condition (amplitude=1) for amplitude normalization. Error bars represent the standard error of the mean (SEM). I/C, binaural stimulation with ipsilateral (relative to recording site) chatter leading contralateral one; ST, simultaneous binaural stimulation; C/I, contralateral chatter leading ipsilateral; I, chatter at ipsilateral ear only; C, chatter at contralateral ear only. \*\* $P<0.01$ , \* $P<0.05$ , repeated-measures ANOVA. (From Du et al., 2009b).



**Fig. 4.** Relative response signal-to-noise ratios (response SNRs) of IC FFRs when the chatter was co-presented with white noise with different ITD disparities ( $ITD_{S+N}$ ). Response SNRs were presented separately for conditions when ipsilateral chatter led contralateral one (left) and conditions when contralateral chatter led ipsilateral one (right). Numbers associated with each bar represent the ITD disparity ( $ITD_{S+N}$ ) value in ms. sSNR: stimulus signal-to-noise ratio. (From Du et al., 2009b with modifications).

LA (Du et al., 2009a). Fig. 4 shows relative response signal-to-noise ratios of IC FFRs when the ipsilateral chatter leads (left panel) or lags behind (right panel) the contralateral one and is co-presented with the noise masker with different ITD disparities. Note that either a 0.1-ms or 0.2-ms ITD disparity between signal and masker is sufficient to enhance the synchrony of phase-locked encoding of signal in the IC.

The results of the animal studies (Du et al., 2009a,b) are generally in agreement with the notion that introducing a difference between signal and masker in binaural configurations improves auditory representations of the signal, as proved by previous reports on binaural/spatial unmasking of single-unit auditory responses in the IC of laboratory animals (e.g., Caird et al., 1991; Jiang et al., 1997; Lane and Delgutte, 2005; Lin and Feng, 2003; Mandava et al., 1996; McAlpine et al., 1996; Palmer et al., 2000; Ratnam and Feng, 1998) and previous reports on binaural unmasking of brainstem FFRs in humans (Wilson and Krishnan, 2005).

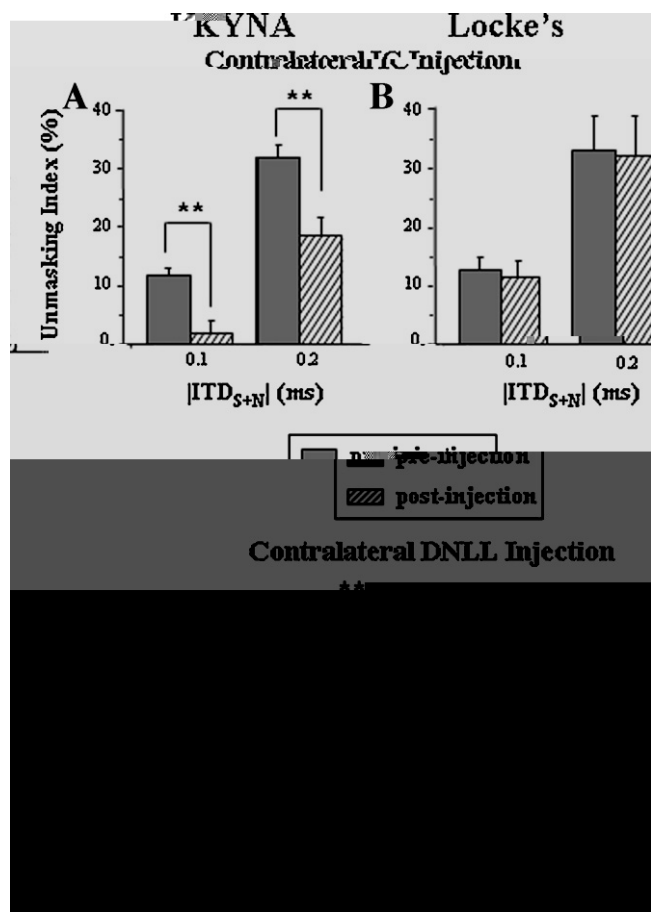
It is of interest to know whether the binaural unmasking of FFRs recorded in the rat's IC shares similar mechanisms with the BMLD as measured in the IC of other species. The BMLD has been demonstrated on single neurons in both the guinea pig's IC (e.g., Caird et al., 1991; Jiang et al., 1997; McAlpine et al., 1996; Palmer and Shackleton, 2002; Palmer et al., 1999, 2000) and the chinchilla's IC (Mandava et al., 1996). In general, the BMLD is considered as a low-frequency phenomenon, because its value has been found efficient when the frequency of the signal is below 1–2 kHz (e.g., Caird et al., 1991; Hirsh, 1948; Mandava et al., 1996). In the Du et al. studies (2009a,b), the  $F_0$  of the chatter was above 2 kHz, suggesting that measurements of binaural unmasking based on synchronized FFRs of a population of neurons exhibit some features that have not been revealed in measurement of BMLD based on single-unit firing counting. Since FFRs to binaural stimulation are ITD dependent, different populations of IC neurons contribute to FFRs differently under different binaural configurations. In other words, when the signal ITD is different from the masker ITD, some IC neurons are driven only by the signal but not by the noise masker, leading to an improvement in FFRs. This population-disparity strategy for unmasking FFRs may be similar to that for BMLD.

However, considering that Lane and Delgutte (2005) have reported that signal-masker spatial separation improves only the population thresholds but not necessarily the single-unit thresholds of IC responses to the noise-masked signal in cats, analyses of

FFRs (which are based on synchronized activities of a population of neurons) in various species are more advantageous than counting numbers of single-unit action potentials in estimating binaural unmasking of IC responses. Particularly, investigation of binaural unmasking of IC FFRs in laboratory animals helps understanding the reports that human brainstem FFRs are both resistant to noise masking (Li and Jeng, 2011; Russo et al., 2004) and unmasked by binaural processing (Wilson and Krishnan, 2005). Binaural unmasking of IC FFRs may also be associated with the benefit in processing target signals by precedence-effect-induced perceived spatial separation between signal and masker (e.g., Freyman et al., 1999; Huang et al., 2008; Li et al., 2004; Wu et al., 2005).

#### 4.3. Mechanisms underlying bottom-up binaural unmasking of IC FFRs

In the rat IC, the majority of auditory neurons are predominantly excited by stimuli at the



**Fig. 5.** Effects of blocking the contralateral IC or the contralateral DNLL with kynurenic acid (KYNA) on binaural unmasking of FFRs when the ipsilateral chatter leads the contralateral one. Unmasking indices (UIs) of FFRs under different ITD disparities are shown before (shaded bars) and after (hatched bars) injection of either KYNA (panels A and C) or Locke's solution (panels B and D) into the contralateral IC (panels A and B) or the contralateral DNLL (panels C and D). Note that the blockade of glutamate receptors in either structure significantly reduced UIs under either 0.1-ms or 0.2-ms ITD disparity between chatter and noise. \* $P < 0.05$ , \*\* $P < 0.01$ , paired-samples  $t$ -tests. (From Du et al., 2009b).

review see Li and Yue, 2002). Clearly, ipsilateral stimulation drives EE neurons in the recorded IC, as well as all the types of neurons in the contralateral DNLL. It has been confirmed that the contralateral DNLL plays a role in suppressing IC FFRs in quiet because IC FFRs were enhanced by blocking the contralateral DNLL when no masker is presented (Du et al., 2009b; Ping et al., 2008). However, when the masker is presented and the ipsilateral chatter leads the contralateral one, binaural unmasking of IC FFRs is significantly reduced by blocking excitatory glutamate transmissions in the contralateral DNLL (Fig. 5C), suggesting that GABAergic projections from the contralateral DNLL play a role in binaurally unmasking IC FFRs.

It has been well known that GABAergic inhibitory inputs to the IC shape binaural responses of individual IC neurons (Burger and Pollak, 2001; Kelly and Li, 1997; Kidd and Kelly, 1996; Li and Kelly, 1992; Van Adel et al., 1999). Also, Lin and Feng (2003) have reported that iontophoretic application of bicuculline, a GABA<sub>A</sub> receptor antagonist, into the frog IC markedly degraded binaural processing involved in spatial unmasking of the IC. Thus, ipsilateral stimulation (relative to the recorded IC) drives the contralateral DNLL, which not only inhibits IC FFRs but also facilitates binaural unmasking of IC FFRs. The unmasking effect may be caused by the

function of the DNLL in both facilitation of binaural responses to the signal and suppression of responses to the noise masker. Some studies (e.g., Klug et al., 2002; Xie et al., 2005) have shown that in the free-tailed bat IC, the neural selectivity to species-specific calls is primarily attributed to local GABAergic inhibition. Thus, the interruption of GABAergic innervations from the contralateral DNLL may also disrupt the response selectivity of IC neurons to the tail-pain chatter, leading to the reduction of FFRs to the chatter against noise masking.

Since both enhancement of signal inputs and suppression of masker inputs can improve the response signal-to-noise ratio in neural representation of acoustic stimuli, the functional integration of excitatory inputs from the contralateral IC and inhibitory inputs from the contralateral DNLL is a critical issue for future studies of binaural unmasking of FFRs.

#### 4.4. Attentional top-down modulation of FFRs

Under “cocktail-party” conditions, listeners with normal hearing are still able to take advantage of certain perceptual/cognitive cues to facilitate their selective attention to target speech and follow the target stream against masker influences. Thus, to determine whether FFRs are useful for studying the “cocktail-party problem”, it is necessary to investigate whether FFRs can be modulated by selective attention.

Auditory selective attention refers to the mental ability to resist distracters and select relevant information from acoustic events (for a review, see Fritz et al., 2007a). In spite of the extensive research on attentional effects at cortical level, the neural basis of top-down attentional control of auditory processing at lower levels such as the auditory brainstem and cochlea is still less investigated. With respect to the brainstem level, a number of early studies recording ABR elicited by brief acoustic clicks have yielded negative results on either within-modal or cross-modal attentional effects (i.e., Picton and Hillyard, 1974; Picton et al., 1981; Woods and Hillyard, 1978). However, a recent functional magnetic resonance imaging (fMRI) study by Rinne et al. (2008) has shown that when a strictly controlled selective-listening paradigm requiring highly focused selective attention throughout the experiment is applied, human IC activation is significantly modulated by auditory selective attention and this modulation depends on where in space attention is directed. The study suggests that auditory processing in the IC is not solely stimulus driven but is also top-down modulated according to behavioral tasks.

Studies of FFRs evoked by pure tones and complex auditory stimuli such as speech syllables have also shown the marked attentional effect on both the FFR amplitude (Galbraith and Arroyo, 1993; Galbraith and Doan, 1995; Galbraith et al., 1998, 2003) and latency (Hoormann et al., 1994, 2000, 2004). For example, Galbraith et al. (2003) have shown that FFR amplitudes are substantially larger when participants direct attention towards evoking tones within the auditory modality than attend visual stimuli. Galbraith et al. (1998) have also shown that FFR amplitudes to the F0 of each vowel are significantly larger when that vowel was attended than ignored. Since the F0 is perceptually salient and also conveys paralinguistic information such as the identity of the speaker, it is conferred that the early attentional effect of evoked activities in human auditory brainstem may differentiate the processing of task-relevant/irrelevant stimuli based on salient paralinguistic cues. Moreover, Hoormann et al. (2000) have shown that significant attentional effects on FFR latency occur when a monotic paired-stimuli paradigm is used, in which the first stimulus serves as the reference for the second one, while no attentional effects are present in a dichotic paradigm with sustained attention to one ear. The authors therefore concluded that auditory attentional effects on brainstem FFRs are evident mainly in unimodal situations with



unilateral stimuli, when attention is highly focused to a restricted time interval to cope with a difficult task.

The primary auditory cortex (A1) is the main cortical source for providing auditory signals to other cortical regions and fore-brain subcortical structures. By measuring regional cerebral blood flows (Hugdahl et al., 2000; OLeary et al., 1997), hemodynamic responses (Jancke et al., 1999; Krumbholz et al., 2007), neuromagnetic fields (Fujiwara et al., 1998; Poghosyan and Ioannides, 2008), or intracranial electrophysiological activities (Bidet-Caulet et al., 2007), studies using human participants suggest that the A1 is involved in auditory attention. Electrophysiological studies using laboratory animals have also shown that the A1 is important for mediating attention in rats (Jaramillo and Zador, 2011; Polley et al., 2006), ferrets (Fritz et al., 2007b), and cats (Lee and Middlebrooks, 2011). Moreover,

older-adult group compared with the younger-adult group, suggesting that the ability of neurons at the brainstem level to phase lock to the components of the stimulus is reduced for older adults. Also, another recent study by Clinard et al. (2010) shows that FFRs recorded in adult participants declined with advancing age from 22 to 77 years old. Thus, the FFR is useful for investigating why older-adult listeners experience the difficulty of understanding speech in “cocktail-party” environments.

## 5. Summary and future studies

Both bottom-up auditory processes, such as binaural unmasking, and higher-level cognitive processes, such as selective attention and language experience, facilitate speech perception in cocktail-party environments. As reviewed in this article, FFRs encode certain critical speech features related to speech intelligibility and exhibit the marked selectivity to various sound sources. Under masking conditions, FFRs to target speech can be binaurally unmasked based on binaural processing in the auditory brainstem and top-down modulated based on selective attention as well. FFRs also exhibit both experience-related and age-related plasticity. Thus, both scalp-recorded FFRs in humans and intracranially recorded FFRs in laboratory animals are useful neurophysiological indices for investigating the “cocktail-party problem”. Here we propose three lines of studies in the future:

- (1) Under adverse listening conditions, human listeners can take advantage of various perceptual/cognitive cues to facilitate their selective attention to target speech against speech masking, leading to an increase of the intelligibility of keywords in target speech. We propose that under noisy conditions, the enhanced representation of target-speech signals in the auditory midbrain contributes to the “cocktail-party problem”. Supportive evidence has been recently reported by Song et al. (in press) that under the six-talker speech-masking condition, FFRs to the F0 during the formant transition of the syllable/da/are correlated with the performance of speech-in-noise (SIN) task. Thus, if the keywords are assigned with particular F0s that are distinctive from those of non-keywords in target speech and those of masking speech, FFRs specific to the keywords would become useful markers for studying how unmasking of target speech in human listeners are achieved by the cues.
- (2) In humans, selective attention to the stimulus enhances FFRs to the stimulus. However, related animal studies are not available in the literature. In the future, appropriate animal models for studying selective attention to acoustic stimuli will be established and FFRs will be recorded in awake laboratory animals under simulated “cocktail-party” conditions. Since the A1 directly mediates neural activities in the IC, the potential corticofugal modulation of FFRs in the IC via its direct projections should be investigated.
- (3) The age-related difficulties in speech recognition under complex listening situations may be due to both age-related bottom-up deficits at the sensory level, including reduced temporal and/or spectral selectivity, and age-related top-down deficits at the cognitive level, including declines in selective attention, working memory, inhibitory control, and general slowing. FFRs will be used in the future for further investigating the age-related bottom-up deficits and top-down deficits.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (30950030; 90920302; 60811140086), the “973”

National Basic Research Program of China (2009CB320901), the Chinese Ministry of Education (20090001110050), and “985” grants from Peking University.

## References

- Aiken, S.J., Picton, T.W., 2008. Envelope and spectral frequency-following responses to vowel sounds. *Hear. Res.* 245, 35–47.
- Akhoun, I., Gallego, S., Moulin, A., Menard, M., Veuille, E., Berger-Vachon, C., Collet, L., Thai-Van, H., 2008. The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme/ba/in normal-hearing adults. *Clin. Neurophysiol.* 119, 922–933.
- Baumann, O., Belin, P., 2010. Perceptual scaling of voice identity: common dimensions for different vowels and speakers. *Psychol. Res.* 74, 110–120.
- Best, V., Ozmeral, E.J., Kopco, N., Shinn-Cunningham, B.G., 2008. Object continuity enhances selective auditory attention. *Proc. Natl. Acad. Sci. U. S. A.* 105, 13174–13178.
- Bidet-Caulet, A., Fischer, C., Besle, J., Aguera, P.E., Giard, M.H., Bertrand, O., 2007. Effects of selective attention on the electrophysiological representation of concurrent sounds in the

- Galbraith, G.C., 1994. Two-channel brain-stem frequency-following responses to pure tone and missing fundamental stimuli. *Electroencephalogr. Clin. Neurophysiol.* 92, 321–330.
- Galbraith, G.C., Arroyo, C., 1993. Selective attention and brainstem frequency-following responses. *Biol. Psychol.* 37, 3–22.
- Galbraith, G.C., Doan, B.Q., 1995. Brainstem frequency-following and behavioral responses during selective attention to pure tone and missing fundamental stimuli. *Int. J. Psychophysiol.* 19, 203–214.
- Galbraith, G.C., Arbagey, P.W., Branski, R., Comerchi, N., Rector, P.M., 1995. Intelligible speech encoded in the human brain stem frequency-following response. *Neuroreport* 6, 2363–2367.
- Galbraith, G.C., Bhuta, S.M., Choate, A.K., Kitahara, J.M., Mullen Jr., T.A., 1998. Human brain stem frequency-following responses to dichotic vowel stimuli during attention. *Neuroreport* 9, 1889–1893.
- Galbraith, G.C., Threadgill, M.R., Hemsley, J., Salour, K., Songdej, N., Ton, J., Cheung, L., 2000. Putative measure of peripheral and brainstem frequency-following in humans. *Neurosci. Lett.* 292, 123–127.
- Galbraith, G.C., Olfman, D.M., Huffman, T.M., 2003. Selective attention affects human brain stem frequency-following response. *Neuroreport* 14, 1–5.
- Galbraith, G., Amaya, E.M., de Rivera, J.M., Donan, N.M., Duong, M.T., Hsu, J.N., Tran, K., Tsang, L.P., 2004. Brain stem evoked response to forward and reversed speech in humans. *Neuroreport* 15, 2057–2060.
- Garcia Lecumberri, M.L., Cooke, M., Cutler, A., 2010. Non-native speech perception in adverse conditions: a review. *Speech Commun.* 52, 864–886.
- Gardi, J., Merzenich, M., McKean, C., 1979. Origins of the scalp recorded frequency-following response in the cat. *Audiology* 18, 358–381.
- Gelfand, S.A., Ross, L., Miller, S., 1988. Sentence reception in noise from one versus two sources: effects of aging and hearing loss. *J. Acoust. Soc. Am.* 83, 248–256.
- Gilkey, R.H., Good, M.D., 1995. Effects of frequency on free-field masking. *Hum. Factors* 37, 835–843.
- Goldberg, J.M., Brownell, W.E., 1973. Discharge characteristics of neurons in the anteroventral and dorsal cochlear nuclei of cat. *Brain Res.* 64, 35–54.
- González-Hernández, T., Mantolán-Sarmiento, B., González-González, B., Pérez-González, H., 1996. Sources of GABAergic input to the inferior colliculus of the rat. *J. Comp. Neurol.* 372, 309–326.
- Grant, K.W., Seitz, P.F., 2000. The use of visible speech cues for improving auditory detection of spoken sentences. *J. Acoust. Soc. Am.* 108, 1197–1208.
- Greenberg, S., Marsh, J.T., Brown, W.S., Smith, J.C., 1987. Neural temporal coding of low pitch. I. Human frequency-following responses to complex tones. *Hear. Res.* 25, 91–114.
- Helfer, K.S., 1992. Aging and the binaural advantage in reverberation and noise. *J. Speech Hear. Res.* 35, 1394–1401.
- Helfer, K.S., Freyman, R.L., 2005. The role of visual speech cues in reducing energetic and informational masking. *J. Acoust. Soc. Am.* 117, 842–849.
- Helfer, K.S., Freyman, R.L., 2008. Aging and speech-on-speech masking. *Ear Hear.* 29, 87–98.
- Helfer, K.S., Freyman, R.L., 2009. Lexical and indexical cues in masking by competing speech. *J. Acoust. Soc. Am.* 125, 447–456.
- Helfer, K.S., Wilber, L.A., 1990. Hearing-loss, aging, and speech-perception in reverberation and noise. *J. Speech Hear. Res.* 33, 149–155.
- Herbert, H., Aschoff, A., Ostwald, J., 1991. Topography of projections from the auditory-cortex to the inferior colliculus in the rat. *J. Comp. Neurol.* 304, 103–122.
- Hernández, O., Rees, A., Malmierca, M.S., 2006. A GABAergic component in the commissure of the inferior colliculus in rat. *Neuroreport* 17, 1611–1614.
- Hickok, G., Poeppel, D., 2004. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* 92, 67–99.
- Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8, 393–402.
- Hine, J.E., Martin, R.L., Moore, D.R., 1994. Free-field binaural unmasking in ferrets. *Behav. Neurosci.* 108, 196–205.
- Hirsh, I.J., 1948. The influence of interaural phase on interaural summation and inhibition. *J. Acoust. Soc. Am.* 20, 536–544.
- Hoormann, J., Falkenstein, M., Hohnsbein, J., 1994. Effect of selective attention on the latency of human frequency-following potentials. *Neuroreport* 5, 1609–1612.
- Hoormann, J., Falkenstein, M., Hohnsbein, J., 2000. Early attention effects in human auditory-evoked potentials. *Psychophysiology* 37, 29–42.
- Hoormann, J., Falkenstein, M., Hohnsbein, J., 2004. Effects of spatial attention on the brain stem frequency-following potential. *Neuroreport* 15, 1539–1542.
- 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. *Hear. Res.* 244, 51–65.
- Huang, Y., Huang, Q., Chen, X., Wu, X.H., Li, L., 2009a. Transient auditory storage of acoustic details is associated with release of speech from informational masking in reverberant conditions. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 1618–1628.
- Huang, Y., Wu, X.H., Li, L., 2009b. Detection of the break in interaural correlation is affected by interaural delay, aging, and center frequency. *J. Acoust. Soc. Am.* 126, 300–309.
- Huang, Y., Xu, L.J., Wu, X.H., Li, L., 2010. The effect of voice cuing on releasing speech from informational masking disappears in older adults. *Ear Hear.* 31, 579–583.
- Hugdahl, K., Law, I., Kyllingsbaek, S., Bronnick, K., Gade, A., Paulson, O.B., 2000. Effects of attention on dichotic listening: an O-15-PET study. *Hum. Brain Mapp.* 10, 87–97.
- Irvine, D.R.F., 1986. The auditory brainstem: a review of the structure and function of auditory brainstem processing mechanisms. In: Autrum, H., Ottosen, D., Perl, E.R., Willis, W.D. (Eds.), *Progress in Sensory Physiology*, vol. 7. Springer-Verlag, Berlin, pp. 142–144.
- Jancke, L., Mirzazade, S., Shah, N.J., 1999. Attention modulates activity in the primary and the secondary auditory cortex: a functional magnetic resonance imaging study in human subjects. *Neurosci. Lett.* 266, 125–128.
- Jaramillo, S., Zador, A.M., 2011. The auditory cortex mediates the perceptual effects of acoustic temporal expectation. *Nat. Neurosci.* 14, 246–251.
- Jiang, D., McAlpine, D., Palmer, A.R., 1997. Detectability index measures of binaural masking level difference across populations of inferior colliculus neurons. *J. Neurosci.* 17, 9331–9339.
- Johnson, D.H., 1980. The relationship between spike rate and synchrony in responses of auditory nerve fibers to single tones. *J. Acoust. Soc. Am.* 68, 1115–1122.
- Johnson, K.L., Nicol, T.G., Kraus, N., 2005. Brain stem response to speech: a biological marker of auditory processing. *Ear Hear.* 26, 424–434.
- Johnson, K.L., Nicol, T., Zecker, S.G., Kraus, N., 2008. Developmental plasticity in the human auditory brainstem. *J. Neurosci.* 28, 4000–4007.
- Jourdan, D., Ardid, D., Chapuy, E., Eschaliier, A., Le Bars, D., 1995. Audible and ultrasonic vocalization elicited by single electrical nociceptive stimuli to the tail in the rat. *Pain* 63, 237–249.
- Keilson, S.E., Richards, V.M., Wyman, B.T., Young, E.D., 1997. The representation of concurrent vowels in the cat anesthetized ventral cochlear nucleus: evidence for a periodicity-tagged spectral representation. *J. Acoust. Soc. Am.* 102, 1056–1071.
- Kelly, J.B., Glenn, S.L., Beaver, C.J., 1991. Sound frequency and binaural response properties of single neurons in rat inferior colliculus. *Hear. Res.* 56, 273–280.
- Kelly, J.B., Li, L., 1997. Two sources of inhibition affecting binaural evoked responses in the rat's inferior colliculus: the dorsal nucleus of the lateral lemniscus and the superior olivary complex. *Hear. Res.* 104, 112–126.
- Kidd, G., Arbogast, T.L., Mason, C.R., Gallun, F.J., 2005a. The advantage of knowing where to listen. *J. Acoust. Soc. Am.* 118, 3804–3815.
- Kidd, G., Mason, C.R., Brughera, A., Hartmann, W.M., 2005b. The role of reverberation in release from masking due to spatial separation of sources for speech identification. *Acust. Acta. Acust.* 91, 526–536.
- Kidd, S.A., Kelly, S.A., 1996. Contribution of the dorsal nucleus of the lateral lemniscus to binaural responses in the inferior colliculus of the rat: interaural time delays. *J. Neurosci.* 16, 7390–7397.
- Klug, A., Bauer, E.E., Hanson, J.T., Hurley, L., Meitzen, J., Pollak, G.D., 2002. Response selectivity for species-specific calls in the inferior colliculus of Mexican free-tailed bats is generated by inhibition. *J. Neurophysiol.* 88, 1941–1954.
- Kraus, N., Nicol, T., 2005. Brainstem origins for cortical “what” and “where” pathways in the auditory system. *Trends Neurosci.* 28, 176–181.
- Kraus, N., Banai, M., 2009. The auditory brainstem: a review of the structure and function of auditory brainstem processing mechanisms. In: Autrum, H., Ottosen, D., Perl, E.R., Willis, W.D. (Eds.), *Progress in Sensory Physiology*, vol. 7. Springer-Verlag, Berlin, pp. 142–144.

- Lin, W.Y., Feng, A.S., 2003. GABA is involved in spatial unmasking in the frog auditory midbrain. *J. Neurosci.* 23, 8143–8151.
- Liu, L.F., Palmer, A.R., Wallace, M.N., 2006. Phase-locked responses to pure tones in the inferior colliculus. *J. Neurophysiol.* 95, 1926–1935.
- Liu, X.-P., Yan, Y.-C., Wang, Y.-L., Yan, J., 2010. Corticofugal modulation of initial neural processing of sound information from the ipsilateral ear in the mouse. *PLoS ONE* 5, e14038.
- Luo, F., Wang, Q.-Z., Kashani, A., Yan, J., 2008. Corticofugal modulation of initial sound processing in the brain. *J. Neurosci.* 28, 11615–11621.
- Malmierca, M.S., Hernández, O., Falconi, A., Lopez-Poveda, E.A., Merchan, M., Rees, A., 2003. The commissure of the inferior colliculus shapes frequency response areas in rat: an in vivo study using reversible blockade with microinjection of kynurenic acid. *Exp. Brain Res.* 153, 522–529.
- Malmierca, M.S., Hernández, O., Rees, A., 2005. Intercollicular commissural projections modulate neuronal responses in the inferior colliculus. *Eur. J. Neurosci.* 21, 2701–2710.
- Malmierca, M.S., Hernández, O., Antunes, F.M., Rees, A., 2009. Divergent and point-to-point connections in the commissural pathway between the inferior colliculi. *J. Comp. Neurol.* 514, 226–239.
- Mandava, P., Rupert, A.L., Moushegian, G., 1996. Inferior colliculus neuronal responses to masking-level-difference stimuli. *Hear. Res.* 99, 79–84.
- Marsh, J.T., Worden, F.G., 1968. Sound evoked frequency-following responses in the central auditory pathway. *Laryngoscope* 78, 1149–1163.
- Marsh, J.T., Worden, F.G., Smith, J.C., 1970. Auditory frequency-following response: neural or artifact? *Science* 169, 1222–1223.
- Marsh, J.T., Brown, W.S., Smith, J.C., 1974. Differential brainstem pathways for the conduction of auditory frequency-following responses. *J. Neurophysiol.* 37, 1149–1163.

- Wallace, M.N., Anderson, L.A., Palmer, A.R., 2007. Phase-locked responses to pure tones in the auditory thalamus. *J. Neurophysiol.* 98, 1941–1952.
- Werff, K.R.V., Burns, K.S., 2011. Brain stem responses to speech in younger and older adults. *Ear Hear.* 32, 168–180.
- Wilson, J.R., Krishnan, A., 2005. Human frequency-following responses to binaural masking level difference stimuli. *J. Am. Acad. Audiol.* 16, 184–195.
- Winter, I.M., Palmer, A.R., 1990. Responses of single units in the anteroventral cochlear nucleus of the guinea pig. *Hear. Res.* 44, 161–178.
- Wong, P.C., Skoe, E., Russo, N.M., Dees, T., Kraus, N., 2007. Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nat. Neurosci.* 10, 420–422.
- Woods, D.L., Hillyard, S.A., 1978. Attention at the cocktail party: Brainstem evoked responses reveal no peripheral gating. In: Otto, D.A. (Ed.), *Multidisciplinary Perspectives in Event-related Brain Potential Research*. U.S. Government Publishing Office, Washington, DC, pp. 230–233.
- Worden, F.G., Marsh, J.T., 1968. Frequency-following (microphonic-like) neural responses evoked by sound. *Electroencephalograp. Clin. Neurophysiol.* 25, 42–52.
- Wu, X.H., Wang, C., Chen, J., Qu, H.W., Li, W.R., Wu, Y.H., Schneider, B.A., Li, L., 2005. The effect of perceived spatial separation on informational masking of Chinese speech. *Hear. Res.* 199, 1–10.
- Xie, R., Meitzen, J., Pollak, G.D., 2005. Differing roles of inhibition in hierarchical processing of species-specific calls in auditory brainstem nuclei. *J. Neurophysiol.* 94, 4019–4037.
- Xu, Y., Krishnan, A., Gandour, J.T., 2006. Specificity of experience-dependent pitch representation in the brainstem. *Neuroreport* 17, 1601–1605.
- Yan, J., Ehret, G., 2002. Corticofugal modulation of midbrain sound processing in the house mouse. *Eur. J. Neurosci.* 16, 119–128.
- Yan, J., Zhang, Y.F., Ehret, G., 2005. Corticofugal shaping of frequency tuning curves in the central nucleus of the inferior colliculus of mice. *J. Neurophysiol.* 93, 71–83.
- Yang, Z.G., Chen, J., Wu, X.H., Wu, Y.H., Schneider, B.A., Li, L., 2007. The effect of voice cuing on releasing Chinese speech from informational masking. *Speech Commun.* 49, 892–904.
- Yonan, C.A., Sommers, M.S., 2000. The effects of talker familiarity on spoken word identification in younger and older listeners. *Psychol. Aging* 15, 88–99.
- Zhang, D.X., Li, L., Wu, S.H., Kelly, J.B., 1998. GABAergic projection from the lateral lemniscus to the inferior colliculus of the rat. *Hear. Res.* 117, 1–12.
- Zurek, P.M., 1993. Binaural advantages and directional effects in speech intelligibility. In: Studebaker, G.A., Hochberg, I. (Eds.), *Acoustical Factors Affecting Hearing Aid Performance*. Allyn and Bacon, Boston, pp. 255–276.