ISSN 0149-7634

Volume 35, Issue 10, November 2011

Neuroscience
& Biobehavioral
Reviews

SPECIAL ISSUE
Wired for Sound

GUEST EDITORS
Jun Yan,
Liang Li

EDITOR-IN-CHIEF
Verity J. Brown

Neuroscience and Biobehavioral Reviews 35 (2011) 2046–2057

Contents lists available at ScienceDirect

Neuroscience and Biobehavioral Reviews

journal homepage: 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”.

Contents

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 (L. Li).

0149-7634/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neubiorev.2011.05.008

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; Sumby 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 differences between the target speaker and masking speakers and the effect of differences in spatial location between target speech and maskers on intelligibility of target speech have been extensively studied.

Human speech, which contains rapidly varying spectrotemporal features, is represented in the central nervous system with a hierarchically organized manner at different processing stages (Hickok and Poeppel, 2004, 2007). To perceptually separate target speech from other disruptive speech inputs (i.e., to parse the merged acoustic waveform generated by multiple speech sources into auditory streams), the auditory system should be able to differentiate target-speech signals from those of other irrelevant sources (Bronkhorst, 2000). In other words, the central auditory system needs to precisely maintain certain information about acoustic details of the speech sources before the stream segregation is achieved. And then, based on the signal representation, both bottom-up (stimulus-driven) and top-down (task-driven) processes are applied to achieve the auditory scene analysis (Bregman, 1990).

Human speech sounds and animal vocalization sounds are generally harmonic or quasi-harmonic: frequencies of the spectral components of these sounds are approximately integer multiples of a common low frequency known as the fundamental frequency (F0). Harmonic sounds are perceived as a single auditory object rather than several concurrent pure tone images. Also, the harmonically related tones induce a pitch that usually corresponds to the F0 and can be perceived even when spectral energy at the F0 of the complex is not present. Since processing F0s of speech sounds is associated with both perceptual grouping of harmonically related components across frequency and time (Brokx and Nooteboom, 1982) and facilitation of speaker identification (Baumann and Belin, 2010), the F0 is critical for speech perception in noisy environments. Indeed, the inharmonicity can provide a cue for segregating concurrent independent sound sources (for a review see Micheyl and Oxenham, 2010) and listeners feel it easier to recognize two concurrent speech sounds when the difference in F0 between the speakers' voices becomes larger (Culling and Darwin, 1993; Du et al., 2011).

On the other hand, it is also known that speech recognition under "cocktail-party" listening conditions is remarkably improved when there is a spatial separation between the target-speech source and interfering-sound sources (for a review see Schneider et al., 2007), a phenomenon that is generally called spatial unmasking. Spatial unmasking partially results from two types of bottom-up processes, i.e., head shadowing (that improves the signal-to-noise ratio at the ear closer to the target) and binaural interaction induced by the disparity between the target and masker in the interaural time difference (ITD, e.g., Shinn-Cunningham et al., 2005; Zurek, 1993). Moreover, spatial unmasking can result from top-down processes by facilitating selectively spatial attention to the target (Freyman et al., 1999, 2001; Huang et al., 2008, 2009a; Kidd et al., 2005b; Li et al., 2004; Rakerd et al., 2006; Wu et al., 2005).

1.2. The four requirements for electrophysiological investigation of the "cocktail-party problem"

Electrophysiological investigation is critical for understanding the neural mechanisms underlying how the F0 of target speech is selected and recognized, and how the processing of target speech is enhanced by binaural integration under "cocktail-party" listening environments. In this review, we propose that to non-intrusively investigate the neurophysiological mechanisms underlying the "cocktail-party problem" in healthy human listeners, at least the following four essential requirements should be satisfied: First, the neurophysiological measure should be able to encode certain critical speech characteristics related to speech intelligibility. In addition, the neurophysiological measure associated with neural responses to the target speech can be reliably distinguished from those to irrelevant stimuli (speech or non-speech). Moreover, under masking conditions, bottom-up binaural unmasking of responses to target speech can be revealed with the neurophysiological measure. Finally, the neurophysiological measure can be used for studying attentional top-down unmasking of target speech.

1.3. The frequency-following responses

Before speech signals reach the cerebral cortex, some critical acoustic properties of speech stimuli are represented in subcortical auditory structures with considerably temporal and/or spectral precision as revealed by the human scalp-recorded frequency-following responses (FFRs) (e.g., Aiken and Picton, 2008; Akhoun et al., 2008; Johnson et al., 2005; Kraus and Nicol, 2005; Krishnan, 1999, 2002; Krishnan and Gandour, 2009; Krishnan and Parkinson, 2000; Russo et al., 2004). FFRs are sustained electrical potentials based on precisely phase-locked responses of neuron populations to low-to-middle-frequency periodical acoustical stimuli (Moushegian et al., 1973; Worden and Marsh, 1968). Thus, the issues on (1) whether FFRs are sufficient to encode certain speech characteristics related to speech intelligibility, (2) whether FFRs to target speech can be differentiated from those to concurrent maskers, (3) whether FFRs to speech stimuli can be binaurally unmasked, and (4) whether FFRs to speech stimuli can be top-down modulated, are all critical for determining whether the FFR is appropriate for studying the "cocktail-party problem".

In this review, we first describe some basic characteristics of scalp-recorded FFRs to speech stimuli in humans and those of intracranially recorded FFRs to vowel-like stimuli in laboratory rats. And then, we summarize both the stimulus selectivity and the noise-resistant trait of FFRs. We also summarize bottom-up binaural unmasking of FFRs in both humans and rats and the recent progresses in investigating the neural mechanisms underlying binaural unmasking of FFRs. Subsequently, since subcortical auditory functions dynamically interact with higher-level cognitive processes (for reviews, see Chandrasekaran and Kraus, 2010; Krishnan and Gandour, 2009; Suga et al., 2002; Suga, 2008), we also review studies of top-down modulation of FFRs by selective attention and experience-dependent plasticity. Finally, we describe the relationship between investigation of auditory aging and that of the "cocktail-party problem", and mention recent studies of FFRs recorded in older adults. We conclude that the FFR to speech is not just a neural "snapshot" of the speech signal, but can be both bottom-up and top-down modulated, making it useful for studying the "cocktail-party problem".

2. Basic characteristics of FFRs

2.1. Subcortical origins of FFRs

In humans, scalp-recorded auditory-brainstem responses (ABRs) to complex sounds such as consonant-vowel speech syllables consist of both transient-onset and sustained-FFR components (e.g., Aiken and Picton, 2008; Akhoun et al., 2008; Johnson et al., 2005; Kraus and Nicol, 2005; Krishnan, 1999, 2002; Krishnan and Gandour, 2009; Krishnan and Parkinson, 2000; Russo et al., 2004; Song et al., in press). Sustained FFRs are characterized by periodic waveforms that follow (synchronize to) periodicities of low-to-middle-frequency sounds, representing temporal structures of harmonic sounds. It has been generally agreed that human scalp-recorded FFRs reflect phase-locked activities in a population of neural elements in the rostral brainstem with an upper limit of frequency around 1000 Hz (Gardi et al., 1979; Stillman et al., 1978). Although it is difficult to find the exact neural generators of scalp-recorded FFRs in humans, several lines of evidence including results from ablation/cooling studies and those from developmental studies suggest a brainstem origin including the inferior colliculus (IC), the lateral lemniscus (LL), and the cochlear nucleus (CN) (for a recent review, see Chandrasekaran and Kraus, 2010). Especially, the IC is regarded as the major neural source of scalp-recorded FFRs. For example, Smith et al. (1975) induced a selective amplitude reduction of the scalp-recorded FFRs in cats following a cryogenic treatment of the IC and regained the original amplitude when the IC was warmed. In humans, scalp-recorded FFRs are absent in participants with lesions confined to the IC (Sohmer et al., 1977). Although Gardi et al. (1979) reported that ablating the CN caused a large reduction (50%) in the amplitude of the scalp-recorded FFRs in the cat, the contradiction is reconciled by the fact that FFRs recorded with the vertical montage that accentuates more rostral brainstem structures (i.e., IC and/or LL) are different from those recorded with the horizontal montage that reflects more peripheral contributions (i.e., auditory nerve and/or CN) (Davis and Britt, 1984; Galbraith, 1994; Galbraith et al., 2000; Møller et al., 1988; Stillman et al., 1978).

Although the human scalp-recorded FFRs provide a non-invasive manner for revealing potential brainstem mechanisms, only intracranial recordings of FFRs in laboratory animals' brainstem structures provide incontrovertible understanding of the nature of FFRs. Around the 1970s, several pioneering intracranial FFR studies were conducted along the ascending auditory pathway in cats with the purpose of ruling out the possibility of cochlear and cortical origins and determining the brainstem generators of FFRs (Marsh and Worden, 1968; Marsh et al., 1970, 1974; Faingold and Caspary, 1979). Recent direct recordings from the IC in rats evidently show robust FFRs to the rat's vowel-like pain call (Du et al., 2009b). Interestingly, both the rat's pain call and tone complex can also elicit vigorous FFRs in the lateral nucleus of the amygdala (LA) (Du et al., 2009a).

2.2. Neural phase locking induces FFRs

Across a variety of species, the upper limit of frequency for phase locking decreases as it ascends the recorded sites in the auditory pathway (Langner, 1992). In the auditory nerve, the upper limit of neural phase locking varies from 3.5 kHz in guinea pigs to over 5 kHz in cats and squirrel monkeys (Johnson, 1980; Palmer and Russell, 1986; Rose et al., 1967). In the ventral CN of guinea pigs, cells can phase lock up to 2–3.5 kHz depending on the neural population (Winter and Palmer, 1990). In the cat dorsal CN, phase locking is limited to frequencies less than 1.5 kHz (Goldberg and Brownell, 1973). While the guinea pig IC, which is regarded as the major neural generator of scalp-recorded FFRs, contains a large proportion (68%) of neurons with phase locking responses, especially in the central nucleus (Liu et al., 2006). Considerable variability also exists about the upper limit of phase-locking frequency in different parts of the IC (Kuwada et al., 1984; Liu et al., 2006). In the cat medial geniculate body, a small proportion of cells ($\sim$2% of units) can phase lock to tones up to 1.5 kHz (Rouiller et al., 1979). While in the medial geniculate body of the guinea pig, the upper limit of phase-locking frequency varies across anatomical divisions from 520 to 1100 Hz (Wallace et al., 2007). At the auditory cortex, neurons are capable of phase locking up to about 250 Hz in anesthetized guinea pigs (Wallace et al., 2005) and 100 Hz in awake monkeys (Steinschneider et al., 2008).

It should be noted that although FFRs are based on phase locking of individual neurons, the upper limit of frequency for phase locking of individual neurons should not confound with that for FFRs of neuron populations. Based on the "volley theory", which proposes that a population of auditory nerve fibers with phase-locked firing at sub-multiples of the stimulating frequency can produce a composite discharge pattern to temporally represent the stimulus (Boudreau, 1965), FFRs are capable of encoding frequencies much higher than the upper limit of phase-locking frequency of individual neurons. For example, the recent study by Ping et al. (2008) has shown that intracranial FFRs recorded in the rat IC can be elicited by presenting pure tone bursts with frequencies of the range from 225 to 4025 Hz. Moreover, one audible and vowel-like component of the rat's vocal responses to tail pain has been called "chatter" and is characterized by an F0 just above 2.0 kHz plus several harmonics (Jourdan et al., 1995). Using this behaviorally relevant call, Du et al. (2009b) found that FFRs to the chatter recorded in the rat IC contain both the F0 (2.1 kHz) and h2 (4.2 kHz) components in all of the 42 rats used (Fig. 1), and even the h3 (6.3 kHz) component in 7 rats, indicating the collective phase-locking effect based on the combination of firings of a neuron population.

3. Representation of critical speech characteristics in FFRs

To determine whether FFRs can be used for investigating the "cocktail-party problem", the first requirement is that FFRs should be able to encode certain critical speech characteristics that are related to speech intelligibility. Indeed, in humans, scalp-recorded FFRs represent some crucial characteristics of speech (e.g., Aiken and Picton, 2008; Akhoun et al., 2008; Johnson et al., 2005; Kraus and Nicol, 2005; Krishnan, 1999, 2002; Krishnan and Gandour, 2009; Krishnan and Parkinson, 2000; Russo et al., 2004). When FFRs elicited by words are "transferred back" and played as acoustical stimuli to human listeners with normal hearing, the listeners are able to correctly identify the words with a marked accuracy (Galbraith et al., 1995), indicating that the acoustic signals associated with speech intelligibility are well represented within FFRs.

More specifically, several studies have indicated that human scalp-recorded FFRs show robust representation of F0 and higher harmonics of speech sounds (e.g., Greenberg et al., 1987; Krishnan et al., 2004, 2005, 2009; Russo et al., 2004; Xu et al., 2006).

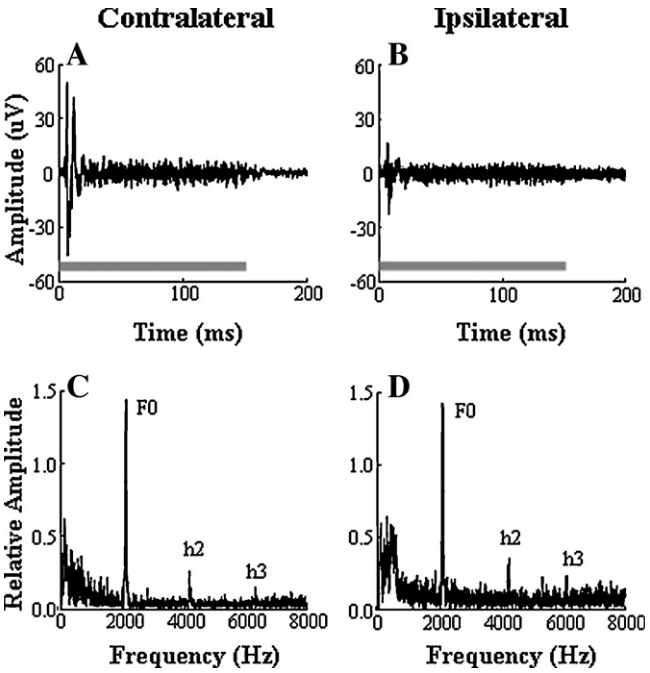

Fig. 1. Typical response waveforms to the chatter presented at the contralateral ear (panel A) or the ipsilateral ear (panel B) and the correspondent fast-Fourier spectral analyses (panels C and D) of FFRs recorded in the inferior colliculus (IC). Note that the recording site contralateral to the stimulated 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 /yi/ with the rising tone remain relatively stable until the signal-to-noise ratio is reduced to 0 dB or lower. The signal-to-noise ratio turning point around 0 dB suggests that the intensity of target stimulus token is recommended to be at least equal to that of the background noise if proper audibility of the pitch is to be ensured.

To further demonstrate the selectivity of FFRs to various periodical-stimulus sources, we diotically presented the mixture of the rat’s pain call (Fig. 2A) with two maskers (M1 and M2). Each of the maskers is a three-tone-harmonic complex (M1: 1.9, 3.8, and 5.7 kHz; M2: 2.3, 4.6, and 6.9 kHz) (Fig. 2B and C) with the SMR of 0 dB at each ear. Fig. 2D shows the spectra of FFRs recorded in the rat IC to the stimulus mixture. Clearly, FFR components to the F0s of the pain call, M1, and M2 can be distinguished. Interestingly, the FFRs to the mixture of the pain call and maskers also contain several low-frequency beats induced by interactions between the tone components.

It is known that neural responses of the auditory system to complex tones undergo a major transformation at the level of the CN. Single units of the auditory nerve and primary-like neurons of the CN in anesthetized guinea pigs (Palmer et al., 1986) and single units of primary-like and chopper neurons of the ventral CN in anesthetized cats (Keilson et al., 1997) generally display synchronized responses to individual components in concurrent vowels or harmonic complexes, but both single chopper neurons of the CN and single neurons of the IC in anesthetized chinchilla exhibit little or no synchronized responses to individual components in harmonic complexes (Sinex, 2008; Sinex and Li, 2007). Thus, one of the advantages of the FFR-recording method that surpass the single-unit recording method in the target specificity is that the FFRs specific to certain components of the target-speech stimulus (e.g., the F0 of the target) can be sufficiently distinguished from those of the co-presented maskers, as long as the target and maskers are different in F0. This unique nature makes FFRs very useful for investigating either bottom-up or top-down modulations of neural responses to target speech (see below).

4.2. Binaural unmasking of FFRs

Spatial unmasking of target stimuli largely depends on binaural processing (e.g., Shinn-Cunningham et al., 2005; Zurek, 1993). Thus, investigation of the brainstem mechanisms underlying binaural unmasking of target speech is critical for understanding the bottom-up processes that enhance the neural representation of the target speech under a typical cocktail-party condition where the target is spatially separated from maskers. Both binaural unmasking and spatial unmasking can be demonstrated in both humans (Gilkey and Good, 1995; Saberi et al., 1991; Shinn-Cunningham et al., 2001) and animals (Dent et al., 1997; Hine et al., 1994). Particularly related to the subject of this review, it has been reported that human brainstem FFRs can be unmasked by binaural processing (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 (SπNo or SoNπ, with a 180° 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

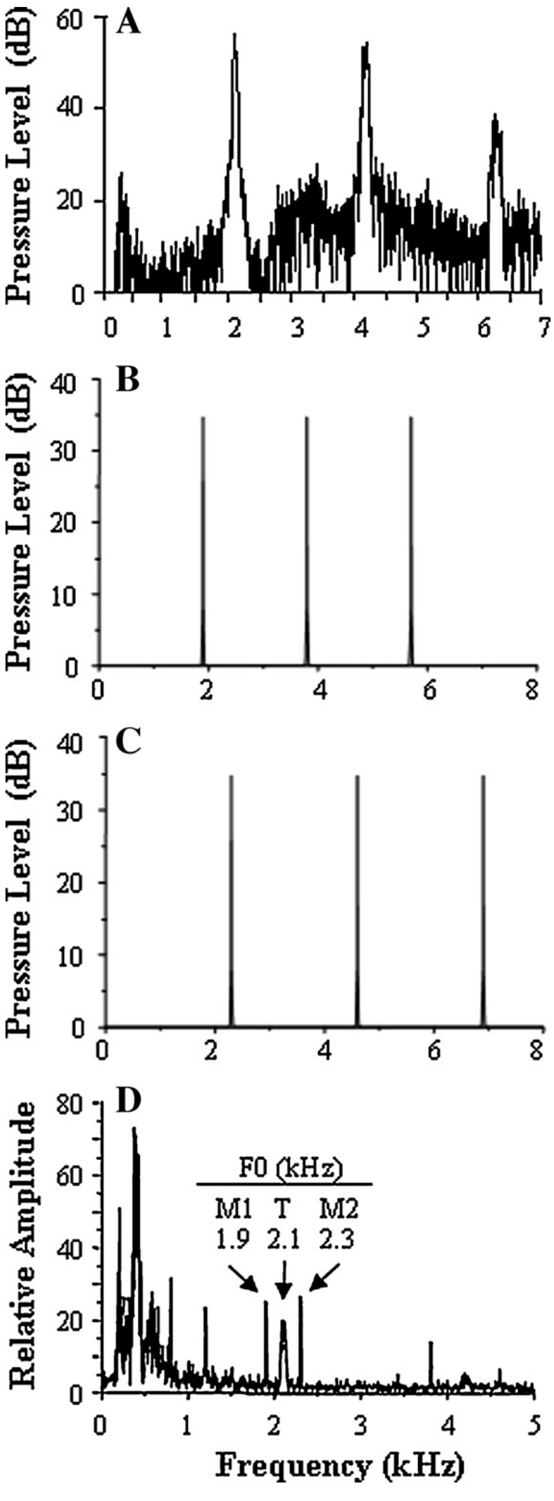

Fig. 2. Panels A, B, and C show the spectra of the rat's pain call (F0 = 2.1 kHz, h2 = 4.2 kHz, h3 = 6.3 kHz), tone-complex masker 1 (M1: 1.9, 3.8, and 5.7 kHz), and tone-complex masker 2 (M2: 2.3, 4.6, and 6.9 kHz), respectively. Panel D shows FFRs recorded in a rat's IC to the diotically presented mixture of the three stimuli with the signal-to-masker ratio (SMR) at each ear being 0 dB. Obviously, FFR components to the F0s of the three stimuli can be distinguished and some low-frequency missing fundamentals occur in the FFRs.

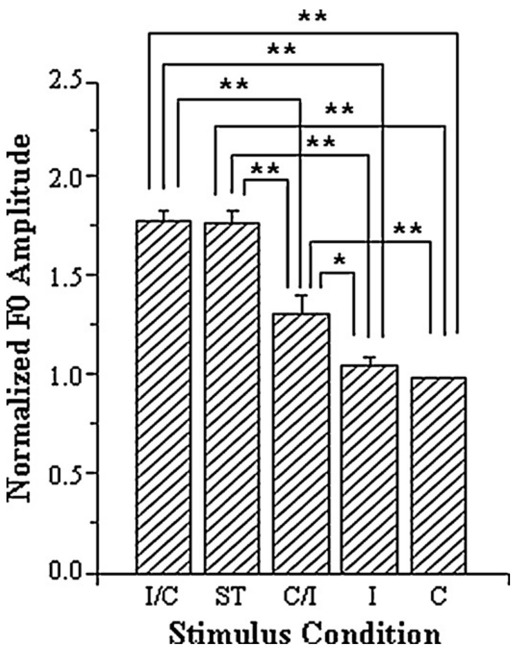

Fig. 3. Mean normalized F0 spectral amplitudes in IC FFRs under various monaural and binaural stimulation conditions. F0 amplitude evoked by contralateral stimulation only (C) served as the baseline condition (amplitude = 1) for amplitude normalization. Error bars represent the standard error of the 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).

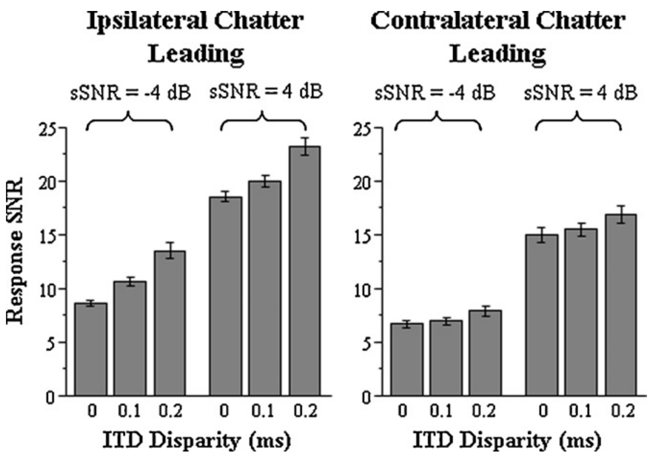

Fig. 4. Relative response signal-to-noise ratios (response SNRs) of IC FFRs when the chatter was co-presented with white noise with different ITD disparities ($|ITD_{S+N}|$). Response SNRs were presented separately for conditions when ipsilateral chatter led contralateral one (left) and conditions when 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 below1-2 kHz (e.g., Caird et al., 1991; Hirsh, 1948; Mandava et al., 1996). In the Du et al. studies (2009a,b), the F0 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 contralateral ear and inhibited by stimuli at the ipsilateral ear, forming the so-called "EI" neurons, and a small portion (about 20%) of neurons are excited by stimuli at either ear, forming the so-called "EE" neurons which are sensitive to ITD (Kelly et al., 1991). It is well known that the IC receives crossed axonal projections from its counterpart, the contralateral IC (Irvine, 1986; González-Hernández et al., 1996; Hernández et al., 2006; Saint Marie, 1996; Zhang et al., 1998), with both divergent and point-to-point wiring patterns (Malmierca et al., 2009). The intercollicular commissure plays a role in modulating both binaural responses and frequency-response areas in the IC (Malmierca et al., 2003, 2005). On the other hand, binaural responses in the IC can also be shaped by GABAergic axonal projections from the contralateral dorsal nucleus of the lateral lemniscus (DNLL) (Burger and Pollak, 2001; Faingold et al., 1993; Kelly and Li, 1997; Kidd and Kelly, 1996; Li and Kelly, 1992; Van Adel et al., 1999; Zhang et al., 1998; for a review see Li and Yue, 2002). As mentioned above, binaural FFRs recorded in the rat IC exhibit a marked ipsilateral predominance (see Fig. 3). Since only EE neurons in the IC exhibit excitatory responses to ipsilateral stimulation, IC EE neurons play the major role in inducing binaural IC FFRs. Also, since stimulation at the ear ipsilateral to the recorded IC activates the contralateral IC and the contralateral DNLL, the ipsilaterally driven IC FFRs must be modulated by projections from the contralateral IC and those from the contralateral DNLL.

For inputs from the contralateral IC, although the existence of a GABAergic projection through the commissure of IC has been described (González-Hernández et al., 1996; Hernández et al., 2006), non-GABAergic projections (Zhang et al., 1998) and strong glutamatergic projections (Saint Marie, 1996) have also been confirmed. Particularly, Malmierca et al. (2005) have reported that auditory responses in the rat IC to either monaural or binaural stimulation are affected by commissural blockade. The Du et al. study (2009b) verifies that the intercollicular connection makes a contribution to the formation of IC FFRs in rats. It is suggested that ipsilateral stimulation drives not only EE neurons in the recorded IC but also EE, EI and EO neurons in the contralateral IC, which, in turn, further activate EE neurons in the recorded IC. In other words, the input from the contralateral IC is one of the sources forming IC FFRs driven by ipsilateral stimulation. The reduction of binaural unmasking of IC FFRs following the chemical blockade of the contralateral IC (Fig. 5A) is due to the reduction of the response signal-to-noise ratio. It would be of interest to know whether the intercollicular connection also contributes to human brainstem FFRs.

On the other hand, IC neurons receive inhibitory (GABAergic) influence from the contralateral DNLL (Burger and Pollak, 2001; Faingold et al., 1993; Kelly and Li, 1997; Kidd and Kelly, 1996; Li and Kelly, 1992; Van Adel et al., 1999; Zhang et al., 1998; for a

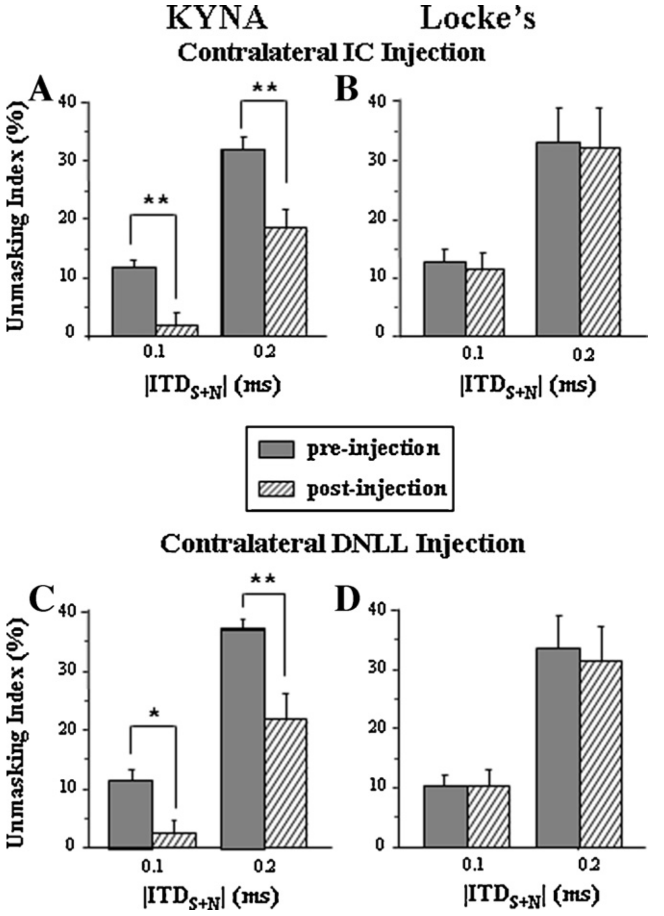

Fig. 5. Effects of blocking the contralateral IC or the contralateral DNLL with kynurenic acid (KYNA) on binaural unmasking of FFRs when the ipsilateral chatter leads the contralateral one. Unmasking indices (UIs) of FFRs under different ITD disparities are shown before (shaded bars) and after (hatched 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 $GABA_A$ 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 forebrain 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, the A1 sends descending axonal projections to the IC (Coomes et al., 2005; Druga et al., 1997; Herbert et al., 1991; Schofield, 2009) and modulates neural activities of the IC (Yan and Ehret, 2002; Yan et al., 2005). The A1 may directly mediate the attentional top-down modulation of FFRs in the IC via its direct projections. Thus, we propose that under noisy conditions, the enhanced representation of target-speech signals in the auditory midbrain contributes to the "cocktail-party problem". Note that in addition to IC, corticofugal modulation (Suga et al., 2002; Suga, 2008) occurs throughout the auditory brainstem including the CN (Liu et al., 2010; Luo et al., 2008). Therefore, the top-down modulation of FFRs may also occur beyond the IC.

Taken together, one important mechanism for top-down attentional control of auditory processing is through enhancing synchronous phase-locked activities of brainstem neurons to behaviorally relevant stimulus. Thus, FFRs are useful for investigating how perceptual/cognitive cues facilitate listeners' selective attention on target speech and improve recognition of target speech against competing speech. It should be noted that one important question which has been neglected by the previous studies mentioned above is whether FFRs to the unattended irrelevant sound are significantly attenuated.

4.5. Experience-dependent plasticity of FFRs

Perceptual training can improve syllable identification in noise (Stecker et al., 2006). It has been suggested that plasticity of the auditory system can also be exploited by studying the interaction between sensory and cognitive processes at the level of the brainstem (Kraus and Banai, 2007). Indeed, FFRs were affected by either shorter-term auditory training (Russo et al., 2005; Song et al., 2008) or longer-term language/musical experience (e.g., Chandrasekaran and Kraus, 2010; Galbraith et al., 2004; Johnson et al., 2008; Krishnan et al., 2005, 2009; Parbery-Clark et al., 2009; Xu et al., 2006). For example, Galbraith et al. (2004) have shown increased FFR amplitudes to forward speech, as compared to reversed speech, indicating that familiar phonetic and prosodic properties of forward speech after lifelong exposure to native language pattern selectively activate brainstem neurons. Also, Krishnan et al. (2005), in a cross-language study, found that FFRs to Mandarin tones exhibit stronger pitch representation and smoother pitch tracking in native versus nonnative listeners, suggesting that long-term experience with linguistic pitch contours enhances pitch representation in the auditory brainstem. Thus, brainstem stages of central processing along the auditory pathway perform computations related to the experience-dependent sensitivity to some linguistically relevant features or dimensions, and the experience-modified change can be revealed by FFRs.

Interestingly, as perception of speech and that of music rely on some shared neural mechanisms, extensive experience in one domain may induce perceptual benefits to the other. As unraveled by recent FFR studies, long-term musical experience not only improves neural timing of the auditory brainstem in processing music (Lee et al., 2009; Musacchia et al., 2007) but also engenders more robust and efficient brainstem representation of speech sounds (Musacchia et al., 2007; Parbery-Clark et al., 2009; Strait et al., 2009; Wong et al., 2007). More importantly, relative to those in non-musicians, FFRs in musicians show the remarkable advantage in speech perception with more resistance to the detrimental effect of background noise (Parbery-Clark et al., 2009). As for the effect of shorter-term experience, Russo et al. (2005) have shown that children with language-based learning problems (i.e., dyslexia) can exhibit a greater timing precision of FFRs to speech syllable and a larger tolerance to the deleterious effects of background noise following a three-month auditory training program.

Due to the role played by prior linguistic experience in speech perception, compared to native listeners, non-native listeners experience more difficulties in recognizing foreign-language speech in adverse conditions (for a recent review, see Garcia Lecumberri et al., 2010). Studies of how linguistic/musical experience affects FFRs in a "cocktail-party" situation will richen our understanding of the nature of the "cocktail-party problem".

4.6. Aging effects on speech recognition in "cocktail-party" situations and FFRs

Our understanding of the nature of the "cocktail-party problem" can also be enriched by studying auditory aging. Older-adult listeners often report that they have difficulties in understanding speech under "cocktail-party" conditions where there is more than one person speaking at the same time (e.g., Cheesman et al., 1995; Gelfand et al., 1988; Helfer and Wilber, 1990; Huang et al., 2008, 2010). Particularly, the age-related difficulty augments when the listening environment is reverberant (Helfer, 1992; Helfer and Wilber, 1990; Huang et al., 2008; Nábělek and Robinson, 1982; Náběelek, 1988). Several lines of research suggest that the age-related difficulty is related to the age-related reduction of the ability to process fine-structure acoustic information. First, under a (simulated) reverberant environment, the primitive auditory memory for transiently storing acoustic details is important for perceptually integrating the direct waveform from a speech source with its reflections, and the perceptual integration plays a role in releasing speech from informational masking (Huang et al., 2009a). Older adults with clinically normal hearing have declined ability to transiently store acoustic details (Huang et al., 2009b; Li et al., 2009) and perform poorly in integrating correlated leading/lagging sound waves for unmasking speech (Huang et al., 2008). In addition, the talker's voice contains speech-content information, talker's identity information and affective information. Knowledge and/or familiarity of the voice of target speech facilitate listeners' selective attention to the vocal characteristics of the target stream, leading to a release of speech from informational masking (Huang et al., 2010; Yang et al., 2007). Compared to younger adults, older adults have reduced abilities to discriminate talkers' voices (Helfer and Freyman, 2008), remember talkers' voices (Yonan and Sommers, 2000), and take advantage of the vocal distinctiveness in target-message identification (Rossi-Katz and Arehart, 2009), and particularly are not able to use the perceptual-level voice-priming cues to unmask speech (Huang et al., 2010).

Since these age-related auditory changes mentioned above are associated with declines in fine-structure processing (e.g., increase of filter bandwidth, reduction of phase locking or synchrony), it is predicted that compared to those recorded in younger adults, FFRs recorded in older adults would decline. Indeed, a recent study by Werff and Burns (2011) shows that the spectral magnitudes of the three harmonic components (F0, the first formant frequencies, and higher frequency harmonics) were all significantly smaller for the 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., Veuillet, 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 human auditory cortex. J. Neurosci. 27, 9252–9261.

Boudreau, J.C., 1965. Neural volleying: Upper frequency limits detectable in the auditory system. Nature 208, 1237–1238.

Bregman, A.S., 1990. Auditory Scene Analysis: The Perceptual Organization of Sound. MIT, Cambridge.

Bronkhorst, A.W., 2000. The cocktail party phenomenon: a review of research on speech intelligibility in multiple-talker conditions. Acta Acus. 86, 117–128.

Brokx, J.P.L., Nooteboom, S.G., 1982. Intonation and the perceptual separation of simultaneous voices. J. Phonetics 10, 23–36.

Brungart, D.S., Simpson, B.D., Ericson, M.A., Scott, K.R., 2001. Informational and energetic masking effects in the perception of multiple simultaneous talkers. J. Acoust. Soc. Am. 110, 2527–2538.

Burger, R.M., Pollak, G.D., 2001. Reversible inactivation of the dorsal nucleus of the lateral lemniscus reveals its role in the processing of multiple sound sources in the inferior colliculus of bats. J. Neurosci. 21, 4830–4843.

Caird, D.M., Palmer, A.R., Rees, A., 1991. Binaural masking level difference effects in single units of the guinea-pig inferior colliculus. Hear. Res. 57, 91–106.

Chandrasekaran, B., Kraus, N., 2010. The scalp-recorded brainstem response to speech: neural origins and plasticity. Psychophysiol. 47, 236–246.

Cheesman, M.F., Hepburn, D., Armitage, J.C., Marshall, K., 1995. Comparison of growth of masking functions and speech discrimination abilities in younger and older adults. Audiology 34, 321–333.

Cherry, C.E., 1953. Some experiments on the recognition of speech, with one and with two ears. J. Acoust. Soc. Am. 25, 975–979.

Clinard, C.G., Tremblay, K.L., Krishnan, A.R., 2010. Aging alters the perception and physiological representation of frequency: evidence from human frequency-following response recordings. Hear. Res. 264, 48–55.

Coomes, D.L., Schofield, R.M., Schofield, B.R., 2005. Unilateral and bilateral projections from cortical cells to the inferior colliculus in guinea pigs. Brain Res. 1042, 62–72.

Culling, J.F., Darwin, C.J., 1993. Perceptual separation of simultaneous vowels: within and across formant grouping by F0. J. Acoust Soc. Am. 93, 3454–3467.

Davis, R.L., Britt, R.H., 1984. Analysis of the frequency following response in the cat. Hear. Res. 15, 29–37.

Dent, M.L., Dooling, R.J., Larsen, O.N., 1997. Free-field binaural unmasking in budgerigars (*Melopsittacus undulatus*). Behav. Neurosci. 111, 590–598.

Druga, R., Syka, J., Rajkowska, G., 1997. Projections of auditory cortex onto the inferior colliculus in the rat. Physiol. Res. 46, 215–222.

Du, Y., Huang, Q., Wu, X.H., Galbraith, G.C., Li, L., 2009a. Binaural unmasking of frequency-following responses in rat amygdala. J. Neurophysiol. 101, 1647–1659.

Du, Y., Ma, T.F., Wang, Q., Wu, X.H., Li, L., 2009b. Two crossed axonal projections contribute to binaural unmasking of frequency-following responses in rat inferior colliculus. Eur. J. Neurosci. 30, 1779–1789.

Du, Y., He, Y., Ross, B., Bardouille, T., Wu, X.-H., Li, L., Alain, C., 2011. Human auditory cortex activity shows additive effects of spectral and spatial cues during speech segregation. Cereb. Cortex 21, 698–707.

Faingold, C.L., Caspary, D.M., 1979. Frequency-following responses in primary auditory and reticular formation structures. Electroencephalogr. Clin. Neurophysiol. 47, 12–20.

Faingold, C.L., Anderson, C.A., Randall, M.E., 1993. Stimulation or blockade of the dorsal nucleus of the lateral lemniscus alters binaural and tonic inhibition in contralateral inferior colliculus neurons. Hear. Res. 69, 98–106.

Freyman, R.L., Helfer, K.S., McCall, D.D., Clifton, R.K., 1999. The role of perceived spatial separation in the unmasking of speech. J. Acoust. Soc. Am. 106, 3578–3588.

Freyman, R.L., Balakrishnan, U., Helfer, K.S., 2001. Spatial release from informational masking in speech recognition. J. Acoust. Soc. Am. 109, 2112–2122.

Fritz, J.B., Elhilali, M., David, S.V., Shamma, S.A., 2007a. Auditory attention-focusing the searchlight on sound. Curr. Opin. Neurobiol. 17, 437–455.

Fritz, J.B., Elhilali, M., Shamma, S.A., 2007b. Adaptive changes in cortical receptive fields induced by attention to complex sounds. J. Neurophysiol. 98, 2337–2346.

Fujiwara, N., Nagamine, T., Imai, M., Tanaka, T., Shibasaki, H., 1998. Role of the primary auditory cortex in auditory selective attention studied by whole-head neuromagnetometer. Cog. Brain Res. 7, 99–109.

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., Comerci, 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., Eschalier, 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, K., 2007. Auditory-processing malleability: focus on language and music. Curr. Direc. Psychol. Sci. 16, 105–110.

Krishnan, A., 1999. Human frequency-following responses to two-tone approximations of steady-state vowels. Audiol. Neurootol. 4, 95–103.

Krishnan, A., 2002. Human frequency-following responses: representation of steady-state synthetic vowels. Hear. Res. 166, 192–201.

Krishnan, A., Parkinson, J., 2000. Human frequency-following response: representation of tonal sweeps. Audiol. Neurootol. 5, 312–321.

Krishnan, A., Xu, Y., Gandour, J.T., Cariani, P.A., 2004. Human frequency following response: representation of pitch contours in Chinese tones. Hear. Res. 189, 1–12.

Krishnan, A., Xu, Y., Gandour, J.T., Cariani, P., 2005. Encoding of pitch in the human brainstem is sensitive to language experience. Brain Res. Cog. Brain Res. 25, 161–168.

Krishnan, A., Gandour, J.T., 2009. The role of the auditory brainstem in processing linguistically-relevant pitch patterns. Brain Lang. 110, 135–148.

Krishnan, A., Swaminathan, J., Gandour, J.T., 2009. Experience-dependent enhancement of linguistic pitch representation in the brainstem is not specific to a speech context. J. Cog. Neurosci. 21, 1092–1105.

Krumbholz, K., Eickhoff, S.B., Fink, G.R., 2007. Feature- and object-based attentional modulation in the human auditory "where" pathway. J. Cog. Neurosci. 19, 1721–1733.

Kuwada, S., Yin, T.C.T., Syka, J., Buunen, T.J., Wickesberg, R.E., 1984. Binaural interaction in low-frequency neurons in inferior colliculus of the cat. IV. Comparison of monaural and binaural response properties. J. Neurophysiol. 51, 1306–1325.

Lane, C.C., Delgutte, B., 2005. Neural correlates and mechanisms of spatial release from masking: single-unit and population responses in the inferior colliculus. J. Neurophysiol. 94, 1180–1198.

Langner, G., 1992. Periodicity coding in the auditory system. Hear. Res. 60, 115–142.

Lee, C.C., Middlebrooks, J.C., 2011. Auditory cortex spatial sensitivity sharpens during task performance. Nat. Neurosci. 14, 108–114.

Lee, K.M., Skoe, E., Kraus, N., Ashley, R., 2009. Selective subcortical enhancement of musical intervals in musicians. J. Neurosci. 29, 5832–5840.

Li, L., Kelly, J.B., 1992. Inhibitory influence of the dorsal nucleus of the lateral lemniscus on binaural responses in the rat's inferior colliculus. J. Neurosci. 12, 4530–4539.

Li, L., Yue, Q., 2002. Auditory gating processes and binaural inhibition in the inferior colliculus. Hear. Res. 168, 113–124.

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? J. Exp. Psychol. Hum. Percept. Perform. 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 Hear. 30, 273–286.

Li, X.M., Jeng, F.C., 2011. Noise tolerance in human frequency-following responses to voice pitch. J. Acoust. Soc. Am. 129, EL21–26.

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. Electroencephalogr. Clin. Neurophysiol. 36, 415–424.

McAlpine, D., Jiang, D., Palmer, A.R., 1996. Binaural masking level differences in the inferior colliculus of the guinea pig. J. Acoust. Soc. Am. 100, 490–503.

Micheyl, C., Oxenham, A.J., 2010. Pitch, harmonicity and concurrent sound segregation: psychoacoustical and neurophysiological findings. Hear. Res. 266, 36–51.

Møller, A.R., Jannetta, P.J., Sekhar, L.N., 1988. Contributions from the auditory nerve to the brain-stem auditory evoked potentials (BAEPs): results of intracranial recording in man. Electroencephalogr. Clin. Neurophysiol. 71, 198–211.

Moushegian, G., Rupert, A.L., Stillman, R.D., 1973. Scalp-recorded early responses in man to frequencies in the speech range. Electroencephalogr. Clin. Neurophysiol. 35, 665–667.

Musacchia, G., Sams, M., Skoe, E., Kraus, N., 2007. Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. Proc. Natl. Acad. Sci. U. S. A. 104, 15894–15898.

Nábělek, A.K., Robinson, P.K., 1982. Monaural and binaural speech perception in reverberation for listeners of various ages. J. Acoust. Soc. Am. 71, 1242–1248.

Nábělek, A.K., 1988. Identification of vowels in quiet, noise, and reverberation: relationships with age and hearing loss. J. Acoust. Soc. Am. 84, 476–484.

Newman, R.S., Evers, S., 2007. The effect of talker familiarity on stream segregation. J. Phon. 35, 85–103.

OLeary, D.S., Andreasen, N.C., Hurtig, R.R., Torres, I.J., Flashman, L.A., Kesler, M.L., Arndt, S.V., Cizadlo, T.J., Ponto, L.L.B., Watkins, G.L., Hichwa, R.D., 1997. Auditory and visual attention assessed with PET. Hum. Brain Mapp. 5, 422–436.

Palmer, A.R., Russell, I.J., 1986. Phase-locking in the cochlear nerve of the guinea pig and its relation to the receptor potential of inner hair cells. Hear. Res. 24, 1–15.

Palmer, A.R., Winter, I.M., Darwin, C.J., 1986. The representation of steady-state vowel sounds in the temporal discharge patterns of the guinea pig cochlear nerve and primarylike cochlear nucleus neurons. J. Acoust. Soc. Am. 79, 100–113.

Palmer, A.R., Jiang, D., McAlpine, D., 1999. Desynchronizing responses to correlated noise: a mechanism for binaural masking level differences at the inferior colliculus. J. Neurophysiol. 81, 722–734.

Palmer, A.R., Jiang, D., McAlpine, D., 2000. Neural responses in the inferior colliculus to binaural masking level differences created by inverting the noise in one ear. J. Neurophysiol. 84, 844–852.

Palmer, A.R., Shackleton, T.M., 2002. The physiological basis of the binaural masking level difference. Acta Acust. United Ac. 88, 312–319.

Parbery-Clark, A., Skoe, E., Kraus, N., 2009. Musical experience limits the degradative effects of background noise on the neural processing of sound. J. Neurosci. 29, 14100–14107.

Picton, T.W., Hillyard, S.A., 1974. Human auditory evoked potentials II: effects of attention. Electroencephalogr. Clin. Neurophysiol. 36, 191–200.

Picton, T.W., Stapells, D.R., Campbell, K.B., 1981. Auditory evoked potentials from the human cochlea and brainstem. J. Otolaryngol. 9 (Suppl.), 1–41.

Ping, J.L., Li, N.X., Galbraith, G.C., Wu, X.H., Li, L., 2008. Auditory frequency-following responses in rat ipsilateral inferior colliculus. Neuroreport 19, 1377–1380.

Plyler, P.N., Ananthanarayan, A.K., 2001. Human frequency-following responses: representation of second formant transitions in normal-hearing and hearing-impaired listeners. J. Am. Acad. Audiol. 12, 523–533.

Poghosyan, V., Ioannides, A.A., 2008. Attention modulates earliest responses in the primary auditory and visual cortices. Neuron 58, 802–813.

Polley, D.B., Steinberg, E.E., Merzenich, M.M., 2006. Perceptual learning directs auditory cortical map reorganization through top-down influences. J. Neurosci. 26, 4970–4982.

Rakerd, B., Aaronson, N.L., Hartmann, W.M., 2006. Release from speech-on-speech masking by adding a delayed masker at a different location. J. Acoust. Soc. Am. 119, 1597–1605.

Ratnam, R., Feng, A.S., 1998. Detection of auditory signals by frog inferior colliculus neurons in the presence of spatially separated noise. J. Neurophysiol. 80, 2848–2859.

Rinne, T., Balk, M.H., Koistinen, S., Autti, T., Alho, K., Sams, M., 2008. auditory selective attention modulates activation of human inferior colliculus. J. Neurophysiol. 100, 3323–3327.

Rose, J.E., Brugge, J.F., Anderson, D.J., Hind, J.E., 1967. Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey. J. Neurophysiol. 30, 769–793.

Rosenblum, L.D., Johnson, J.A., Saldana, H.M., 1996. Point-light facial displays enhance comprehension of speech in noise. J. Speech. Hear. Res. 39, 1159–1170.

Rossi-Katz, J., Arehart, K.H., 2009. Message and talker identification in older adults: Effects of task, distinctiveness of the talkers' voices, and meaningfulness of the competing message. J. Speech Lang. Hear. Res. 52, 435–453.

Rouiller, E., de Ribaupierre, Y., de Ribaupierre, F., 1979. Phase-locked responses to low frequency tones in the medial geniculate body. Hear. Res. 1, 213–226.

Rudmann, D.S., McCarley, J.S., Kramer, A.F., 2003. Bimodal displays improve speech comprehension in environments with multiple speakers. Hum. Factors 45, 329–336.

Russo, N.M., Nicol, T.G., Musacchia, G., Kraus, N., 2004. Brainstem responses to speech syllables. Clin. Neurophysiol. 115, 2021–2030.

Russo, N.M., Nicol, T.G., Zecker, S.G., Hayes, E.A., Kraus, N., 2005. Auditory training improves neural timing in the human brainstem. Behav. Brain Res. 156, 95–103.

Russo, N.M., Skoe, E., Trommer, B., Nicol, T., Zecker, S., Bradlow, A., et al., 2008. Deficient brainstem encoding of pitch in children with autism spectrum disorders. Clin. Neurophysiol. 119, 1720–1731.

Saberi, K., Dostal, L., Sadralodabai, T., Bull, V., Perrott, D.R., 1991. Free-field release from masking. J. Acoust. Soc. Am. 90, 1355–1370.

Saint Marie, R.L., 1996. Glutamatergic connections of the auditory midbrain: selective uptake and axonal transport of D-[3H]aspartate. J. Comp. Neurol. 373, 255–270.

Schneider, B.A., Li, L., Daneman, M., 2007. How competing speech interferes with speech comprehension in everyday listening situations. J. Am. Acad. Audiol. 18, 559–572.

Schofield, B.R., 2009. Projections to the inferior colliculus from layer VI cells of auditory cortex. Neuroscience 159, 246–258.

Shinn-Cunningham, B.G., Schickler, J., Kopco, N., Litovsky, R., 2001. Spatial unmasking of nearby speech sources in a simulated anechoic environment. J. Acoust. Soc. Am. 110, 1118–1129.

Shinn-Cunningham, B.G., Ihlefeld, A., Larson, E., 2005. Bottom-up and top-down influences on spatial unmasking. Acta Acust. United. Ac. 91, 967–979.

Sinex, D.G., 2008. Responses of cochlear nucleus neurons to harmonic and mistuned complex tones. Hear. Res. 238, 39–48.

Sinex, D.G., Li, H., 2007. Responses of inferior colliculus neurons to double harmonic tones. J. Neurophysiol. 98, 3171–3184.

Smith, J.C., Marsh, J.T., Brown, W.S., 1975. Far-field recorded frequency-following responses: evidence for the locus of brainstem sources. Electroencephalogr. Clin. Neurophysiol. 39, 465–472.

Sohmer, H., Pratt, H., Kinarti, R., 1977. Sources of frequency following responses (FFR) in man. Electroencephalogr. Clin. Neurophysiol. 42, 656–664.

Song, J.H., Skoe, E., Wong, P.C., Kraus, N., 2008. Plasticity in the adult human auditory brainstemfollowing short-term linguistic training. J. Cog. Neurosci. 10, 1892–1902.

Song, J., Skoe, E., Banai, K., Kraus, N., in press. Perception of speech in noise: Neural correlates. J. Cogn. Neurosci.

Stecker, G.C., Bowman, G.A., Yund, E.W., Herron, T.J., Roup, C.M., Woods, D.L., 2006. Perceptual training improves syllable identification in new and experienced hearing aid users. J. Rehab. Res. Dev. 43, 537–551.

Steinschneider, M., Fishman, Y.I., Arezzo, J.C., 2008. Spectrotemporal analysis of evoked and induced electroencephalographic responses in primary auditory cortex (A1) of the awake monkey. Cereb. Cortex 18, 610–625.

Stillman, R.D., Crow, G., Moushegian, G., 1978. Components of the frequency-following potential in man. Electroencephalogr. Clin. Neurophysiol. 44, 438–446.

Strait, D.L., Skoe, E., Kraus, N., Ashley, R., 2009. Musical experience and neural efficiency: effects of training on subcortical processing of vocal expressions of emotion. Eur. J. Neurosci. 29, 661–668.

Suga, N., 2008. Role of corticofugal feedback in hearing. J. Comp. Physiol. A: Neuroethol., Sens., Neural, Behav. Physiol. 194, 169–183.

Suga, N., Xiao, Z., Ma, X., Ji, W., 2002. Plasticity and corticofugal modulation for hearing in adult animals. Neuron 36, 9–18.

Sumby, W.H., Pollack, I., 1954. Visual contribution to speech intelligibility in noise. J. Acoust. Soc. Am. 26, 212–215.

Summerfield, A.Q., 1979. Use of visual information for phonetic processing. Phonetica 36, 314–331.

Summerfield, A.Q., 1992. Lipreading and audio-visual speech perception. Philos. Trans. R. Soc. Lond. B Biol. Sci. 335, 71–78.

Van Adel, B.A., Kidd, S.A., Kelly, J.B., 1999. Contribution of the commissure of Probst to binaural evoked responses in the rat's inferior colliculus: interaural time differences. Hear. Res. 130, 115–130.

Wallace, M.N., Shackleton, T., Anderson, L.A., Palmer, A.R., 2005. Representation of the purr call in the guinea pig primary auditory cortex. Hear. Res. 204, 115–126.

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.