



## Research Paper

# Neural representations of concurrent sounds with overlapping spectra in rat inferior colliculus: Comparisons between temporal-fine structure and envelope

Lu Luo<sup>a</sup>, Qian Wang<sup>a, c</sup>, Liang Li<sup>a, b, d, \*</sup><sup>a</sup> Department of Psychology and Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing, China<sup>b</sup> Speech and Hearing Research Center, Key Laboratory on Machine Perception (Ministry of Education), Peking University, Beijing, China<sup>c</sup> Beijing Key Laboratory of Epilepsy, Epilepsy Center, Department of Functional Neurosurgery, Sanbo Brain Hospital, Capital Medical University, Beijing, China<sup>d</sup> Beijing Institute for Brain Disorders, Beijing, China

## ARTICLE INFO

## Article history:

Received 19 February 2017

Received in revised form

21 May 2017

Accepted 12 June 2017

Available online 17 June 2017

## Keywords:

Binaural unmasking

Envelope

Frequency-following response

Inferior colliculus

Sound segregation

Temporal fine structure

## ABSTRACT

Perceptual segregation of multiple sounds, which overlap in both time and spectra, into individual auditory streams is critical for hearing in natural environments. Some cues such as interaural time disparities (ITDs) play an important role in the segregation, especially when sounds are separated in space. In this study, we investigated the neural representation of two uncorrelated narrowband noises that shared the identical spectrum in the rat inferior colliculus (IC) using frequency-following-response (FFR) recordings, when the ITD for each noise stimulus was manipulated. The results of this study showed that recorded FFRs exhibited two distinctive components: the fast-varying temporal fine structure (TFS) component (FFR<sub>TFS</sub>) and the slow-varying envelope component (FFR<sub>ENV</sub>). When a single narrowband noise was presented alone, the FFR<sub>TFS</sub>, but not the FFR<sub>ENV</sub>, was sensitive to ITDs. When two narrowband noises were presented simultaneously, the FFR<sub>TFS</sub> took advantage of the ITD disparity that was associated with perceived spatial separation between the two concurrent sounds, and displayed a better linear synchronization to the sound with an ipsilateral-leading ITD. However, no effects of ITDs were found on the FFR<sub>ENV</sub>. These results suggest that the FFR<sub>TFS</sub> and FFR<sub>ENV</sub> represent two distinct types of signal processing in the auditory brainstem and contribute differentially to sound segregation based on spatial cues: the FFR<sub>TFS</sub> is more critical to spatial release from masking.

© 2017 Elsevier B.V. All rights reserved.

## 1. Introduction

In natural environments, both humans and animals often listen to multiple-source sounds with both temporal and spectral overlaps. The ability to perceptually segregate mixed acoustic waves into different auditory streams, a process known as “auditory scene analysis” (Bregman, 1990), is crucial for survival. Several cues are employed by the auditory system to achieve this streaming, including fundamental frequency ( $F_0$ ), onset time, and a few spatial cues such as interaural time difference (ITD) (Moore and Gockel, 2002).

The improvement of signal detection/perception when

(perceived) spatial separation exists between multiple sound sources has been a long-studied psychophysical phenomenon, often referred to as “spatial release from masking” (SRM) (Hirsh, 1950). Previous behavioral studies have provided evidences demonstrating the beneficial effects of (perceived) spatial separation for both humans (Bronkhorst and Plomp, 1988; Freyman et al., 2001; Kidd et al., 1998; Li et al., 2004; Shinn-Cunningham et al., 2001) and animals (e.g., Bee, 2008; Du et al., 2009, 2012; Hine et al., 1994; Schmidt and Römer, 2011). Listeners with impaired hearing usually show reduced effects of SRM (Best et al., 2011; Ching et al., 2011; Peissig and Kollmeier, 1997). Therefore, it is of great interest to understand the underlying neural mechanisms.

To date, most electrophysiological studies in this line of research have focused on some non-spatial segregation cues such as  $F_0$  (Fishman et al., 2014; Keilson et al., 1997), difference frequency (i.e. beats) (Bodnar and Bass, 1999, 2001a, 2001b), sound level

\* Corresponding author. School of Psychological and Cognitive Sciences, Peking University, Beijing, 100080, China.

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

(Nakamoto et al., 2010; Sinex and Li, 2007), and onset synchrony (Bidet-Caulet et al., 2007). Notably, the acoustic stimuli used in these studies had distinctive spectral features that tagged their sources readily, including pure tones with different frequencies (Bidet-Caulet et al., 2007), different harmonic tones (Fishman et al., 2014; Sinex and Li, 2007), vowels with different periodic components (Keilson et al., 1997; Sayles et al., 2016), and narrowband noises (NBNs) with different frequency ranges (Witten et al., 2010). It remains an open question how mixed sounds with overlapping spectra are resolved by auditory neurons (Day et al., 2012; Keller and Takahashi, 2005), which is often the case in natural environments.

When complex sounds reach the ear, the peripheral auditory system filters the sound waves into narrowband channels through a series of band-pass filters. For each of the narrowband channels, the output signals are further decomposed into the fast fluctuating temporal fine structures (TFSs) and the slowly varying envelopes (Moore, 2008). The contributions of TFS and envelope in auditory processing have long been debated. One notion suggests that while the envelope alone is sufficient for hearing content of speech sounds in quiet, the TFS information plays a crucial role in speech recognition under noisy situations (e.g., Apoux et al., 2013; Lorenzi et al., 2006; Smith et al., 2002; Zeng et al., 2004). However, how the TFS and envelope components contribute to the segregation of concurrent sounds is still poorly understood, especially when interaural integration is involved (Swaminathan et al., 2016).

Frequency following responses (FFRs) are sustained electrical potentials of neuron populations that synchronize to periodicities of low- and middle-frequency sounds (Chandrasekaran and Kraus, 2010; Du et al., 2011; Marsh and Worden, 1969; Moushegian et al., 1973; Weinberger et al., 1970; Worden and Marsh, 1968). Human scalp-recorded FFRs show a response limit up to 1.5 kHz (Glaser et al., 1976), and intracranially recorded FFRs in rats show a response limit up to 4 kHz (Ping et al., 2008). In humans, scalp-recorded FFRs can be evoked by a variety of sounds, from simple tones (Chimento and Schreiner, 1990; Galbraith, 1994) to complex sounds such as speech syllables (Aiken and Picton, 2008; Akhoun et al., 2008; Krishnan, 2002; Russo et al., 2004; Song et al., 2008; Wong et al., 2007) and music (Musacchia et al., 2007). FFRs have been shown to be behaviorally and cognitively relevant, and subject to attentional modulation (Du et al., 2012; Galbraith et al., 1998; Hairston et al., 2013; Hoormann et al., 2004; Lehmann and Schönwiesner, 2014), short-term training (Skoe et al., 2013; Song et al., 2008), and long-term experience (Chandrasekaran and Kraus, 2010; Musacchia et al., 2007; Wong et al., 2007). Release from masking based on binaural spatial cues for FFR has been reported (Du et al., 2009, 2012; Wilson and Krishnan, 2005). Interestingly, both human (Lehmann and Schönwiesner, 2014) and animal studies (Du et al., 2011) have confirmed that FFRs precisely represent concurrent periodical-stimulus sources, with components of different frequency being clearly separated in the FFR spectra. Also, previous reports have shown that FFRs can represent the spectral information conveyed in both the TFS and the envelope components of narrowband noises (Wang and Li, 2015) and speech sounds (Aiken and Picton, 2008; Ananthkrishnan et al., 2016; Skoe and Kraus, 2010). Therefore, FFRs are useful for investigating not only stream segregation that occurs at the brainstem level but also specific representations of the TFS and envelope components (Du et al., 2011).

Intracranial-recorded FFRs have been reported recently in inferior colliculus (IC) of anesthetized rats (Du et al., 2009; Ping et al., 2008; Wang and Li, 2015) and amygdala of awake rats (Du et al., 2012). Particularly, the IC, which is the brainstem hub for auditory processing from lower nuclei converge (Schreiner and Winer, 2005), is generally considered as the major source of

scalp-recorded FFRs (Chandrasekaran and Kraus, 2010; Smith et al., 1975; Sohmer et al., 1977).

To examine the brainstem representation of concurrent sounds with overlapping spectra and the effect of ITD on source segregation of the concurrent sounds, in this study we used pairs of uncorrelated NBNs with the identical center frequency and bandwidth to evoke FFRs in rat IC when the ITD of each NBN was manipulated separately. The TFS and envelope components of FFRs were extracted and investigated separately, focusing on their sensitivities to the ITD.

## 2. Materials and methods

### 2.1. Animal preparation

Sixteen young-adult male Sprague-Dawley rats (270–359 g, purchased from the Vital River Experimental Animal Company, Beijing) were used in this study. They were anesthetized with 10% chloral hydrate (400 mg/kg i.p.). The state of anesthesia was monitored throughout the experiment and maintained by supplemental injection of the same anesthetic. Stainless steel electrodes (10–20 k $\Omega$ ) insulated by silicon tubes (0.3 mm in diameter) except at the 0.25-mm-diameter tip (Du et al., 2009, 2012; Wang and Li, 2015) were inserted in the left or right central nucleus of the IC (left:  $n = 8$ ; right:  $n = 8$ ). Based on the stereotaxic coordinates of Paxinos and Watson (1997), the coordinates of the aimed IC site referenced to Bregma were: AP,  $-8.8$  mm; ML,  $\pm 1.5$  mm; DV,  $-4.5$  to  $-5.0$  mm.

The treatments of animals in this study were in accordance with the Guidelines of the Beijing Laboratory Animal Center. All experimental procedures were approved by the Committee for Protecting Human and Animal Subjects in the School of Psychology and Cognitive Sciences at Peking University.

### 2.2. Apparatus and stimuli

Acoustic stimuli were processed through a Tucker-Davis Technology RZ6 auditory processor and presented by two MF1 loudspeakers. For the close-field sound delivery, two 10-cm PVC tubes were connected to the tips of the loudspeakers, with the other ends being inserted into each of the rat's ear canals. All stimuli were calibrated using a Larson Davis Audiometer Calibration and Electroacoustic Testing System (AUDit and System 824), and the sound pressure level (SPL) was 70 dB for each loudspeaker.

For each rat, a pair of uncorrelated NBNs (24414-Hz sampling rate, 16-bit amplitude quantization) with a center frequency of 1200 Hz and a bandwidth of 400 Hz were drawn from a pool of 500 NBN pairs, which were generated by passing Gaussian white noises through a 512-point filter with MATLAB (MathWorks, Natick, MA) (see Fig. 1A for an example). Since 16 rats were used in this study, a total of 16 different pairs of NBNs were used. These NBN pairs were chosen for their low mutual coherences (mean =  $0.070 \pm 0.016$  SD). To avoid the spectral dominance from one single sound of a NBN pair, the coherence between each individual sounds and the summation of the NBN pair was checked to make sure that the coherence difference within each pair was minimum (mean coherence difference =  $0.067 \pm 0.056$  SD). The stimulus duration was 150 ms with 5-ms Hanning onset/offset ramps. The (offset-onset) inter-stimulus interval was 250 ms.

Under the single-source condition, only one sound from the NBN pair was presented binaurally with the interaural time difference (ITD) of either  $+0.16$  ms (ipsilateral-leading, referenced to the recording site) or  $-0.16$  ms (contralateral-leading, referenced to the recording site). The ITDs of  $\pm 0.16$  ms were chosen because this is not only the natural ITD limit of low-frequency fine structures for

adult rats as measured by Koka et al. (2008), but also below one-quarter of the period of the center frequency (0.21 ms) allowed by the 24414-Hz-sampling-rate recording system.

Normally, when the ITD is +0.16 ms, the stimulus image would be perceived as from the ipsilateral ear; when ITD is −0.16 ms, the stimulus image would be perceived as from the contralateral ear (Fig. 1B). Under the double-source condition, each of the two sounds in a NBN pair was presented binaurally, with an ITD of either +0.16 ms or −0.16 ms, leading to 4 laterality configurations between the two NBNs. For the sake of clarification, the NBNs used in both single- and double-source conditions were named as “source A”, while the NBNs only used in double-source conditions were named as “source B”. Therefore, when source A and source B share the same ITD, they are perceived as from the same location (“perceived co-location”); when the ITD of source A is different from the ITD of source B, they are perceived as from different locations, with one at the ipsilateral ear and the other at the contralateral ear (“perceived spatial separation”) (Fig. 1B).

Acoustically evoked potentials were recorded in a sound-attenuating chamber, amplified by a TDT RA16P Medusa preamplifier, and filtered through a 5- to 10,000-Hz online band-pass filter (with a 50-Hz notch). For each condition, stimuli were presented either in the original polarity or in the inverted polarity, and the neural responses associated with each of the stimulation polarities were averaged 100 times. All the stimulation conditions were presented in random order for each rat. Online recordings were processed with TDT Biosig software, digitized at 24414 Hz, and stored in a disk for off-line analyses.

### 2.3. Data analyses

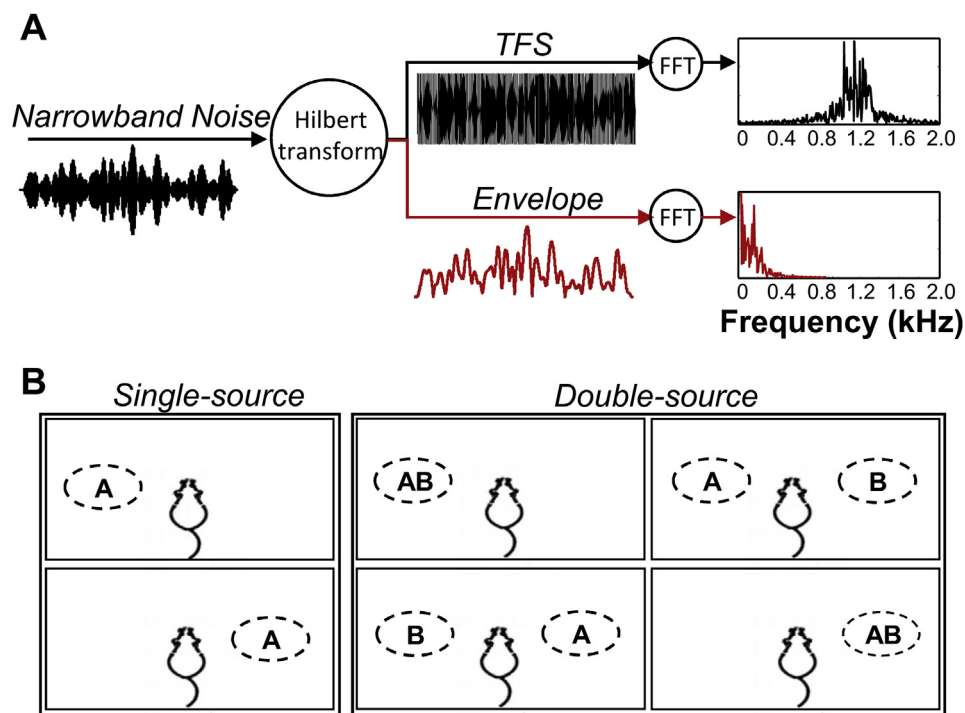
Inverting the polarity of NBN waves reversed the stimulus fine structures, while leaving the envelope unchanged. Therefore, the envelope component of FFRs ( $FFR_{ENV}$ ) could be extracted by

averaging the summation of FFRs to stimulus with opposite polarities, and the fine structure components of FFRs ( $FFR_{TFS}$ ) could be extracted by subtracting response to the inverted stimulus from that to the original stimulus and dividing the results of subtraction by two (Aiken and Picton, 2008; Joris, 2003; Skoe and Kraus, 2010).

The response latency under each condition was determined by the onset peak using Matlab function *findpeaks* and further confirmed manually. The onset peak was defined as the first positive peak in  $FFR_{ENV}$  that was two standard deviations above the mean amplitude of the 150-ms baseline activity recorded in quiet before the stimulus onset in each stimulation condition. For every  $FFR_{ENV}$  and  $FFR_{TFS}$  component, a 150-ms fragment starting from the response onset was used for the following analyses.

To examine the synchronization between FFRs and NBN stimuli, the stimulus-response coherence (S-R coherence) was calculated. The coherence between two signals  $x$  and  $y$  ( $C_{xy}$ ) is a function of frequency ( $f$ ), and defined as  $C_{xy} = |P_{xy}(f)|^2 / P_{xx}(f)P_{yy}(f)$ , where  $P_{xy}(f)$  represents the cross spectral density of  $x$  and  $y$ , and  $P_{xx}(f)$  and  $P_{yy}(f)$  represent the power spectral densities of  $x$  and  $y$ , respectively (Middleton et al., 2006; Weiss and Mueller, 2003, 2005). The TFS and envelope components of the NBN stimuli were extracted separately using Hilbert transform (Smith et al., 2002), the S-R coherence for the TFS component was calculated between  $FFR_{TFS}$  and the stimulus TFS, and the S-R coherence for the envelope component was calculated between the  $FFR_{ENV}$  and the stimulus envelope.

In this study, the frequency resolution of the coherence analysis was 6.7 Hz, and a moving Hanning window with a length of 512 data points and 75% overlap was chosen. The resulting coherence values range from 0 to 1, and measure the linear similarity between the FFRs and the stimuli. More specifically, a coherence of 1 represents perfect linear synchronization, which requires constant phase shift and amplitude ratio between the two signals at a particular frequency. Meanwhile, a coherence of 0 indicates



**Fig. 1.** Schematic of the stimulus and conditions. **A:** An example of narrowband noise used in this study, which is decomposed into TFS component and envelope component with Hilbert transform. The FFT spectrums of each component were also shown. **B:** A diagram of the stimulus conditions. The perceived locations of each sounds was marked by dashed circles.

complete lack of linear synchronization (Levy et al., 2000; Middleton et al., 2006; Rosenberg et al., 1989). For comparisons across different conditions, the S-R coherence value for each  $FFR_{TFS}$  and that for each  $FFR_{ENV}$  were averaged across the frequency respectively. Based on the spectrum of the NBN stimuli, the frequency range of averaging was 1000–1400 Hz for the TFS component, and 0 to 400 Hz for the envelope component.

Moreover, to compare the synchronization to each NBN sound in the double-source conditions,  $\Delta Coh$  was calculated as the normalized difference between the S-R coherence with source A and source B:  $\Delta Coh = \frac{coherence_A - coherence_B}{coherence_A + coherence_B}$ . Therefore,  $\Delta Coh$  could be interpreted as the ability to convey information of individual sound by FFRs. A higher  $\Delta Coh$  suggests more exclusive representation of either source A or B, and a  $\Delta Coh$  of 0 suggests equal contribution of the two sources.

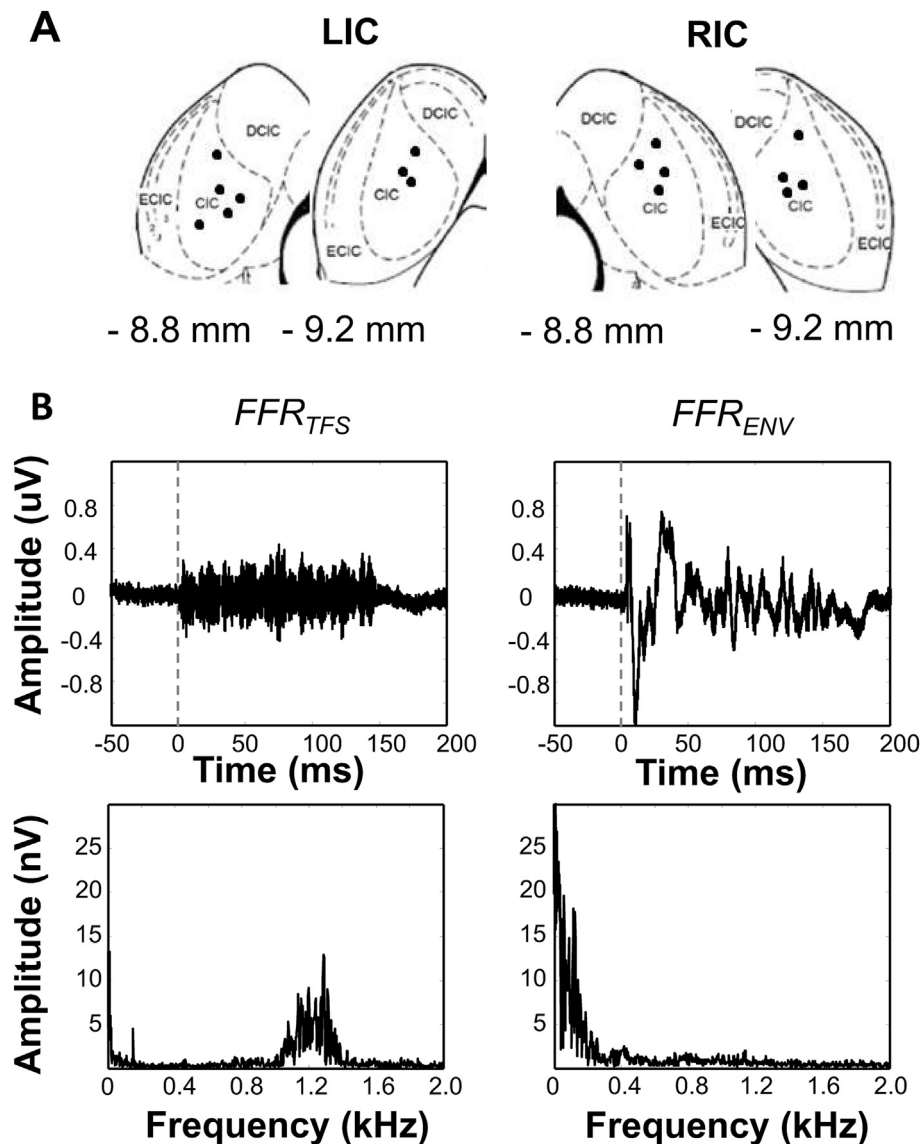
#### 2.4. Statistical analyses

Statistical analyses were performed with IBM SPSS Statistics 20 software (Chicago, IL, USA). To evaluate the effects of different

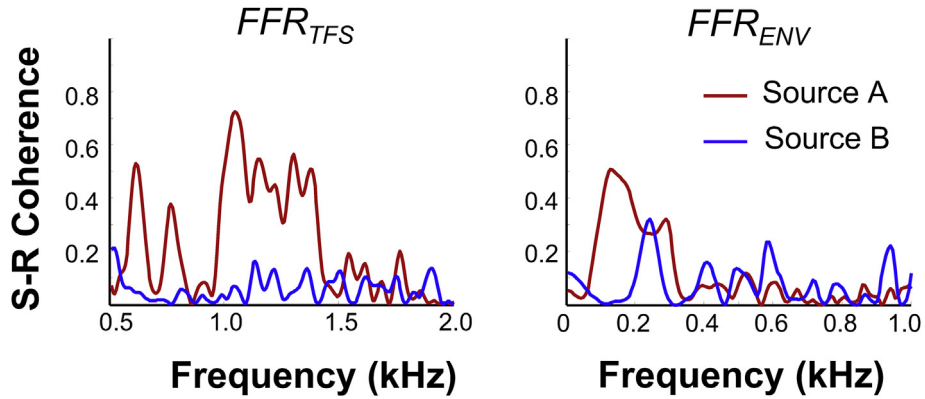
stimulus conditions on S-R coherence, within-subjects repeated measures analyses of variance (ANOVAs), paired *t*-tests, Student's *t*-tests, and Pearson correlation tests were conducted. The coherence values were Fisher-z transformed before being submitted to statistical tests, while the results presented in figures were based on the original coherence values. Since no significant differences were found between the two IC recording sites (left IC, right IC), data from both sides of the IC were pooled together for all the analyses. All *p*-values of multiple comparisons were adjusted using Bonferroni correction.

#### 2.5. Histology

After all recording sessions were completed, rats were sacrificed with an overdose of chloral hydrate. The recording sites were marked with a DC current (500  $\mu$ A for 10 s) via the inserted electrodes. The brains were removed and stored in 10% formalin with 30% sucrose and sectioned at 55  $\mu$ m in the frontal plane in a cryostat ( $-20$  °C). Sections were examined to determine locations of recording electrodes.



**Fig. 2.** Histology results and FFR examples. **A:** Histological examination of recording sites in 16 rats showed that 16 of 16 electrodes were precisely located in the central nucleus of IC (CIC), with 8 electrodes inserted in the left CIC and 8 electrodes inserted in the right CIC. DCIC, dorsal cortex of IC; ECIC, external cortex of IC. **B:** An example of the extracted  $FFR_{TFS}$  (left column) and  $FFR_{ENV}$  (right column). Both the waveforms (upper row) and the FFT spectrums (lower row) were shown.



**Fig. 3.** An example of S-R coherence as functions of frequency for  $FFR_{TFS}$  (left) and  $FFR_{ENV}$  (right) from a recording site under single-source condition with a contralateral-leading sound named as source A. The S-R coherence with source A was plotted in red, and the S-R coherence with an uncorrelated source B was plotted in blue.

**3. Results**

According to the histological examination, all the 16 electrodes were located precisely in the center nucleus of IC (Fig. 2A), and descriptions and statistical analyses were based on the data from all the recording sites.

**3.1. The reliability of representations of noise stimuli by  $FFR_{TFS}$  and  $FFR_{ENV}$**

NBN stimulus evoked robust IC FFRs under each of the stimulation conditions, and the extracted  $FFR_{TFS}$  and  $FFR_{ENV}$  showed similar spectra with the noise-stimulus TFS (around the center frequency of 1200 Hz) and the noise-stimulus envelope (between 0 and 400 Hz), respectively (see Fig. 2B for examples).

Under the single-source conditions, only one NBN sound (i.e. source A) was presented. The averaged stimulus-response coherence (S-R coherence) between FFR and the corresponding source A was  $0.455 \pm 0.060$  SD for the TFS component, and  $0.316 \pm 0.108$  SD for the envelope component. To further confirm that the S-R coherence analysis could reveal the stimulus specific information conveyed by FFRs, the S-R coherence values between FFR and source A were compared with the S-R coherence values between

the same FFR segment and the unrepresented, uncorrelated source B. Fig. 3 shows an example from one recording site under the condition with an ipsilateral-leading stimulus. The S-R coherence of TFS and that of envelope were plotted as functions of frequency. The results showed that while the S-R coherence with source B remained low across all frequencies, the S-R coherence with source A exhibited increased power in the TFS spectra range for  $FFR_{TFS}$  and envelope spectra range for  $FFR_{ENV}$ . Paired-*t* tests further confirmed this observation, showing that for both TFS and envelope, S-R coherence with source A in the corresponding spectra range was significantly higher than the S-R coherence with source B (for all recording sites and both single-source conditions,  $p < 0.001$ ).

**3.2. ITD sensitivity of  $FFR_{TFS}$  and  $FFR_{ENV}$  under the single-source conditions**

For comparisons between different single-source conditions,  $FFR_{TFS}$  and  $FFR_{ENV}$  exhibited different sensitivity to the ITD (Fig. 4). For  $FFR_{TFS}$ , a paired-*t* test showed significant higher S-R coherence with the ipsilateral-leading stimulus (mean =  $0.448 \pm 0.060$  SD) than that with the contralateral-leading stimulus (mean =  $0.399 \pm 0.058$  SD) ( $t_{15} = 3.726, p = 0.002$ ).

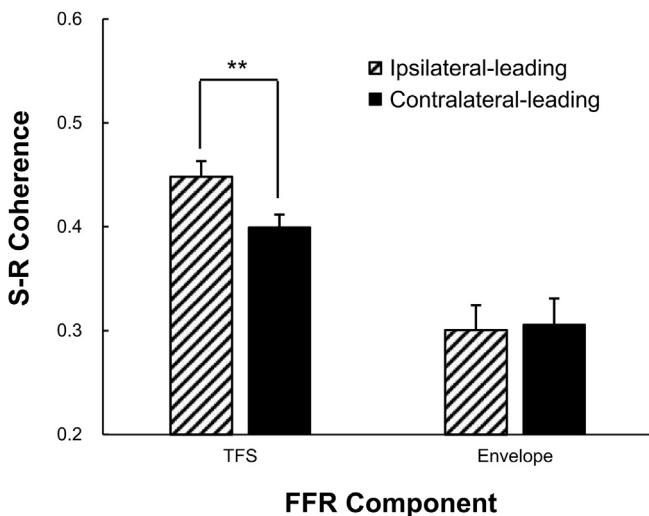
However, for  $FFR_{ENV}$ , no significant difference occurred between the S-R coherence with the ipsilateral-leading stimulus (mean =  $0.301 \pm 0.096$  SD) and that with the contralateral-leading stimulus (mean =  $0.306 \pm 0.101$  SD) ( $t_{15} = 0.902, p = 0.381$ ). Therefore, the linear synchronization between the  $FFR_{TFS}$  and the stimulus TFS was affected by the stimulus ITD with a preference towards the ipsilateral-leading sounds, and the linear synchronization between the  $FFR_{ENV}$  and the stimulus envelope appeared to be independent of the stimulus ITD.

Pearson correlation test was conducted between the S-R coherence of  $FFR_{TFS}$  and the S-R coherence of  $FFR_{ENV}$ , and no significant correlation between these two components was found ( $r = -0.229, p = 0.208, n = 32$ ).

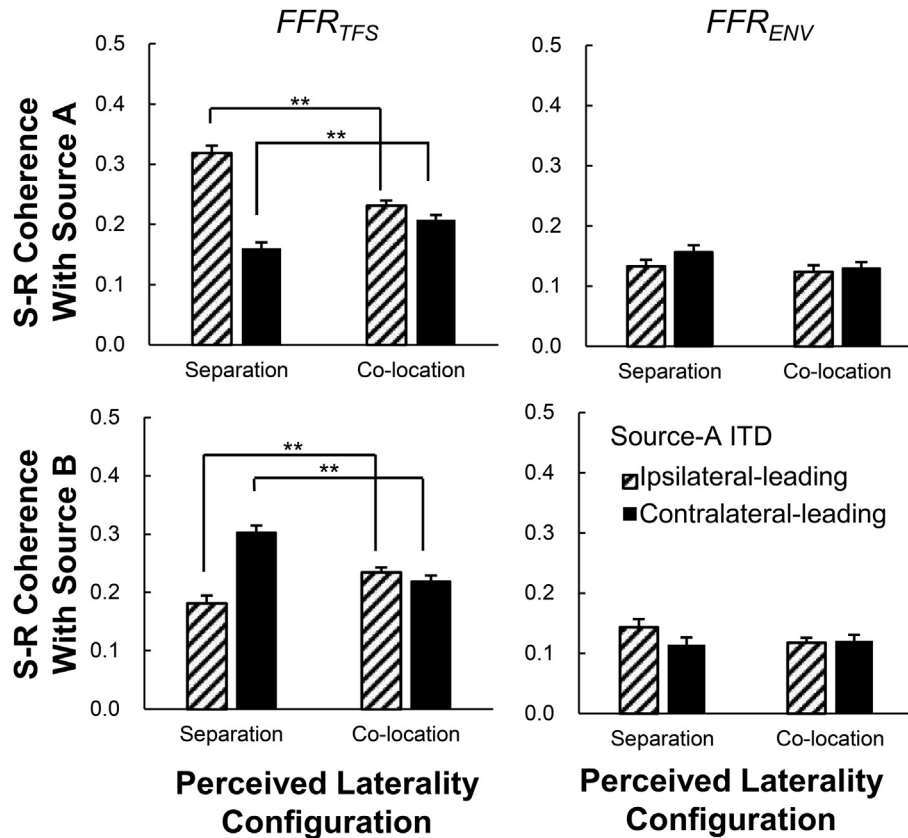
**3.3. Effects of concurrent sounds**

As a concurrent sound was introduced, both the averaged source-A S-R coherence values of  $FFR_{TFS}$  and those of  $FFR_{ENV}$  decreased compared with single source conditions (paired *t*-tests, all  $p < 0.001$ ), indicating a degradation in the accuracy of target (source-A) synchronization.

For the double-source  $FFR_{TFS}$ , a  $2 \times 2$  (source-A ITD: ipsilateral leading, contralateral leading; associated laterality configuration: perceived spatial separation, perceived co-location) repeated-



**Fig. 4.** Comparisons of S-R coherence values across different single-source conditions for  $FFR_{TFS}$  and  $FFR_{ENV}$ . Error bars: SE; \*\* $p < 0.01$ .



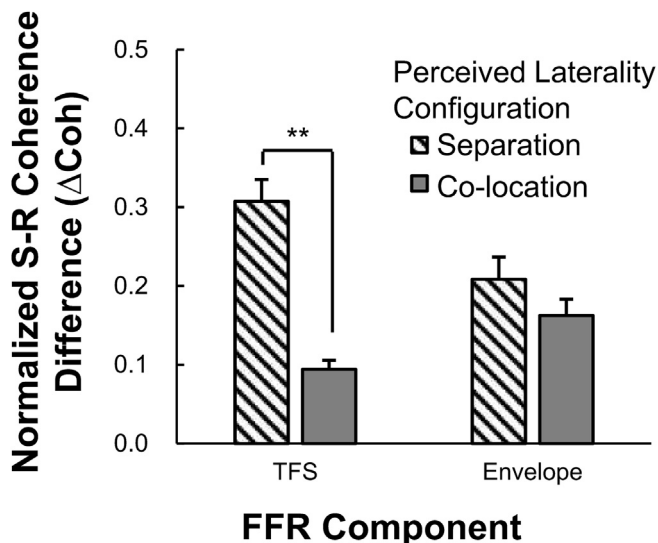
**Fig. 5.** Comparisons of source-A S-R coherence values (upper row) and source-B S-R coherence values (lower row) across double-source conditions for  $FFR_{TFS}$  (left column) and  $FFR_{ENV}$  (right column). Error bars: SE; \*\* $p < 0.01$ .

measures ANOVA demonstrated a significant main effect of the source-A ITD ( $F_{1, 15} = 89.190, p < 0.001$ ) and the laterality configuration ( $F_{1, 15} = 9.485, p = 0.008$ ), and, most importantly, significant interaction ( $F_{1, 15} = 68.801, p < 0.001$ ) on the S-R coherence with source A. Additional paired  $t$ -tests showed that the S-R coherence with an ipsilateral leading source A was significantly higher under

the perceived-spatial-separation conditions (mean =  $0.319 \pm 0.049$  SD) compared to that under the perceived-co-location conditions (mean =  $0.231 \pm 0.033$  SD) ( $t_{15} = 6.702, p < 0.001$ ). Opposite effect of laterality configuration was found for the S-R coherence with a contralateral-leading source A, which was significantly decreased when ITD disparity was introduced between source A and source B (mean =  $0.161 \pm 0.037$  SD) compared with that under the perceived-co-location conditions (mean =  $0.207 \pm 0.032$  SD) ( $t_{15} = 9.014, p < 0.001$ ) (Fig. 5).

Since source A and source B were interchangeable between the two NBNS, the same analyses were also conducted for the S-R coherence with source B. As expected, similar patterns of changes across conditions were observed. A repeated-measures ANOVA also found significant main effects of source-A ITD and laterality configuration and their interaction (main effect of source-A ITD:  $F_{1, 15} = 40.589, p < 0.001$ ; main effect of laterality configuration:  $F_{1, 15} = 5.143, p = 0.039$ ; interaction effect:  $F_{1, 15} = 33.097, p < 0.001$ ). The S-R coherence with the ipsilateral-leading source B also benefited from ITD disparity and showed markedly larger values when source A was contralateral-leading, and the S-R with the contralateral-leading source B was decreased by the presence of a ipsilateral-leading source A (paired  $t$ -tests, both  $p < 0.001$ ) (Fig. 5).

The source-A S-R coherence and the source-B S-R coherence of the  $FFR_{TFS}$  under different double-source conditions were further examined. Under the perceived spatial separation condition, the S-R coherence with the ipsilateral-leading sounds were stronger than those with the contralateral-leading sounds, whether the ipsilateral-leading sounds were source A ( $t_{15} = 5.781, p < 0.001$ ) or source B ( $t_{15} = 9.648, p < 0.001$ ). Meanwhile, under the perceived co-location condition, no significant differences were found



**Fig. 6.** Comparisons of the normalized differences between the source-A S-R coherence and the source-B S-R coherence ( $\Delta\text{Coh}$ ) in the perceived-spatial-separation conditions and the perceived-co-location conditions. Error bars: SE; \*\* $p < 0.01$ .

between the S-R coherence with source A and those with source B (ipsilateral co-location condition:  $t_{15} = 0.241$ ,  $p = 0.813$ ; contralateral co-location condition:  $t_{15} = 1.051$ ,  $p = 0.310$ ).  $\Delta\text{Coh}$  were further introduced by calculating the normalized difference between the source-A S-R coherence and the source-B S-R coherence (for details see Materials and methods). A two-tailed Student's  $t$ -test showed that the  $\Delta\text{Coh}$  under the perceived spatial separation condition (mean =  $0.307 \pm 0.157$  SD) was significantly larger than that under the perceived co-location condition (mean =  $0.094 \pm 0.064$  SD) ( $t_{62} = 6.886$ ,  $p < 0.001$ ), indicating that the extraction of individual sounds benefited from perceived spatial separation between source A and source B based on ITD disparity, and the TFS information of the preferred source could be represented with a relatively better “signal-to-noise ratio” in IC (Fig. 6).

However, for  $\text{FFR}_{\text{ENV}}$ , two-way repeated-measures ANOVAs did not show significant main effects of source-A ITD ( $F_{1, 15} = 1.133$ ,  $p = 0.304$ ) and laterality configuration ( $F_{1, 15} = 3.884$ ,  $p = 0.067$ ) on the S-R coherence with envelope of source A. The interaction was not significant ( $F_{1, 15} = 2.370$ ,  $p = 0.145$ ). The same results was also demonstrated for source-B S-R coherence (main effect of ITD:  $F_{1, 15} = 2.543$ ,  $p = 0.132$ ; main effect of laterality configuration:  $F_{1, 15} = 0.643$ ,  $p = 0.435$ ; interaction effect:  $F_{1, 15} = 0.960$ ,  $p = 0.343$ ) (Fig. 5). The S-R coherence with source A and that with source B showed no difference across all four double-source conditions (paired- $t$  tests, all  $p > 0.05$ ). Also, no difference in  $\Delta\text{Coh}$  was found between the perceived-spatial-separation condition and the perceived-co-location condition ( $t_{62} = 1.350$ ,  $p = 0.182$ ) (Fig. 6). In summary, unlike the  $\text{FFR}_{\text{TFS}}$ , the  $\text{FFR}_{\text{ENV}}$  was not sensitive to the ITD cue.

## 4. Discussion

### 4.1. Faithful representations of noise stimuli by $\text{FFR}_{\text{TFS}}$ and $\text{FFR}_{\text{ENV}}$

In agreement with previous studies (e.g., Aiken and Picton, 2008; Ananthakrishnan et al., 2016; Skoe and Kraus, 2010; Wang and Li, 2015), the results of this study showed that both the  $\text{FFR}_{\text{TFS}}$  and  $\text{FFR}_{\text{ENV}}$  accurately conveyed the TFS and envelope information of the noise stimulus, respectively. Specifically, both the  $\text{FFR}_{\text{TFS}}$  and  $\text{FFR}_{\text{ENV}}$  showed markedly stronger coherence with the actually presented NBN relative to that with an irrelevant, uncorrelated NBN. Since the coherence analysis measures linear synchrony and requires both constant phase shift and amplitude ratio between the two signals to achieve a high coherence value (Levy et al., 2000; Middleton et al., 2006; Rosenberg et al., 1989), the neural representation of spectral features of NBNs by  $\text{FFR}_{\text{TFS}}$  and  $\text{FFR}_{\text{ENV}}$  described here can be considered as being largely relied on linear processing.

### 4.2. Sensitivity of $\text{FFR}_{\text{TFS}}$ and $\text{FFR}_{\text{ENV}}$ to ITD under the single-source condition

In this study, the  $\text{FFR}_{\text{TFS}}$  and  $\text{FFR}_{\text{ENV}}$  exhibited different sensitivity to ITD-cued perceived stimulus location. For the  $\text{FFR}_{\text{TFS}}$ , the ability to synchronize with the “target” sound was significantly dependent on the ITD, with the ipsilateral-leading sound eliciting markedly stronger S-R coherence than the contralateral-leading sound. Similarly, in previous reports, the amplitude of  $\text{FFR}_{\text{TFS}}$  to pain chatter (Du et al., 2009), harmonic complex (Ping et al., 2008), and pure tone bursts (Wilson and Krishnan, 2005) were ipsilaterally predominant. However, the IC  $\text{FFR}_{\text{ENV}}$  remained largely unchanged across different ITD conditions.

The lack of ITD tuning in  $\text{FFR}_{\text{ENV}}$  seems to be contrary to previous psychophysical studies showing that human listeners are sensitive

to the envelopes of high-frequency carriers (Henning, 1974; McFadden and Pasanen, 1976) and the fact that envelope-ITD-sensitive neurons have been found in animal IC (e.g. Batra et al., 1989; Batra et al., 1993; Griffin et al., 2005; Smith et al., 2008). However, behavioral studies have also revealed that this sensitivity is not only highly listener-dependent (Macpherson and Middlebrooks, 2002; Trahiotis and Bernstein, 1986), but also typically weaker comparing with the sensitivity to the ITD cues in low-frequency stimuli (Bernstein and Trahiotis, 1982; Wightman and Kistler, 1992; Yost et al., 1971). For example, the lateralization of high-frequency NBN has been shown to be relied on ITD cues conveyed by the low-frequency component of the stimulus more than the envelope ITD of the high-frequency sound itself (Bernstein and Trahiotis, 1982). The relatively poorer ability to use envelope ITD cues observed in psychophysical studies is supported by some neurophysiological studies showing that in IC neurons the tuning to envelope ITDs is much less sharp comparing with that to TFS ITDs (Griffin et al., 2005; Smith et al., 2008). One exception has been reported for the so-called “transposed tones”, with which IC neurons showing an envelope-ITD sensitivity that is comparable to the TFS-ITD sensitivity (Griffin et al., 2005). This type of sounds has a distinct “off period” in the envelope (i.e. sufficiently long intervals with minimum envelope amplitude in each period), which elicits phase-locking responses to the stimulus envelope better than the traditional amplitude-modulated tones (Griffin et al., 2005). Moreover, a model simulation suggests that the intrinsic envelope of high-frequency Gaussian noise is not salient enough to convey ITD information (Macpherson and Middlebrooks, 2002). According to these previously reported findings, the envelope ITD cues in the NBN sounds may be too weak for the IC neurons to be detected. Thus, the results of this study suggest that TFS signals and envelope signals of NBN sounds, along with the embedded ITD cues, are processed separately in the IC. It is worth mentioning that the stimulus-response-coherence metric used in this study can only reflect the neural coding based on linear mapping of stimulus spectral features (Middleton et al., 2006). Non-linear processing of envelope ITD cues may also be involved, but cannot be revealed by the stimulus-response-coherence analysis. It is of interest to compare the envelope sensitivity of higher-order auditory centers with that of the IC.

The differences between  $\text{FFR}_{\text{TFS}}$  and  $\text{FFR}_{\text{ENV}}$  in sensitivity to ITD also suggest possibly different neural origins of these two FFR components. Previous studies (Ruggles et al., 2012; Shinn-Cunningham et al., 2013) have suggested that the tonotopic peripheral channels may have different contributions to FFRs: the  $\text{FFR}_{\text{TFS}}$  is mainly driven by frequency channels that are tuned to the stimulus frequency, while the  $\text{FFR}_{\text{ENV}}$  is mainly driven by mid- to high-frequency channels. This hypothesis is further supported by the findings that the phase-locking strength of scalp-recorded  $\text{FFR}_{\text{TFS}}$  and  $\text{FFR}_{\text{ENV}}$  in humans are not correlated with each other (Ruggles et al., 2012), in agreement with our results showing a lack of correlation between the S-R coherence of  $\text{FFR}_{\text{TFS}}$  and that of  $\text{FFR}_{\text{ENV}}$ . Evidences from computational models and neurophysiological recordings (Carney et al., 2015; Wang and Li, in press) have suggested the involvement of IC neurons with different temporal processing properties in the representation of acoustic signals. More specifically, Carney's model shows that neurons with band-reject or low-pass modulation transfer functions (MTFs) tune to quickly fluctuated vowel formants, while neurons with bandpass MTFs tune to slowly fluctuated amplitude modulation (Carney et al., 2015). A recent FFR model based on this hypothesis has successfully separated  $\text{FFR}_{\text{TFS}}$  and  $\text{FFR}_{\text{ENV}}$ , and the simulated neural activity can well predict the  $\text{FFR}_{\text{TFS}}$  and  $\text{FFR}_{\text{ENV}}$  to NBN sounds recorded in rat IC (Wang and Li, in press). It is of importance for future studies to examine the roles of different types of IC neurons

in the formation of FFRs and binaural processing.

#### 4.3. Binaural release from masking for FFR<sub>TFS</sub>

The results of this study provide evidence showing that, when perceived spatial separation is established through the introduction of ITD disparity between the concurrent sounds, the FFR<sub>TFS</sub> benefits from the spatial cues and displays better synchronization selectively to the sound that is perceived from the ipsilateral ear.

TFS information has been shown to play an important role in speech perception (Hopkins and Moore, 2010; Lorenzi et al., 2006), especially when target speech is presented with temporally fluctuating background noises (Gnansia et al., 2008; Hopkins and Moore, 2009; Hopkins et al., 2008; for a review, see Moore, 2008). In a recent study examining the human scalp-recorded FFRs to noise-degraded speech, stronger response-stimulus correlation was found for the TFS component than that for the envelope component (Bidelman, 2016). Deficit in excessing TFS cues due to excessive encoding of envelope information has also been proposed to cause difficulties in understanding target sounds in noise for listeners with hearing loss (Anderson et al., 2013; Kale and Heinz, 2010). The better representation of individual sounds by FFR<sub>TFS</sub> when sounds are spatially separated may lead to improved perceptual performance, and hence serve as a “bottom-up” neural basis for the behavioral phenomenon general known as “spatial release from masking” (SRM).

Consistent with our findings, previous studies in both rats (Du et al., 2009) and humans (Wilson and Krishnan, 2005) have reported that FFRs can be unmasked by binaural spatial cues. Further, Lane and Delgutte (2005) have also reported the SRM effect in the population neural responses of cat IC. However, there is one fundamental difference in stimulus choosing when comparing these studies to ours: the maskers used in these previous studies are all broadband noises, hence lack the pronounced periodicity possessed by the target signals, which turned out to be able to significantly influence the neural responses (Lane and Delgutte, 2005). In this study, the two concurrent sounds were both NBNs with the same spectrum range, therefore the cues introduced by physical differences were strictly limited, making it very difficult to resolve individual sounds. Nevertheless, the results of this study indicate that IC neurons can achieve the segregation of sounds with the identical spectra based on ITD disparity.

Keller and Takahashi (2005) have shown that when two amplitude-modulated broadband noises with completely overlapping spectra were presented spatially separated at the same time, the neurons in the external nucleus of IC of barn owls could separate the concurrent sounds using binaural cues even when the two sounds shared the identical envelope. Therefore, they proposed that source separation at the level of midbrain is mostly based on TFS information. However, whether and how the detailed TFS information is preserved is not clear. From this perspective, we have extended the previous understanding of sound segregation at the brainstem level, and demonstrated that the TFS component of FFRs precisely conveys TFS information of acoustic signals with a relatively high signal-to-noise ratio when concurrent sounds are in perceived spatial separation conditions. Thus, FFRs are useful for investigating the neural representation of complex auditory scenes including the presentation of concurrent sounds with overlapping spectra.

#### 4.4. Concurrent sound segregation is reflected by FFR<sub>TFS</sub> and FFR<sub>ENV</sub>

The results of this study indicate that FFR<sub>TFS</sub>, but not FFR<sub>ENV</sub>, reflects the unmasking effect of ITD cues for concurrent sounds. Therefore, our study for the first time provides evidence suggesting

that FFR<sub>TFS</sub> and FFR<sub>ENV</sub> reflect differentially the coding of concurrent sounds in the rat IC. To our knowledge, only few reports have investigated the differences between TFS and envelope in sound segregation based on spatial localization. As mentioned above, Keller and Takahashi (2005) have shown that source segregation of spatially separated sounds at the level of midbrain is relied on differences in TFS while no evidence on envelope-based grouping is observed. Meanwhile, in a behavioral study employing a speech identification task with competing speech maskers, Swaminathan et al. (2016) have shown that the absence of TFS information in low-frequency channels (below 1500 Hz) results in a decreased beneficial effect of SRM, even though the spatial cues in envelope are fully preserved. These findings, along with those of this study, suggest a critical role of TFS in the segregation of concurrent sounds when spatial processing are involved; meanwhile the envelope cues may be less useful in the segregation. A computational model simulating the peripheral auditory system has demonstrated that, comparing with envelope cues, TFS cues are more robust in noisy environment in the early stage of auditory processing, represented by the less degraded phase-locking to the stimulus TFS in the auditory nerve (Shamma and Lorenzi, 2013). Thus, central auditory system may give more weight to the noise-resistant TFS cues in the processing of concurrent sounds.

On the other hand, considering the importance of envelope cues to speech intelligibility both in quiet (Shannon et al., 1995; Smith et al., 2002) and in noise (Swaminathan and Heinz, 2012), perceiving target-speech signals with concurrent masker presentation may require that envelope information of individual sounds is successfully resolved. Although in this study, no signs of interactions between TFS and envelope processing was found at the level of brainstem, it is of great interest to ask whether and how higher auditory centers can use the TFS-based sound segregation processing to facilitate the retrieval of intelligibility information conveyed by envelope. Ding et al. (2014) have provided imaging evidence showing that the cortical entrainment to speech envelope is not a simple representation of envelope, but a collective representation of auditory features that rely on the intactness of TFS information, indicating an integration of TFS and envelope information at the cortical level. Further studies are needed to investigate how the processing of TFS and that of envelope are integrated and the associated auditory pathway.

## 5. Conclusions

When two or more sounds with overlapping spectra are presented at the same time, spatial information serves as an important cue for auditory segregation. This study used uncorrelated NBN pairs to simulate this scenario and revealed both the TFS and envelope components of the phase-lock-based FFRs that were elicited in the rat auditory midbrain IC. The FFR<sub>TFS</sub>, but not the FFR<sub>ENV</sub>, shows the sensitivity to ITD and better synchronizes with the sound that is perceived from a preferred location when perceived spatial separation between the concurrent sounds is established due to the ITD disparity.

## Acknowledgements

This study was supported by the National Natural Science Foundation of China (31470987) and the “985” Project of 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., Gallégo, S., Moulin, A., Ménard, 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.
- Ananthakrishnan, S., Krishnan, A., Bartlett, E., 2016. Human frequency following response: neural representation of envelope and temporal fine structure in listeners with normal hearing and sensorineural hearing loss. *Ear Hear.* 37, e91–e103.
- Anderson, S., Parbery-Clark, A., White-Schwoch, T., Dreihobl, S., Kraus, N., 2013. Effects of hearing loss on the subcortical representation of speech cues. *J. Acoust. Soc. Am.* 133, 3030–3038.
- Apoux, F., Yoho, S.E., Youngdahl, C.L., Healy, E.W., 2013. Role and relative contribution of temporal envelope and fine structure cues in sentence recognition by normal-hearing listeners. *J. Acoust. Soc. Am.* 134, 2205–2212.
- Batra, R., Kuwada, S., Stanford, T.R., 1989. Temporal coding of envelopes and their interaural delays in the inferior colliculus of the unanesthetized rabbit. *J. Neurophysiol.* 61, 257–268.
- Batra, R., Kuwada, S., Stanford, T.R., 1993. High-frequency neurons in the inferior colliculus that are sensitive to interaural delays of amplitude-modulated tones: evidence for dual binaural influences. *J. Neurophysiol.* 70, 64–80.
- Bee, M.A., 2008. Finding a mate at a cocktail party: spatial release from masking improves acoustic mate recognition in grey treefrogs. *Anim. Behav.* 75, 1781–1791.
- Bernstein, L.R., Trahiotis, C., 1982. Detection of interaural delay in high-frequency noise. *J. Acoust. Soc. Am.* 71, 147–152.
- Best, V., Mason, C.R., Kidd Jr., G., 2011. Spatial release from masking in normally hearing and hearing-impaired listeners as a function of the temporal overlap of competing talkers. *J. Acoust. Soc. Am.* 129, 1616–1625.
- Bidelman, G.M., 2016. Relative contribution of envelope and fine structure to the subcortical encoding of noise-degraded speech. *J. Acoust. Soc. Am.* 140, EL358–EL363.
- Bidet-Caulet, A., Fischer, C., Bauchet, F., Aguera, P.-E., Bertrand, O., 2007. Neural substrate of concurrent sound perception: direct electrophysiological recordings from human auditory cortex. *Front. Hum. Neurosci.* 1, 1–11.
- Bodnar, D.A., Bass, A.H., 1999. Midbrain combinatorial code for temporal and spectral information in concurrent acoustic signals. *J. Neurophysiol.* 81, 552–563.
- Bodnar, D.A., Bass, A.H., 2001a. Coding of concurrent vocal signals by the auditory midbrain: effects of stimulus level and depth of modulation. *J. Acoust. Soc. Am.* 109, 809–825.
- Bodnar, D.A., Bass, A.H., 2001b. Coding of concurrent vocal signals by the auditory midbrain: effects of duration. *J. Comp. Physiol. A* 187, 381–391.
- Bregman, A.S., 1990. *Auditory Scene Analysis*. MIT Press, Cambridge MA.
- Bronkhorst, A., Plomp, R., 1988. The effect of head-induced interaural time and level differences on speech intelligibility in noise. *J. Acoust. Soc. Am.* 83, 1508–1516.
- Carney, L.H., Li, T., McDonough, J.M., 2015. Speech coding in the brain: representation of vowel formants by midbrain neurons tuned to sound fluctuations. *eNeuro* 2, 4–15.
- Chandrasekaran, B., Kraus, N., 2010. The scalp-recorded brainstem response to speech: neural origins and plasticity. *Psychophysiology* 47, 236–246.
- Chimento, T., Schreiner, C., 1990. Selectively eliminating cochlear microphonic contamination from the frequency-following response. *Electroencephalogr. Clin. Neurophysiol.* 75, 88–96.
- Ching, T.Y., van Wanrooy, E., Dillon, H., Carter, L., 2011. Spatial release from masking in normal-hearing children and children who use hearing aids. *J. Acoust. Soc. Am.* 129, 368–375.
- Day, M.L., Koka, K., Delgutte, B., 2012. Neural encoding of sound source location in the presence of a concurrent, spatially separated source. *J. Neurophysiol.* 108, 2612–2628.
- Ding, N., Chatterjee, M., Simon, J.Z., 2014. Robust cortical entrainment to the speech envelope relies on the spectro-temporal fine structure. *NeuroImage* 88, 41–46.
- Du, Y., Ma, T., Wang, Q., Wu, X., Li, L., 2009. Two crossed axonal projections contribute to binaural unmasking of frequency-following responses in rat inferior colliculus. *Eur. J. Neurosci.* 30, 1779–1789.
- Du, Y., Kong, L., Wang, Q., Wu, X., Li, L., 2011. Auditory frequency-following response: a neurophysiological measure for studying the “cocktail-party problem”. *Neurosci. Biobehav. Rev.* 35, 2046–2057.
- Du, Y., Wang, Q., Zhang, Y., Wu, X., Li, L., 2012. Perceived target-masker separation unmasks responses of lateral amygdala to the emotionally conditioned target sounds in awake rats. *Neuroscience* 225, 249–257.
- Fishman, Y.I., Steinschneider, M., Micheyl, C., 2014. Neural representation of concurrent harmonic sounds in monkey primary auditory cortex: implications for models of auditory scene analysis. *J. Neurosci.* 34, 12425–12443.
- Freyman, R.L., Balakrishnan, U., Helfer, K.S., 2001. Spatial release from informational masking in speech recognition. *J. Acoust. Soc. Am.* 109, 2112–2122.
- Galbraith, G.C., 1994. Two-channel brain-stem frequency-following responses to pure tone and missing fundamental stimuli. *Electroencephalogr. Clin. Neurophysiol. Evoked Potentials Sect.* 92, 321–330.
- Galbraith, G.C., Bhuta, S.M., Choate, A.K., Kitahara, J.M., Mullen Jr., T.A., 1998. Brain stem frequency-following response to dichotic vowels during attention. *Neuroreport* 9, 1889–1893.
- Gnansia, D., Jourdes, V., Lorenzi, C., 2008. Effect of masker modulation depth on speech masking release. *Hear. Res.* 239, 60–68.
- Glaser, E., Suter, C., Dasheiff, R., Goldberg, A., 1976. The human frequency-following response: its behavior during continuous tone and tone burst stimulation. *Electroencephalogr. Clin. Neurophysiol.* 40, 25–32.
- Griffin, S.J., Bernstein, L.R., Ingham, N.J., McAlpine, D., 2005. Neural sensitivity to interaural envelope delays in the inferior colliculus of the Guinea pig. *J. Neurophysiol.* 93, 3463–3478.
- Hairston, W.D., Letowski, T.R., McDowell, K., 2013. Task-related suppression of the brainstem frequency following response. *PLoS ONE* 8, e55215.
- Henning, G.B., 1974. Detectability of interaural delay in high-frequency complex waveforms. *J. Acoust. Soc. Am.* 55, 84–90.
- Hine, J.E., Martin, R.L., Moore, D.R., 1994. Free-field binaural unmasking in ferrets. *Behav. Neurosci.* 108, 196–205.
- Hirsh, I.J., 1950. The relation between localization and intelligibility. *J. Acoust. Soc. Am.* 22, 196–200.
- Hoormann, J., Falkenstein, M., Hohnsbein, J., 2004. Effects of spatial attention on the brain stem frequency-following potential. *Neuroreport* 15, 1539–1542.
- Hopkins, K., Moore, B.C., 2009. The contribution of temporal fine structure to the intelligibility of speech in steady and modulated noise. *J. Acoust. Soc. Am.* 125, 442–446.
- Hopkins, K., Moore, B.C., 2010. The importance of temporal fine structure information in speech at different spectral regions for normal-hearing and hearing-impaired subjects. *J. Acoust. Soc. Am.* 127, 1595–1608.
- Hopkins, K., Moore, B.C., Stone, M.A., 2008. Effects of moderate cochlear hearing loss on the ability to benefit from temporal fine structure information in speech. *J. Acoust. Soc. Am.* 123, 1140–1153.
- Joris, P.X., 2003. Interaural time sensitivity dominated by cochlea-induced envelope patterns. *J. Neurosci.* 23, 6345–6350.
- Kale, S., Heinz, M.G., 2010. Envelope coding in auditory nerve fibers following noise-induced hearing loss. *J. Assoc. Res. Otolaryngol.* 11, 657–673.
- 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.
- Keller, C.H., Takahashi, T.T., 2005. Localization and identification of concurrent sounds in the Owl's auditory space map. *J. Neurosci.* 25, 10446–10461.
- Kidd Jr., G., Mason, C.R., Rohtla, T.L., Deliwala, P.S., 1998. Release from masking due to spatial separation of sources in the identification of nonspeech auditory patterns. *J. Acoust. Soc. Am.* 104, 422–431.
- Koka, K., Read, H.L., Tollin, D.J., 2008. The acoustical cues to sound location in the rat: measurements of directional transfer functions. *J. Acoust. Soc. Am.* 123, 4297–4309.
- Krishnan, A., 2002. Human frequency-following responses: representation of steady-state synthetic vowels. *Hear. Res.* 166, 192–201.
- 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, 1180–1195.
- Lehmann, A., Schönwiesner, M., 2014. Selective attention modulates human auditory brainstem responses: relative contributions of frequency and spatial cues. *PLoS ONE* 9, e85442.
- Levy, R., Hutchison, W.D., Lozano, A.M., Dostrovsky, J.O., 2000. High-frequency synchronization of neuronal activity in the subthalamic nucleus of parkinsonian patients with limb tremor. *J. Neurosci.* 20, 7766–7775.
- 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.
- Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., Moore, B.C., 2006. Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proc. Natl. Acad. Sci.* 103, 18866–18869.
- Macpherson, E.A., Middlebrooks, J.C., 2002. Listener weighting of cues for lateral angle: the duplex theory of sound localization revisited. *J. Acoust. Soc. Am.* 111, 2219–2236.
- Marsh, J.T., Worden, F.G., 1969. Some factors modulating neural activities in peripheral auditory centers. *Brain Res.* 12, 99–111.
- McFadden, D., Pasanen, E.G., 1976. Lateralization at high frequencies based on interaural time differences. *J. Acoust. Soc. Am.* 59, 634–639.
- Middleton, J.W., Longtin, A., Benda, J., Maler, L., 2006. The cellular basis for parallel neural transmission of a high-frequency stimulus and its low-frequency envelope. *Proc. Natl. Acad. Sci.* 103, 14596–14601.
- Moore, B.C., 2008. The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people. *J. Assoc. Res. Otolaryngol.* 9, 399–406.
- Moore, B.C., Gockel, H., 2002. Factors influencing sequential stream segregation. *Acta Acust. United Acust.* 88, 320–333.
- 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.* 104, 15894–15898.
- Nakamoto, K.T., Shackleton, T.M., Palmer, A.R., 2010. Responses in the inferior colliculus of the Guinea pig to concurrent harmonic series and the effect of inactivation of descending controls. *J. Neurophysiol.* 103, 2050–2061.
- Paxinos, G., Watson, C., 1997. *The Rat Brain in Stereotaxic Coordinates*. Academic Press, San Diego.
- Peissig, J., Kollmeier, B., 1997. Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners. *J. Acoust. Soc. Am.* 101, 1660–1670.

- Ping, J., Li, N., Galbraith, G.C., Wu, X., Li, L., 2008. Auditory frequency-following responses in rat ipsilateral inferior colliculus. *Neuroreport* 19, 1377–1380.
- Rosenberg, J., Amjad, A., Breeze, P., Brillinger, D., Halliday, D., 1989. The Fourier approach to the identification of functional coupling between neuronal spike trains. *Prog. Biophys. Mol. Biol.* 53, 1–31.
- Ruggles, D., Bharadwaj, H., Shinn-Cunningham, B.G., 2012. Why middle-aged listeners have trouble hearing in everyday settings. *Curr. Biol.* 22, 1417–1422.
- Russo, N., Nicol, T., Musacchia, G., Kraus, N., 2004. Brainstem responses to speech syllables. *Clin. Neurophysiol.* 115, 2021–2030.
- Sayles, M., Stasiak, A., Winter, I.M., 2016. Neural Segregation of Concurrent Speech: Effects of Background Noise and Reverberation on Auditory Scene Analysis in the Ventral Cochlear Nucleus, Physiology, Psychoacoustics and Cognition in Normal and Impaired Hearing. Springer, pp. 389–397.
- Schmidt, A.K., Römer, H., 2011. Solutions to the cocktail party problem in insects: selective filters, spatial release from masking and gain control in tropical crickets. *PLoS ONE* 6, e28593.
- Schreiner, C., Winer, J.A., 2005. *The Inferior Colliculus*. Springer, New York.
- Shamma, S., Lorenzi, C., 2013. On the balance of envelope and temporal fine structure in the encoding of speech in the early auditory system. *J. Acoust. Soc. Am.* 133, 2818–2833.
- Shannon, R.V., Zeng, F.-G., Kamath, V., Wygonski, J., Ekelid, M., 1995. Speech recognition with primarily temporal cues. *Science* 270, 303–304.
- Shinn-Cunningham, B., Ruggles, D.R., Bharadwaj, H., 2013. How Early Aging and Environment Interact in Everyday Listening: from Brainstem to Behavior through Modeling. *Basic Aspects of Hearing*. Springer, pp. 501–510.
- Shinn-Cunningham, B.G., Schickler, J., Kopčo, N., Litovsky, R., 2001. Spatial unmasking of nearby speech sources in a simulated anechoic environment. *J. Acoust. Soc. Am.* 110, 1118–1129.
- Sinex, D.G., Li, H., 2007. Responses of inferior colliculus neurons to double harmonic tones. *J. Neurophysiol.* 98, 3171–3184.
- Skoe, E., Kraus, N., 2010. Auditory brainstem response to complex sounds: a tutorial. *Ear Hear.* 31, 302–324.
- Skoe, E., Krizman, J., Spitzer, E., Kraus, N., 2013. The auditory brainstem is a barometer of rapid auditory learning. *Neuroscience* 243, 104–114.
- 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.
- Smith, Z.M., Delgutte, B., Oxenham, A.J., 2002. Chimaeric sounds reveal dichotomies in auditory perception. *Nature* 416, 87–90.
- Smith, Z.M., Delgutte, B., 2008. Sensitivity of inferior colliculus neurons to interaural time differences in the envelope versus the fine structure with bilateral cochlear implants. *J. Neurophysiol.* 99, 2390–2407.
- 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 brainstem following short-term linguistic training. *J. Cogn. Neurosci.* 20, 1892–1902.
- Swaminathan, J., Heinz, M.G., 2012. Psychophysiological analyses demonstrate the importance of neural envelope coding for speech perception in noise. *J. Neurosci.* 32, 1747–1756.
- Swaminathan, J., Mason, C.R., Streeter, T.M., Best, V., Roverud, E., Kidd, G., 2016. Role of binaural temporal fine structure and envelope cues in cocktail-party listening. *J. Neurosci.* 36, 8250–8257.
- Trahiotis, C., Bernstein, L.R., 1986. Lateralization of bands of noise and sinusoidally amplitude-modulated tones: effects of spectral locus and bandwidth. *J. Acoust. Soc. Am.* 79, 1950–1957.
- Wang, Q., Li, L., 2015. Auditory midbrain representation of a break in interaural correlation. *J. Neurophysiol.* 114, 2258–2264.
- Wang, Q., Li, L., 2017. Modelling Envelope and Temporal Fine Structure Components of Frequency-following Responses in Rat Inferior Colliculus. *SCIENCE CHINA Technological Sciences*. <http://dx.doi.org/10.1007/s11431-016-9044-5>.
- Weinberger, N., Kitzes, L., Goodman, D., 1970. Some characteristics of the 'auditory neurophonic'. *Experientia* 26, 46–48.
- Weiss, S., Mueller, H.M., 2003. The contribution of EEG coherence to the investigation of language. *Brain Lang.* 85, 325–343.
- Weiss, S., Mueller, H.M., Schack, B., King, J.W., Kutas, M., Rappelsberger, P., 2005. Increased neuronal communication accompanying sentence comprehension. *Int. J. Psychophysiol.* 57, 129–141.
- Wightman, F.L., Kistler, D.J., 1992. The dominant role of low-frequency interaural time differences in sound localization. *J. Acoust. Soc. Am.* 91, 1648–1661.
- Wilson, J.R., Krishnan, A., 2005. Human frequency-following responses to binaural masking level difference stimuli. *J. Am. Acad. Audiol.* 16, 184–195.
- Witten, I.B., Knudsen, P.F., Knudsen, E.L., 2010. A dominance hierarchy of auditory spatial cues in barn owls. *PLoS ONE* 5, e10396.
- 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.
- Worden, F., Marsh, J., 1968. Frequency-following (microphonic-like) neural responses evoked by sound. *Electroencephalogr. Clin. Neurophysiol.* 25, 42–52.
- Yost, W.A., Wightman, F.L., Green, D.M., 1971. Lateralization of filtered clicks. *J. Acoust. Soc. Am.* 50, 1526–1531.
- Zeng, F.-G., Nie, K., Liu, S., Stickney, G., Del Rio, E., Kong, Y.-Y., Chen, H., 2004. On the dichotomy in auditory perception between temporal envelope and fine structure cues. *J. Acoust. Soc. Am.* 116, 1351–1354.