

Research Paper

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

Lu Luo ^a, Qian Wang ^{a, c}, Liang Li ^{a, b, d, *}

^a Department of Psychology and Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing, China

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

^c Beijing Key Laboratory of Epilepsy, Epilepsy Center, Department of Functional Neurosurgery, Sanbo Brain Hospital, Capital Medical University, Beijing, China

^d 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_{TFS}) and the slow-varying envelope component (FFR_{ENV}). When a single narrowband noise was presented alone, the FFR_{TFS}, but not the FFR_{ENV}, was sensitive to ITDs. When two narrowband noises were presented simultaneously, the FFR_{TFS} 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_{ENV}. These results suggest that the FFR_{TFS} and FFR_{ENV} represent two distinct types of signal processing in the auditory brainstem and contribute differentially to sound segregation based on spatial cues: the FFR_{TFS} 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

(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; Ananthakrishnan 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 Ω) 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, ± 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 10.8(at10n)-291(oc)52.(a)-3633

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

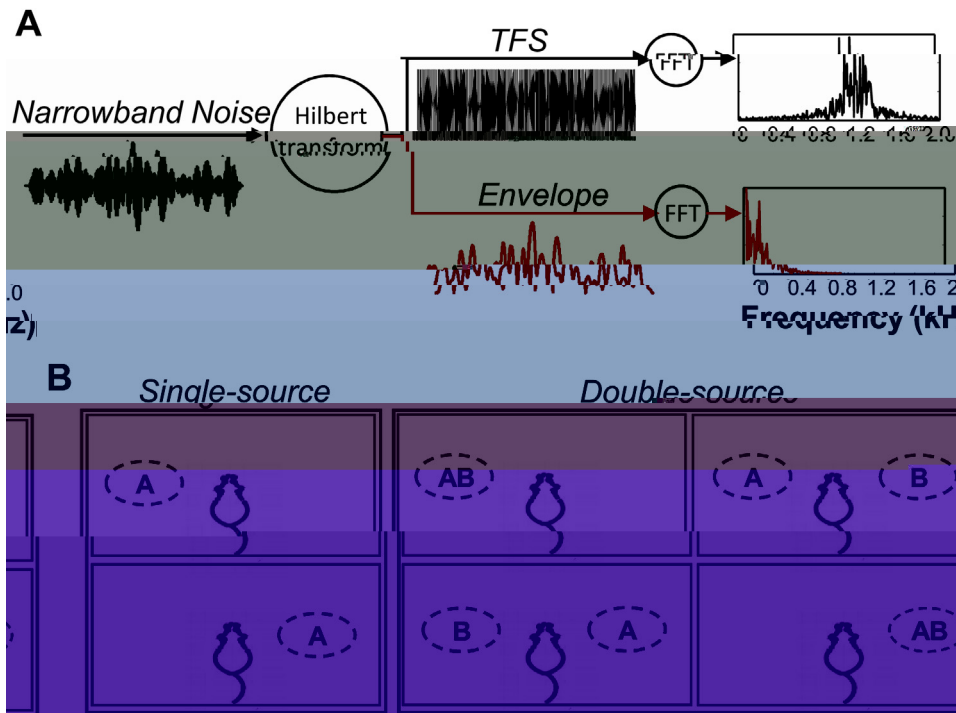


Fig. 1. Schematic of the stimulus and conditions. A: An example of narrowband noise used in this study, which is decomposed into TFS component and envelope component with Hilbert transform. The FFT spectrums of each component were also shown. B: A diagram of the stimulus conditions. The perceived locations of each sound were 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, ΔCoh 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, ΔCoh could be interpreted as the ability to convey information of individual sound by FFRs. A higher ΔCoh suggests more exclusive representation of either source A or B, and a ΔCoh 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 μ A for 10 s) via the inserted electrodes. The brains were removed and stored in 10% formalin with 30% sucrose and sectioned at 55 μ m in the frontal plane in a cryostat (-20 °C). Sections were examined to determine locations of recording electrodes.

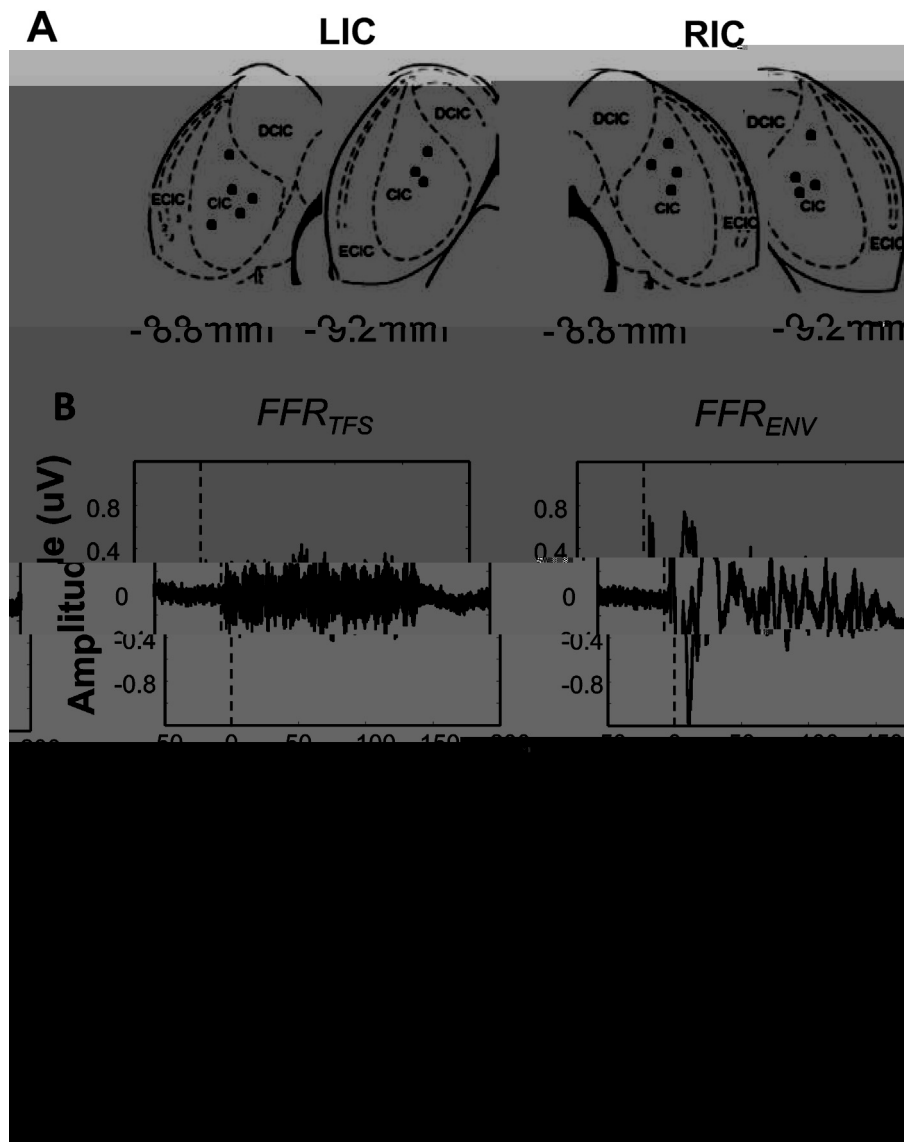


Fig. 2. Histology results and FFR examples. **A:** Histological examination of recording sites in 16 rats showed that 16 of 16 electrodes were precisely located in the central nucleus of IC (CIC), with 8 electrodes inserted in the left CIC and 8 electrodes inserted in the right CIC. DCIC, dorsal cortex 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.

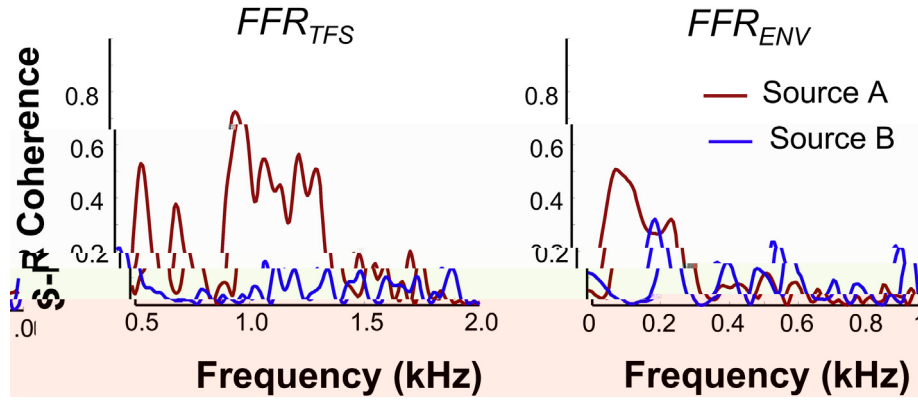


Fig. 3. An example of S-R coherence as functions of frequency for FFR_{TFS} (left) and FFR_{ENV} (right) from a recording site under single-source condition with a contralateral-leading sound named as source A. The S-R coherence with source A was plotted in red, and the S-R coherence with an uncorrelated 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 ± 0.060 SD for the TFS component, and 0.316 ± 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 ± 0.060 SD) than that with the contralateral-leading stimulus (mean = 0.399 ± 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 ± 0.096 SD) and that with the contralateral-leading stimulus (mean = 0.306 ± 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 2×2 (source-A ITD: ipsilateral leading, contralateral leading; associated laterality configuration: perceived spatial separation, perceived co-location) repeated-

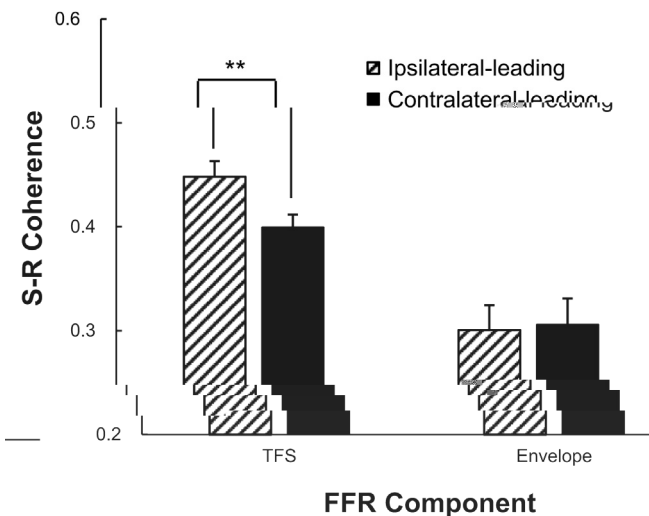


Fig. 4. Comparisons of S-R coherence values across different single-source conditions for FFR_{TFS} and FFR_{ENV} . Error bars: SE; ** $p < 0.01$.

measures ANOVA demonstrated a significant main effect of the source-A ITD ($F_{1, 15} = 89.190, p < 0.001$) and the laterality configuration ($F_{1, 15} =$

between the S-R coherence with source A and those with source B
(ipsilateral co-location condition: $t_{15} = 0.241$, $p = 0.813$; contra-
lateral co-location condition: t_{15}

in the formation of FFRs and binaural processing.

4.3. Binaural release from masking for FFR_{TFS}

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_{TFS} 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_{TFS} 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_{TFS} and FFR_{ENV}

The results of this study indicate that FFR_{TFS}, but not FFR_{ENV}, reflects the unmasking effect of ITD cues for concurrent sounds. Therefore, our study for the first time provides evidence suggesting

that FFR_{TFS} and FFR_{ENV} 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_{TFS}, but not the FFR_{ENV}, 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., Veillet, E., Berger-Vachon, C., Collet, L., Thai-Van, H., 2008. The temporal relationship between speech audi-

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