Research Paper

Neural representation of interaural correlation in human auditory brainstem: Comparisons between temporal-fine structure and envelope

Qian Wang [a, b], Hao Lu [a], Zhemeng Wu [a], Liang Li [a, c, d, *]

[a] *School of Psychological and Cognitive Sciences, Beijing Key Laboratory of Behavior and Mental Health, Peking University, Beijing, 100080, China*
[b] *Beijing Key Laboratory of Epilepsy, Epilepsy Center, Department of Functional Neurosurgery, Sanbo Brain Hospital, Capital Medical University, Beijing, 100093, China*
[c] *Speech and Hearing Research Center, Key Laboratory on Machine Perception (Ministry of Education), Peking University, Beijing, 100871, China*
[d] *Beijing Institute for Brain Disorders, Beijing, 100096, China*

ABSTRACT

Central processing of interaural correlation (IAC), which depends on the precise representation of acoustic signals from the two ears, is essential for both localization and recognition of auditory objects. A complex soundwave is initially filtered by the peripheral auditory system into multiple narrowband waves, which are further decomposed into two functionally distinctive components: the quickly-varying temporal-fine structure (TFS) and the slowly-varying envelope. In rats, a narrowband noise can evoke auditory-midbrain frequency-following responses (FFRs) that contain both the TFS component (FFR_{TFS}) and the envelope component (FFR_{Env}), which represent the TFS and envelope of the narrowband noise, respectively. These two components are different in sensitivity to the interaural time disparity. In human listeners, the present study investigated whether the FFR_{TFS} and FFR_{Env} components of brainstem FFRs to a narrowband noise are different in sensitivity to IAC and whether there are potential brainstem mechanisms underlying the integration of the two components. The results showed that although both the amplitude of FFR_{TFS} and that of FFR_{Env} were significantly affected by shifts of IAC between 1 and 0, the stimulus-to-response correlation for FFR_{TFS}, but not that for FFR_{Env}, was sensitive to the IAC shifts. Moreover, in addition to the correlation between the binaurally evoked FFR_{TFS} and FFR_{Env}, the correlation between the IAC-shift-induced change of FFR_{TFS} and that of FFR_{Env} was significant. Thus, the TFS information is more precisely represented in the human auditory brainstem than the envelope information, and the correlation between FFR_{TFS} and FFR_{Env} for the same narrowband noise suggest a brainstem binding mechanism underlying the perceptual integration of the TFS and envelope signals.

© 2018 Elsevier B.V. All rights reserved.

1. Introduction

It is astonishing to know how the brain is able to selectively track target-sound steams when multiple sounds are heard (the "cocktail party problem", Cherry, 1953). To achieve a successful sound selection, localization and recognition, a critical central process is to compute the similarity of acoustic signals at the two ears (i.e., the interaural correlation (IAC), Jeffress and Robinson, 1962). The processing of IAC also plays a role in both sound localization (Coffey et al., 2006; Franken et al., 2014; Soeta and Nakagawa, 2006) and target-object detection/recognition in noisy environments (Durlach et al., 1986; Palmer et al., 1999).

To achieve the processing of IAC, the auditory system must precisely represent dynamic sound signals. For example, depending on the bandwidth, fluctuations of both interaural phase and interaural level of narrowband noises are the important cues for processing IAC (including the detection of interaural incoherence, Goupell and Hartmann, 2006, 2007a,b). In the peripheral auditory system, a complex sound is initially filtered into multiple narrowband waves, and then each narrowband wave is decomposed into both quickly-varying temporal fine structures (TFSs) and slowly-varying envelopes (Moore, 2008; Rosen, 1992). Therefore, steady-state narrowband noises are naturally useful for examining the

* Corresponding author. School of Psychological and Cognitive Sciences, Peking University, Beijing, 100080, China.
E-mail address: liangli@pku.edu.cn (L. Li).

central representations of TFS and envelope signals. Whether the TFS and envelope components are functionally different remains debated (Apoux et al., 2013; Hopkins et al., 2008; Hopkins and Moore, 2009, 2010; Lorenzi et al., 2006; Shamma and Lorenzi, 2013; Smith et al., 2002; Swaminathan et al., 2016; Zeng et al., 2004). Also, some studies have emphasized the mutual facilitation between TFS and envelope (Moon et al., 2014; Swaminathan and Heinz, 2012). If binaural processing is critical for sound localization/recognition and the TFS is functionally different from the envelope, it is of interest to know whether the neural representation of TFS signals and that of envelope signals are different in the sensitivity to IAC.

Theoretically, a steady-state Gaussian narrowband noise with a center frequency of c Hz and a bandwidth of b Hz has not only the TFS energy around c Hz, but also the envelope energy within the frequency range between 0 and b Hz (Longtin et al., 2008). Thus, steady-state narrowband noises are very useful for extracting the TFS and envelope components when the IAC value is modulated artificially.

Scalp-recorded frequency-following responses (FFRs) are sustained neuro-electrical potentials representing the periodicity of acoustic stimuli (Worden and Marsh, 1968) with the origin site in the auditory midbrain, including the inferior colliculus (IC, Bidelman, 2015; Chandrasekaran and Kraus, 2010; Du et al., 2009; Luo et al., 2017; Marsh et al., 1974; Ping et al., 2008; Smith et al., 1975; Sohmer et al., 1977; Wang and Li, 2015, 2017, 2018; Weinberger et al., 1970). FFRs can encode both the sound TFS (e.g., Chandrasekaran and Kraus, 2010; Du et al., 2011; Galbraith, 1994; Krishnan, 2002; Krishnan and Gandour, 2009; Luo et al., 2017; Russo et al., 2004; Wang and Li, 2015, 2017, 2018) and envelope components (also called envelope-following response) (e.g., Aiken and Picton, 2006, 2008; Dolphin and Mountain, 1992, 1993; Hall, 1979; Luo et al., 2017; Shinn-Cunningham et al., 2013; Supin and Popov, 1995; Wang and Li, 2015, 2017, 2018; Zhu et al., 2013).

Some studies have suggested that these two components are different in response patterns (Luo et al., 2017; Shinn-Cunningham et al., 2017; Wang and Li, 2015, 2017, 2018). Particularly, in rats, narrowband-noise-evoked IC FFRs contain both the TFS component (FFR$_{TFS}$) and the envelope component (FFR$_{Env}$), representing the TFS and envelope of the narrowband noise, respectively (Luo et al., 2017; Wang and Li, 2015, 2017, 2018). FFR$_{TFS}$ and FFR$_{Env}$ are different in the sensitivity to the interaural time disparity (Luo et al., 2017). To date, however, it is not clear in humans whether the brainstem FFR$_{TFS}$ and FFR$_{Env}$ are different in the sensitivity to IAC. It is important to investigate whether the brainstem FFR$_{TFS}$ and FFR$_{Env}$ are different in the sensitivity to IAC, because this line of research can improve our understanding of how the spatial and non-spatial features of an auditory object are represented in the auditory brainstem, especially under noisy listening conditions.

More importantly, according to the "Binding Theory" (Treisman and Gelade, 1980), the formation of a unified perceptual object depends on certain linking mechanisms for integrating various physiologically decomposed features. Thus, there must be certain central mechanisms underlying the binding of central representation of TFSs and that of envelopes to form a unified sound percept. However, this "binding problem" has not been solved: How are FFR$_{TFS}$ and FFR$_{Env}$ bound to induce perceptual integration of TFS and envelope features?

In this study, binaurally evoked FFRs to narrowband noises were recorded from normal-hearing human participants under either the diotic (IAC = 1) or the dichotic (IAC = 0) condition. The two main issues of this study include: (1) whether FFR$_{TFS}$ and FFR$_{Env}$ are different in the sensitivity to IAC; (2) whether FFR$_{TFS}$ and FFR$_{Env}$ are correlated with a source specificity.

2. Materials and methods

2.1. Participants

Twenty-five university students (12 females and 13 males; mean age = 20.7 years, SD = 2.4 years) participated in the study. The all had symmetrical hearing (no more than 15-dB difference between the two ears) and normal pure-tone hearing thresholds (no more than 25 dB HL at each ear) between 0.125 and 8 kHz (ANSI-S3.6, 2004). All the participants provided informed consent and received stipends for their participation. The experimental procedures were approved by the Committee for Protecting Human and Animal Subjects in the School of Psychological and Cognitive Sciences, Peking University.

2.2. Acoustic stimuli

Two uncorrelated (independent) Gaussian white noises with the duration of 200 ms (including the 5-ms rise/fall periods) were generated with MATLAB (Math Works Inc., Natick, Massachusetts, USA) at the sampling rate of 20 kHz with 16-bit amplitude quantization. The noises were then filtered with a 512-point band-pass FIR filters to obtain two uncorrelated narrowband noises (sound A and sound B) with the center frequency of 500 Hz and the bandwidth of 1/3 octaves (Fig. 1B). After the filtering, the actual correlation coefficient between the two uncorrelated narrowband noises was −0.041, and both the correlation coefficient for the TFS component and that for the envelope component were less than 0.1. In this study, only the single polarity was used. The TFS and envelope signals were separated by band-pass filters during data analyses.

The noise signals were transferred using a Creative Sound Blaster (Creative SB X-Fi Surround 5.1 Pro, Creative Technology Ltd,
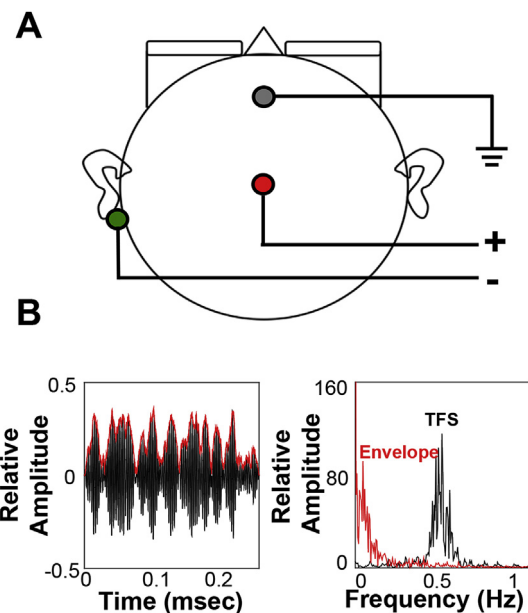

Fig. 1. *Panel A*: The illustration of the electrode positions for recording human frequency-following responses (FFRs). The active electrode (*red dot*) was placed at the vertex, the reference electrode was at the left mastoid (*green dot*), and the ground electrode (*gray dot*) was on the forehead. *Panel B*: Two temporal components of a narrowband noise stimulus (sound A, 500-Hz center frequency, 1/3-octave bandwidth). Both the waveforms (*left subpanel*) and the spectra (*right subpanel*) of the acoustic temporal fine structure (TFS, *black curves*) and the acoustic envelope (*red curves*) are presented.

Singapore) and presented to participants with insert earphones (ER-3, Etymotic Research, Elk Grove Village, IL) at the intensity of 75 dB SPL. All noises were calibrated using a Larson Davis Audiometer Calibration and Electroacoustic Testing System (AUDitTM and System 824, Larson Davis, USA).

2.3. FFR recordings

FFRs can be recorded under passive listening conditions (Skoe and Kraus, 2010). In this study, participants were instructed to watch a quiet movie of their choice during the recording sessions when they listened to acoustic stimuli in a sound attenuated chamber (EMI Shielded Audiometric Examination Acoustic Suite). Acoustically evoked potentials were recorded using a NeuroScan SynAmp system (Compumedics Limited, Victoria, Australia). Recorded neural responses were digitized at the rate of 20 kHz and collected with a 30–3000 Hz online bandpass filter. The active electrode was placed at the vertex (Cz), with reference at left mastoid. The ground electrode was placed on the forehead between Fp1 and Fp2 (Fig. 1A).

In total 4000 sweeps were presented in two blocks, one for the condition with the IAC of 1 and the other for the condition with the IAC of 0. The presentation order of the two blocks was balanced across participants.

2.4. Data analyses

The response waveforms were off-line segmented into epochs from −50 to 250 ms after the noise-stimulus onset and then baseline-corrected against the pre-stimulus level (−50 to 0 ms). Epochs exceeding $\pm 50\,\mu V$ were rejected as artifacts, and the remaining ones were averaged. Both the FFR_{TFS} and FFR_{Env} were extracted from original response potentials using 512-order low-pass (below 200 Hz) and band-pass (400–600 Hz) FIR filters (designed in MATLAB using the *fir2* function), respectively. Then the FFR_{TFS} and FFR_{Env} components were processed by long-term fast Fourier transform (FFT) to calculate the spectral amplitude Amp_f as a function of frequency f. The response latency under each condition was defined as the time of the first primary peak after the sound onset (e.g., the first positive peak shown in Fig. 3 panel A), and automatically determined by the max function in Matlab within a time window from 0 to 20 msec.

As mentioned in the Introduction, a steady-state Gaussian narrowband noise with a center frequency of c Hz and a bandwidth of b Hz has the TFS energy around c Hz and the envelope energy within the frequency range between 0 and b Hz (Longtin et al., 2008). Thus, for a narrowband noise with bandwidth b, the TFS energy distributes from the low-cut (f_{lc}) to the high-cut (f_{hc}) frequencies, and the envelope energy emerges below the frequency b. The normalized amplitude of FFR_{TFS} can be calculated by the following function (see also in Wang and Li, 2015):

$$FFR_{TFS}\_normalized\_amplitude = \sum_{l=f_{lc}}^{f_{hc}} Amp_l \Big/ \sum_{n=2}^{5000} Amp_n \qquad (1)$$

The normalized amplitude of FFR_{Env} can be calculated by the following function:

$$FFR_{Env}\_normalized\_amplitude = \sum_{l=2}^{b} Amp_l \Big/ \sum_{n=2}^{5000} Amp_n \qquad (2)$$

where the denominator represents the level of noise floor ranging from 2 to 5000 Hz while the numerator represents the spectral region of interest.

To determine the neural fidelity of acoustic inputs, the stimulus-to-response (S-R) correlation was estimated by a cross-correlation function (CCF). Both the acoustic TFS component and the envelope component of the narrowband stimulus were extracted by the Hilbert transform (Hilbert, 1912). On the other hand, the FFR_{TFS} and FFR_{Env} components were also separated by band-pass and low-pass filters described above. The value with the optimal delay (which was associated with the maximum S-R correlation) was used to access the S-R correlation coefficient (flow-process diagram see in Fig. 2).

To analyze the correlation between the IAC-sensitive TFS components and the IAC-sensitive envelope components for human scalp FFRs, both ΔFFR_{TFS} and ΔFFR_{Env} were examined. Generally, ΔFFR was defined as the relative difference between the relative amplitude of FFR when the IAC was 1 and that when the IAC was 0, as shown by function (3):

$$\Delta FFR = (FFR_{IAC\,=\,1} - FFR_{\,IAC\,=\,0}) / FFR_{\,IAC\,=\,1} \qquad (3)$$

Statistical analyses were performed with IBM SPSS Statistics 20 (SPSS Inc., Chicago, Illinois 60606). Within-subjects, repeated-measures analyses of variance (ANOVAs), t-tests, Pearson correlation, and Bonferroni *post-hoc* tests were conducted to assess differences between two stimulation conditions. The null-hypothesis rejection level was set at 0.05.

3. Results

3.1. Transient subcortical responses and interaural correlation

The results of this study showed that the steady-state narrowband noise efficiently evoked human scalp-recorded field potentials (Fig. 3A). By measuring the latency and amplitude of the sound-evoked first-peak potential for individual participants, the results showed that for the first-peak latency, no significant differences were found between the interaurally correlated condition (i.e., the diotic condition, IAC = 1) (mean = 9.02 msec, SD = 0.62 msec) and the interaurally uncorrelated condition (i.e., the dichotic condition, IAC = 0) (mean = 9.11 msec, SD = 0.52 msec) (paired t-test, $t_{24} = 1.256$, $p = 0.221$). For the first peak amplitude, no significant difference was found between the interaurally correlated condition (mean = 0.62 μV, SD = 0.19 μV) and the interaurally uncorrelated condition (mean = 0.61 μV, SD = 0.15 μV) (paired t-test, $t_{24} = 0.110$, $p = 0.913$). These results indicate that the early transient auditory-brainstem response was not sensitive to IAC.

3.2. Sensitivity of FFR_{TFS} and FFR_{Env} to interaural correlation

Fourier transform analyses of noise-evoked human scalp-recorded FFRs (Fig. 3A) clearly showed that the narrowband-noise-evoked FFRs contained both a TFS component (FFR_{TFS}) and an envelope component (FFR_{Env}). For these two components, their normalized amplitudes were calculated as a signal-to-noise ratio value (see details in the Methods). Both the FFR_{TFS} amplitude ($t_{24} = 2.186$, $p = 0.039$) and the FFR_{Env} amplitude ($t_{24} = 3.079$, $p = 0.005$) were significantly larger under the interaurally correlated condition (IAC = 1) than those under the interaurally uncorrelated condition (IAC = 0, Fig. 3B), suggesting that both the brainstem representation of the TFS component and that of the envelope component of the noise stimulus are sensitive to IAC.
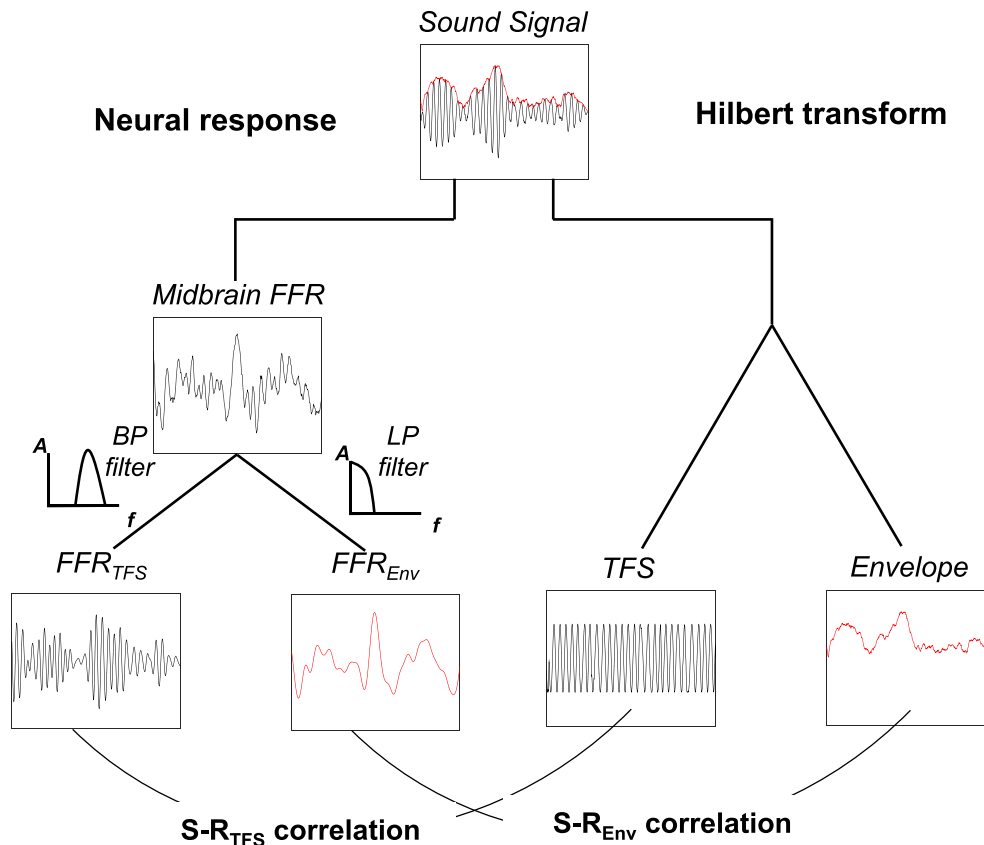
Fig. 2. The diagram showing how the stimulus-response (S-R) correlation function is obtained. The acoustic TFS and envelope components were extracted by Hilbert transformation (right flow) while the neural TFS and envelope components were separated by the filtering processes (left flow, details see the Methods). Cross-correlation functions are used to calculate the best S-R correlation coefficients for the TFS and envelope components. *TFS: temporal fine structure; Env: envelope; HP: band pass; LP: low pass.*

3.3. Sensitivity of stimulus-to-response (S-R) correlation for FFR_{TFS} and FFR_{Env} to interaural correlation

To further examine how faithfully the FFR_{TFS} and FFR_{Env} could represent the two acoustic features (TFS and envelope) of the narrowband-noise stimulus, the stimulus-to-response (S-R) correlation (between the reference noise (e.g., noise A) and FFRs) was calculated using the cross-correlation function (Fig. 2). As the examples showed in Fig. 4A, Pearson correlation tests showed that the S-R correlation between the acoustic TFS and the (neural) FFR_{TFS} was significant (all $p < 0.01$); the S-R correlation between the acoustic envelope and the (neural) FFR_{Env} was also significant (all $p < 0.01$).

Moreover, to test whether the S-R correlation was vulnerable to IAC, as shown in Fig. 4B, the S-R correlation for FFR_{TFS} to binaurally presented sound A under the interaurally correlated condition (IAC = 1, sound A was presented at each ear, indicated as "sound A + sound A", the reference stimulus for the correlation calculation was sound A) was also compared with that under the interaurally uncorrelated condition (IAC = 0, sound A was presented at one ear and sound B at the other ear, indicated as "sound A + sound B").

A 2 (reference acoustic stimulus: sound A, sound B) by 2 (stimulated condition: interaurally correlated, interaurally uncorrelated) ANOVA showed that the main effect of reference stimulus on the S-R correlations for FFR_{TFS} was significant ($F_{1,24} = 27.443$, $p < 0.001$), the main effect of stimulated condition was not significant ($F_{1,24} = 0.003$, $p = 0.958$), and the interaction effect was significant ($F_{1,24} = 18.438$, $p < 0.001$). The results of *post hoc* tests (corrected with *Bonferroni* adjustment) showed that when the reference stimulus was sound A, the S-R correlation for FFR_{TFS} to

sound A (but not that to sound B) under the interaurally correlated condition (sound A presented to both ears) was significantly larger than that under the interaurally uncorrelated condition (sound A presented to one ear and sound B presented to the other ear) ($p = 0.001$), indicating that introducing an interfering sound (i.e., sound B in the binaural stimulation condition) affected the neural fidelity of FFR_{TFS} to the reference sound A.

In addition, when the reference stimulus was also sound A for the correlation calculation, the S-R correlation for FFR_{TFS} to (binaural) sound A was significantly larger than that to (binaural) sound B under the interaurally correlated condition ($p < 0.001$), indicating that the FFR_{TFS} entrainment of sound A was stimulus-specific. Interestingly, also with the reference stimulus being sound A, the S-R correlation for FFR_{TFS} to sound B under the interaurally correlated condition (sound B presented to the two ears) was significantly smaller than S-R correlation for FFR_{TFS} under the interaurally uncorrelated condition ($p = 0.002$, sound A presented to one ear and sound B presented to the other ear), confirming that the FFR_{TFS} representation of sound B was also stimulus-specific even under noisy environments (interfered by sound A). Also, under the interaurally uncorrelated condition, no significant difference was found between the S-R correlation for FFR_{TFS} when the reference stimulus was sound A and that when the reference stimulus was sound B ($p = 0.104$).

The same tests were conducted for the S-R correlations for FFR_{Env}, but no significant results were obtained (all $p > 0.05$) (Fig. 4B), suggesting that the FFR_{TFS}, but not FFR_{Env} could specifically represent the acoustic information in the human auditory brainstem even under listening conditions with interfering sources.
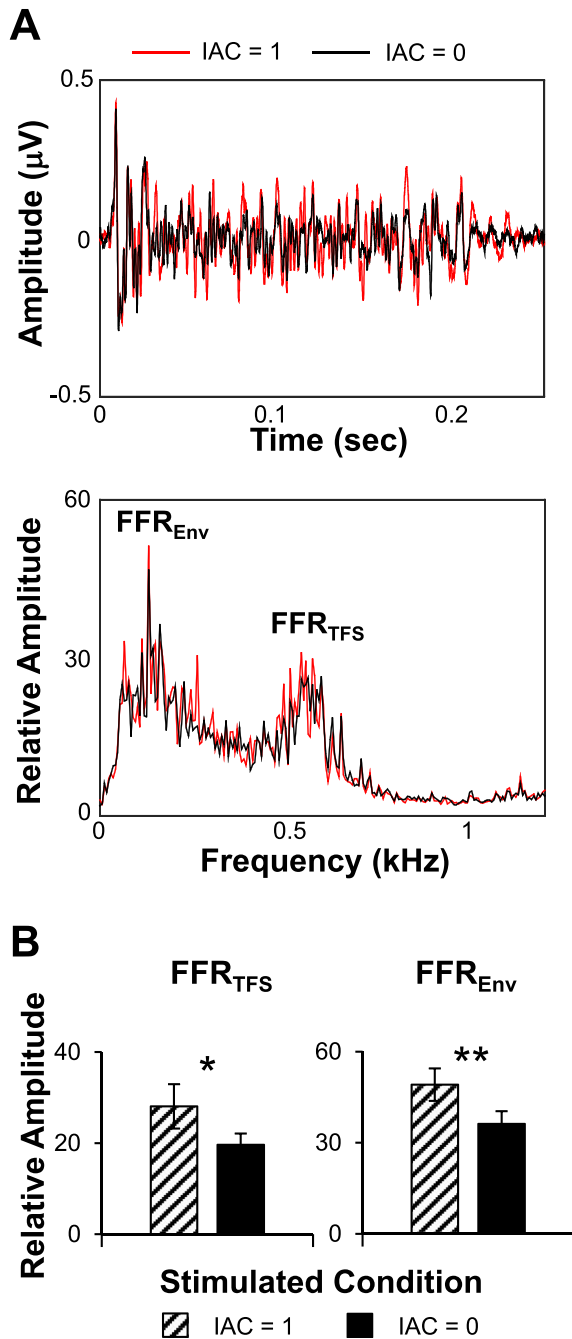
Fig. 3. *Panel A*: Waveforms of the human scalp-recorded FFRs at either the condition with the IAC of 1 (*red curves*) or the condition with the IAC of 0 (*black curves*). Group averaged waveforms (*top subpanel*) and spectra (*bottom subpanel*) of the FFRs to the narrowband noise were obtained across 25 human participants. *Panel B*: Comparison of the mean relative amplitude of the FFR_{TFS} and that of the FFR_{Env} between the condition with the IAC of 1 (*striped bars*) and the condition with the IAC of 0 (*black bars*). IAC: interaural correlation. Error bars represent the standard error of the mean. *, $p < 0.05$; **, $p < 0.01$.

3.4. Correlation between the amplitude of FFR_{TFS} and that of FFR_{Env}

To estimate whether the correlation between TFS and envelope signals are source dependent (i.e., the potentially source-specific temporal binding in the auditory brainstem), Pearson correlation tests were conducted between the normalized amplitude of FFR_{TFS} and that of FFR_{Env}. The results showed that the correlation was significant under the interaurally correlated stimulation condition

($r_{24} = 0.480$, $p = 0.015$) but marginally significant under the interaurally uncorrelated stimulation condition ($r_{24} = 0.392$, $p = 0.052$), suggesting that there existed a feature binding of the TFS and envelope components in the brainstem (Fig. 5).

Moreover, Pearson correlation tests between ΔFFR_{TFS} and ΔFFR_{Env}, which represented the IAC-sensitive TFS component and the IAC-sensitive envelope component (ΔFFR was defined as the relative difference between the relative amplitude of FFR when the IAC was 1 and that when the IAC was 0, see the Methods). The results showed a significant correlation between ΔFFR_{TFS} and ΔFFR_{Env} ($r_{24} = 0.727$, $p < 0.001$) (Fig. 6), further implying a potential functional binding between FFR_{TFS} and FFR_{Env} at the brainstem level.

Moreover, the correlation between FFR_{TFS} under the interaurally correlated condition ("sound A + sound A") and FFR_{Env} under the interaurally uncorrelated condition ("sound A + sound B") was not significant ($r_{24} = 0.157$, $p = 0.453$); the correlation between FFR_{TFS} under the interaurally uncorrelated condition ("sound A + sound B") and FFR_{Env} under the interaurally correlated condition ("sound A + sound A") was also not significant ($r_{24} = 0.332$, $p = 0.105$).

4. Discussion

As mentioned in the Introduction, in the peripheral auditory system a broadband soundwave is bandpass-filtered into a series of narrowband waves, and then each narrowband wave is further decomposed into the quickly varying TFS component and the slowly varying envelope component (Moore, 2014; Rosen, 1992). It has been debated whether these two components contribute to different perceptual performances (Apoux et al., 2013; Hopkins et al., 2008; Hopkins and Moore, 2009, 2010; Lorenzi et al., 2006; Shamma and Lorenzi, 2013; Smith et al., 2002; Swaminathan et al., 2016; Zeng et al., 2004).

Encoding of sounds in the auditory system with high fidelity is critical for survival (Dallos et al., 1996; Yost and Sheft, 1993). The neural strategies for precisely representing acoustic features of stimuli may depend on the hierarchically functional organization of the auditory system (for reviews see Nelken and Bar-Yosef, 2008; Shamma et al., 2011). Recently, it has been confirmed that both the TFS component and the envelope component of a narrowband noise can be represented in brainstem FFRs (Luo et al., 2017; Wang and Li, 2015, 2017, 2018), showing that narrowband noises are particularly useful in this line of research.

Although a narrowband noise is aperiodic, it contains both TFS and envelope acoustic components (Longtin et al., 2008), and particularly, can evoke both the TFS and envelope components of FFRs recorded in the IC (Luo et al., 2017; Wang and Li, 2015, 2017, 2018). More importantly, the usage of narrowband noises allows manipulations of the similarity (correlation) between acoustic stimuli presented at the two ears (Goupell and Hartmann, 2006, 2007a,b; Jeffress and Robinson, 1962), thereby being particularly applicable to the investigation of the IAC processing in the auditory system.

4.1. Sensitivity to the IAC

Using scalp-recorded onset responses and FFRs to either diotic (IAC = 1) or dichotic (IAC = 0) narrowband noises, this study examined in humans whether the onset, TFS, and envelope components of the stimulus can be precisely represented by onset responses, FFR_{TFS}, and FFR_{Env} respectively, because the sensitivity of these ERP components to IAC depends on how precisely these acoustic components are represented in the human auditory brainstem.

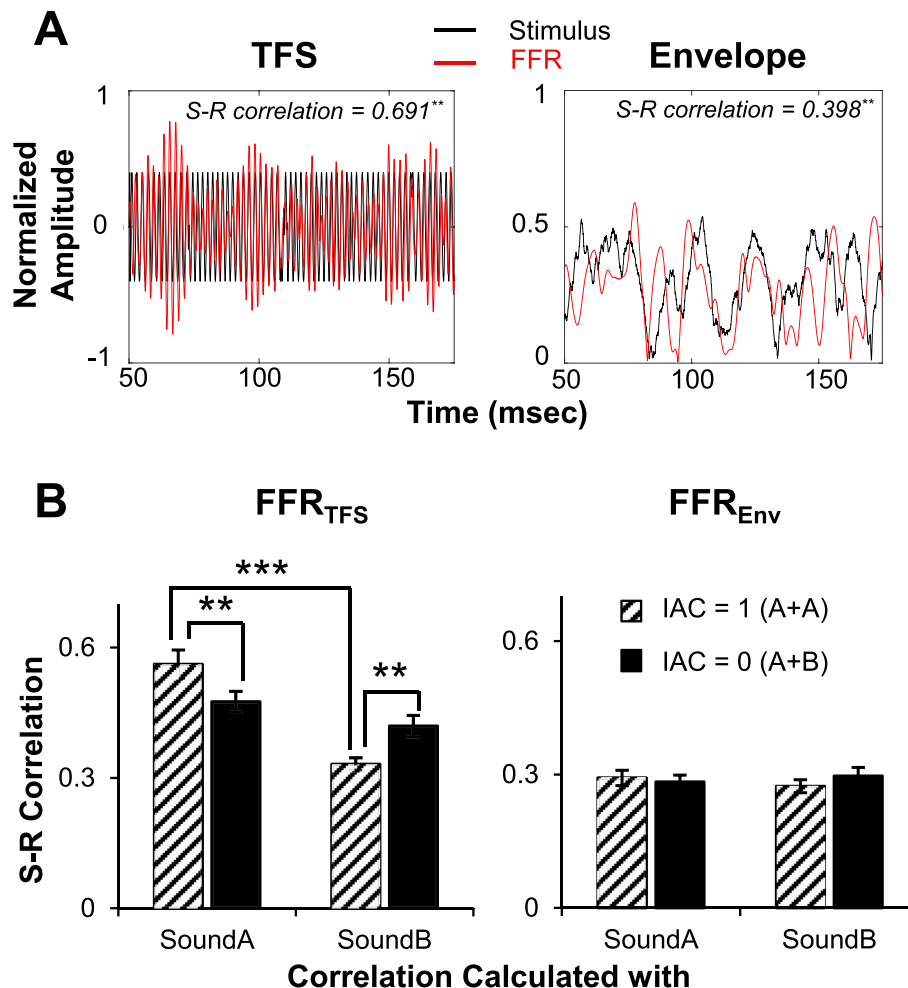Previous studies have shown that click-evoked auditory

Fig. 4. *Panel A*: Examples for illustrating the stimulus-response (S–R) correlation for the temporal fine structure (TFS) and that for the envelope by exhibiting comparisons between the stimulus waveform (*black curves*) and the FFR waveform (*red curves*). Pearson *r* tests were conducted to estimate whether the S–R correlation value was significant. **Panel B**: Comparisons of the S–R correlations for the FFR$_{TFS}$ and those for the FFR$_{Env}$ between the two listening conditions (IAC = 1 or 0). The S–R correlations were calculated between the neural response with either sound A or sound B being the reference stimulus. The horizontal axis indicates the reference stimulus for the comparisons (either sound A or sound B). IAC: interaural correlation. Error bars represent the standard error of the mean. **, p < 0.01; ***, p < 0.001.

brainstem responses were sensitive to interaural time difference (Fowler, 2004; Wrege and Starr, 1981). The results of this study showed that neither the latency nor the amplitude of the onset response to the noise stimulus was affected by the IAC, indicating that without a sufficiently temporal buildup of interaural processing, the early transient (onset) auditory-brainstem response was not sensitive to IAC. On the other hand, however, both FFR$_{TFS}$ and FFR$_{Env}$ were significantly sensitive to IAC, suggesting that both the TFS signal and the envelope signal of the narrowband-noise stimulus are represented by FFRs in the auditory brainstem. However, the results of this study further showed that the accuracy of the FFR$_{TFS}$ in representing acoustic features is much better than that of the FFR$_{Env}$ (see below).

In this study, to further examine how faithfully the TFS and envelope components of the noise stimulus can be represented by the FFR$_{TFS}$ and FFR$_{Env}$, respectively, the S–R correlation was calculated using the cross-correlation function. The results showed that the S–R correlation between the acoustic TFS and the (neural) FFR$_{TFS}$ was significant, and the S–R correlation between the acoustic envelope and the (neural) FFR$_{Env}$ was also significant (all *p* < 0.01). Moreover, the S–R correlation between the acoustic TFS of a reference noise (e.g., noise A) and the binaurally induced FFR$_{TFS}$ was significantly larger when the reference noise was presented at each

ear (under the diotic stimulation condition, IAC = 1) than when the reference noise was presented at one ear and an independent noise (e.g., noise B) was presented at the other ear (under the dichotic stimulation condition, IAC = 0). Thus, the S–R correlation for FFR$_{TFS}$ is sensitive to the IAC. Surprisingly, this sensitivity of S–R correlation to IAC does not occur for FFR$_{Env}$.

The sensitivity of the S–R correlation for FFR$_{TFS}$ to IAC was further confirmed by the results of this study that the S–R correlation between the acoustic TFS of a reference noise (e.g., noise A) and the binaurally induced FFR$_{TFS}$ was significantly larger when the reference noise was presented at one ear and another independent noise (e.g., noise B) was presented at the other ear (under the dichotic stimulation condition, IAC = 0) than when the other different noise (i.e., noise B) was presented at each ear (under the diotic stimulation condition, IAC = 1). Also, this sensitivity of S–R correlation to IAC did not occur for FFR$_{Env}$.

As mentioned in the Introduction, whether the TFS and envelope components contribute to different perceptual performances has been a long-term debate (Apoux et al., 2013; Hopkins et al., 2008; Hopkins and Moore, 2009, 2010; Lorenzi et al., 2006; Shamma and Lorenzi, 2013; Smith et al., 2002; Swaminathan et al., 2016; Zeng et al., 2004). The results of this study showed that the S–R correlation for FFR$_{TFS}$, but not that for FFR$_{Env}$, was sensitive to IAC,
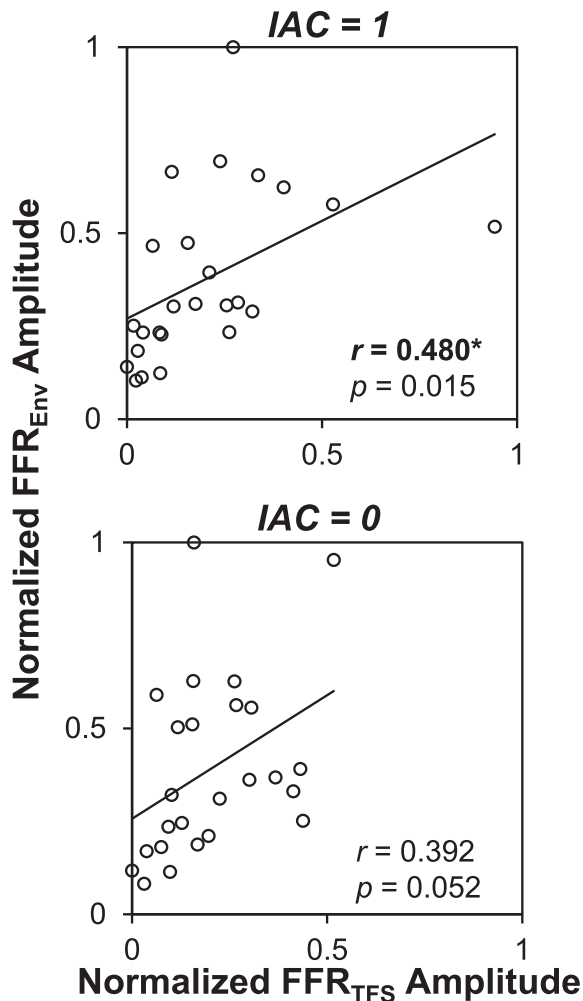
Fig. 5. Correlations between the relative amplitude of FFR_{TFS} and the FFR_{Env} under either the interaurally correlated condition (IAC = 1, *top panel*) or the interaurally uncorrelated condition (IAC = 0, *bottom panel*). IAC: interaural correlation. *, $p < 0.05$.

Fig. 6. Correlation between the IAC-sensitive ΔFFR_{TFS} and the IAC-sensitive ΔFFR_{Env}. ΔFFR was defined as the relative difference between the FFR relative amplitude when the IAC was 1 and that when the IAC was 0 as the following function: $\Delta FFR = (FFR_{IAC = 1} - FFR_{IAC = 0})/FFR_{IAC = 1}$. Individual participants' ΔFFR_{TFS} values (presented along the abscissa) were significantly correlated to their ΔFFR_{Env} values (presented along the ordinate) across 25 participants ($r = 0.784$, $p < 0.001$). IAC: interaural correlation. ***, $p < 0.001$.

confirming that the accuracy of the FFR_{TFS} in representing acoustic features is much higher than that of the FFR_{Env}, thereby supporting the functional dichotomy between FFR_{TFS} and FFR_{Env} (also see Luo et al., 2017; Wang and Li, 2015, 2017, 2018). Compared to those of envelope signals, the neural representations of TFS signals may play a more important role in the perceptual segregation of concurrent sound sources, because the FFR_{TFS} benefits more from binaural cues.

4.2. Correlations between FFR_{TFS} and FFR_{Env}

Theoretically, any perceptual systems must intrinsically organize physical features of the outside world into integrated perceptual objects (Griffiths and Warren, 2004; Treisman and Gelade, 1980). However, this binding mechanism is largely unknown. According to the "Binding Theory" (Treisman and Gelade, 1980; Spence, 2011; Burwick, 2014), the formation of a unified perceptual object depends on certain linking mechanisms for integrating various physiologically decomposed features (such as higher spatial-frequency textures and lower spatial-frequency contours of a single visual image). Although there has been a long-stand debate as to whether TFS and envelope are separately associated with different perceptual functions (e.g., Apoux et al., 2013; Rosen, 1992; Seeber and Hafter, 2011; Smith et al., 2002;
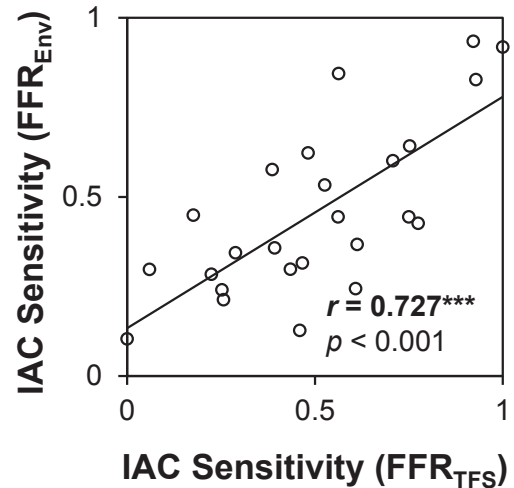
Srinivasan and Zahorik, 2014; Swaminathan and Heinz, 2012; Xu and Pfingst, 2003; Zeng et al., 2004), it is clear that listeners do not perceive the TFS and envelope components separately, and various auditory/speech perceptions must be based on the combined processing of TFS and envelope signals (Gnansia et al., 2009; Huang et al., 2011; Hopkins and Moore, 2009; Moon et al., 2014). Thus, there must be certain mechanisms underlying the integration between the TFSs and envelopes that belong to a certain sound source, leading to that listeners only perceive unified auditory objects. Indeed, it has been reported that the integration processing of TFS and envelope signals is essential to various auditory/speech perceptions (e.g., Gnansia et al., 2009; Huang et al., 2011; Hopkins and Moore, 2009; Moon et al., 2014; Schimmel et al., 2008). Recent psychoacoustic studies have also emphasized a mutual facilitation when TFS and envelope are highly coherent (Swaminathan and Heinz, 2012; Moon et al., 2014), particularly in a noisy multiple-source environment, suggesting a feature-specific binding between the TFS and envelope components.

To estimate potential mechanisms underlying the binding of TFS and envelope signals in the auditory brainstem, in this study Pearson correlation tests were conducted between the normalized amplitude of FFR_{TFS} and that of FFR_{Env}. The results showed that the correlation between FFR_{TFS} and FFR_{Env} was significant with the source specificity particularly under the interaurally correlated listening condition. Moreover, Pearson correlation between ΔFFR_{TFS} and ΔFFR_{Env}, (ΔFFR, the relative difference between the FFR relative amplitude when the IAC is 1 and that when the IAC is 0) was also significant.

In a "cocktail-party" listening condition with multiple sound sources, if each of the narrow-band-signal series in the peripheral auditory system for an individual source is decomposed into both quickly-varying TFSs and slowly-varying envelopes, how are listeners able to perceive individual sound images? If binaural processing is critical for sound localization/recognition under such adverse listening conditions and the brainstem representation of TFS signals (i.e. FFR_{TFS}) is more sensitive to changes in IAC than the brainstem representation of envelope signals (i.e., FFR_{Env}), the functional dichotomy theory is not sufficient to elucidate how

distinctive auditory objects are able to form under the multiple-source conditions. Thus, the results of this study will encourage further studies in this line of research.

4.3. Limitations

It should be noted that the scalp-recorded FFRs may also be affected by the phase cancelling effect on volume conducted signals under the out-of-phase condition, particularly when the binaural sounds are uncorrelated. Since FFR_{TFS} fluctuates faster than FFR_{Env}, the phase cancelling effect is likely more prevalent for FFR_{TFS} than FFR_{Env}. Therefore, the difference in IAC sensitivity between FFR_{TFS} and FFR_{Env} should be further investigated in the future.

5. Conclusions

(1) The TFS and envelope components of a narrowband noise are faithfully represented in the human auditory brainstem by the FFR_{TFS} and FFR_{Env}, respectively.
(2) Both sustained FFR_{TFS} and FFR_{Env}, but not the transient brainstem onset responses, are sensitive to IAC, and the FFR_{TFS} is more sensitive to IAC than the FFR_{Env}.
(3) The brainstem FFR_{TFS} and FFR_{Env} are correlated to each other with the source specificity.
(4) Both the highly precise representation of the TFS information in the human auditory brainstem and the source-specific correlation between FFR_{TFS} and FFR_{Env} indicate a brainstem-processing strategy underlying feature integrations for auditory perception.

Acknowledgments

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.heares.2018.05.015.

References

Aiken, S.J., Picton, T.W., 2006. Envelope following responses to natural vowels. Audiol. Neurotol. 11, 213—232.
Aiken, S.J., Picton, T.W., 2008. Envelope and spectral frequency-following responses to vowel sounds. Hear. Res. 245, 35—47.
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.
Bidelman, G.M., 2015. Multichannel recordings of the human brainstem frequency-following response: scalp topography, source generators, and distinctions from the transient ABR. Hear. Res. 323, 68—80.
Burwick, T., 2014. The binding problem. Wiley interdisciplinary reviews. Cognit. Sci. 5, 305—315.
Chandrasekaran, B., Kraus, N., 2010. The scalp-recorded brainstem response to speech: neural origins and plasticity. Psychophysiology 47, 236—246.
Cherry, E.C., 1953. Some experiments on the recognition of speech, with one and with two ears. J. Acoust. Soc. Am. 25, 975—979.
Coffey, C.S., Ebert, C.S., Marshall, A.F., Skaggs, J.D., Falk, S.E., Crocker, W.D., Fitzpatrick, D.C., 2006. Detection of interaural correlation by neurons in the superior olivary complex, inferior colliculus and auditory cortex of the unanesthetized rabbit. Hear. Res. 221, 1—16.
Dallos, P., Popper, A.N., Fay, R.R., 1996. The Cochlea, Vol. 8 of Springer Handbook of Auditory Research.
Dolphin, W.F., Mountain, D.C., 1992. The envelope following response: scalp potentials elicited in the Mongolian gerbil using sinusoidally AM acoustic signals. Hear. Res. 58, 70—78.
Dolphin, W.F., Mountain, D.C., 1993. The envelope following response (EFR) in the Mongolian gerbil to sinusoidally amplitude-modulated signals in the presence of simultaneously gated pure tones. J. Acoust. Soc. Am. 94, 3215—3226.
Du, Y., Kong, L.Z., Wang, Q., Wu, X.H., Li, L., 2011. Auditory frequency-following response: a neurophysiological measure for studying the "cocktail-party problem". Neurosci. Biobehav. Rev. 35, 2046—2057.
Du, Y., Ma, T.F., Wang, Q., Wu, X.H., 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.
Durlach, N.I., Gabriel, K.J., Colburn, H.S., Trahiotis, C., 1986. Interaural correlation discrimination II Relation to binaural unmasking. J. Acoust. Soc. Am. 79, 1548—1557.
Fowler, C.G., 2004. Electrophysiological evidence for binaural processing in auditory evoked potentials: the binaural interaction component. Semin. Hear. 25, 39—49.
Franken, T.P., Bremen, P., Joris, P.X., 2014. Coincidence detection in the medial superior olive: mechanistic implications of an analysis of input spiking patterns. Front. Neural Circ. 8.
Galbraith, G.C., 1994. Two-channel brain-stem frequency-following responses to pure tone and missing fundamental stimuli. Electroencephalogr. Clin. Neurophysiol. 92, 321—330.
Gnansia, D., Péan, V., Meyer, B., Lorenzi, C., 2009. Effects of spectral smearing and temporal fine structure degradation on speech masking release. J. Acoust. Soc. Am. 125, 4023—4033.
Goupell, M.J., Hartmann, W.M., 2006. Interaural fluctuations and the detection of interaural incoherence: bandwidth effects. J. Acoust. Soc. Am. 119, 3971—3986.
Goupell, M.J., Hartmann, W.M., 2007a. Interaural fluctuations and the detection of interaural incoherence. II. Brief duration noises. J. Acoust. Soc. Am. 121, 2127—2136.
Goupell, M.J., Hartmann, W.M., 2007b. Interaural fluctuations and the detection of interaural incoherence. III. Narrowband experiments and binaural models. J. Acoust. Soc. Am. 122, 1029—1045.
Griffiths, T.D., Warren, J.D., 2004. What is an auditory object? Nat. Rev. Neurosci. 5, 887—892.
Hall, J.W., 1979. Auditory brainstem frequency following responses to waveform envelope periodicity. Science 205, 1297—1299.
Hilbert, D., 1912. Grundzüge einer Allgemeinen Theorie der Linearen Integralgleichungen Leipzig. BG Teubner.
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.
Huang, Y., Li, J.Y., Zou, X.F., Qu, T.S., Wu, X.H., Mao, L.H., Wu, Y.H., Li, L., 2011. Perceptual fusion tendency of speech sounds. J. Cognit. Neurosci. 23, 1003—1014.
Jeffress, L.A., Robinson, D.E., 1962. Formulas for the coefficient of interaural correlation for noise. J. Acoust. Soc. Am. 34, 10.
Krishnan, A., 2002. Human frequency-following responses: representation of steady-state synthetic vowels. Hear. Res. 166, 192—201.
Krishnan, A., Gandour, J.T., 2009. The role of the auditory brainstem in processing linguistically-relevant pitch patterns. Brain Lang. 10, 135—146.
Longtin, A., Middleton, J.W., Cieniak, J., Maler, L., 2008. Neural dynamics of envelope coding. Math. Biosci. 214, 87—99.
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. Unit. States Am. 103, 18866—18869.
Luo, L., Wang, Q., Li, L., 2017. Neural representations of concurrent sounds with overlapping spectra in rat inferior colliculus: comparisons between temporal-fine structure and envelope. Hear. Res. 353, 87—96.
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.
Moon, I.J., Won, J.H., Park, M.H., Ives, D.T., Nie, K., Heinz, M.G., Lorenzi, C., Rubinstein, J.T., 2014. Optimal combination of neural temporal envelope and fine structure cues to explain speech identification in background noise. J. Neurosci. 34, 12145—12154.
Moore, B.C., 2014. Auditory Processing of Temporal fine Structure: Effects of Age and Hearing Loss. World Scientific.
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.
Nelken, I., Bar-Yosef, O., 2008. Neurons and objects: the case of auditory cortex. Front. Neurosci. 2, 107.
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.
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.
Rosen, S., 1992. Temporal information in speech: acoustic, auditory and linguistic

aspects. Phil. Trans. Roy. Soc. B 336, 367–373.

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

Schimmel, O., van de Par, S., Breebaart, J., Kohlrausch, A., 2008. Sound segregation based on temporal envelope structure and binaural cues. J. Acoust. Soc. Am. 124, 1130–1145.

Seeber, B.U., Hafter, E.R., 2011. Failure of the precedence effect with a noise-band vocoder. J. Acoust. Soc. Am. 129, 1509–1521.

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.

Shamma, S.A., Elhilali, M., Micheyl, C., 2011. Temporal coherence and attention in auditory scene analysis. Trends Neurosci. 34, 114–123.

Shinn-Cunningham, B., Ruggles, D.R., Bharadwaj, H., 2013. How early aging and environment interact in everyday listening: from brainstem to behavior through modeling. In: Basic Aspects of Hearing: Physiology and Perception. Springer, New York, pp. 501–510.

Shinn-Cunningham, B., Varghese, L., Wang, L., Bharadwaj, H., 2017. Individual differences in temporal perception and their implications for everyday listening. In: The Frequency-following Response. Springer International Publishing, pp. 159–192.

Skoe, E., Kraus, N., 2010. Auditory brainstem response to complex sounds: a tutorial. Ear Hear. 31, 302–324.

Smith, Z.M., Delgutte, B., Oxenham, A.J., 2002. Chimaeric sounds reveal dichotomies in auditory perception. Nature 416, 87–90.

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.

Soeta, Y., Nakagawa, S., 2006. Auditory evoked magnetic fields in relation to interaural time delay and interaural correlation. Hear. Res. 220, 106–115.

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

Spence, C., 2011. Crossmodal correspondences: a tutorial review. Atten. Percept. Psychophys. 73, 971–995.

Srinivasan, N.K., Zahorik, P., 2014. Enhancement of speech intelligibility in reverberant rooms: role of amplitude envelope and temporal fine structure. J. Acoust. Soc. Am. 135, 239–245.

Supin, A.Y., Popov, V.V., 1995. Envelope-following response and modulation transfer function in the dolphin's auditory system. Hear. Res. 92, 38–46.

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 Jr., G., 2016. Role of binaural temporal fine structure and envelope cues in cocktail-party listening. J. Neurosci. 36, 8250–8257.

Treisman, A., Gelade, G., 1980. A feature-integration theory of attention. Cognit. Psychol. 12, 97–136.

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. Sci. China Technol. Sci. 60, 966–973.

Wang, Q., Li, L., 2018. Differences between auditory frequency-following responses and onset responses: intracranial evidence from rat inferior colliculus. Hear. Res. 357, 25–32.

Weinberger, N.M., Kitzes, L.M., Goodman, D.A., 1970. Some characteristics of the 'auditory neurophonic'. Experientia 26, 46–48.

Worden, F.G., Marsh, J.T., 1968. Frequency-following (microphonic-like) neural responses evoked by sound. Electroencephalogr. Clin. Neurophysiol. 25, 42–52.

Wrege, K.S., Starr, A., 1981. Binaural interaction in human auditory brainstem evoked potentials. Arch. Neurol. 38, 572–580.

Xu, L., Pfingst, B.E., 2003. Relative importance of temporal envelope and fine structure in lexical-tone perception (L). J. Acoust. Soc. Am. 114, 3024–3027.

Yost, W.A., Sheft, S., 1993. Auditory perception. In: Yost, W.A., et al. (Eds.), Human Psychophysics. Springer, New York, pp. 193–236.

Zeng, F.G., Nie, K.B., Liu, S., Stickney, G., Del Rio, E., Kong, Y.Y., Chen, H.B., 2004. On the dichotomy in auditory perception between temporal envelope and fine structure cues (L). J. Acoust. Soc. Am. 116, 1351–1354.

Zhu, L., Bharadwaj, H., Xia, J., Shinn-Cunningham, B., 2013. A comparison of spectral magnitude and phase-locking value analyses of the frequency-following response to complex tones. J. Acoust. Soc. Am. 134, 384–395.