

# Relative contribution of envelope and fine structure to the subcortical encoding of noise-degraded speech

Gavin M. Bidelman<sup>a)</sup>

*School of Communication Sciences & Disorders, University of Memphis, Memphis, Tennessee 38152, USA*  
*g.bidelman@memphis.edu*

**Abstract:** Brainstem frequency-following responses (FFR) were elicited to the speech token /ama/ in noise containing only envelope (ENV) or fine structure (TFS) cues to assess the relative contribution of these temporal features to the neural encoding of degraded speech. Successive cue removal weakened FFRs with noise having the most deleterious effect on TFS coding. Neuro-acoustic and response-to-response correlations revealed speech-FFRs are dominated by stimulus ENV for clean speech, with TFS making a stronger contribution in moderate noise levels. Results suggest that the relative weighting of temporal ENV and TFS cues to the neural transcription of speech depends critically on the degree of noise in the soundscape.

© 2016 Acoustical Society of America

[DDO]

**Date Received:** May 11, 2016    **Date Accepted:** October 5, 2016

## 1. Introduction

The temporal amplitude characteristics of speech can be decomposed into slowly varying envelope (ENV) fluctuations and rapidly varying temporal fine structure (TFS).<sup>1,2</sup> While not completely orthogonal,<sup>3</sup> the ENV-TFS dichotomy of speech acoustics is useful in understanding the relative role of temporal cues in speech perception as well as their susceptibility to acoustic degradations (e.g., noise). Indeed, perceptual studies demonstrate that listeners exploit temporal ENV cues for robust comprehension in unadulterated (i.e., clean) listening environments.<sup>1,2,4</sup> For example, whispered speech maintains the speech ENV but contains TFS that is entirely noise, yet human listeners still achieve robust speech recognition.<sup>5</sup> Robust intelligibility is also achievable under cases of severe ENV distortion (e.g., peak clipping<sup>6</sup>) or when only fine structure is present (e.g., sine-wave speech),<sup>7</sup> suggesting that TFS cues also contribute to successful spoken word recognition. TFS is thought to be especially important in noisy environments,<sup>4,8–10</sup> degradations to the speech ENV (occluded by noise) are at least partly counteracted by exploiting TFS cues to aid degraded speech recognition.<sup>4,8–10</sup>

Despite their differential importance to normal (e.g., voiced vs unvoiced<sup>5</sup>) and impaired (e.g., cochlear implant) speech perception, surprisingly few studies have evaluated the neurophysiological basis of temporal ENV and TFS coding. Current understanding of temporal coding rests mainly in single-unit studies in noise-exposed animal models.<sup>11,12</sup> Recent reports show that in cases of sensorineural hearing loss, auditory nerve fiber responses show an overall enhancement of ENV coding for clean signals containing no noise.<sup>11</sup> On the other hand, deficits in auditory nerve TFS phase-locking (with hearing loss) emerge primarily in noisy listening conditions.<sup>12</sup> Collectively, animal studies imply that the relative weighting of ENV and TFS cues carried by auditory nerve fibers might change depending on the listening condition (i.e., clean vs noise) and audiometric status (i.e., normal vs impaired hearing). However, surprisingly few studies have examined ENV and TFS coding in normal-hearing humans, despite the importance of these cues to normal and degraded speech communication.

The brainstem frequency-following response (FFR) provides an early window into the neural encoding of speech in human listeners at subcortical levels of auditory processing.<sup>13,14</sup> The FFR reflects sustained, subcortical neural phase-locking to the time-frequency characteristics of acoustic signals including speech.<sup>13</sup> Previous FFR studies demonstrate weakened and delayed brainstem responses to speech in the presence of acoustic interferences (e.g., noise and reverberation).<sup>15</sup> While these studies

---

<sup>a)</sup>Also at: Institute for Intelligent Systems, University of Memphis, Memphis, Tennessee 38152, USA.

reveal noise-related changes in the *aggregate* subcortical transcription of speech, it is still unclear how noise alters the individual temporal constituents of the speech waveform (i.e., ENV vs TFS information).

To this end, we measured brainstem FFRs in human listeners elicited by speech sounds containing only ENV or TFS cues in order to assess the relative contribution of these temporal features to the neural encoding of noise-degraded speech. Based on animal<sup>11,12</sup> and human psychophysical studies,<sup>8–10</sup> we hypothesized that noise would have a differential effect on the subcortical encoding of ENV and TFS properties of speech with ENV cues dominating neural representations for clean speech and TFS making larger contributions to speech coding in the presence of noise.

## 2. Methods

### 2.1 Participants

Fifteen young adults (age: 23–34 years; 7 male, 8 female) participated in the experiment. All participants were native speakers of English, right-handed, had normal hearing (i.e., audiometric thresholds  $\leq 25$  dB hearing level; 500–4000 Hz), and reported no previous history of neuropsychiatric illnesses. Musical training can improve degraded speech processing.<sup>15,16</sup> As such, all participants were required to have minimal ( $< 5$  years) formal musical training. All were paid and gave written informed consent in compliance with a protocol approved by the IRB of The University of Memphis.

### 2.2 TFS and ENV speech stimuli

FFRs were elicited by the 300 ms (10 ms rise/fall) /vCv/ speech token /ama/ used in our previous studies on noise-degraded speech coding.<sup>17</sup> The token was a natural production recorded from a male speaker whose time-varying pitch prosody fell gradually over its duration (i.e., 120 to 88 Hz; see Ref. 17 for additional details). This vCv contains only voiced elements and is thus optimal for eliciting brainstem FFRs which require low pitched, periodic stimuli.<sup>13</sup> Speech tokens were processed to extract the temporal envelope and fine structure using the Hilbert transform<sup>2</sup> (Fig. 1). This resulted in three stimulus conditions: the original signal (i.e., ENV + TFS), the envelope (ENV) component, and fine structure (TFS) signal of the original speech signal. The ENV signal was then lowpass filtered ( $f_c = 100$  Hz)<sup>1</sup> and multiplied by broadband noise to create a speech stimulus that preserved the amplitude envelope but contained no informative spectral cues (Fig. 1, “ENV”). Note that this bandwidth preserves the so-called “periodicity envelope” of speech.<sup>18</sup> Contrastively, the TFS signal contained spectral details of the original token but did not contain prominent envelope fluctuations that are characteristic of speech (Fig. 1, “TFS”).

The three stimulus conditions were then degraded with noise babble<sup>17</sup> at various signal-to-noise ratios (SNRs) to examine parametric changes in the FFR coding of speech ENV and TFS as a function of acoustic interference. In addition to the “clean” stimuli (SNR =  $\infty$  dB), noise-degraded speech was created at +10 and +5 dB SNRs. SNR was manipulated by changing the level of the masker rather than the level of the signal, ensuring the noise interference was inversely correlated with overall sound level.<sup>17</sup> The babble noise was presented continuously such that it was not time-locked to the stimulus presentation.

Listeners heard 2000 repetitions of each of the nine tokens (3 speech conditions  $\times$  3 SNRs) presented with fixed (i.e., rarefaction) polarity and delivered binaurally through ER-30 insert earphones (Etymotic Research) at 80 dB sound pressure level (ISI = 10 ms). Extended acoustic tubing of these headphones (20 ft) eliminated stimulus artifact from overlapping neural responses.<sup>17,19</sup> Stimulus presentation was controlled by MATLAB<sup>®</sup> (The MathWorks, Inc.) routed to a TDT RP2 interface (Tucker-Davis Technologies).

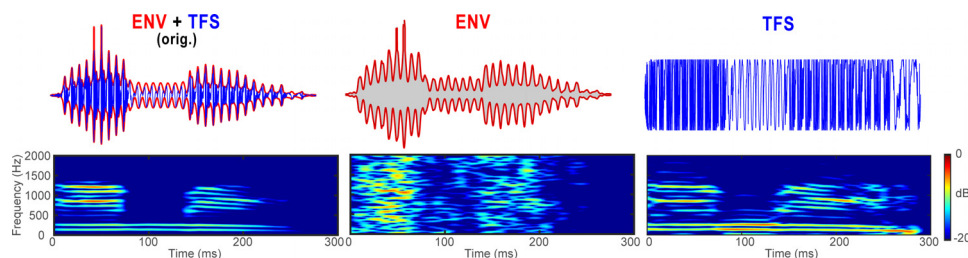


Fig. 1. (Color online) Stimulus waveforms and spectrograms. Speech stimuli were processed to contain only the acoustic envelope (ENV) or temporal fine structure (TFS).

### 2.3 FFR recording and analysis

Neuroelectric activity was recorded differentially between Ag/AgCl disk electrodes placed on the scalp at the high forehead (~Fpz) referenced to linked mastoids (A1/A2) (mid-forehead electrode = ground). Interelectrode impedance was  $\leq 3$  k $\Omega$ . EEGs were digitized at 10 kHz (SynAmps RT amplifiers; Neuroscan) using an online passband of DC–4000 Hz. EEGs were then epoched (0–310 ms window) and averaged in the time domain to derive FFRs for each condition. Sweeps exceeding  $\pm 50$   $\mu$ V were rejected as artifacts prior to averaging. FFRs were then bandpass filtered (90–2500 Hz) for response visualization and quantification.

From each FFR spectrum, we measured the fundamental (F0) and first formant (F1) frequency to quantify voice “pitch” and “timbre” coding<sup>14,16</sup> for each noise and stimulus condition. Fast Fourier transforms (FFTs) were computed from the steady-state portion of brainstem time-waveforms (average of [20–120 ms] and [180–240 ms] segments). We estimated the magnitude of the response at F0 and F1, taken as the maximum spectral energy (on a linear scale) in the frequency ranges between 117 and 125 Hz (F0) and 685–950 Hz (F1), respectively. These ranges were determined based on the expected F0/F1 range from the input stimulus (see Fig. 1). F0 and F1 magnitudes parsimoniously describe how noise interference degrades the overall brainstem representation of voice “pitch” (F0) and “timbre” (F1) elements of speech.<sup>14–16</sup>

### 2.4 Relative contribution of ENV and TFS to speech-FFRs

To clarify the relative role of ENV and TFS cues to the brainstem encoding of speech in noise, we examined two sets of correlations at each noise level: (i) stimulus-to-response correlations between the *acoustic* ENV and TFS and the intact FFR (i.e., FFR<sub>orig</sub>: ENV+TFS); (ii) response-to-response correlations between FFR<sub>orig</sub> and clean FFR<sub>ENV</sub> and FFR<sub>TFS</sub> responses. Acoustic ENV and TFS was extracted from the speech token via the Hilbert transform. For neuro-acoustic correlations, a cross-correlation was also performed to shift the FFR up to 10 ms so as to align it with the acoustic signal and account for the transmission delay to the brainstem.<sup>20</sup> The Fisher *r*-to-*z* transform was then used to directly compare the correlation strengths between ENV and TFS contributions in order to assess the relative weighting/importance of each temporal cue to the speech-FFR as a function of noise.

## 3. Results

FFRs waveforms and spectra are shown for each of the speech conditions (original, ENV, and TFS) across SNRs in Figs. 2(A) and 2(B). FFRs showed two bursts of phase-locked energy corresponding to the two vowel portions of the /ama/ stimulus (see Fig. 1). Comparisons across conditions suggested more robust encoding for speech containing both ENV and TFS cues which weakened systematically with the removal of each temporal cue and the addition of noise (i.e., decreasing SNR). Response spectra contained

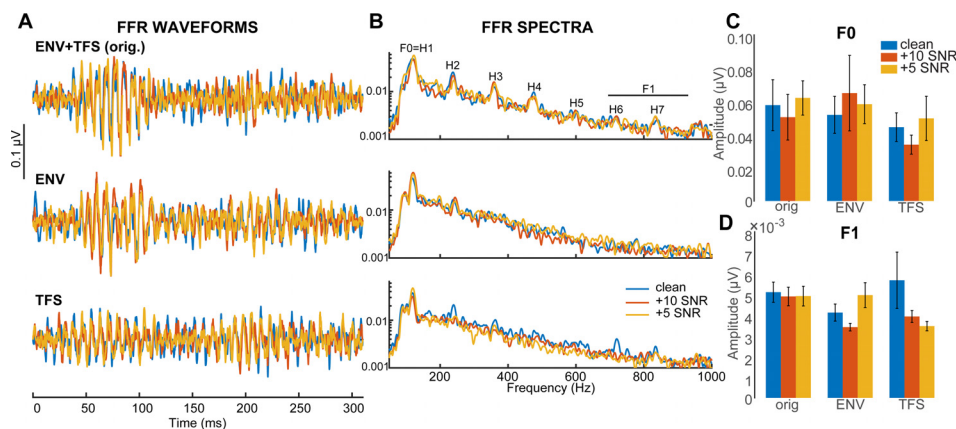


Fig. 2. (Color online) Brainstem FFR waveforms (A) and spectra (B). Neural responses reveal energy at the voice fundamental (F0) and integer-related harmonics up to  $\sim 1000$  Hz (H1–H7). F1, first formant range. Removal of temporal cues systematically weakens speech FFRs in both the time- and frequency-domains (orig > ENV > TFS). Decreasing SNR similarly weakens responses. Although there is no explicit envelope for TFS speech tokens (see Fig. 1), FFRs show an envelope-like modulation which likely reflect “recovered envelope” cues from the stimulus (Refs. 3, 4) (C), (D) Brainstem encoding of voice “pitch” (F0) and “timbre” (F1) as a function of the delivered speech cue and SNR. FFR encoding of F0 shows no systematic change in SNR across speech cues (i.e., orig = ENV = TFS), whereas F1 shows a cue  $\times$  SNR interaction (linear effect of SNR in the TFS condition only). Error bars =  $\pm 1$  s.e.m.

energy at the F0 and integer-related multiples up to the upper limit of brainstem phase-locking ( $\sim 1000$  Hz).<sup>21</sup> Also apparent in the time waveforms is the observation that TFS speech evoked a quasi-modulated response in FFRs which mirrored the stimulus envelope. This suggests that while acoustic ENV cues were not explicit in the TFS stimulus itself, brainstem responses captured so-called “recovered envelope” cues resulting from narrowband cochlear filtering.<sup>3,4</sup>

Comparisons of FFR F0 (“pitch”) and F1 (“timbre”) coding are shown in Figs. 2(C) and 2(D). A two-way mixed model analysis of variance with stimulus cue (3 levels: orig, ENV, TFS) and SNR (3 levels: clean, +10, +5 dB SNR) as fixed factors (subjects = random effect) revealed that FFR encoding at F0 was invariant to manipulations of SNR [ $F_{2,96} = 1.72$ ,  $p = 0.19$ ] and stimulus cue [ $F_{2,96} = 2.54$ ,  $p = 0.08$ ] with no interaction [ $F_{4,96} = 0.31$ ,  $p = 0.87$ ] [Fig. 2(C)]. In contrast, we found a stimulus cue  $\times$  SNR interaction on the neural encoding of F1 [ $F_{4,96} = 3.34$ ,  $p = 0.0132$ ] [Fig. 2(D)]. *Post hoc* Tukey-Kramer adjusted contrasts (i.e., linear effect of SNR) revealed this interaction was attributable to a strong SNR effect in the TFS condition [ $t_{96} = 3.29$ ,  $p = 0.0042$ ]. That is, FFR F1-timbre coding was sensitive to noise but only when TFS information was provided. This effect was not observed in neural responses to the original [ $t_{96} = 0.30$ ,  $p = 1.0$ ] or ENV speech [ $t_{96} = -1.38$ ,  $p = 0.51$ ].

Neuro-acoustic correlations between the speech ENV, TFS, and the  $\text{FFR}_{\text{orig}}$  are shown in Fig. 3(A) and response-to-response correlations (i.e., clean  $\text{FFR}_{\text{ENV}}$  and clean  $\text{FFR}_{\text{TFS}}$  vs  $\text{FFR}_{\text{orig}}$  at each SNR) in Fig. 3(B). Comparing the relative strength of correlations, FFRs showed better correspondence with the stimulus ENV than TFS for clean speech ( $z = 2.12$ ,  $p = 0.034$ ). The reverse was true for +10 dB SNR, where TFS cues dominated the FFRs ( $z = -2.24$ ,  $p = 0.025$ ). ENV and TFS made equal contributions in the +5 dB SNR condition ( $z = 0.97$ ,  $p = 0.33$ ).

For response-to-response correlations,  $\text{FFR}_{\text{orig}}$  showed higher correspondence with envelope coding relative to fine structure responses (i.e.,  $\text{FFR}_{\text{ENV}} > \text{FFR}_{\text{TFS}}$ ;  $z = 10.65$ ,  $p < 0.0001$ ) for clean speech. Again, the reverse was found for moderately (+10 dB) degraded speech, where the aggregate FFR was more similar to the TFS response (i.e.,  $\text{FFR}_{\text{TFS}} > \text{FFR}_{\text{ENV}}$ ;  $z = -2.33$ ,  $p = 0.019$ ). In the noisiest condition (+5 dB SNR), the aggregate  $\text{FFR}_{\text{orig}}$  showed equal correspondence with the  $\text{FFR}_{\text{ENV}}$  and  $\text{FFR}_{\text{TFS}}$  responses. Collectively, these findings indicate that speech coding is dominated by contributions from the stimulus ENV for clean speech, TFS makes a stronger contribution for moderate levels of noise, and the two cues become equally important in the case of severely degraded speech.

#### 4. Discussion

By measuring brainstem FFRs elicited by vCv speech stimuli containing only ENV or TFS cues in various amounts of noise our findings suggest the relative importance between ENV and TFS cues changes dependent on speech SNR. Whereas ENV is most important to the brainstem encoding of clean speech, TFS cues become increasingly dominant at moderate SNRs and make an equal contribution to the ENV in the most severe levels of noise.

With the addition of noise, FFR coding of the sustained F0 periodicity was well-maintained at decreasing (poorer) SNRs. This finding is consistent with previous

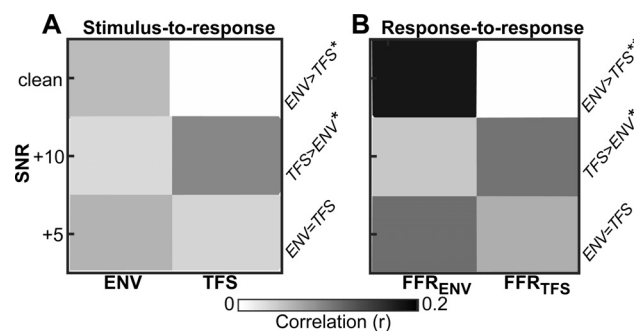


Fig. 3. Stimulus-to-response (A) and response-to-response (B) correlations as a function of SNR. Statistical comparisons to the right of each matrix denote the row contrasts comparing correlations between the aggregate FFR and individual cues of the acoustic signal (A) or FFR itself (B). Clean speech responses are dominated by the acoustic ENV whereas TFS dominates noise degraded speech FFRs. Similarly, the aggregate FFR ( $\text{FFR}_{\text{orig}}$ ) shows higher correspondence with its ENV than TFS constituent ( $\text{FFR}_{\text{ENV}} > \text{FFR}_{\text{TFS}}$ ) for clean speech but this pattern is reversed at moderate (+10 dB SNR) noise levels ( $\text{FFR}_{\text{TFS}} > \text{FFR}_{\text{ENV}}$ ) (B). Both neuro-acoustic and neural-neural correlations show an equal contribution of ENV and TFS at the most severe noise levels. \* $p < 0.05$ ; \*\*\* $p < 0.001$ .

FFR studies showing a maintenance or even facilitation of the F0 response with additive noise.<sup>22</sup> Physiologically, the robustness of F0 coding to noise could reflect the engagement of low-frequency “tails” of basal, high-frequency neurons that phase-lock to the common F0 across cochlear channels.<sup>15</sup> Multiple responses to F0 across the cochlear array would tend to reinforce one another at the population level and consequently offer some resilience/redundancy in pitch cues observed here in scalp-recorded FFRs (for discussion of other mechanisms of F0 enhancement see Refs. 15, 22).

Contrastively, we found a systematic degradation in the neural encoding of “timbre”-related F1 harmonics, which were more susceptible to noise manipulation. The resilience of the brainstem FFR at F0 (but not higher F1 harmonics) in the presence of noise has been noted by a number of investigators<sup>22</sup> and suggests that neural synchronization at the F0 is relatively robust to acoustic interference. However, it remains possible that stimuli with more dynamically changing F0 than those used here would have produced stronger noise-related changes in pitch coding.<sup>23</sup> In speech perception, F0 provides a correlate of voice pitch, a robust cue for stream segregation, and identifying the number of sources in complex auditory scenes.<sup>24</sup> Lower susceptibility of FFR F0 to noise is consistent with the notion that pitch remains a robust cue for segregating target speech from a sound mixture despite substantial signal degradation.<sup>16,24</sup> In contrast, higher spectral components captured by the FFR (e.g., formant-related harmonics) are systematically degraded. This dissociation between “pitch” (F0) and “timbre” (F1) related components of speech observed here at the neural level corroborates findings from several behavioral studies which similarly show an independence between F0 and F1 properties of speech.<sup>16,25</sup> However, analysis of F0 and F1 is limited, as these are two isolated features of the FFR spectrum and may not provide a complete picture of the differential contribution of the entire ENV and/or TFS signal (themselves broadband) to the neural encoding of speech.

Comparisons between FFRs to isolated ENV and TFS speech allowed us to tease apart the relative contribution of these temporal cues to the brainstem encoding of noise-degraded vCv stimuli. Paralleling animal work,<sup>11,12</sup> recent FFR studies have demonstrated a disruption in the balance of ENV and TFS coding in hearing impaired listeners.<sup>26</sup> Our study replicates and extends these previous findings by demonstrating that even in normal hearing listeners, noise alters the balance between the contributions of ENV and TFS to the neural encoding of speech. This was evident in neural-acoustic correlations (Fig. 3): brainstem FFRs were dominated by the acoustic speech ENV for clean signals, whereas TFS cues began to dominate neural responses at moderate noise levels. This suggests that while ENV and TFS are perceptually important for robust speech perception, the encoding of TFS becomes more important to the neural representation of speech under noisy listening conditions—consistent with human psychophysical reports.<sup>8–10</sup>

Interestingly, the cue dominance observed here in human brainstem FFRs (Fig. 3) also parallels that observed in auditory nerve, where ENV cues drive neural speech representations in quiet listening conditions with TFS taking a more significant role in degraded speech coding.<sup>4</sup> Consistent with single-unit recordings in animals,<sup>4</sup> we similarly find that ENV and TFS each contribute to the brainstem speech-FFR at severe SNRs and that periodicity ENV (i.e., F0) is surprisingly resilient to noise manipulations [Fig. 2(C)]. Important signal transformations between the peripheral and central auditory nervous system are thought to occur during concurrent sound processing.<sup>27</sup> Future studies examining simultaneous responses from subcortical and cortical structures<sup>27</sup> could help clarify the hierarchy of signal transformations between peripheral and central aspects of the auditory neuroaxis during degraded speech coding. Additionally, it would be worthwhile to extend current results to a broader range of speech tokens to assess generalizability and possible differential effects of noise and temporal cues on different elements of speech acoustics.

### Acknowledgments

This work was supported by grants from the American Hearing Research Foundation (AHRF) and American Academy of Audiology (AAA) Foundation awarded to G.M.B.

### References and links

<sup>1</sup>R. V. Shannon, F. G. Zeng, V. Kamath, J. Wygonski, and M. Ekelid, “Speech recognition with primarily temporal cues,” *Science* **270**, 303–304 (1995).

<sup>2</sup>Z. M. Smith, B. Delgutte, and A. J. Oxenham, “Chimaeric sounds reveal dichotomies in auditory perception,” *Nature* **416**, 87–90 (2002).

- <sup>3</sup>F.-G. Zeng, K. Nie, S. Liu, G. Stickney, E. Del Rio, Y.-Y. Kong, and H. Chen, "On the dichotomy in auditory perception between temporal envelope and fine structure cues," *J. Acoust. Soc. Am.* **116**, 1351–1354 (2004).
- <sup>4</sup>J. Swaminathan and M. G. Heinz, "Psychophysiological analyses demonstrate the importance of neural envelope coding for speech perception in noise," *J. Neurosci.* **32**, 1747–1756 (2012).
- <sup>5</sup>V. C. Tartter, "What's in a whisper?," *J. Acoust. Soc. Am.* **86**, 1678–1683 (1989).
- <sup>6</sup>J. C. R. Licklider and I. Pollack, "Effects of differentiation, integration and infinite peak clipping upon the intelligibility of speech," *J. Acoust. Soc. Am.* **20**, 42–52 (1948).
- <sup>7</sup>R. E. Remez, P. E. Rubin, D. B. Pisoni, and T. D. Carrell, "Speech perception without traditional speech cues," *Science* **212**, 947–949 (1981).
- <sup>8</sup>C. Lorenzi, G. Gilbert, H. Carn, S. Garnier, and B. C. Moore, "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," *Proc. Natl. Acad. Sci. U.S.A.* **103**, 18866–18869 (2006).
- <sup>9</sup>K. Hopkins and B. C. J. Moore, "The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise," *J. Acoust. Soc. Am.* **130**, 334–349 (2011).
- <sup>10</sup>M. K. Qin and A. J. Oxenham, "Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification," *Ear Hear.* **26**, 451–460 (2005).
- <sup>11</sup>S. Kale and M. G. Heinz, "Envelope coding in auditory nerve fibers following noise-induced hearing loss," *J. Assoc. Res. Oto.* **11**, 657–673 (2010).
- <sup>12</sup>K. S. Henry and M. G. Heinz, "Diminished temporal coding with sensorineural hearing loss emerges in background noise," *Nat. Neurosci.* **15**, 1362–1364 (2012).
- <sup>13</sup>G. M. Bidelman, "Multichannel recordings of the human brainstem frequency-following response: Scalp topography, source generators, and distinctions from the transient ABR," *Hear. Res.* **323**, 68–80 (2015).
- <sup>14</sup>K. Banai, J. Hornickel, E. Skoe, T. Nicol, S. Zecker, and N. Kraus, "Reading and subcortical auditory function," *Cereb. Cortex* **19**, 2699–2707 (2009).
- <sup>15</sup>G. M. Bidelman, "Communicating in challenging environments: Noise and reverberation," in *Springer Handbook of Auditory Research: The Frequency-Following Response: A Window into Human Communication*, edited by N. Kraus, S. Anderson, T. White-Schwoch, R. R. Fay, and A. N. Popper (Springer Nature, New York, in press), Vol. 61.
- <sup>16</sup>G. M. Bidelman and A. Krishnan, "Effects of reverberation on brainstem representation of speech in musicians and non-musicians," *Brain Res.* **1355**, 112–125 (2010).
- <sup>17</sup>G. M. Bidelman and M. Howell, "Functional changes in inter- and intra-hemispheric auditory cortical processing underlying degraded speech perception," *Neuroimage* **124**, 581–590 (2016).
- <sup>18</sup>S. Rosen, "Temporal information in speech: Acoustic, auditory, and linguistic aspects," *Philos. Trans. R. Soc. London.* **336**, 367–373 (1992).
- <sup>19</sup>T. Campbell, J. R. Kerlin, C. W. Bishop, and L. M. Miller, "Methods to eliminate stimulus transduction artifact from insert earphones during electroencephalography," *Ear Hear.* **33**, 144–150 (2012).
- <sup>20</sup>G. M. Bidelman, J. W. Villafuerte, S. Moreno, and C. Alain, "Age-related changes in the subcortical-cortical encoding and categorical perception of speech," *Neurobiol. Aging* **35**, 2526–2540 (2014).
- <sup>21</sup>L. F. Liu, A. R. Palmer, and M. N. Wallace, "Phase-locked responses to pure tones in the inferior colliculus," *J. Neurophysiol.* **95**, 1926–1935 (2006).
- <sup>22</sup>F. Prevost, M. Laroche, A. M. Marcoux, and H. R. Dajani, "Objective measurement of physiological signal-to-noise gain in the brainstem response to a synthetic vowel," *Clin. Neurophysiol.* **124**, 52–60 (2013).
- <sup>23</sup>T. White-Schwoch, E. C. Davies, E. C. Thompson, K. Woodruff Carr, T. Nicol, A. R. Bradlow, and N. Kraus, "Auditory-neurophysiological responses to speech during early childhood: Effects of background noise," *Hear. Res.* **328**, 34–47 (2015).
- <sup>24</sup>P. F. Assmann and Q. Summerfield, "Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies," *J. Acoust. Soc. Am.* **88**, 680–697 (1990).
- <sup>25</sup>C. Liu and D. Kewley-Port, "Formant discrimination in noise for isolated vowels," *J. Acoust. Soc. Am.* **116**, 3119–3129 (2004).
- <sup>26</sup>S. Anderson, A. Parbery-Clark, T. White-Schwoch, S. Drehobl, and N. Kraus, "Effects of hearing loss on the subcortical representation of speech cues," *J. Acoust. Soc. Am.* **133**, 3030–3038 (2013).
- <sup>27</sup>G. M. Bidelman and C. Alain, "Hierarchical neurocomputations underlying concurrent sound segregation: Connecting periphery to percept," *Neuropsychologia* **68**, 38–50 (2015).