JARO
Journal of the Association for Research in Otolaryngology

Quantifying Envelope and Fine-Structure Coding in Auditory Nerve Responses to Chimaeric Speech

MICHAEL G. HEINZ[1,2] AND JAYAGANESH SWAMINATHAN[1]

[1]*Department of Speech, Language, and Hearing Sciences, Purdue University, 500 Oval Drive, West Lafayette, IN 47907, USA*
[2]*Weldon School of Biomedical Engineering, Purdue University, West Lafayette, IN 47907, USA*

ABSTRACT

Any sound can be separated mathematically into a slowly varying envelope and rapidly varying fine-structure component. This property has motivated numerous perceptual studies to understand the relative importance of each component for speech and music perception. Specialized acoustic stimuli, such as auditory chimaeras with the envelope of one sound and fine structure of another have been used to separate the perceptual roles for envelope and fine structure. Cochlear narrowband filtering limits the ability to isolate fine structure from envelope; however, envelope recovery from fine structure has been difficult to evaluate physiologically. To evaluate envelope recovery at the output of the cochlea, neural cross-correlation coefficients were developed that quantify the similarity between two sets of spike-train responses. Shuffled auto- and cross-correlogram analyses were used to compute separate correlations for responses to envelope and fine structure based on both model and recorded spike trains from auditory nerve fibers. Previous correlogram analyses were extended to isolate envelope coding more effectively in auditory nerve fibers with low center frequencies, which are particularly important for speech coding. Recovered speech envelopes were present in both model and recorded responses to one- and 16-band speech fine-structure chimaeras and were significantly greater for the one-band case, consistent with perceptual studies. Model predictions suggest that cochlear recovered envelopes are reduced following sensori-neural hearing loss due to broadened tuning associated with outer-hair cell dysfunction. In addition to the within-fiber cross-stimulus cases considered here, these neural cross-correlation coefficients can also be used to evaluate spatiotemporal coding by applying them to cross-fiber within-stimulus conditions. Thus, these neural metrics can be used to quantitatively evaluate a wide range of perceptually significant temporal coding issues relevant to normal and impaired hearing.

Keywords: neural cross correlation, temporal coding, recovered envelopes, cochlear implants

INTRODUCTION

Numerous perceptual studies have addressed fundamental questions about the relative contributions of the slowly varying *envelope* and rapidly varying *fine-structure* components of speech and music (Smith et al. 2002; Xu and Pfingst 2003; Zeng et al. 2005b). Envelope information is important for speech perception and supports robust speech identification in quiet when provided in as few as four frequency bands (Shannon et al. 1995). This finding has important implications for cochlear implants, which currently only provide envelope information over a small number (eight to 16) of electrodes and is consistent with the observation that many cochlear-implant patients understand speech remarkably well in quiet (Wilson et al. 1991). The relative roles of envelope and fine structure have recently been evaluated using specialized acoustic stimuli called auditory chimaeras, which have the envelope of one

Correspondence to: Michael G. Heinz · Department of Speech, Language, and Hearing Sciences · Purdue University · 500 Oval Drive, West Lafayette, IN 47907, USA. Telephone: +1-765-4966627; fax: +1-765-4940771; email: mheinz@purdue.edu

sound and the fine structure of another (Smith et al. 2002). Chimaeric speech constructed from two sentences is generally perceived as the sentence that provided envelope, whereas chimaeric music is perceived as the melody that contributed fine structure. The perceptual salience of acoustic fine structure for music perception and sound localization (Smith et al. 2002), lexical-tone perception (Xu and Pfingst 2003), and speech perception in noise (Qin and Oxenham 2003; Lorenzi et al. 2006) has been given as motivation for efforts to develop cochlear-implant strategies to provide fine structure in addition to envelope cues (e.g., Rubinstein et al. 1999; Nie et al. 2005).

Interpretation of perceptual studies that utilize auditory chimaeras relies on the assumption that envelope and fine structure can be isolated. However, signal-processing theorems state that the envelope and fine structure of band-limited signals are not independent, and information about one can be recovered mathematically from the other, e.g., envelope can be recovered from fine-structure by narrowband filtering (e.g., Voelcker 1966; Rice 1973; Logan 1977). Thus, narrowband cochlear filtering imposes constraints on the ability to isolate a sound's fine structure from its envelope (Ghitza 2001, also see Saberi and Hafter 1995). Although envelope is clearly more salient than fine-structure for eight- and 16-band speech chimaeras, a reversal occurs for one- and two-band chimaeras for which fine structure supports robust speech recognition rather than envelope (Smith et al. 2002). Perceptual studies have suggested that recovered envelopes at the output of the cochlea may explain the reversal in these conditions for which the chimaeric analysis bands were much broader than cochlear filters (Zeng et al. 2004; Gilbert and Lorenzi 2006). However, these considerations were limited to perceptually based filter-bank models, which capture the basic effects of cochlear filtering but exclude many physiological factors (e.g., adaptation, phase-locking roll-off, two-tone suppression) that may affect envelope and fine-structure coding in neural responses to complex sounds.

The present study provides physiological evidence for the presence of recovered envelopes in auditory nerve (AN) responses to chimaeric speech. Neural cross-correlation coefficients were developed to quantify the similarity between envelope (or fine structure) components in two sets of spike-train responses. Auto- and cross-correlogram analyses were used to separate the contributions of envelope and fine structure (Joris 2003). These neural metrics can also be used to evaluate fundamental questions related to *across-fiber* temporal coding, which was recently hypothesized to be involved in the difficulties that hearing-impaired listeners have in understanding speech in complex acoustic backgrounds (Lorenzi et al. 2006; Moore 2008).

METHODS

Auditory nerve model

Spike-train data from a computational model of AN responses (Zilany and Bruce 2006, 2007) was used to evaluate systematically the dependence of neural cross correlation on both stimulus parameters (e.g., number of chimaeric analysis bands) and AN-fiber parameters (e.g., characteristic frequency (CF), the frequency at which the fiber responds to the lowest sound level). The phenomenological AN model represents the most recent extension of a well-established model that has been rigorously tested against physiological AN responses to both simple and complex stimuli, including tones, broadband noise, and speech-like sounds (Carney 1993; Heinz et al. 2001a; Zhang et al. 2001; Bruce et al. 2003; Tan and Carney 2003; Zilany and Bruce 2006, 2007). Model threshold tuning curves have been well fit to the CF-dependent variation in bandwidth for normal-hearing cats (Miller et al. 1997), which is comparable to that of chinchillas (Shera et al. 2007; Temchin et al. 2008a, b). Many of the physiological properties associated with nonlinear cochlear tuning are captured by this model, including compression, suppression, and broadened tuning and best-frequency shifts with increases in sound level. The stochastic nature of AN responses is accounted for by a nonhomogenous Poisson process that was modified to include the effects of both absolute and relative refractory periods. Although the Poisson-based model does not capture all of the detailed stochastic properties of AN responses (e.g., Heil et al. 2007), the major statistical properties relevant to this work are captured by this model (e.g., Young and Barta 1986). Although the Zilany and Bruce (2006, 2007) model was chosen for this study, the results presented here do not depend on this choice, and several other AN models exist that would be expected to produce similar results (reviewed by Lopez-Poveda 2005).

The AN-model input is the sound stimulus waveform, while the output is a set of spike times for a single AN fiber with a specified CF. All model simulations were for high-spontaneous-rate (50 spikes/s) fibers, for which this AN model was designed and tested. Similar results were obtained for both broadband noise and speech when the model was extended to simulate higher-threshold, low-spontaneous rate fibers (not shown). Stimuli were re-sampled to 100 kHz prior to presentation to the model.

Surgical procedures and neurophysiological recording techniques

Several model predictions were verified by computing neural cross-correlation coefficients from spike trains recorded from 28 AN fibers during other experiments

in the lab. All methods of animal care and use were approved by the Purdue Animal Care and Use Committee. Single-unit recordings were made from AN fibers from four adult chinchillas using standard techniques (e.g., Kiang et al. 1965; Heinz and Young 2004). All four chinchillas had AN-fiber thresholds within normal limits (i.e., lowest thresholds were 5–15 dB SPL). The animals weighed 400–600 g and were initially anesthetized with xylazine (1–1.5 mg/kg im) and ketamine (50–65 mg/kg im). Supplemental doses of fluids and barbiturate anesthesia (sodium pentobarbital, ~7.5 mg/kg/h iv) were given to maintain an areflexic state. Rectal temperature was maintained between 37°C and 38°C. A tracheotomy was performed to facilitate quiet breathing, and the bulla was vented with a polyethylene tube to equalize middle-ear pressure.

During the recordings, the animals were held in place with a stereotaxic apparatus within a double-walled, sound-attenuating chamber (Industrial Acoustics, Bronx, NY, USA). Sound was delivered monaurally through a custom closed-field acoustic system, with dynamic speakers (DT-48, Beyer Dynamic, Farmingdale, NY, USA) connected to a hollow ear bar inserted into the right ear canal to deliver calibrated acoustic stimuli near the tympanic membrane. The acoustic system was calibrated at the beginning of the experiment using a probe-tube microphone (ER-7C, Etymōtic, Elk Grove Village, IL, USA) placed within a few millimeters of the tympanic membrane. Single-unit recordings were made with 10–30 MΩ glass micropipettes filled with 3 M NaCl. The electrode signal was amplified (Dagan, Minneapolis, MN, USA) and filtered prior to timing (10-μs resolution) the action potentials based on a time–amplitude window discriminator (Bak Electronics, Mount Airy, MD, USA). Synchronous presentation of acoustic stimuli and data recording was controlled by custom software running in MATLAB (The Mathworks, Natick, MA, USA) integrated with custom and commercial hardware (Tucker-Davis Technologies, Alachua, FL, USA; National Instruments, Austin, TX, USA). A broadband noise search stimulus was used to isolate AN fibers. Fibers were characterized using an automated tuning-curve algorithm that was used to determine the fiber's CF, threshold, and Q_{10} (Chintanpalli and Heinz 2007), as well as by CF-tone rate-level functions and PST histograms.

Stimuli

Two independently generated frozen noise waveforms (A and B) were used for testing basic properties of the neural cross-correlation metrics. Noises A and B were both Gaussian and broadband, with a 2-s duration for model data and a 1.7-s duration for AN-fiber data

(10-ms rise-fall times in both cases). Chimaeric speech stimuli were created from the original speech utterance "A boy fell from the window," which had a duration of 1.7 s (Shen et al. 2001). Chimaeric stimuli were created from this utterance and a spectrally matched broadband noise using the chimaerizing algorithm and MATLAB code developed by Smith and colleagues (Shen et al. 2001; Smith et al. 2002). A variable number of FIR band-pass filters were equally spaced on a cochlear frequency map and spanned the frequency range from 80 to 8820 Hz. Envelope and fine structure were separated from the output of each filter for each sound using a Hilbert transform. The envelope from the first sound was multiplied by the fine structure from the second sound within each filter, prior to adding the resulting individual band-pass signals to create the final chimaera. The FIR group delay resulting from the chimaeric filtering was manually compensated for prior to presentation of the chimaeric stimuli. The *speech fine-structure chimaera* consisted of a combination of the fine structure from the original speech token and the envelope from the noise, whereas the *speech envelope chimaera* was created from the opposite contributions.

For every stimulus condition, spike trains were obtained in response to the original stimulus (A+) and its polarity-inverted pair (A−). The polarity inversion introduces a 180° phase shift of all frequency components, thereby inverting the fine structure of the stimulus, while not affecting the stimulus envelope (Joris 2003). Model and AN-fiber responses were obtained for 16–25 repetitions of each stimulus, which was sufficient to collect roughly 3,500 spikes for each stimulus condition. All stimuli presented to AN fibers were 1.7 s in duration, and a new stimulus was presented every 2.5 s. All stimuli within a given set (e.g., noises A+, A−, B+, and B−) were presented in an interleaved manner until the desired number of repetitions was completed. For all analyses, spikes within the first 50 ms of the response were excluded to avoid onset effects.

For both types of stimuli, and for both model and AN-fiber data, sound levels were chosen for each AN fiber to maximize the number of spikes, while minimizing the degradation in envelope coding at high levels due to saturation (e.g., Joris and Yin 1992; Louage et al. 2004). Data from model simulations were collected typically at the best modulation level for each stimulus type (i.e., noise A, or the original speech token). Best modulation level was determined for each model fiber as the sound level that produced the maximum amount of envelope coding, as quantified by the sumcor peak height (defined below; also see Louage et al. 2004). The sound levels used in the neurophysiological experiments were chosen typically to be within the upper one third of the fiber's

dynamic range based on a measured rate-level function for each stimulus type. Although these criteria were slightly different, the sound levels relative to fiber threshold were only slightly higher for the recorded AN data than for the model data. Larger differences in absolute sound level between model simulations and experimental data (up to 30 dB) arose due to higher (~10 dB) AN thresholds in chinchillas relative to cats (Miller et al. 1997; Temchin et al. 2008b) and because model thresholds were designed to match the lowest AN fiber threshold within the span of AN thresholds (~40 dB) at each CF, rather than the mean (Miller et al. 1997; Zilany and Bruce 2006). Noise spectrum levels ranged from −27 to −7 dB for model data, and from 2 to 26 dB for recorded AN data. Overall speech levels were 35 dB SPL for model data (normal hearing) and ranged from 62 to 72 dB SPL for recorded AN data.

Envelope and fine-structure cross-correlation metrics computed from spike trains

The neural cross-correlation coefficients (ρ_{ENV} and ρ_{TFS}) developed in this study provide metrics ranging from 0 to 1 that represent the degree of similarity between the temporal envelope (ENV) or fine structure (TFS) of responses to two different conditions (e.g., one neuron responding to two different stimuli, or two neurons responding to one stimulus). Each neural cross-correlation coefficient represents the degree of *cross-correlation* between two different responses *relative to* the strength of temporal coding within each individual response (as represented by the *autocorrelation*). Computation of the auto- and cross-correlations was based on the shuffled correlogram analyses developed by Joris and colleagues, which were designed to separate the quantification of envelope and fine-structure coding (Joris 2003; Louage et al. 2004; Joris et al. 2006a, 2008b). Figure 1 illustrates the complete set of computations involved in computing the neural cross-correlation coefficients for temporal fine-structure (ρ_{TFS}) and envelope (ρ_{ENV}) for a simple example (one AN fiber responding to two uncorrelated noises).

The first two columns of Figure 1 illustrate the computations involved in quantifying *within-fiber* temporal coding for noises A and B. Shuffled autocorrelograms (SACs, thick lines in panels A and B) were computed from single sets of spike trains recorded from a chinchilla AN fiber (CF=827 Hz) responding to repeated presentations of the same stimulus (Louage et al. 2004). The shuffling involved comparing spike times between all possible pairs of stimulus repetitions within a given set of spike trains, i.e., for N repetitions, there are $N \times (N-1)$ pairs. For each pair, intervals between every spike in the first spike train
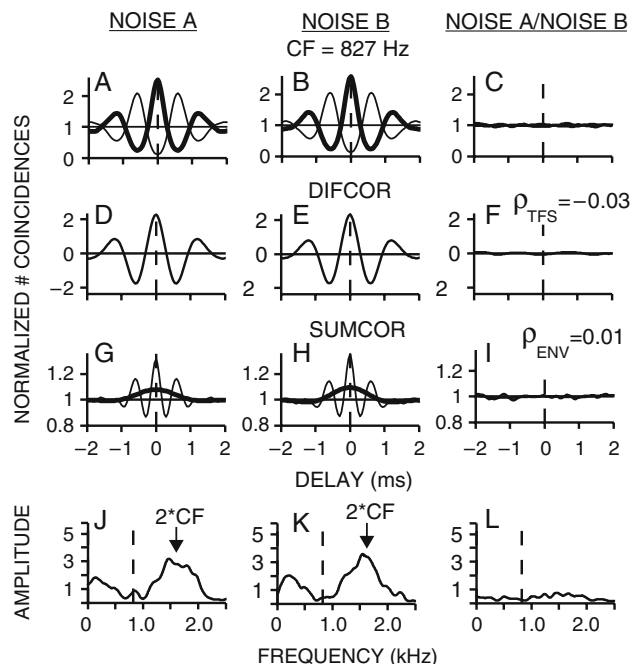

FIG. 1. Neural cross-correlation coefficients for fine-structure and envelope were computed based on a normalized comparison of within-stimulus (columns 1 and 2) and cross-stimulus (column 3) temporal coding. Spike trains were recorded from a chinchilla AN fiber responding to two uncorrelated broadband noises (A and B). **A–B** Normalized shuffled autocorrelograms [*thick line*, e.g., SAC(A+)] and cross-polarity correlograms [*thin line*, e.g., SCC(A+,A−)]. **C** Shuffled cross-stimulus correlogram [*thick line*, e.g., SCC(A+,B+)] and cross-stimulus, cross-polarity correlogram [*thin line*, e.g., SCC(A+,B−)]. **D–F** Difcors emphasize fine-structure coding and are computed by subtracting the cross-polarity correlogram from the auto- or cross-stimulus correlogram. ρ_{TFS} was computed (Eq. 1) by comparing the peak heights of difcor(AB) to difcor(A) and difcor(B) at a characteristic delay of zero. **G–I** Sumcors, which emphasize envelope coding, were the average of the cross-polarity and auto- or cross-stimulus correlograms (*thin lines*). *Thick lines* represent corrected sumcors that eliminate fine-structure leakage by including only energy at frequencies below CF (see "Methods" section). **J–L** Envelope power spectral density (PSD; **J**, **K**) and cross-spectral density (CSD; **L**) computed as the Fourier transform of the original sumcors. Leaked fine-structure energy occurs near 2×CF (*arrow*); dashed line at CF. ρ_{ENV} was computed (Eq. 2) from the corrected sumcors by comparing peak heights at a characteristic delay of zero. AN fiber CF=827 Hz; threshold=15 dB SPL; Q_{10}=1.4; spontaneous rate=86 spikes/s. Noise spectrum level=26 dB SPL.

and every spike in the second spike train were tallied. The SAC is the shuffled all-order interval histogram obtained by tallying all intervals from all pairs with a 50-μs binwidth.

By computing inter-spike intervals across repetitions (rather than within), shuffled correlograms provide a more robust analysis of temporal responses than the all-order interval histograms that have often been used to estimate autocorrelation functions from neural responses to complex sounds (e.g., Ruggero 1973; Cariani and Delgutte 1996a, b). Temporal correlations at very small delays are accurately cap-

tured by shuffling because the deleterious effects of refractoriness are avoided. Also, shuffling produces significantly smoother estimates of correlation functions based on the same amount of data because the number of computed intervals is proportional to N^2, rather than N. Previous studies of the temporal coding of envelope and fine structure have often used periodic stimuli (e.g., Johnson 1980; Joris and Yin 1992). However, the shuffled correlogram analyses are applicable to any repeatable stimulus (Joris 2003; Louage et al. 2004), e.g., broadband noise or chimaeric speech.

SACs are typically normalized [by $N(N-1)r^2\Delta\tau D$, where r is the average discharge rate, $\Delta\tau$ is the binwidth, and D is the duration of the response window] to allow a more intuitive interpretation of temporal coding. With this normalization, a baseline value of 1 represents the absence of any temporal correlation (envelope or fine structure). A value greater than 1 represents positive correlation, and a value less than 1 represents negative correlation. SACs are plotted as function of delay (or inter-spike interval) and are therefore much like autocorrelation functions. As such, the SACs in Figures 1A, B have a peak at zero delay, with a symmetric damped oscillatory shape similar to an autocorrelation function for band-limited noise. The prominent oscillation occurs with a period equal to the reciprocal of CF, and the rate of decay of the side lobes is determined by the fiber's bandwidth (Louage et al. 2004; Joris et al. 2008a). Because the statistical properties of the noise do not change with stimulus polarity, the SACs shown in Figure 1 are the average of the SACs for both polarities [e.g., SAC(A+) and SAC(A−)].

Joris and colleagues have demonstrated that the coding of fine structure and envelope can be separated by considering the neural responses to a stimulus and its polarity-inverted pair (Joris 2003; Louage et al. 2004; Joris et al. 2006a). The polarity inversion acts to invert the fine-structure of the stimulus, while not affecting the stimulus envelope. To facilitate the separation of envelope and fine-structure coding, Joris and colleagues computed what they referred to as a cross-stimulus autocorrelogram (XAC). They used the term *auto* to indicate the same fiber and the term *cross-stimulus* applied to the original stimulus and its polarity-inverted pair. To allow a general consideration of auto- and cross-correlation analyses, the term *autocorrelogram* was restricted in the present study to correlograms computed from a single set of spike trains, for which classic autocorrelation properties are expected (e.g., a symmetric shape with a peak at 0 delay). The term *cross-correlogram* was used for any correlogram computed across two separate sets of spike trains, for which classic autocorrelation properties are not expected (including XACs). Shuffled

cross-correlograms [e.g., SCC(A,B)] were computed by comparing all possible repetition pairs ($N_A \times N_B$) between two separate sets of spike trains in response to conditions A and B. Similar to the SAC computation, all-order intervals between every spike in the first spike train and every spike in the second spike train were tallied, with the SCC representing the complete all-order interval histogram based on all possible pairs (normalized by $N_A N_B r_A r_B \Delta\tau D$). In this generalized notation, the XAC (Joris 2003; Louage et al. 2004) used to facilitate the separation of envelope and fine structure is a cross-polarity correlogram [e.g., SCC(A+, A−)]. Each thin line in Figures 1A, B is the average of two statistically identical cross-polarity correlograms [e.g., SCC(A+,A−) and SCC(A−,A+)].

To emphasize the contribution of fine structure to temporal coding, *difcors* (second row of Fig. 1) were computed as the difference between the SAC (original envelope, original fine structure) and the cross-polarity correlogram [e.g., SCC(A+,A−), with original envelope and inverted fine structure]. By taking the difference, the common envelope contributions are minimized, thus emphasizing the fine-structure coding. Temporal fine-structure coding within a single AN fiber was quantified as the peak height of the difcor (Joris 2003; Louage et al. 2004). Note that while difcors clearly emphasize fine-structure coding, the isolation from envelope coding is not perfect (e.g., the difcor of a low-level sinusoidally amplitude-modulated (SAM) tone contains small oscillations at the modulation frequency superimposed on the prominent oscillations at the carrier frequency). However, the influence of envelope coding on difcor estimates of fine-structure coding can be argued to be small based on the small effect of sound level on difcor peak heights. While AN-fiber envelope coding typically degrades significantly as sound level increases (Joris and Yin 1992; Louage et al. 2004), difcor peak heights are only slightly reduced (Louage et al. 2004). Some of this decrease in difcor peak height with level represents a true decrease in phase locking (Johnson 1980). Nonetheless, this small drop provides an upper limit on the contribution of envelope coding to difcor peak height. For the well-defined SAM-tone stimulus, a more precise measure of fine-structure coding would be the synchrony coefficient at the carrier frequency; however, this analysis requires knowledge of the stimulus. For more complex and nonperiodic stimuli, such as broadband noise or speech, the difcor provides a useful estimate of the neural coding of fine structure that does not require knowledge of the stimulus.

To emphasize envelope coding, *sumcors* were computed as the average of the SAC and the cross-polarity correlogram [e.g., SCC(A+,A−)]. By taking the average, the common contribution of envelope coding

was emphasized, and the contribution of fine-structure coding was minimized. Joris and colleagues quantified temporal envelope coding within a single fiber as the peak height of the sumcor (Joris 2003; Louage et al. 2004). Their quantification works well at high CFs, where fine structure is minimal and envelope is the primary source of temporal coding. However, at low CFs, responses to broadband stimuli have significant fine structure and envelope contributions, and the fine-structure contributions do not always cancel out completely in the sumcor (thin lines, row 3 of Fig. 1; also see Fig. 1D in Joris 2003). This leakage of fine structure into the sumcor reflects distortion that arises from rectification associated with neural responses. Thus, sumcors for low-CF fibers have undesirable high-frequency oscillations that are not associated with the slow envelope response. The fine-structure leakage was more easily isolated in the spectral domain, where it was apparent in the Fourier transform of the sumcor as a high-frequency spectral peak centered at $2 \times CF$ (row 4, Fig. 1). The Fourier transform of the sumcor was taken using a 25-ms rectangular window centered on zero delay, with the sumcor baseline value of 1 subtracted prior to computing the Fourier transform. To correct for the triangular shape imposed on shuffled correlograms by the limited stimulus duration, an inverted triangular compensation ranging from 1 to 0 was added to the SACs and SCCs prior to computation of the sumcors.

Because the sumcor is meant to represent the autocorrelation function corresponding to the temporal envelope response, the magnitude of the sumcor Fourier transform can be thought of as the envelope power spectral density. The true envelope power spectrum for AN responses to noise is roughly triangular and limited to low frequencies (Figs. 1 J,K), consistent with the narrowband output of cochlear filters (Dau et al. 1999). The undesirable contribution of fine-structure coding to the sumcor was eliminated by considering only the envelope spectra below CF (vertical dashed line, row 4 of Fig. 1). The choice of CF as the cutoff frequency represents a compromise that eliminates the energy locus near $2 \times CF$ due to fine-structure leakage, while including as much potential low-frequency envelope energy as possible. Corrected sumcors (thick lines, row 3 of Fig. 1) were computed as the inverse Fourier transform of the envelope spectra below CF. This correction produces a sumcor with a smaller peak height and a single-lobe shape that is consistent with sumcors from high-CF fibers, where fine structure is negligible (Joris 2003; Louage et al. 2004). Figure 2 illustrates the effect of fine-structure leakage as a function of CF by comparing original (squares) and corrected (circles) sumcor peak heights from broadband noise responses. Fine-structure leakage into sumcors pro-
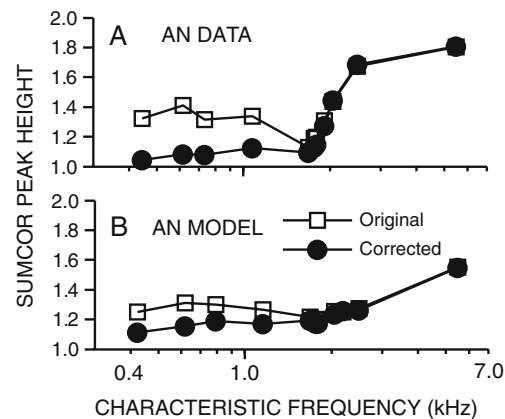

FIG. 2. Fine-structure leakage affects sumcor estimates of low-frequency envelope coding. **A** Data from chinchilla AN fibers. **B** Data from AN model. *Open symbols* represent peak heights of the standard sumcors (thin lines in row 3 of Fig. 1), which represent the average of the auto- and cross-polarity correlograms, e.g., SAC(A+) and SCC(A+,A−). *Filled symbols* represent peak heights of corrected sumcors (thick lines in row 3 of Fig. 1), which were computed by eliminating spectral energy above CF that results from fine-structure leakage into the sumcor (see "Methods" section). Responses were to broadband noise.

duced overestimates of envelope coding at low CFs in both recorded (top panel) and model (bottom panel) AN responses. For CFs above 2 kHz, original and corrected sumcor peak heights were nearly identical, consistent with decreased AN phase locking to fine structure at high frequencies (Johnson 1980). Thus, the corrected sumcors provide a much less confounded representation of low-frequency envelope coding, which is of particular interest in the present study.

The third column of Figure 1 illustrates the use of shuffled cross-correlograms to evaluate the similarity between two recorded spike trains from one AN fiber responding to two different stimuli (uncorrelated noises A and B); however, the same analyses are applicable to the condition of two different fibers responding to the same stimulus. From the spike-train data collected for A+, A−, B+, and B−, two statistically identical SCCs [i.e., SCC(A+,B+) and SCC(A−,B−)] were computed and averaged to obtain the cross-stimulus correlogram (thick line in Fig. 1C). Cross-stimulus, cross-polarity correlograms [e.g., SCC(A+,B−)] were computed to facilitate the separation of fine-structure and envelope cross-correlations by using difcors and sumcors, respectively. The thin line in Figure 1C represents the average of SCC(A+,B−) and SCC(A−,B+). The flat cross-stimulus correlograms (Figs. 1C, F, and I) indicate that there were very few common temporal fluctuations between the responses to these two uncorrelated noises.

More generally, shuffled correlograms quantify cross-correlation as a function of delay and demonstrate a peak at the characteristic delay (CD) between

the two responses. Nonzero characteristic delays can arise both for *within-fiber, cross-stimulus* conditions [e.g., due to a delay or phase shift between two stimuli, or a change in stimulus level (Joris et al. 2008b)] and for *cross-fiber, within-stimulus* conditions [e.g., due to a traveling wave delay between different CFs (Joris et al. 2006b)]. Thus, the CDs in SCCs provide relative phase information across conditions and therefore avoid a commonly cited limitation of SAC analysis, i.e., the elimination of phase information.

Neural cross-correlation coefficients were computed in a manner similar to a statistical correlation coefficient between two random variables A and B, where the covariance between A and B is divided by the standard deviation of A and the standard deviation of B to obtain a normalized absolute value that ranges between 0 and 1. For the neural data, the degree of similarity (as indicated by the cross-correlogram, column 3 of Fig. 1) was normalized by the degree of temporal coding to each stimulus individually (as indicated by the autocorrelograms, columns 1 and 2 of Fig. 1). To emphasize the coding of temporal fine structure, the correlograms used were the difcors (row 2 of Fig. 1). The cross-correlation coefficient for fine structure (ρ_{TFS}) was computed from the difcors as:

$$\rho_{TFS} = \frac{\text{difcor}_{AB}}{\sqrt{\text{difcor}_A \times \text{difcor}_B}} \quad (1)$$

where each of the difcors were evaluated at the corresponding CD. For all conditions, the CD of difcor(A) and difcor(B) was zero, by definition for autocorrelation functions. For the cross-correlograms, the CD was taken as zero when a single fiber was responding to two stimuli without a defined delay between one another. Otherwise, the CD was the delay at which the difcor peak occurred. For the single-fiber responses in Figure 1, the computed value of ρ_{TFS} was −0.03, which was very close to 0 as expected for two uncorrelated noises. Likewise, the neural cross correlation coefficient for envelope (ρ_{ENV}) was computed from the corrected sumcors (after subtracting the baseline value of 1) as:

$$\rho_{ENV} = \frac{(\text{sumcor}_{AB} - 1)}{\sqrt{(\text{sumcor}_A - 1) \times (\text{sumcor}_B - 1)}} \quad (2)$$

where the sumcor values were taken at the CD for the given condition. The value of ρ_{ENV}=0.01 obtained for the data in Figure 1 was also quite close to zero for these two uncorrelated noises.

A limitation of these stationary cross-correlation analyses for the present study (and similar perceptual approaches, e.g., Sheft et al. 2008) is that they only estimate the overall strength of recovered envelopes averaged across time-varying speech stimuli, i.e., they preclude the analysis of temporal variations in recov-

ered envelopes that may occur. Short-time analyses could be developed to explore the nonstationarity of recovered envelopes, but such analyses would be limited by the large amount of data required to provide robust estimates of auto- and cross-correlations within short temporal windows. Alternative approaches could be explored, such as the compound PST histogram (e.g., Pfeiffer and Kim 1972), which is similar to the difcor in that it combines responses to stimuli of opposite polarity and could be extended to estimate envelope in addition to fine-structure coding. However, the reliability and noise floor of such an approach would need to be compared to the present shuffled-correlogram-based approach, which benefits from efficient smoothing provided by the shuffling procedure.

RESULTS

Spike trains were obtained from 28 chinchilla AN fibers as well as from a computational AN model to test the robustness of the neural cross-correlation metrics across a variety of conditions. The dynamic range of the neural cross-correlation coefficients was tested first in several conditions involving broadband noise. The neural cross-correlation metrics were used to provide physiological evidence for the recovery of speech envelope cues from speech fine structure at the output of the cochlea. Model responses were used to evaluate systematically the dependence of temporal envelope recovery on the number of analysis bands used to generate chimaeric speech and to predict the effects of sensorineural hearing loss (SNHL) on envelope recovery. In all cases, separate cross-correlation coefficients were computed for temporal envelope and fine-structure responses.

Separating envelope and fine-structure cross-correlations in broadband noise responses

Broadband noise does not have a particularly interesting envelope or fine-structure pattern; however, when passed through a narrowband cochlear filter, well-defined temporal envelope and fine-structure waveforms are created that depend on the characteristics of the filter (Ruggero 1973; Ghitza 2001; Joris 2003; Louage et al. 2004). Here, broadband noise was used as a stimulus for the development of neural cross-correlation coefficients for envelope and fine structure. Figure 1 illustrates the computations involved in computing these metrics using two sets of spike trains recorded from a single chinchilla AN fiber responding to two uncorrelated broadband noises (see "Methods" section). The neural cross-correlation coefficients (ρ_{TFS} and ρ_{ENV}) were computed as the

amount of common temporal coding between the two responses (column 3) relative to the degree of temporal coding within each response individually (columns 1 and 2). The neural cross-correlation coefficients for fine structure ($\rho_{TFS}=-0.03$), computed from the difcors (Eq. 1), and for envelope ($\rho_{ENV}=0.01$), computed from the corrected sumcors (Eq. 2), were both very close to zero for these responses to two uncorrelated noises.

The dynamic range of the neural cross-correlation coefficients was tested as a function of CF for AN responses to uncorrelated and correlated broadband noises. Figure 3 shows values of ρ_{TFS} (top row) and ρ_{ENV} (bottom row) computed from both chinchilla (left column) and model (right column) AN spike trains. Values of ρ_{TFS} and ρ_{ENV} for uncorrelated noises (triangles) were consistently below 0.1 (dashed lines) for both recorded and model AN spike trains. Thus, a value of 0.1 was used throughout as an estimate of the noise floor for uncorrelated conditions. The opposite end of the correlation continuum was represented by considering the cross-correlation between two separate sets of spike trains obtained in response to the same noise waveform (A/A). In this case, the main source of variability between the two sets of spike
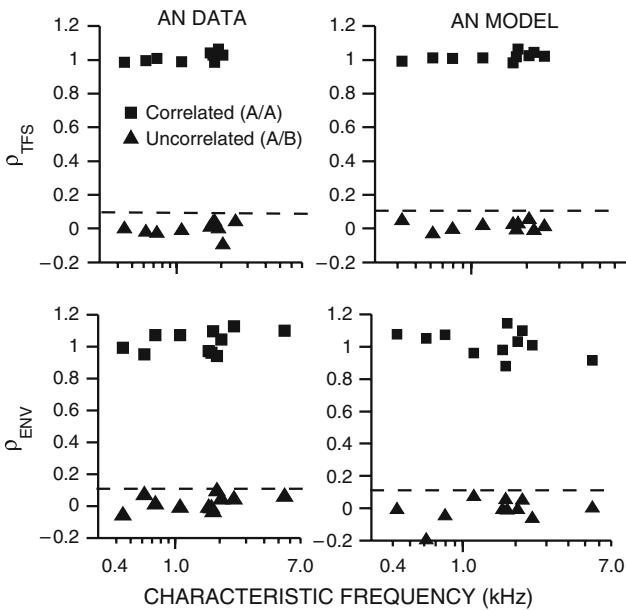

FIG. 3. Neural cross-correlation coefficients ranged from near zero for uncorrelated conditions to near one for correlated conditions. Cross-correlation coefficients between AN spike-train responses to two broadband noises (A and B) are shown for temporal fine-structure (ρ_{TFS}, *top row*) and for temporal envelope (ρ_{ENV}, *bottom row*). *Squares* represent correlations between responses to the same noise (A/A). *Triangles* represent responses to two uncorrelated noises (A/B). *Left column*: Data from chinchilla AN fibers. *Right column*: Data from AN model. *Dashed lines* at $\rho=0.1$ in all panels represent the noise-floor estimate for uncorrelated conditions used throughout. Values of ρ_{TFS} were not computed for CF > 5 kHz due to very low autocorrelogram difcor peak heights.

trains came from the Poisson-like randomness associated with AN responses. For both model and recorded AN spike trains, cross-correlation coefficients computed for both envelope and fine structure were near 1 (squares, Fig. 3), as desired for this condition in which the temporal responses to an identical stimulus should be correlated. Note that some computed values (particularly for ρ_{ENV}) were above 1 due to the stochastic nature of neural responses, which can produce slight mismatches in the three sumcor peaks heights used to compute this normalized metric (Eq. 2). Another demonstration (not shown) of their dynamic range is that both ρ_{TFS} and ρ_{ENV} were systematically reduced as the CF separation increased between two AN fibers responding to the same broadband noise, starting near 1 for no CF separation and decaying to low asymptotic values (near 0.1) for large CF separations (Heinz and Swaminathan 2008). Although this is a different type of cross-correlation (i.e., cross-fiber within-stimulus, rather than within-fiber cross-stimulus), the computations were identical to those illustrated in Figure 1.

Cross-stimulus correlations in chimaeric speech responses

The recovery of speech-envelope cues in cochlear responses to speech fine-structure stimuli has been predicted using perceptually based models and a general cross-correlation framework (Zeng et al. 2004; Gilbert and Lorenzi 2006; Sheft et al. 2008). In this framework, the presence of recovered speech-envelope cues was indicated by a large cross-correlation between the envelope responses (extracted from the model output waveforms) to the original speech and to the corresponding chimaeric stimulus made up of the fine-structure from speech and the envelope from a spectrally matched noise. The neural cross-correlation coefficients developed here were applied in the same general framework to provide a physiologically based evaluation of recovered envelope cues in both recorded and model AN spike trains.

Figure 4 shows the within- and across-stimulus envelope and fine-structure coding in spike trains recorded from a chinchilla AN fiber (CF=490 Hz) responding to the original speech token and to a one-band speech fine-structure chimaera. Both fine-structure and envelope coding were observed in the responses to each stimulus individually for this low-CF fiber. There was significant cross-correlation between the fine-structure responses to the chimaera and to the original speech (Fig. 4F; $\rho_{TFS}=0.69$) because the chimaeric stimulus was created with the speech fine structure. The more interesting result was that there was also significant cross-correlation between the envelope responses to the chimaeric stimulus and the original speech (Fig. 4I; $\rho_{ENV}=$
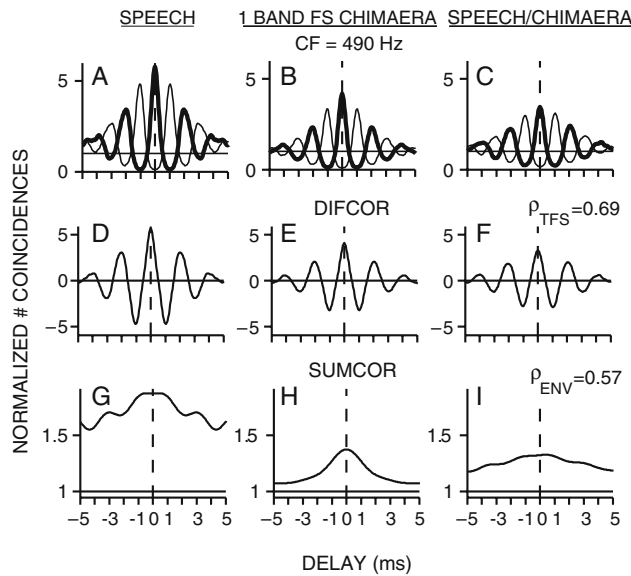
FIG. 4. Recovered speech-envelope cues were present in neural responses to a one-band speech fine-structure (FS) chimaera. Spike trains were recorded from a single chinchilla AN fiber responding to the original speech token (stimulus A, "A boy fell from the window") and to the corresponding chimaera (stimulus B) that had the fine structure of the original speech and the envelope of a spectrally matched noise. A–I same format as in Figure 1. Recovery of speech-envelope cues was demonstrated by a large envelope correlation (ρ_{ENV}=0.57) between the responses to the original speech and to the speech fine-structure chimaera. AN fiber CF=490 Hz; threshold= 45 dB SPL; Q_{10}=1.8; spontaneous rate=0.8 spikes/s. Overall speech sound level was 67 dB SPL.

0.57), even though the chimaera envelope came entirely from noise. The computed value of ρ_{ENV} was well above the noise floor (0.1), which indicated significant recovery of speech envelope cues for the 1-band speech fine-structure chimaera. The same analyses of within- and across-stimulus envelope and fine-structure coding were applied to spike train responses recorded from the same AN fiber in response to the 16-band speech fine-structure chimaera (Fig. 5). In agreement with perceptual modeling (Gilbert and Lorenzi 2006), there was much less recovery of speech envelope cues for the 16-band speech fine-structure chimaera (Fig. 5I; ρ_{ENV}=0.22). As for the one-band case, there was significant envelope coding to each stimulus individually in the 16-band case. However, the similarity between the envelope responses to each stimulus was greatly reduced in the 16-band case. In contrast to the reduction in envelope correlation in the 16-band case, the fine-structure correlation remained high in the 16-band case (ρ_{TFS}=0.58).

The dependence of envelope recovery on CF was examined in a set of chinchilla AN fibers from which spike train responses were measured to one- and 16-band speech fine-structure chimaeras. Figure 6 shows the computed fine-structure (panel A) and envelope

(panel B) neural cross-correlation coefficients as a function of AN-fiber CF. Envelope recovery was most prominent (largest ρ_{ENV}) for the one-band fine-structure chimaera in CFs below 500 Hz. Above 700 Hz, envelope recovery was greatly reduced for the one-band chimaera, as indicated by most values of ρ_{ENV} being near the noise floor. In contrast, for the 16-band chimaera, a small amount of envelope recovery existed (ρ_{ENV} above the noise floor) in most AN fibers across CF. The values of ρ_{ENV} were more similar between the one- and 16-band chimaeras for the higher CFs. Speech fine structure was encoded equally well for the one- and 16-band chimaeras for CFs below 1 kHz. Above 1 kHz, ρ_{TFS} for the one-band chimaera decreased below the values for the 16-band chimaera. These results from chinchilla AN fibers indicate that there are significant speech envelope cues, in addition to fine-structure cues, in low-CF neural responses to one-band fine-structure chimaeras. The presence of envelope cues was decreased in neural responses, but not eliminated, for the 16-band fine-structure chimaera.

The effectiveness of auditory chimaeras in isolating fine-structure and envelope cues at the output of the cochlea was investigated more systematically using spike trains from the computational AN model. Figure 7 shows computed values of ρ_{ENV} and ρ_{TFS} as a function of model CF for one- and 16-band speech-fine-structure and speech-envelope chimaeras. The only case for which a one-band chimaera provided stronger encoding of speech cues than the corresponding 16-band chimaera was for the encod-
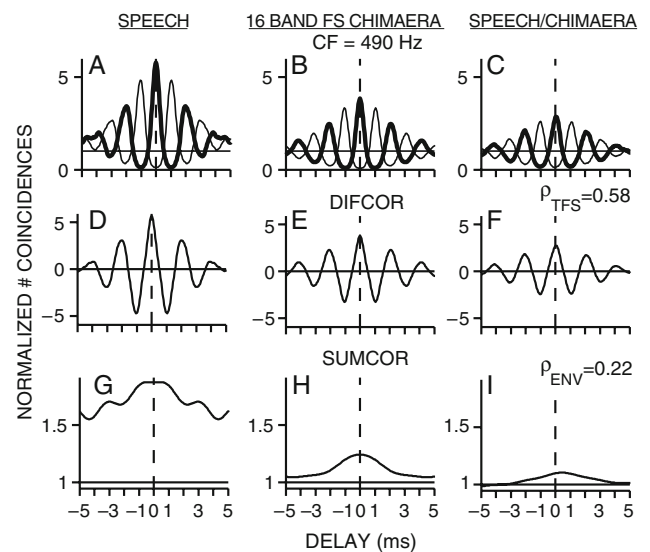

FIG. 5. Recovered speech-envelope cues were reduced in neural responses to a 16-band speech fine-structure (FS) chimaera. Envelope correlation (ρ_{ENV}=0.22) between neural responses to the original speech stimulus and the 16-band chimaera was lower than for the one-band chimaera responses shown in Figure 4. All stimulus and AN-fiber details identical to Figure 4 A–I.
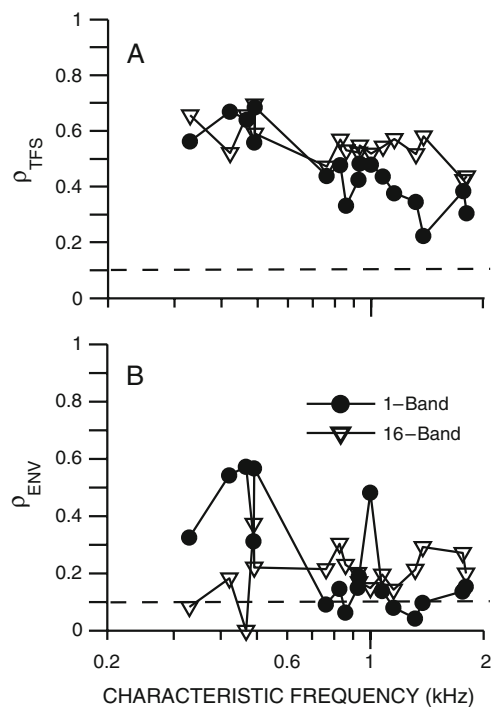
FIG. 6. Recovered envelopes were observed at low CFs in chinchilla AN-fiber responses to speech fine-structure chimaeras. Neural cross-correlation coefficients (ρ_{TFS}, **A**; ρ_{ENV}, **B**) quantify the similarity between single-fiber spike-train responses to a speech token ("A boy fell from the window") and the corresponding chimaera with fine structure from speech and envelope from a spectrally matched noise. *Closed circles*: chimaera generated with one analysis band; *open triangles*: 16-band chimaera. *Dashed lines* at $\rho = 0.1$ represent the estimated noise floor for uncorrelated conditions (see Fig. 3). Overall speech level was chosen to be in the upper one third of each fiber's dynamic range for the original speech token, and ranged from 62 to 72 dB SPL.

ing of speech envelope cues for the speech-fine-structure chimaera (Fig. 7C). Consistent with the recorded AN data for the tested speech utterance (Fig. 6B), the enhanced recovery of speech envelope cues for the one-band chimaera occurred only for CFs up to 500 Hz. Despite the difference between one- and 16-band chimaeras, recovered envelopes were predicted not to be eliminated completely at any CF for either the one- or 16-band chimaera. Significant recovered envelopes were also predicted to exist in AN responses (not shown) for both TFS speech (with no competing acoustic envelope, Gilbert and Lorenzi 2006) and for speech–speech chimaeras (with a meaningful acoustic envelope taken from a different sentence, Smith et al. 2002).

The physiological model predictions suggest that recovered envelopes are greater at low CFs than at high CFs for both the one- and 16-band conditions. This result is consistent with recent perceptual modeling predictions for 16-band TFS speech and has been suggested to result from narrower cochlear filters at low CFs or from the presence of fundamen-

tal-frequency information at low CFs (Sheft et al. 2008). However, perceptual modeling predictions for one-band TFS speech showed recovered envelopes were small at low CFs and maximal around 1 kHz (Gilbert and Lorenzi 2006), which is inconsistent with the physiological results (Figs. 6 and 7). It was noted that the peak in recovered envelopes for the one-band condition corresponded with the maximum acoustic energy being near 1 kHz for the set of VCV stimuli used in the perceptual study; however, this correspondence did not exist for the 16-band condition (Gilbert and Lorenzi 2006). The sentence used in the present study had maximum spectral energy near 550 Hz.

Model predictions of fine-structure coding (Fig. 7A) were also consistent with the recorded AN data (Fig. 6A). Fine-structure coding was similar between both fine-structure chimaeras for low CFs, with ρ_{TFS} dropping for the one-band case at higher CFs. The other "cross-over" condition (i.e., speech-fine-structure coding for the speech-envelope chimae-
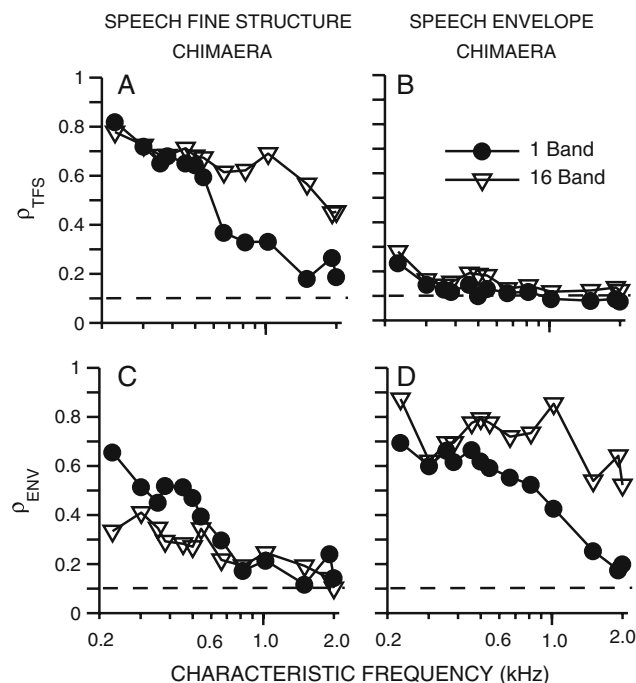

FIG. 7. A–D Neural cross-correlation coefficients were used to predict the effectiveness of auditory chimaeras in isolating fine-structure and envelope cues at the output of the cochlea as a function of CF. Spike trains were obtained from AN-model responses. ρ_{TFS} (*top row*) and ρ_{ENV} (*bottom row*) quantify the similarity between single-fiber spike-train responses to a speech token ("A boy fell from the window") and corresponding speech–noise chimaeras. *Left*: chimaera with fine structure from speech and envelope from a spectrally matched noise. *Right*: chimaera with speech envelope and noise fine structure. *Closed circles*: chimaera generated with one analysis band; *open triangles*: 16-band chimaera. *Dashed lines* at $\rho = 0.1$ represent the estimated noise floor for uncorrelated conditions (see Fig. 3). The overall speech level (35 dB SPL) was chosen to match most closely to the best-modulation levels for the original speech token for all CFs.

ra, Fig. 7B) generally showed minimal fine-structure coding at all CFs for both chimaeras. Speech envelope coding was robust for the 16-band speech envelope chimaera at all CFs and for the one-band chimaera for CFs up to 800 Hz, above which point ρ_{ENV} decreased (Fig. 7D). Thus, auditory chimaeras were predicted to be effective in isolating speech envelope cues in AN responses to both one- and 16-band chimaeras, whereas the isolation of fine-structure cues was difficult to achieve, particularly in the one-band case.

The predicted effect of the number of analysis bands on the fine-structure and envelope coding of chimaeric speech is shown in Figure 8 for a 550-Hz CF model AN fiber. The 550-Hz CF was chosen to match the frequency at which the speech token had maximal energy. The recovery of speech envelope cues, as indicated by the values of ρ_{ENV} for the speech fine-structure chimaeras (squares), was reduced from ~0.45 to ~0.2 as the number of analysis bands increased from 1 to 16. These physiological predictions are consistent with previous perceptually based modeling, where recovered envelopes for TFS speech were reduced most significantly (but not eliminated) for eight and 16 bands (Gilbert and Lorenzi 2006). A similar dependence on the number of analysis bands was predicted with TFS speech stimuli using the neural metrics (Heinz and Swaminathan 2008). In contrast to ρ_{ENV}, the values of ρ_{TFS} for the fine-structure chimaeras remained consistent as the number of bands varied from 1 to 16. For the speech-envelope chimaeras (triangles), which have their fine structure
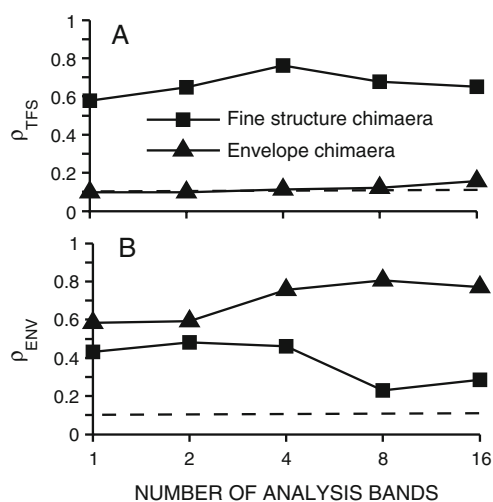

FIG. 8. Recovered envelope cues in neural responses are predicted to decrease as the number of chimaeric analysis bands increases. Spike trains were obtained from one model AN-fiber (CF=550 Hz). Same stimuli were used as in Figure 7. *Squares*: chimaeras with fine structure from speech and envelope from a spectrally matched noise. *Triangles*: chimaera with speech envelope and noise fine structure. *Dashed lines* at ρ=0.1 represent the estimated noise floor for uncorrelated conditions (see Fig. 3). Overall speech level was 35 dB SPL.

derived from noise, ρ_{TFS} values were near the noise floor for all conditions. Speech envelope was well encoded for all of the speech-envelope chimaeras. The value of ρ_{ENV} was lowest for the one-band condition (~0.6), but reached an asymptotic value (~0.8) for as few as four bands. These results suggest that the most significant effect of increasing the number of analysis bands for generating auditory chimaeras from 1 to 16 is that the recovery of envelope cues from speech fine structure is diminished.

The dependence of envelope recovery on the relative bandwidth of cochlear and chimaeric-analysis filters (Gilbert and Lorenzi 2006; Sheft et al. 2008) suggests that the reduced frequency selectivity often associated with SNHL (Liberman and Dodds 1984a; Glasberg and Moore 1986) may degrade the recovery of envelope cues from speech fine structure. Thus, predicted neural cross-correlation coefficients for speech fine-structure chimaeras were compared between normal-hearing and two hearing-impaired versions of the computational AN model (Fig. 9). Outer-hair-cell (OHC) damage was modeled as reducing the gain of the cochlear amplifier, thus reducing cochlear compression, frequency selectivity, and suppression (Zilany and Bruce 2006, 2007). Inner-hair-cell (IHC) damage was modeled as reducing the transduction slope of the IHC, which raised threshold without affecting cochlear nonlinearity, e.g., frequency selectivity was not directly degraded. This implementation of IHC damage produced shallower rate-level functions with shapes that were consistent with those observed following acoustic trauma and furosemide administration (Liberman and Kiang 1984; Sewell 1984; Heinz and Young 2004; Zilany and Bruce 2006). The reduction in spontaneous rate associated with IHC damage (Liberman and Dodds 1984b) was not modeled; however, similar results were observed for both high- and low-spontaneous-rate fibers. Potential temporal effects of IHC damage and/or disrupted AN activity have been implicated in the perceptual effects associated with auditory neuropathy (Zeng et al. 2005a); however, these potential effects were not modeled here due to the lack of thorough single-unit data characterizing these temporal effects.

The impaired predictions shown in Figure 9 represent a 30-dB threshold elevation due to either selective OHC or IHC damage. Overall sound level was increased by 25 dB for the impaired conditions in order to match the best-modulation level of the impaired 550-Hz CF model fibers. Both neural envelope (ρ_{ENV}) and fine-structure (ρ_{TFS}) coding were predicted to be degraded in the case of OHC damage (open triangles), but not in the case of IHC damage (open circles) relative to the normal-hearing predictions (filled squares). The degree of degrada-
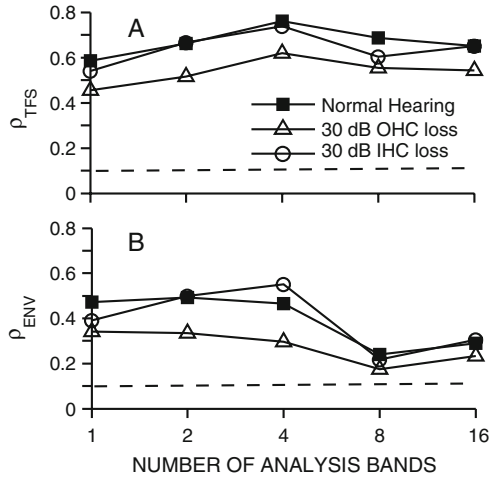
FIG. 9. Sensorineural hearing loss is predicted to degrade both fine-structure and envelope coding of speech in neural responses to psychoacoustic stimuli that contain only speech fine structure. Spike trains were obtained from model AN fibers with CF=550 Hz. Responses are compared from three versions of the AN model: normal hearing (*filled squares*), 30-dB hearing loss due to selective outer-hair-cell (OHC) damage (*open triangles*), and 30-dB selective inner-hair-cell (IHC) hearing loss (*open circles*). Stimuli were the same as used in Figure 8. *Dashed lines* at $\rho=0.1$ represent the estimated noise floor for uncorrelated conditions (see Fig. 3). The overall speech levels (35 dB SPL for normal hearing, 60 dB SPL for impaired conditions) were chosen based on the AN-fiber best-modulation level for the original speech for each model condition.

tion was larger for recovered envelopes (for the 1–4 band conditions) than for fine-structure coding, representing a ~38% decrease in envelope coding compared to a 23% decrease in fine-structure coding. The lack of a predicted degradation for the 30-dB IHC loss provided a useful control to suggest that threshold shift alone does not account for these predicted degradations. Rather, it is likely that the reduction in cochlear nonlinearity that occurs for OHC damage and not for IHC damage is likely to be the basis for the predicted degradations in recovered envelope cues and fine-structure coding of speech fine-structure chimaeras.

DISCUSSION

The neural cross-correlation coefficients developed in this study have broad applications to studying temporal coding in that they provide a general framework for computing the similarity between either envelope or fine-structure components of two sets of spike-train responses. These metrics have a wide dynamic range in both within-fiber and across-fiber conditions, ranging from near 0 for uncorrelated conditions to near 1 for correlated conditions. Although this study focused on temporal coding at the output of the cochlea, the neural cross-correlation coefficients are generally

applicable to auditory spike-train responses from any location within the auditory pathway. More generally, these analyses may be useful for studying the perceptual relevance and neural coding of stimulus information across different time scales in various sensory modalities (e.g., Gamzu and Ahissar 2001; Lu et al. 2001; Vickers et al. 2001; van Boxtel et al. 2006).

A normalized representation of neural cross-correlation

The neural cross-correlation metrics developed here represent an extension of shuffled auto- and cross-correlogram analyses recently developed by Joris and colleagues (Joris 2003; Louage et al. 2004, 2006; Joris et al. 2006a, b). These neural metrics provide normalized representations of correlated temporal coding computed as the degree of common temporal fluctuations in two-spike trains (cross-correlograms) relative to the degree of temporal coding within each spike-train response individually (autocorrelograms). Although this normalization is beneficial for a similarity metric, it eliminates the overall degree of temporal coding and can produce misleading results if used when there is very little baseline temporal coding (e.g., ρ_{TFS} was not computed in Fig. 3 for fiber CFs>5 kHz due to the rolloff in phase locking).

The most significant benefit of this within-fiber normalization is reduced variability in quantifying cross-correlation based on neural responses. The degree of envelope and fine-structure coding (sumcor and difcor peak heights) varies greatly across neurons (e.g., different CFs) and across stimulus conditions (e.g., sound levels; Louage et al. 2004). A population study that quantified correlation based solely on cross-correlogram peak heights (i.e., without normalization) would likely have too much variability to accurately quantify the small correlations identified with the neural cross-correlation coefficients.

Quantifying envelope coding in low-CF neural responses

An extension of previous correlogram analyses (Joris 2003; Louage et al. 2004) was also needed to improve quantification of envelope coding of chimaeric speech for low CFs. Sumcors, which nominally represent envelope coding as the common temporal aspects of responses to a stimulus and its polarity-inverted pair, do not perfectly eliminate fine structure (Fig. 1G) at the low frequencies of primary interest for speech. A more accurate isolation of envelope information was obtained by eliminating fine-structure contributions from a spectral representation of the sumcor (Fig. 2). The Fourier transform of the sumcor (Figs. 1J–L) estimates the envelope spectral

density (or cross-spectral density), which could in fact be analyzed in more detail if certain envelope frequencies were of particular interest.

Envelope coding in the AN model

Although current AN models capture the important qualitative trends in responses to amplitude-modulated sounds, they quantitatively underestimate the degree of envelope coding (Nelson and Carney 2004). The underestimation is due to limitations of synapse adaptation in the model (Nelson and Carney 2004; Zhang and Carney 2005). Most computational AN models of this type would likely have the same limitation (Smith and Brachman 1980; Hewitt and Meddis 1991). This limitation can be seen in high-CF model sumcor peak heights (Fig. 2B), which were lower than corresponding values from AN data (e.g., Fig. 2A; also see Fig. 15A in Louage et al. 2004). However, qualitative trends in model sumcors matched very well with those from AN data. The effect of underestimated envelope coding was likely minimized by the normalized nature of the neural cross-correlation coefficients, as supported by similar findings from model and recorded responses (Figs. 3, 6, 7). Nonetheless, limitations in model predictions further motivate the usefulness of developing neural cross-correlation coefficients that can be applied directly to spike trains recorded from normal-hearing or hearing-impaired animals.

Quantifying envelope recovery at the output of the cochlea

Proper interpretation of the perceptual salience of TFS cues must include consideration of the fact that acoustic TFS not only produces true neural TFS cues, but also recovered envelope cues (Fig. 10). The potential for recovered neural envelope cues to contribute to the perceptual salience of acoustic TFS has important implications for the design of auditory prostheses. Any perceptual benefit of acoustic TFS that arises from recovered envelopes in normal-hearing listeners (sharp cochlear tuning) will not be restored with auditory prostheses designed to enhance TFS coding in listeners with SNHL (broad cochlear tuning), but could be achieved through strategies to restore normal neural envelope coding. The neural metrics developed here provide a general framework in which both true and recovered temporal coding can be quantified at the output of the cochlea for both TFS and envelope cues (Fig. 10). Similar metrics based on waveform responses have been used with perceptual models (Sheft et al. 2008).

The present results based on ρ_{ENV} provide physiological evidence for recovered envelopes in AN
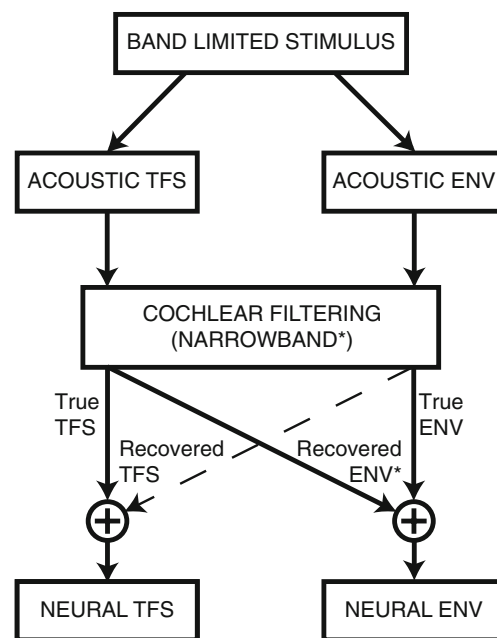

FIG. 10. Acoustic TFS can produce recovered envelope (ENV) cues at the output of the cochlea in addition to true neural TFS cues. Recovered envelopes arise due to narrowband cochlear filtering (*asterisk*; Ghitza 2001) and therefore are reduced or not present in listeners with sensorineural hearing loss or in cochlear-implant patients. Thus, caution is required in applying perceptual results from normal-hearing listeners to the design of auditory prostheses. Note that recovered TFS cues can also arise (Zeng et al. 2004); however, they typically occur with a high number of analysis bands (e.g., 48 or 64) and thus were not apparent in the present study (e.g., Fig. 7).

responses to speech-noise chimaeras (with noisy true envelope cues created by cochlear filtering). Model predictions confirmed that recovered envelopes also exist for TFS speech (without true envelope cues) and speech–speech chimaeras (with meaningful true envelope cues prior to cochlear filtering). Thus, salient physiological recovered envelopes can exist in a variety of conditions with different types of true-envelope cues (Fig. 10).

The existence of physiological recovered envelopes is consistent with perceptual studies demonstrating intelligible recovered envelope cues at the output of gammatone filterbank models (Zeng et al. 2004; Gilbert and Lorenzi 2006). The finding that physiological recovered envelopes were larger for one-band than for 16-band chimaeras, but were not completely eliminated for eight- or 16-band conditions, is also consistent with perceptual results (Gilbert and Lorenzi 2006). Although generally consistent, some important differences and caveats must be considered, since these perceptual results were interpreted as suggesting that recovered envelopes were "essentially abolished" for eight or more analysis bands (Gilbert and Lorenzi 2006; Lorenzi et al. 2006). The model of cat AN tuning likely underestimates envelope recovery in

humans, for which tuning was estimated to be two to three times sharper than cats (Shera et al. 2002; but see Ruggero and Temchin 2005). Also, the prominence of physiological recovered envelopes at low CFs (below ~500 Hz) was consistent with perceptual modeling for 16-band TFS speech (below ~340 Hz; Sheft et al. 2008), but not for one-band conditions (Gilbert and Lorenzi 2006). The exact CF limit for recovered envelopes likely depends on both the species and specific stimuli used; however, it is unclear why the CF dependence for 1- and 16-band conditions would differ in perceptual predictions and not in physiological predictions.

More recent perceptual modeling has confirmed that it is extremely difficult to completely eliminate recovered envelopes even with complex signal processing schemes (Sheft et al. 2008). However, the lack of a significant correlation between the degree of predicted recovered envelopes and measured speech identification across a variety of conditions was taken by Sheft et al. as evidence that recovered envelopes do not contribute substantially to TFS-speech perception. High-pass filtered (at 340 Hz) TFS speech eliminated the predicted prominent recovered envelopes in the fundamental-frequency region. Equivalent identification performance and phonetic-feature reception were obtained for high-pass and unfiltered TFS speech; however, physiological predictions suggest that significant recovered envelopes exist above 340 Hz. Identification for 32-band TFS speech was lower than for 16 bands, which was inconsistent with higher predicted recovered envelopes for 32-band than for 16-band conditions for many CFs below 1,000 Hz. However, predicted physiological recovered envelopes for 32-band chimaeras (not shown) were not different than 16-band conditions (Fig. 7). Further evidence that TFS speech perception is not solely based on recovered envelopes was provided by Gilbert et al. (2007), who demonstrated a larger effect of modulation masking on 16-band envelope speech than on 16-band TFS speech. Although these studies suggest recovered envelopes are not the basis for TFS-speech perception, the discrepancies and caveats discussed suggest a better integration of physiological metrics with perceptual studies would be beneficial.

Implications for the effects of sensorineural hearing loss on TFS cues

Recent demonstrations that listeners with SNHL have a reduced ability to use TFS cues (Lorenzi et al. 2006; Hopkins et al. 2008; Moore 2008) have motivated the idea that novel hearing-aid amplification strategies should be developed to enhance TFS coding. Given that acoustic TFS produces two types of neural cues (Fig. 10), it is important to consider the effects of

SNHL on recovered envelope cues as well as true TFS cues. Outer-hair-cell damage, and associated loss of cochlear nonlinearities, was predicted to degrade recovered envelope cues (Fig. 9), which if perceptually relevant could contribute to an acoustic-TFS deficit. Recent perceptual studies have suggested that reduced frequency selectivity is not the only cause of degraded TFS processing. Listeners with high-frequency SNHL, but with normal low-frequency thresholds (and assumed normal low-CF frequency selectivity), were unable to identify low-pass filtered TFS speech (Lorenzi et al. 2009). Also, listeners with SNHL and only mildly degraded low-CF frequency selectivity had significant deficits in TFS processing for binaural pitch perception (Santurette and Dau 2007). However, the physiological bases for these perceptual deficits remain unknown.

There are several physiological mechanisms hypothesized to underlie the perceptual TFS deficit in SNHL listeners (Moore 2008). Conflicting evidence exists as to whether within-fiber encoding of fine-structure (i.e., phase locking) is degraded following SNHL (Harrison and Evans 1979; Woolf et al. 1981; Miller et al. 1997). Alternatively, a significant effect of SNHL on fine-structure coding may occur in terms of across-fiber encoding. The neural metric ρ_{TFS} provides an intuitive representation of across-fiber fine-structure coding. Across-fiber correlation decreases as the CF separation increases between two AN fibers. Broader tuning from SNHL was predicted to increase the range of CF separations over which correlated activity existed (Heinz and Swaminathan 2008). This degradation would functionally reduce the number of independent information channels in the AN-fiber population. Broader tuning was also predicted to reduce the traveling-wave delay between different CFs, which was quantified as the characteristic delay in across-fiber cross-correlograms. Thus, SNHL is predicted to degrade normal spatiotemporal response patterns, which have been hypothesized to provide robust neural cues for a range of perceptual phenomena, including speech, pitch, and intensity perception, and tone detection in noise (Loeb et al. 1983; Shamma 1985; Heinz et al. 2001b; Carney et al. 2002; Cedolin and Delgutte 2007; Heinz 2007).

Cochlear implants and other applications

In addition to addressing fundamental neural coding questions about normal and impaired hearing, the neural cross-correlation metrics have direct applications in a number of other areas. Recent perceptual findings suggesting an important role for fine structure have led to an effort to provide fine-structure information to cochlear-implant patients, in addition to envelope information currently supplied (Rubinstein

et al. 1999; Litvak et al. 2001; Nie et al. 2005). Neural cross-correlation coefficients are useful metrics for evaluating novel cochlear-implant stimulation (or hearing-aid amplification) strategies because they provide a quantitative physiological framework to test the ability to deliver complex-stimulus-related envelope and/or fine-structure information to the AN. Likewise, audio-coding strategies designed to compress the representation of sound without affecting perception could be evaluated in a physiological framework using neural cross-correlation coefficients. Depending on the acoustic material to be compressed (e.g., speech or music), a varying emphasis on envelope or fine-structure coding could be applied based on neural responses. Thus, the ability of neural cross-correlation coefficients to quantify recovery of stimulus-related temporal cues can be applied in cases for which recovery is undesirable (e.g., psychoacoustic stimuli to isolate fine structure or envelope) and in cases where recovery is desirable (e.g., cochlear implants or audio coding).

ACKNOWLEDGMENTS

REFERENCES

BRUCE IC, SACHS MB, YOUNG ED. An auditory-periphery model of the effects of acoustic trauma on auditory nerve responses. J. Acoust. Soc. Am. 113:369–388, 2003.

CARIANI PA, DELGUTTE B. Neural correlates of the pitch of complex tones. I. Pitch and pitch salience. J. Neurophysiol. 76:1698–1716, 1996a.

CARIANI PA, DELGUTTE B. Neural correlates of the pitch of complex tones. II. Pitch shift, pitch ambiguity, phase invariance, pitch circularity, rate pitch, and the dominance region for pitch. J. Neurophysiol. 76:1717–1734, 1996b.

CARNEY LH. A model for the responses of low-frequency auditory-nerve fibers in cat. J. Acoust. Soc. Am. 93:401–417, 1993.

CARNEY LH, HEINZ MG, EVILSIZER ME, GILKEY RH, COLBURN HS. Auditory phase opponency: a temporal model for masked detection at low frequencies. Acustica-Acta Acustica. 88:334–347, 2002.

CEDOLIN L, DELGUTTE B. Spatio–temporal representation of the pitch of complex tones in the auditory nerve. In: Kollmeier B, Klump G, Hohmann V, Langemann U, Mauermann M, Uppenkamp S

and Verhey J (eds) Hearing—From Sensory Processing to Perception. Springer-Verlag, Berlin, pp. 61–70, 2007.

CHINTANPALLI A, HEINZ MG. The effect of auditory-nerve response variability on estimates of tuning curves. J. Acoust. Soc. Am. 122: EL203–EL209, 2007.

DAU T, VERHEY J, KOHLRAUSCH A. Intrinsic envelope fluctuations and modulation-detection thresholds for narrow-band noise carriers. J. Acoust. Soc. Am. 106:2752–2760, 1999.

GAMZU E, AHISSAR E. Importance of temporal cues for tactile spatial-frequency discrimination. J. Neurosci. 21:7416–7427, 2001.

GHITZA O. On the upper cutoff frequency of the auditory critical-band envelope detectors in the context of speech perception. J. Acoust. Soc. Am. 110:1628–1640, 2001.

GILBERT G, LORENZI C. The ability of listeners to use recovered envelope cues from speech fine structure. J. Acoust. Soc. Am. 119:2438–2444, 2006.

GILBERT G, BERGERAS I, VOILLERY D, LORENZI C. Effects of periodic interruptions on the intelligibility of speech based on temporal fine-structure or envelope cues. J. Acoust. Soc. Am. 122:1336, 2007.

GLASBERG BR, MOORE BC. Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments. J. Acoust. Soc. Am. 79:1020–1033, 1986.

HARRISON RV, EVANS EF. Some aspects of temporal coding by single cochlear fibres from regions of cochlear hair cell degeneration in the guinea pig. Arch. Otorhinolaryngol. 224:71–78, 1979.

HEIL P, NEUBAUER H, IRVINE DR, BROWN M. Spontaneous activity of auditory-nerve fibers: insights into stochastic processes at ribbon synapses. J. Neurosci. 27:8457–8474, 2007.

HEINZ MG. Spatiotemporal encoding of vowels in noise studied with the responses of individual auditory nerve fibers. In: Kollmeier B, Klump G, Hohmann V, Langemann U, Mauermann M, Uppenkamp S and Verhey J (eds) Hearing—From Sensory Processing to Perception. Springer-Verlag, Berlin, pp. 107–115, 2007.

HEINZ MG, SWAMINATHAN J. Neural cross-correlation metrics to quantify envelope and fine-structure coding in auditory-nerve responses. J. Acoust. Soc. Am. 123(A):3056, 2008.

HEINZ MG, YOUNG ED. Response growth with sound level in auditory-nerve fibers after noise-induced hearing loss. J. Neurophysiol. 91:784–795, 2004.

HEINZ MG, COLBURN HS, CARNEY LH. Auditory nerve model for predicting performance limits of normal and impaired listeners. Acoust. Res. Lett. Online 2:91–96, 2001a.

HEINZ MG, COLBURN HS, CARNEY LH. Rate and timing cues associated with the cochlear amplifier: Level discrimination based on monaural cross-frequency coincidence detection. J. Acoust. Soc. Am. 110:2065–2084, 2001b.

HEWITT MJ, MEDDIS R. An evaluation of eight computer models of mammalian inner hair-cell function. J. Acoust. Soc. Am. 90:904–917, 1991.

HOPKINS K, MOORE BC, STONE MA. 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, 2008.

JOHNSON DH. The relationship between spike rate and synchrony in responses of auditory-nerve fibers to single tones. J. Acoust. Soc. Am. 68:1115–1122, 1980.

JORIS PX. Interaural time sensitivity dominated by cochlea-induced envelope patterns. J. Neurosci. 23:6345–6350, 2003.

JORIS PX, YIN TC. Responses to amplitude-modulated tones in the auditory nerve of the cat. J. Acoust. Soc. Am. 91:215–232, 1992.

JORIS PX, LOUAGE DH, CARDOEN L, VAN DER HEIJDEN M. Correlation index: a new metric to quantify temporal coding. Hear. Res. 216–217:19–30, 2006a.

JORIS PX, VAN DE SANDE B, LOUAGE DH, VAN DER HEIJDEN M. Binaural and cochlear disparities. Proc. Natl. Acad. Sci. USA. 103:12917–12922, 2006b.

Joris PX, Louage DH, van der Heijden M. Temporal damping in response to broadband noise. II. Auditory nerve. J. Neurophysiol. 99:1942–1952, 2008a.

Joris PX, Michelet P, Franken TP, McLaughlin M. Variations on a Dexterous theme: peripheral time-intensity trading. Hear. Res. 238:49–57, 2008b.

Kiang NYS, Watanabe T, Thomas EC, Clark LF. Discharge patterns of single fibers in the cat's auditory nerve. Cambridge, MA, MIT, 1965.

Liberman MC, Dodds LW. Single-neuron labeling and chronic cochlear pathology. III. Stereocilia damage and alterations of threshold tuning curves. Hear. Res. 16:55–74, 1984a.

Liberman MC, Dodds LW. Single-neuron labeling and chronic cochlear pathology. II. Stereocilia damage and alterations of spontaneous discharge rates. Hear. Res. 16:43–53, 1984b.

Liberman MC, Kiang NYS. Single-neuron labeling and chronic cochlear pathology. IV. Stereocilia damage and alterations in rate- and phase-level functions. Hear. Res. 16:75–90, 1984.

Litvak L, Delgutte B, Eddington D. Auditory nerve fiber responses to electric stimulation: modulated and unmodulated pulse trains. J. Acoust. Soc. Am. 110:368–379, 2001.

Loeb GE, White MW, Merzenich MM. Spatial cross-correlation—a proposed mechanism for acoustic pitch perception. Biol. Cybernetics. 47:149–163, 1983.

Logan BF, Jr. Information in the zero crossings of bandpass signals. Bell Syst. Tech. J. 56:487–510, 1977.

Lopez-Poveda EA. Spectral processing by the peripheral auditory system: facts and models. Int. Rev. Neurobiol. 70:7–48, 2005.

Lorenzi C, Gilbert G, Carn H, Garnier S, Moore BC. 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.

Lorenzi C, Debruille L, Garnier S, Fleuriot P, Moore BC. Abnormal processing of temporal fine structure in speech for frequencies where absolute thresholds are normal. J. Acoust. Soc. Am. 125:27–30, 2009.

Louage DH, van der Heijden M, Joris PX. Temporal properties of responses to broadband noise in the auditory nerve. J. Neurophysiol. 91:2051–2065, 2004.

Louage DH, Joris PX, van der Heijden M. Decorrelation sensitivity of auditory nerve and anteroventral cochlear nucleus fibers to broadband and narrowband noise. J. Neurosci. 26:96–108, 2006.

Lu T, Liang L, Wang X. Temporal and rate representations of time-varying signals in the auditory cortex of awake primates. Nat. Neurosci. 4:1131–1138, 2001.

Miller RL, Schilling JR, Franck KR, Young ED. Effects of acoustic trauma on the representation of the vowel /ε/ in cat auditory nerve fibers. J. Acoust. Soc. Am. 101:3602–3616, 1997.

Moore BC. 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, 2008.

Nelson PC, Carney LH. A phenomenological model of peripheral and central neural responses to amplitude-modulated tones. J. Acoust. Soc. Am. 116:2173–2186, 2004.

Nie KB, Stickney G, Zeng FG. Encoding frequency modulation to improve cochlear implant performance in noise. IEEE. Trans. Biomed. Eng. 52:64–73, 2005.

Pfeiffer RR, Kim DO. Response patterns of single cochlear nerve fibers to click stimuli: descriptions for cat. J. Acoust. Soc. Am. 52:1669–1677, 1972.

Qin MK, Oxenham AJ. Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers. J. Acoust. Soc. Am. 114:446–454, 2003.

Rice SO. Distortion produced by band limitation of an FM wave. Bell Syst. Tech. J. 52:605–626, 1973.

Rubinstein JT, Wilson BS, Finley CC, Abbas PJ. Pseudospontaneous activity: stochastic independence of auditory nerve fibers with electrical stimulation. Hear. Res. 127:108–118, 1999.

Ruggero MA. Response to noise of auditory nerve fibers in the squirrel monkey. J. Neurophysiol. 36:569–587, 1973.

Ruggero MA, Temchin AN. Unexceptional sharpness of frequency tuning in the human cochlea. Proc. Natl. Acad. Sci. U. S. A. 102:18614–18619, 2005.

Saberi K, Hafter ER. A common neural code for frequency- and amplitude-modulated sounds. Nature 374:537–539, 1995.

Santurette S, Dau T. Binaural pitch perception in normal-hearing and hearing-impaired listeners. Hear. Res. 223:29–47, 2007.

Sewell WF. Furosemide selectively reduces one component in rate-level functions from auditory-nerve fibers. Hear. Res. 15:69–72, 1984.

Shamma SA. Speech processing in the auditory system. I: The representation of speech sounds in the responses of the auditory nerve. J. Acoust. Soc. Am. 78:1612–1621, 1985.

Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. Science. 270:303–304, 1995.

Sheft S, Ardoint M, Lorenzi C. Speech identification based on temporal fine structure cues. J. Acoust. Soc. Am. 124:562–575, 2008.

Shen C, Smith ZM, Oxenham AJ, Delgutte B. Auditory Chimera Demo. http://research.meei.harvard.edu/Chimera/, 2001.

Shera CA, Guinan JJ, Jr., Oxenham AJ. Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. Proc. Natl. Acad. Sci. U. S. A. 99:3318–3323, 2002.

Shera CA, Guinan JJ, Oxenham AJ. Otoacoustic estimates of cochlear tuning: validation in the chinchilla. Assoc Res Otolaryngol Abs 30:519, 2007.

Smith RL, Brachman ML. Response modulation of auditory-nerve fibers by AM stimuli: effects of average intensity. Hear. Res. 2:123–133, 1980.

Smith ZM, Delgutte B, Oxenham AJ. Chimaeric sounds reveal dichotomies in auditory perception. Nature 416:87–90, 2002.

Tan Q, Carney LH. A phenomenological model for the responses of auditory-nerve fibers. II. Nonlinear tuning with a frequency glide. J. Acoust. Soc. Am. 114:2007–2020, 2003.

Temchin AN, Rich NC, Ruggero MA. Threshold tuning curves of chinchilla auditory-nerve fibers. I. Dependence on characteristic frequency and relation to the magnitudes of cochlear vibrations. J. Neurophysiol. 100:2889–2898, 2008a.

Temchin AN, Rich NC, Ruggero MA. Threshold tuning curves of chinchilla auditory nerve fibers. II. Dependence on spontaneous activity and relation to cochlear nonlinearity. J. Neurophysiol. 100:2899–2906, 2008b.

van Boxtel JJ, van Ee R, Erkelens CJ. A single system explains human speed perception. J. Cogn. Neurosci. 18:1808–1819, 2006.

Vickers NJ, Christensen TA, Baker TC, Hildebrand JG. Odour-plume dynamics influence the brain's olfactory code. Nature 410:466–470, 2001.

Voelcker HB. Towards a unified theory of modulation. I. Phase–envelope relationships. Proc. IEEE. 54:340–354, 1966.

Wilson BS, Finley CC, Lawson DT, Wolford RD, Eddington DK, Rabinowitz WM. Better speech recognition with cochlear implants. Nature 352:236–238, 1991.

Woolf NK, Ryan AF, Bone RC. Neural phase-locking properties in the absence of cochlear outer hair cells. Hear. Res. 4:335–346, 1981.

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

Young ED, Barta PE. Rate responses of auditory nerve fibers to tones in noise near masked threshold. J. Acoust. Soc. Am. 79:426–442, 1986.

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

ZENG FG, KONG YY, MICHALEWSKI HJ, STARR A. Perceptual consequences of disrupted auditory nerve activity. J. Neurophysiol. 93:3050–3063, 2005a.

ZENG FG, NIE K, STICKNEY GS, KONG YY, VONGPHOE M, BHARGAVE A, WEI C, CAO K. Speech recognition with amplitude and frequency modulations. Proc. Natl. Acad. Sci. U. S. A. 102:2293–2298, 2005b.

ZHANG X, CARNEY LH. Analysis of models for the synapse between the inner hair cell and the auditory nerve. J. Acoust. Soc. Am. 118:1540–1553, 2005.

ZHANG X, HEINZ MG, BRUCE IC, CARNEY LH. A phenomenological model for the responses of auditory-nerve fibers: I. Nonlinear tuning with compression and suppression. J. Acoust. Soc. Am. 109:648–670, 2001.

ZILANY MSA, BRUCE IC. Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery. J. Acoust. Soc. Am. 120:1446–1466, 2006.

ZILANY MSA, BRUCE IC. Representation of the vowel /ɛ/ in normal and impaired auditory nerve fibers: model predictions of responses in cats. J. Acoust. Soc. Am. 122:402–417, 2007.