Predicted effects of sensorineural hearing loss on across-fiber envelope coding in the auditory nerve^{a)}

Jayaganesh Swaminathan^{b)}

Department of Speech, Language, and Hearing Sciences, Purdue University, West Lafayette, Indiana 47907-2038

Michael G. Heinz

Weldon School of Biomedical Engineering, Purdue University, West Lafayette, Indiana 47907-2032

(Received 27 May 2010; revised 23 December 2010; accepted 30 March 2011)

Cross-channel envelope correlations are hypothesized to influence speech intelligibility, particularly in adverse conditions. Acoustic analyses suggest speech envelope correlations differ for syllabic and phonemic ranges of modulation frequency. The influence of cochlear filtering was examined here by predicting cross-channel envelope correlations in different speech modulation ranges for normal and impaired auditory-nerve (AN) responses. Neural cross-correlation coefficients quantified across-fiber envelope coding in syllabic (0–5 Hz), phonemic (5–64 Hz), and periodicity (64–300 Hz) modulation ranges. Spike trains were generated from a physiologically based AN model. Correlations were also computed using the model with selective hair-cell damage. Neural predictions revealed that envelope cross-correlation decreased with increased characteristic-frequency separation for all modulation ranges (with greater syllabic-envelope correlation than phonemic or periodicity). Syllabic envelope was highly correlated across many spectral channels, whereas phonemic and periodicity envelopes were correlated mainly between adjacent channels. Outer-hair-cell impairment increased the degree of cross-channel correlation for phonemic and periodicity ranges for speech in quiet and in noise, thereby reducing the number of independent neural information channels for envelope coding. In contrast, outer-hair-cell impairment was predicted to decrease cross-channel correlation for syllabic envelopes in noise, which may partially account for the reduced ability of hearing-impaired listeners to segregate speech in complex backgrounds.

© 2011 Acoustical Society of America. [DOI: 10.1121/1.3583502]

PACS number(s): 43.71.Qr, 43.71.Ky, 43.71.Gv, 43.64.Sj [KWG]

Pages: 4001–4013

I. INTRODUCTION

Speech is perhaps the most ecologically important acoustic stimulus to humans. The speech waveform is often characterized as a sum of amplitude-modulated signals contained within a set of narrow frequency channels distributed across the acoustic spectrum (e.g., [Flanagan, 1980](#)). In this view, the output of each channel is described as a carrier signal that specifies the waveform fine structure and a modulating signal that specifies its temporal envelope. Envelope information has been shown to be important for speech perception and supports robust speech identification in quiet when provided in as few as four spectral channels ([Shannon et al., 1995](#)). The acoustic speech envelope is complex and can be considered as made up of several ranges of modulation frequencies ([Rosen, 1992](#)). Previous psychophysical studies have suggested that there are differential contributions of low-, mid-, and high-rate envelope fluctuations across spectral channels for speech identification in quiet and in noisy listening conditions ([Apoux and Bacon, 2004](#);

[2008](#)). Moreover, cross-channel envelope correlations have been hypothesized to influence speech intelligibility, particularly in adverse conditions ([Crouzet and Ainsworth, 2001](#)).

Acoustical analyses indicate that for very low-rate modulations (<5 Hz), the correlation of envelope information between most spectral channels is very high (>0.8) ([Crouzet and Ainsworth, 2001](#); see also [Steeneken and Houtgast, 1999](#) and [Vickers et al., 2009](#)). This high correlation reflects the syllabic rate in speech (i.e., syllable onsets and offsets are similar across the speech spectrum) ([Arai and Greenberg, 1997](#)). For higher-rate envelope modulations (>8 Hz), the correlation remains significant only for adjacent spectral channels. This suggests that temporal envelopes extracted from remote spectral regions do not necessarily convey the same information and may be independent. Results from recent psychophysical studies have found that normal-hearing (NH) listeners take advantage of such independence between the envelopes from different spectral channels for speech perception in adverse listening conditions ([Healy and Bacon, 2002](#); [Healy et al., 2005](#); [Baskent, 2006](#)). Using multichannel envelope vocoders, [Baskent \(2006\)](#) found that in conditions with background noise, the number of vocoder channels at which performance reached an asymptote was 1.5 to 2 times larger for NH listeners than for hearing-impaired (HI) listeners. The ability of NH listeners to benefit from more spectral channels was linked to sharp tuning of the auditory filters, i.e., better frequency selectivity as

^{a)}Portions of this research were presented at the 33rd Midwinter Meeting of the Association for Research in Otolaryngology, Anaheim, CA, February 2010.

^{b)}Author to whom correspondence should be addressed. Current address: Research Laboratory of Electronics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Room 36-757, Cambridge, MA 02139. Electronic mail: jswamy@mit.edu

compared to HI listeners. The main cause for reduced frequency selectivity with impairment is thought to arise from damage to the outer hair cells (OHCs) (Liberman and Dodds, 1984). OHCs are believed to be an essential part of the active mechanism in the cochlea that gives rise to several cochlear nonlinearities. When there is damage in the cochlea, particularly in the OHCs, the nonlinearities are reduced and the auditory filters are broadened (Liberman and Dodds, 1984; Glasberg and Moore, 1986; Patuzzi *et al.*, 1989; Ruggero and Rich, 1991).

Most of the results linking the limited use of across-channel envelope cues by HI listeners to reduced frequency selectivity have emerged from perceptual studies (Healy and Bacon, 2002; Healy *et al.*, 2005; Baskent, 2006). However, the neural correlates of these perceptual observations remain unknown. The goal of the present study was to evaluate quantitatively the influence of cochlear filtering by predicting cross-fiber envelope correlations in different modulation-frequency ranges for normal and impaired auditory-nerve (AN) responses. Spike train responses to complex stimuli were obtained from a phenomenological model of the AN. Part of the difficulty in evaluating the neural bases for perceptual effects of sensorineural hearing loss (SNHL) stems from the absence of appropriate metrics that quantify cross-channel envelope correlations from neural spike train responses. To address this limitation, neural cross-correlation coefficients were developed to quantify across-fiber similarity in envelope responses within different ranges of modulation frequency. Furthermore, a hearing-impaired version of the AN model was used with these neural metrics to evaluate quantitatively the effect of selective hair cell damage on cross-fiber envelope correlation.

II. ACROSS-FIBER ENVELOPE CODING OF BROADBAND NOISE

Broadband noise (BBN) creates a well defined temporal envelope when passed through a narrowband filter (Rice, 1954). Thus, neural responses to BBN demonstrate significant envelope coding due to the narrowband nature of cochlear filters (Joris, 2003). The purpose of this first section of the study was to use BBN as a control stimulus to examine the predicted effects of selective hair-cell damage on across-fiber envelope coding when the envelope responses were primarily cochlear induced. These predictions provide a useful comparison to more complex stimuli, such as speech, where both stimulus- and cochlear-induced envelope responses are present (considered in Sec. III).

A. Methods

1. Auditory nerve model

Spike trains were obtained from a computational AN model that has been tested extensively against neurophysiological data obtained from normal-hearing and hearing-impaired cats (Zilany and Bruce, 2006; 2007b). The phenomenological AN model represents an 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 speechlike sounds (Carney, 1993; Heinz *et al.*, 2001a; Zhang *et al.*, 2001; Bruce *et al.*, 2003; Tan and Carney, 2003; Zilany and Bruce, 2006; 2007b). Model threshold tuning curves have been well fit to the CF dependent variation in bandwidth for normal-hearing cats (Miller *et al.*, 1997). 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 (Heinz, 2010). 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).

A key feature of the model is that the functionality of OHCs and inner hair cells (IHCs) is specified by two distinct model parameters that take a value within a range of 1 (normal function) to 0 (complete hair-cell loss). Intermediate values result in partial IHC or OHC damage. OHC damage was modeled as reducing the gain of the cochlear amplifier, thus reducing cochlear compression, frequency selectivity, and suppression (Zilany and Bruce, 2006; 2007b). Figure 1 shows the dB loss due to selective OHC impairment for all CFs used in this study. For CFs above 1 kHz, values of C_{OHC} were chosen between 0 and 1 to approximate a fixed 30-dB loss. However, for CFs below 1 kHz, the maximum gain provided by OHCs (and thus the maximum loss due to OHC damage) is less than 30 dB due to reduced cochlear-amplifier gain at low CFs in the AN model (Zilany and Bruce, 2006). Thus, for low CFs for which 30 dB of OHC loss was not able to be obtained, C_{OHC} was set to zero to achieve the maximum OHC loss for those CFs. However, for all CFs the

FIG. 1. Model threshold elevations for the selective OHC impairment conditions in the present study. Data points represent the CFs used in this study. For CFs above 1 kHz, a fixed 30-dB loss was approximated by choosing appropriate values of C_{OHC} between 0 and 1 in the model. Threshold shifts less than 30 dB for CFs < 1 kHz represent the maximum amount of OHC loss (i.e., complete OHC loss corresponding to $C_{\text{OHC}} = 0$) produced by the AN model, which has reduced cochlear gain at lower CFs.

tuning was significantly broader in the predictions for selective OHC impairment as compared to NH (e.g., ~ 1.4 to 2.3 times broader, depending on CF). IHC damage was modeled as reducing the transduction slope of the IHC (i.e., the slope of the function that transforms BM motion into IHC potential), which raises threshold without significantly affecting cochlear nonlinearity such that frequency selectivity is not directly degraded (Bruce *et al.*, 2003). 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 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. Stimuli were resampled to 100 kHz prior to presentation to the model.

2. Across-fiber envelope cross-correlation metrics computed from spike trains

Neural cross-correlation coefficients have been developed to provide a general framework for computing the similarity between either temporal fine-structure or envelope components of two sets of spike-train responses (Heinz and Swaminathan, 2009). The neural cross-correlation coefficients provide metrics ranging from 0 to 1 that represent the degree of similarity between the temporal envelope or fine structure of responses to two different conditions [e.g., one neuron responding to different stimuli (Heinz and Swaminathan, 2009), or two neurons responding to the same stimulus (present study; also see Heinz *et al.*, 2010)].

Shuffled correlogram analyses (Louage *et al.*, 2004; Joris *et al.*, 2006) were used to quantify within- and across-CF envelope coding from single AN-fiber responses to broadband noise and speech. Figure 2 shows correlogram analyses for two CFs separated by 0.4 octaves based on spike trains recorded from a normal-hearing chinchilla in response to broadband noise (Heinz and Swaminathan, 2009). Within-fiber temporal coding was evaluated based on normalized shuffled auto correlograms [(SACs, thick lines, Figs. 2(A) and 2(B)], which were computed by comparing spike times between all possible pairs of stimulus presentations for a given CF separation. For each pair, intervals between every spike in the first spike train and every spike in the second spike train were tallied with a 50- μ sec binwidth to create a shuffled all-order interval histogram. SACs are typically normalized [by $N(N-1)r^2\Delta\tau D$, where N is the number of stimulus repetitions, 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 a function of delay (or interspike interval), and are therefore much like autocorrelation functions. As such,

FIG. 2. Correlogram analyses of envelope coding *within* (columns 1 and 2) and *across* (column 3) spike trains from two chinchilla AN fibers [(A) $CF_A=827$ Hz; (B) $CF_B=618$ Hz; 0.4-octave separation] responding to the same broadband noise. (A),(B) Normalized shuffled autocorrelograms [thick line, e.g., $SAC(A+)$] and cross-polarity correlogram [thin line, e.g., $SCC(A+,A-)$]. (C) Shuffled cross-fiber correlogram [thick line, e.g., $SCC(A+,B+)$] and cross-fiber, cross-polarity correlogram [thin line, e.g., $SCC(A+,B-)$]. The X denotes characteristic delay ($CD_{SCC}=400$ μ s), which occurs due to the traveling-wave delay. (D)–(F) Corrected sumcors (see text), which emphasize envelope coding, were based on the average of the cross-polarity and auto- or cross-fiber correlograms. Sumcor peak height [in (D) and (E)] quantifies within-fiber envelope coding. Across-CF envelope coding is quantified with a neural cross-correlation coefficient [ρ_{ENV} , Eq. (1)] by comparing the peak heights of sumcor(AB) to sumcor(A) and sumcor(B). Spike train data from Heinz and Swaminathan (2009).

the SACs in Figs. 2(A) and 2(B) have a peak at zero delay, with a symmetric damped oscillatory shape similar to an autocorrelation function for band-limited noise.

Responses to positive and negative polarity versions of each stimulus were recorded to facilitate the isolation of envelope (and fine-structure coding, which was not considered in this study). Cross-polarity correlograms [$SCC(A+,A-)$, thin lines, Figs. 2(A) and 2(B)] were computed by tallying intervals between all spikes in response to the positive and negative polarity versions of the stimulus. Polarity inversion acts to invert the stimulus fine-structure while not affecting the stimulus envelope. Thus, sumcors computed by averaging the $SCC(A+,A-)$ with the SAC emphasize envelope coding, which was significant (sumcor peak greater than baseline value of 1) for both CFs in Fig. 2. Corrected sumcors were computed by eliminating potential leakage of fine structure information into the sumcors, which can occur for envelope modulation frequencies above CF (see Fig. 1J in Heinz and Swaminathan, 2009). This leakage reflects distortion that arises from rectification associated with neural responses and can produce overestimates of envelope coding for low-CF AN fibers (see Fig. 2 in Heinz and Swaminathan, 2009). Rectification creates undesirable high-frequency oscillations (roughly at $2\times CF$) in the sumcors that are not associated with the slow envelope response. Fine-structure leakage is isolated most easily in the spectral domain, where it is apparent in the Fourier transform of the sumcor as a high-frequency spectral peak centered at $2\times CF$. The undesirable contribution of fine-structure coding to the sumcor was eliminated by considering only the envelope spectra below CF. The choice of CF as the cutoff frequency

represents a compromise that eliminates the energy locus near $2 \times \text{CF}$ due to fine-structure leakage, while including as much low-frequency envelope energy as possible. Corrected sumcors were computed as the inverse Fourier transform of the envelope spectra below CF.

Across-CF envelope coding was evaluated based on shuffled cross-fiber correlograms [e.g., $\text{SCC}(\text{A}+, \text{B}+)$, thick line in Fig. 2C] and cross-fiber, cross-polarity correlograms [e.g., $\text{SCC}(\text{A}+, \text{B}-)$, thin line in Fig. 2C], which were computed by comparing spike trains across a pair of CFs. For each CF pair, the cross-correlogram sumcor was used to evaluate across-CF envelope coding. For BBN responses, the contributions of all envelope fluctuations up to the CF of the AN fiber were considered in the sumcor. A neural cross-correlation coefficient (ρ_{ENV}) was used to represent the degree of similarity between the two spike-train responses, and was computed as the ratio of the peak height of the cross-correlogram sumcor [Fig. 2(F)] to the geometric mean of the autocorrelogram sumcor peak heights [Figs. 2(D)–2(E)]. The cross-correlation coefficient for envelope was computed from the corrected sumcors (after subtracting the baseline value of 1) as

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

Values of ρ_{ENV} range from near zero for uncorrelated conditions (noise floor ≈ 0.01) to near one for correlated conditions (Heinz and Swaminathan, 2009). A significant benefit of this self-normalized similarity metric is that the degree of cross correlation is evaluated relative to the strength of envelope coding within each fiber, which varies with differences in CF, spontaneous rate, and stimulus level (Louage *et al.*, 2004). The computed value of $\rho_{\text{ENV}} = 0.42$ in Fig. 2 indicates that there is significant common temporal envelope in these responses for CFs separated by 0.4 octaves.

3. Procedures

AN model spike trains were obtained in response to frozen Gaussian broadband noise which had a duration of 1.7 s. 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 responses were obtained for 16–20 repetitions of each stimulus, which was sufficient to collect roughly 3500 spikes for each stimulus condition. For all analyses, spikes within the first 50 ms of the response were excluded to avoid onset effects (Louage *et al.*, 2004). The sound levels were chosen for each AN fiber to maximize the number of spikes, while minimizing the degradation in envelope coding that occurs at high levels due to saturation (Joris and Yin, 1992; Louage *et al.*, 2004). For predictions corresponding to both normal and impaired cases, data were collected at the best modulation level (BML) for each stimulus type. BML was determined for each model fiber as the sound level (based on 5-dB steps) that produced the maximum amount of envelope coding, as quantified by the sumcor peak height. The predictions for OHC and IHC damage were made based on a 30-dB threshold elevation in each AN fiber (or maximum OHC loss for CFs < 1 kHz; see Fig. 1). Typically, the predicted BMLs for the impaired cases with 30-dB loss were 20–25 dB above that for NH. Although BML was chosen for convenience of presentation, this choice is not critical for the results in the present study due to the normalized nature of the ρ_{ENV} metric.

B. Results and discussion

Figure 3 illustrates the predicted decay of across-fiber envelope coding with CF separation for broadband noise.

FIG. 3. Model predictions of the effects of SNHL on across-CF coding of envelope for CFs centered at 500 Hz [panel (A)] and 3500 Hz [panel (B)]. Stimuli: Broadband noise; duration: 1.7 s. ρ_{ENV} is plotted as a function of CF separation. Each data point represents mean ρ_{ENV} computed over twenty repetitions. Filled circles: Predictions for normal hearing at best modulation levels (BML = 45 dB SPL for 500 Hz and 30 dB SPL for 3500 Hz). Filled diamonds: Predictions for normal hearing at the same sound level as the predictions for the impaired cases (65 dB SPL for 500 Hz and 60 dB SPL for 3500 Hz). Open triangles: OHC damage (21 dB loss for 500 Hz, see Fig. 1; 30 dB loss for 3500 Hz); open squares: 30-dB IHC damage. BML for the impaired case was 65 dB SPL for 500 Hz and 60 dB SPL for 3500 Hz. The smallest CF separation (ΔCF in octaves) at which ρ_{ENV} dropped to 0.7 is shown using arrows and the values are inset. The NH value in parenthesis represents the NH prediction for the equal-SPL condition.

Data are shown for NH fibers at two sound levels and from hearing-impaired fibers with either selective OHC or IHC damage. Neural cross-correlation coefficients (ρ_{ENV}) were computed for 21 CF pairs with CF separation ranging from 0 to 1 octaves for CFs geometrically centered at 500 Hz (panel A) and 3500 Hz (panel B). Predicted envelope cross-correlation (ρ_{ENV}) decreased monotonically with increasing CF separation for NH and selective OHC and IHC impairment.

For predictions in both the NH and HI cases, it is apparent that cross-fiber envelope correlation decays more rapidly for the higher CF [3500 Hz; panel (B)] than for the lower CF [500 Hz; panel (A)] condition. For example, ρ_{ENV} in the NH case reached the noise floor ($\rho_{\text{ENV}} \sim 0.1$) at a CF separation of roughly 0.25 octaves for the 3500-Hz CF, in contrast to the 500-Hz CF, where ρ_{ENV} did not reach the noise floor until a CF separation of roughly 0.55 octaves. The more rapid decay of cross-fiber correlation is consistent with the sharper tuning that is typically observed at higher CFs. For example, AN tuning curves are roughly 2.5 times sharper in cats for CF = 3500 Hz (mean $Q_{10} \sim 5$) than for CF = 500 Hz (mean $Q_{10} \sim 2.1$), where Q_{10} equals the ratio of CF to tuning-curve bandwidth 10 dB above threshold (Miller *et al.*, 1997). It should also be noted that the model of cat AN tuning likely underestimates the degree of tuning in humans, for whom tuning has been estimated to be two to three times sharper than for cats (Shera *et al.*, 2002; however, also see Ruggero and Temchin, 2005). Nonetheless, the general effects predicted here are expected to be the same for human neural responses.

The width of the correlated region was quantified as the CF separation at which ρ_{ENV} first dropped to 0.7. Predictions for selective OHC impairment demonstrated a much wider (2–3 times) CF range of correlated activity (0.35 octaves for 500 Hz and 0.26 octaves for 3500 Hz) as compared to predictions for the NH case (0.16 octaves for 500 Hz and 0.07 octaves for 3500 Hz). IHC damage was also predicted to produce a broader correlated region (0.20 octaves for 500 Hz and 0.15 octaves for 3500 Hz), but to a much lesser degree than OHC damage. However, the broadening with IHC damage appears to be due primarily to the higher sound level necessary to overcome the 30-dB IHC loss, which leads to broader tuning based on normal OHC function (Bruce *et al.*, 2003). The correlated widths for the normal-hearing fibers in the equal-SPL conditions were 0.17 octaves for 500 Hz and 0.14 octaves for 3500 Hz, which were quite similar to the predictions for IHC impairment. Note that the maximal OHC impairment produced by the model for CF = 500 Hz was ~ 21 dB (Fig. 1), which is somewhat less than the flat 30-dB IHC loss. Hence the estimated relative increase in width of correlated activity between selective OHC and IHC impairment was somewhat conservative at the 500-Hz CF.

Broadened tuning associated with OHC impairment was predicted to increase the range of CF separations over which correlated activity exists. The predicted degree of degradation for the selective IHC impairment provided a useful control to suggest that threshold shift alone does not account for the predicted across-fiber degradations in envelope correlations. Rather, it is likely that the reduction in

cochlear nonlinearity that occurs for OHC damage and not for IHC damage is the basis for the predicted increase in width of the correlated activity for the across-fiber envelope correlations.

III. ACROSS-FIBER ENVELOPE CODING OF SPEECH

The purpose of this second part of the study was to measure the predicted effects of selective hair-cell damage on across-fiber envelope coding in responses to a speech stimulus. Unlike BBN, speech has externally imposed acoustic envelopes such that AN responses to speech include both stimulus- and cochlear-induced envelopes. Moreover, the acoustic envelope of speech can be considered as being comprised of several ranges of modulation frequencies (Rosen, 1992). Thus the neural representations of several modulation frequency ranges were analyzed separately.

A. Methods

1. Across-fiber envelope correlations computed for different speech modulation frequency ranges

The methodological details for computing the across-fiber envelope correlations for speech were similar to those for broadband noise except that the neural envelopes (represented by sumcors) were segregated into three ranges of modulation frequencies: 0–5 Hz envelope fluctuations, mainly representing *syllabic* information (Arai and Greenberg, 1997); 5–64 Hz, mainly representing *phonemic* information; and 64–300 Hz, mainly representing *periodicity* information (Rosen, 1992). 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. Syllabic, phonemic and periodicity sumcors were computed as the inverse Fourier transform of the envelope spectra from 0–5, 5–64, and 64–300 Hz, respectively. The overall sumcor comprising syllabic, phonemic, and periodicity information was also computed. The across-fiber envelope cross-correlation coefficient was then computed for each modulation frequency range. The syllabic, phonemic, periodicity, and overall envelope correlation coefficients will be denoted as $\rho_{\text{ENV } 0-5 \text{ Hz}}$, $\rho_{\text{ENV } 5-64 \text{ Hz}}$, $\rho_{\text{ENV } 64-300 \text{ Hz}}$, and $\rho_{\text{ENV } 0-300 \text{ Hz}}$, respectively.

2. Procedures

AN model spike trains were obtained in response to the speech utterance “A boy fell from the window” (Nilsson *et al.*, 1994), which had a duration of 1.72 s. All other procedures were similar to those for the first part of this study (Sec. II).

B. Results and discussion

The across-fiber envelope coding for responses to speech from a NH fiber at two sound levels and from an impaired fiber with either selective OHC or IHC damage is

FIG. 4. Model predictions of the effects of SNHL on across-CF coding of speech envelopes for CFs geometrically centered at 500 Hz. Stimuli: Speech (“A boy fell from the window”), duration: 1.72 s. ρ_{ENV} is plotted as a function of CF separation for overall envelope [panel (A) 0–300 Hz], syllabic envelope [panel (B) 0–5 Hz], phonemic envelope [panel (C) 5–64 Hz], and periodicity envelope [panel (D) 64–300 Hz]. Each data point represents mean ρ_{ENV} computed over twenty repetitions. Filled circles: Predictions for NH at BML (40 dB SPL). Filled diamonds: Predictions for NH at the same sound level as the predictions for the impaired cases (65 dB SPL). Open triangles: maximal (21 dB) hearing loss due to OHC damage; Open squares: 30-dB IHC damage. BML for the impaired case was 65 dB SPL. The smallest CF separation (ΔCF in octaves) at which ρ_{ENV} dropped to 0.7 is shown using arrows and the values are inset. The NH value in parenthesis represents the NH prediction for the equal-SPL condition.

illustrated in Fig. 4. Values of ρ_{ENV} were computed for 21 CF pairs with CF separation ranging from 0 to 1 octaves in 0.05 octave steps for AN fibers with CFs geometrically centered at 500 Hz. Each data point represents mean ρ_{ENV} com-

puted over twenty sets of spike trains in response to the same speech stimulus.

Evaluation of the effect of CF separation revealed that ρ_{ENV} decreased with increasing CF separation for predictions from the NH, selective OHC, and selective IHC cases for all modulation frequency ranges. However, the trend was different across the modulation frequency ranges, with the drop in correlated activity being steeper for ρ_{ENV} 64–300 Hz [panel (D)] than for ρ_{ENV} 5–64 Hz [panel (C)] and ρ_{ENV} 0–5 Hz [panel (B)]. For example, for the predictions from the NH case ρ_{ENV} 0–300 Hz decreased to ~ 0.6 [panel (A)], ρ_{ENV} 0–5 Hz dropped to 0.8 [panel (B)], ρ_{ENV} 5–64 Hz dropped to ~ 0.45 [panel (C)], and ρ_{ENV} 64–300 Hz dropped to ~ 0.4 [panel (D)] at the maximum CF separation of 1 octaves.

The width of the correlated activity for each modulation frequency range was quantified as the CF separation at which ρ_{ENV} reached 0.7. For the predictions from the NH case at BML, this separation was 0.54 octaves for ρ_{ENV} 0–300 Hz [panel (A)], greater than 1 octave for ρ_{ENV} 0–5 Hz [panel (B)], 0.56 octaves for ρ_{ENV} 5–64 Hz [panel (C)], and 0.22 octaves for ρ_{ENV} 64–300 Hz [panel (D)]. The width of the correlated activity for predictions from the NH case at higher sound levels was similar to that for NH at BML for ρ_{ENV} 0–300 Hz (0.52 vs 0.54 octaves). However, the width of correlated activity at higher SPL was slightly broader for ρ_{ENV} 0–5 Hz (both > 1 octave), ρ_{ENV} 5–64 Hz (0.75 vs 0.56 octaves), and ρ_{ENV} 64–300 Hz (0.26 vs 0.22 octaves). This increase in width is likely to be due to the higher sound level leading to broader tuning based on normal OHC function.

Predictions for the selective OHC impairment case (open triangles) demonstrated a much wider CF range of correlated activity for all modulation frequency ranges (> 1 octaves). The width of the correlated activity with selective IHC damage was predicted to be similar to that for NH for ρ_{ENV} 0–300 Hz (0.54 octaves) and ρ_{ENV} 0–5 Hz (> 1 octaves). However, the roll-off was steeper for ρ_{ENV} 5–64 Hz (0.45 octaves) and shallower for ρ_{ENV} 64–300 Hz (0.31 octaves). The lack of a consistent trend (relative to NH) in the predicted degradations for selective IHC damage across different modulation frequency ranges suggests that the predicted effects of SNHL are complex and different across each modulation frequency range. An analysis based only on the overall envelope (e.g., ρ_{ENV} 0–300 Hz) can thus be misleading and could mask such differential effects of SNHL on each modulation frequency range.

IV. SNHL EFFECTS ON ACROSS-CHANNEL ENVELOPE CODING OF SPEECH IN QUIET AND IN NOISE

The purpose of this third section of the study was to measure the predicted effects of SNHL on across-channel envelope coding in responses to speech in quiet and in noise. Here, across-channel effects consider much broader spectral comparisons than the across-fiber comparisons in Secs. II and III, which were limited to CF separations of an octave or less. The neural metrics developed in Sec. III were used to make across-channel predictions for syllabic, phonemic, and periodicity modulation frequency ranges.

A. Methods

1. Across-channel envelope correlations computed for different speech modulation frequency ranges

The neural metrics developed in Sec. III were used to evaluate across-channel envelope correlations in different speech modulation frequency ranges. Eight AN model fibers with CFs of 132, 277, 548, 965, 1605, 2590, 4102, and 6427 Hz were selected as in the study by [Crouzet and Ainsworth \(2001\)](#). Each AN fiber acted as a separate spectral channel. Across-channel envelope correlations were computed between each pair of spectral channels for speech in quiet and in noise.

2. Procedure

The speech sentence “A boy fell from the window” ([Nilsson et al., 1994](#)), which had a duration of 1.72 s, was used. In addition to the original utterance, envelope vocoded speech was created to relate the findings from the present study to previous psychophysical studies (e.g., [Baskent, 2006](#)). Noise-band or tonal envelope vocoders have been used in psychophysical studies to force listeners to use primarily temporal envelope cues by reducing spectral and temporal-fine-structure cues. Here, the speech sentence was initially band-pass filtered into 16 bands using third-order Butterworth filters. Forward and backward filtering was used to cancel phase delays. The analysis filter bands used for vocoder synthesis spanned the frequency range 80–8020 Hz and followed a logarithmic spacing within this nominal bandwidth. The cutoff frequencies and bandwidths of the analysis filters were similar to those used in previous psychophysical studies ([Gilbert and Lorenzi, 2006](#); [Lorenzi et al., 2006](#)). Hilbert envelopes were extracted in each band and lowpass filtered at 500 Hz with a sixth-order Butterworth filter. The resulting envelopes were used to amplitude modulate sine-wave carriers with frequencies at the center frequency of the analysis filters. These modulated signals were then band-pass filtered into the same frequency bands as used initially. The modulated signals were finally summed over all frequency bands. Steady speech shaped noise was added to the intact speech and vocoded speech. Speech shaped noise was generated by taking the inverse Fourier transform of the magnitude spectrum of the original speech sentence (or its vocoded version) combined with a random phase spectrum. For each physiological channel, AN model spike trains were obtained in response to the speech sentence and its vocoded version in quiet and at three different signal-to-noise ratios (5, 0, –5 dB SNR). Predictions for OHC damage were made based on the threshold elevations shown in Fig. 1 for each AN channel. For each channel, sound levels were chosen at the BML for each stimulus type in quiet. Typically, the BMLs for the impaired cases were 20 dB above those for the NH cases. Note that the predictions in noise were made at the BML selected for in-quiet conditions for the respective channels. All other procedures were similar to those for the first and second parts of this study (Secs. II and III).

B. Results and discussion

Tables I and II give the ρ_{ENV} values for speech in quiet for the NH and selective OHC impairment cases, respectively. Each cell represents ρ_{ENV} between two physiological channels. For example, 0.94, 0.77, and 0.01 are the values for ρ_{ENV} 0–5 Hz, ρ_{ENV} 5–64 Hz, and ρ_{ENV} 64–300 Hz, respectively, for speech in quiet between the two normal-hearing channels with CFs of 132 and 277 Hz (Table I). Values in bold correspond to coefficients that are above the noise floor (~ 0.01). Thus, in this example there was significant syllabic and phonemic envelope cross-correlation between the two lowest-CF channels. In contrast, the periodicity envelope correlation was at the noise floor. For these lowest-CF channels, the higher-rate (periodicity) envelopes were reduced by the cochlear filters because the maximum modulation frequency in a band-limited signal cannot exceed half the spectral bandwidth ([Lawson and Uhlenbeck, 1950](#); [Apoux and Bacon, 2008](#)). Hence, ρ_{ENV} 64–300 Hz for the two

TABLE I. Predicted cross-channel envelope correlations for different modulation frequency ranges (A: ρ_{ENV} 0–5 Hz, syllabic; B: ρ_{ENV} 5–64 Hz, phonemic; C: ρ_{ENV} 64–300 Hz, periodicity) for speech in quiet and normal hearing. Each cell represents mean ρ_{ENV} between channels computed over five repetitions. Values printed in bold correspond to coefficients that were above the noise floor. The degree of cross correlation (DCC) is inset for each modulation frequency range, and represents the fraction of cells with significant correlations.

NORMAL HEARING: SPEECH IN QUIET							
ρ_{ENV} 0–5 Hz							
A	2 (277)	3 (548)	4 (965)	5 (1605)	6 (2590)	7 (4102)	8 (6427)
1 (132)	0.94	0.60	0.41	0.01	0.19	0.01	0.01
2 (277)		0.55	0.35	0.01	0.16	0.01	0.01
3 (548)			0.85	0.32	0.15	0.18	0.01
4 (965)				0.35	0.01	0.22	0.01
5 (1605)		(DCC = 19/28 = 0.68)			0.80	0.84	0.51
6 (2590)						0.70	0.53
7 (4102)							0.76
ρ_{ENV} 5–64 Hz							
B	2 (277)	3 (548)	4 (965)	5 (1605)	6 (2590)	7 (4102)	8 (6427)
1 (132)	0.77	0.50	0.28	0.01	0.01	0.01	0.01
2 (277)		0.48	0.27	0.01	0.01	0.01	0.01
3 (548)			0.52	0.37	0.24	0.06	0.01
4 (965)				0.31	0.07	0.12	0.01
5 (1605)		(DCC = 18/28 = 0.64)			0.45	0.43	0.32
6 (2590)						0.44	0.30
7 (4102)							0.37
ρ_{ENV} 64–300 Hz							
C	2 (277)	3 (548)	4 (965)	5 (1605)	6 (2590)	7 (4102)	8 (6427)
1 (132)	0.01	0.01	0.01	0.01	0.01	0.01	0.01
2 (277)		0.19	0.26	0.01	0.01	0.01	0.01
3 (548)			0.49	0.01	0.01	0.01	0.01
4 (965)				0.20	0.01	0.01	0.01
5 (1605)		(DCC = 9/28 = 0.32)			0.29	0.07	0.01
6 (2590)						0.30	0.05
7 (4102)							0.14

TABLE II. Predicted cross-channel envelope correlations for speech in quiet with selective OHC impairment. Table layout is the same as for Table I.

SELECTIVE OHC IMPAIRMENT: SPEECH IN QUIET								
$\rho_{ENV\ 0-5\ Hz}$								
A	2 (277)	3 (548)	4 (965)	5 (1605)	6 (2590)	7 (4102)	8 (6427)	
1 (132)	0.96	0.80	0.60	0.47	0.34	0.32	0.06	
2 (277)		0.90	0.69	0.61	0.50	0.50	0.25	
3 (548)			0.91	0.83	0.67	0.62	0.46	
4 (965)				0.93	0.70	0.63	0.57	
5 (1605)		(DCC = 28/28 = 1)			0.86	0.83	0.74	
6 (2590)						0.96	0.88	
7 (4102)							0.94	
$\rho_{ENV\ 5-64\ Hz}$								
B	2 (277)	3 (548)	4 (965)	5 (1605)	6 (2590)	7 (4102)	8 (6427)	
1 (132)	0.78	0.48	0.19	0.01	0.01	0.01	0.01	
2 (277)		0.67	0.37	0.17	0.01	0.01	0.01	
3 (548)			0.76	0.61	0.47	0.37	0.25	
4 (965)				0.87	0.71	0.55	0.42	
5 (1605)		(DCC = 21/28 = 0.75)			0.82	0.77	0.64	
6 (2590)						0.81	0.63	
7 (4102)							0.77	
$\rho_{ENV\ 64-300\ Hz}$								
C	2 (277)	3 (548)	4 (965)	5 (1605)	6 (2590)	7 (4102)	8 (6427)	
1 (132)	0.01	0.01	0.01	0.01	0.01	0.01	0.01	
2 (277)		0.22	0.20	0.10	0.01	0.01	0.01	
3 (548)			0.63	0.20	0.01	0.01	0.01	
4 (965)				0.71	0.31	0.06	0.01	
5 (1605)		(DCC = 14/28 = 0.5)			0.73	0.47	0.33	
6 (2590)						0.76	0.57	
7 (4102)							0.75	

lowest-CF channels was insignificant because there were no periodicity envelopes at the output of these physiological channels.

The spread of correlated envelope activity across all physiological channels for each modulation frequency range was quantified by the degree of cross correlation (DCC). The DCC was computed as the number of significant coefficients (i.e., $\rho_{ENV} > 0.01$) divided by the total number of coefficients considered. DCC values range between 0 and 1, with a value of 0 representing no correlated activity across chan-

nels and 1 representing complete spread of correlated envelope activity across all channels. For the NH case, syllabic-envelope predictions [Table I(A)], had 19 of the overall 28 values of $\rho_{ENV\ 0-5\ Hz}$ above the noise floor, yielding a DCC of 0.68. The observed correlation pattern differed for $\rho_{ENV\ 5-64\ Hz}$ (DCC = 0.64) and especially for $\rho_{ENV\ 64-300\ Hz}$ (DCC = 0.32), with correlated activity being more limited to nearby channels.

Compared to predictions for NH, the predictions for selective OHC impairment (Table II) showed that the spread of correlated activity was broader for all modulation frequency ranges (DCC = 1 for $\rho_{ENV\ 0-5\ Hz}$, DCC = 0.75 for $\rho_{ENV\ 5-64\ Hz}$, and DCC = 0.5 for $\rho_{ENV\ 64-300\ Hz}$). Broadened tuning associated with OHC damage increases the overlap between spectral channels thereby increasing the cross-channel envelope correlations, as expected.

Figure 5 shows the DCC computed for speech in quiet and in noise for each modulation frequency range. The DCC decreased consistently with the addition of noise for all ranges of modulation frequency for the predictions in the NH case (filled symbols). The DCC dropped by 54 percentage points (0.68 in quiet to 0.14 at -5 dB SNR) for $\rho_{ENV\ 0-5\ Hz}$, by 60 percentage points (0.64 in quiet to 0.04 at -5 dB SNR) for $\rho_{ENV\ 5-64\ Hz}$, and by 28 percentage points (0.32 in quiet to 0.04 at -5 dB SNR) for $\rho_{ENV\ 64-300\ Hz}$. A similar trend was observed with vocoded speech (Fig. 6), where the DCC dropped by 64, 60, and 28 percentage points for $\rho_{ENV\ 0-5\ Hz}$, $\rho_{ENV\ 5-64\ Hz}$, and $\rho_{ENV\ 64-300\ Hz}$, respectively. Thus, for normal-hearing fibers (with sharp tuning), the addition of noise decorrelated the common envelope between physiological channels in all modulation frequency ranges.

Similar to the predictions for NH, decreased SNR reduced the DCC for OHC impairment (open symbols) in all modulation frequency ranges by de-correlating the responses across physiological channels for both intact (Fig. 5) and vocoded speech (Fig. 6). For intact speech, the syllabic DCC decreased by 100 percentage points across the SNR conditions considered here, whereas the decorrelation was less pronounced for phonemic (43 percentage points) and periodicity (14 percentage points) envelopes. In contrast to the predictions for NH, the phonemic-envelope DCC did not continue to decrease for $SNR \leq 0\ dB$ and the periodicity-envelope DCC was constant for $SNR \leq 5\ dB$. These results suggest that OHC impairment has a larger effect on cross-channel correlations (phonemic and periodicity) at degraded SNRs than for the in-quiet condition.

FIG. 5. Degree of cross correlation (DCC) is plotted as a function of SNR for speech in quiet and in noise for each modulation frequency range. (A) syllabic (0-5 Hz); (B) phonemic (5-64 Hz); (C) periodicity (64-300 Hz). Filled symbols represent predictions for NH and open symbols represent predictions for selective OHC impairment.

FIG. 6. DCC as a function of SNR for envelope vocoded speech in quiet and in noise for each modulation frequency range. Figure layout is the same as for Fig. 5.

Thus, although the addition of noise generally reduced the predicted cross-channel correlation for both NH and OHC-impaired cases, differences in the degree of decorrelation were observed across the modulation frequency ranges and across SNRs.

Differences were also observed between the modulation frequency ranges in terms of the relative effects of OHC impairment (increased cross-channel correlation) and the addition of noise (decreased cross-channel correlation). Similar to the predictions for in-quiet conditions (Tables I and II), OHC impairment produced an increase in the DCC relative to NH for both the periodicity and phonemic modulation frequency ranges for all three SNRs [Figs. 5(B)–5(C) and 6(B)–6(C)]. In contrast, the predicted DCC values for OHC impairment for syllabic modulations were above the NH values only for the in-quiet condition, but were reduced relative to NH for all three in-noise conditions. Thus, for the syllabic modulation frequency range, the decorrelating effect of noise was predicted to outweigh the increase in correlation due to OHC impairment.

V. DISCUSSION

A. Quantifying neural envelope coding in different speech modulation frequency ranges

Previous correlogram analyses have mainly focused on neural coding of broadband envelopes of complex stimuli like broadband noise (Joris, 2003; Louage *et al.*, 2004) and speech (Heinz and Swaminathan, 2009). However, results from psychophysical studies have shown that there are differential contributions of low-, mid-, and high-rate envelope fluctuations for speech perception in quiet and in noise (Silipo *et al.*, 1999; Apoux and Bacon, 2008). Hence, an extension was needed to improve quantification of neural envelope coding of speech. In this study, neural envelopes were quantified separately for syllabic, phonemic, and periodicity modulation frequency ranges to better understand the neural correlates behind these perceptual observations. This kind of segregation should help in fine-grained analyses to study the effects of hearing impairment and/or the degradations of speech by noise on the coding of envelopes in different speech modulation frequency ranges. Results from this study provide evidence that speech-envelope coding within these ranges of modulation frequencies (in terms of cross-fiber correlations) differed for both hearing-impaired and noise-degraded conditions.

B. Effects of SNHL on across-fiber temporal coding for complex sounds

Predictions for OHC impairment showed an increase in the across-fiber envelope correlation for broadband noise and speech. However, the increase in across-CF envelope correlation following OHC impairment was more severe for speech (Fig. 4) than it was for noise (Fig. 3). Similar findings have been observed for across-fiber temporal-fine-structure coding (Heinz *et al.*, 2010; Swaminathan, 2010). These results suggest that data for simple stimuli may underestimate the degree to which SNHL affects temporal coding for complex stimuli such as speech.

Results from physiological data collected from chinchillas showed that across-fiber coding of fine structure was degraded with noise-induced hearing loss in terms of both an increase in cross-CF correlation (consistent with envelopes) and a decrease in characteristic delay (CD) between CFs (Heinz *et al.*, 2010). Similar characterizations could be made on the CD measured from sumcors representing the envelopes. However, due to the broad nature of the sumcor peaks, the delays were not well defined and hence were not quantified in this study. Across-CF fine structure predictions were also made using the AN model for comparison with envelope predictions for most of the conditions considered in the present study (Swaminathan, 2010). For predictions in the NH case, the width of the correlated activity for temporal fine structure was greater than that for envelope when both were computed from the same set of spike-train responses to broadband noise or speech. This finding suggests that there may be more independent channels of neural information conveyed for envelope than for fine structure, i.e., envelope information appears to be less redundant across overlapping physiological channels.

Findings from psychophysical studies suggest that listeners with SNHL have a reduced ability to use temporal fine-structure cues and that this deficit is correlated with their reduced understanding of speech in complex backgrounds (Lorenzi *et al.*, 2006; Hopkins and Moore, 2007; 2009). Recent neurophysiological evidence showed that while the strength of within-fiber phase locking was not degraded with noise-induced hearing loss, across-fiber coding of fine-structure (i.e., spatiotemporal coding) was degraded (Heinz *et al.*, 2010). Spatiotemporal coding has been hypothesized to be relevant for speech, pitch, and intensity perception (Loeb *et al.*, 1983; Shamma, 1985; Heinz *et al.*, 2001b; Carney *et al.*, 2002). Complementing the neurophysiological

findings for temporal fine structure, the present study suggests that degradations also occur in across-fiber envelope coding following SNHL. Envelope cues have been shown to be very important for speech perception (e.g., [Shannon et al., 1995](#)), and thus degradations in across-fiber envelope coding may also have implications for poor speech perception with impairment.

C. Across-channel effects of SNHL differ across modulation frequency ranges for speech in noise

The predicted effects of OHC impairment on across-channel correlations were similar for the phonemic and periodicity modulation frequency ranges (>5 Hz), which have been shown to play an important role for speech identification ([Drullman et al., 1994a,b](#); [Apoux and Bacon, 2008](#)). In [Drullman et al. \(1994a,b\)](#), an attempt was made to assess the amount by which higher-rate temporal modulations can be eliminated through envelope low-pass filtering without affecting performance in a phoneme identification task. Results showed that NH listeners can only partially understand speech in quiet when the amplitude fluctuations are limited to less than ~ 4 Hz, whereas performance was hardly affected by temporal envelope smoothing characterized by cutoff frequencies higher than 16 Hz. These results were interpreted as suggesting that envelope fluctuations between ~ 4 and 16 Hz were most important and that fluctuations above ~ 16 Hz contributed minimally to speech identification. However, the results from this study have to be interpreted with caution as [Drullman et al. \(1994a,b\)](#) used fine structure from the original speech to create their envelope vocoded signals. [Ghitza \(2001\)](#) showed that if the envelope of a critical-band signal is temporally smoothed while the instantaneous phase information (fine structure) remains unaltered, the resulting speech signal evokes cochlear envelope signals that are not necessarily smoothed as significantly as the acoustic signals. A consequence of this result is that the low-pass filtered (or smoothed) acoustic envelopes used in the [Drullman et al. \(1994a,b\)](#) studies may not have completely abolished the higher-rate envelope modulations post cochlear filtering. Using a multichannel vocoder, [Apoux and Bacon \(2008\)](#) performed a study to assess the role of high-rate envelopes for consonant perception. In their study, consonant perception was measured for envelope cutoff frequencies ranging from 4 to 400 Hz. By using tone carriers instead of the original fine structure to create their envelope vocoded signals, potential recovered envelopes from the phase information were essentially abolished. The results from their study show that consonant perception was poorer when only envelope fluctuations less than 4 Hz were retained (consistent with [Drullman et al., 1994a,b](#)), but improved substantially as the envelope cutoff frequency increased up to about 160 Hz. The results from their study suggest that higher-rate envelope fluctuations (at least up to 160 Hz) can play a critical role in consonant identification. Thus, it is important to consider the predictions from the present study for both phonemic (5–64 Hz) and periodicity (64–300 Hz) modulations, which differ in their predicted variation across SNR (Figs. 5 and 6).

The findings from the present study suggest that OHC damage produces a broader range of spectral channels with

correlated phonemic and periodicity envelope responses for both speech in quiet and in noise, as represented by higher values of the DCC [Figs. 5(B)–5(C)]. This increase in cross-channel correlation is intuitive based on broadened tuning that results from OHC damage. For example, the rapid fluctuations ($>\sim 5$ Hz) that occur for phonemes (e.g., those associated with formant transitions) are relatively narrowband and thus the across-channel correlations associated with these features would be expected to be increased by broadened cochlear filters. In terms of overall speech intelligibility, an increase in DCC due to OHC impairment for higher-rate modulations would be expected to result in reduced performance since this effect represents a reduction in the number of spectral channels providing independent neural envelope information for speech perception. Thus, the present results may (at least partially) account for the reduced ability of HI (and cochlear-implant) listeners to take advantage of additional spectral channels in vocoded speech compared to NH listeners ([Friesen et al., 2001](#); [Baskent, 2006](#)). In fact, the greater effect of OHC impairment on the DCC for SNRs ≤ 0 dB predicted in the present data is consistent with this perceptual deficit being most significant for SNRs ≤ 0 dB ([Baskent, 2006](#)).

The effects of OHC impairment on syllabic envelope modulations in noise were different than those predicted for phonemic and periodicity envelopes. In contrast to the increased DCC for higher-rate modulations, OHC impairment was predicted to decrease the DCC for syllabic modulations in noise. This difference may result from different sources of cross-channel correlation for slow and fast modulations. Unlike the relatively narrowband nature of faster fluctuations (e.g., for phonemes), syllabic envelopes ($<\sim 5$ Hz) reflect the slow average syllable rate ([Arai and Greenberg, 1997](#); [Greenberg, 1999](#)) and are generally characterized by broader spectral events (e.g., syllable onsets and offsets). Thus, syllabic cross-channel correlation arises primarily from broadband stimulus features and thus is not expected to be as dependent on cochlear filter bandwidth. In the case of OHC damage, this appears to allow the decorrelating effect from noise to reduce the syllabic speech modulation DCC values despite broader cochlear filters.

As discussed above, syllabic modulations have been shown to contribute much less to speech intelligibility than phonemic and periodicity envelopes ([Drullman et al., 1994a,b](#); [Apoux and Bacon, 2008](#)). Rather, these very slow modulations may be more relevant for source segregation or grouping because they represent syllable onsets and offsets, which provide common features across broad spectral regions ([Crouzet and Ainsworth, 2001](#)). Thus, in terms of segregation or grouping, the predicted decrease in the syllabic DCC in noise following OHC impairment would be expected to lead to reduced performance and may partially account for the reduced ability of HI listeners to segregate and understand speech in complex acoustic backgrounds.

D. Potential implications for improving speech intelligibility metrics and hearing aids

The present approach addresses physiological correlates of the perceptual role of correlated envelope fluctuations

across spectral channels for speech identification in normal-hearing (Apoux and Bacon, 2008) and hearing-impaired listeners (Healy and Bacon, 2002; Healy *et al.*, 2005; Baskent, 2006). This approach provides a quantitative physiological framework that may be beneficial for expanding current methods for predicting speech intelligibility (e.g., the speech transmission index, STI, Steeneken and Houtgast, 1980; Houtgast and Steeneken, 1985) by considering physiological envelope responses (Elhilali *et al.*, 2003), including the effects of cross-channel correlation (Steeneken and Houtgast, 1999; Vickers *et al.*, 2009).

The STI measures the overall reduction in modulations present in the intensity envelope of speech (or any complex modulated signal) resulting from acoustic degradations. This metric was first developed using simple test signals like sinusoidal intensity-modulated noise (Houtgast and Steeneken, 1985) and has been extended to more complex modulated signals like speech (Payton and Braida, 1999). In its current calculation, only the contributions of modulation frequencies up to ~ 12.5 Hz are considered. Within each spectral channel, the contributions from individual modulation frequencies are averaged together. The present results suggest that it may be beneficial to consider the contributions of higher modulation frequencies (as also recommended by Apoux and Bacon, 2008) and to treat the syllabic, phonemic, and periodicity modulation frequency ranges separately. Furthermore, inclusion of the mutual dependence of acoustic envelopes between adjacent octave bands has been shown to improve STI predictions (Steeneken and Houtgast, 1999), and the present approach provides a physiological framework for computing these cross-channel envelope correlations based on neural spike-train responses.

Neural based STI metrics (Elhilali *et al.*, 2003; Zilany and Bruce, 2007a) have been proposed for designing/testing hearing-aid signal processing strategies (Dinath and Bruce, 2008). Such an approach may benefit from attempting to restore channel independence (Baskent, 2006) and normal spatiotemporal patterns (Shamma, 1985; Carney *et al.*, 2002; Heinz, 2007; Heinz *et al.*, 2010). The cross-channel neural envelope correlation metrics developed in the present study may be beneficial for expanding current approaches to evaluating the efficacy of signal processing algorithms aimed at improving hearing-aid design by restoring channel independence.

One limitation of STI-based approaches is that only *overall* speech intelligibility is predicted, i.e., these approaches are not currently able to predict specific confusion patterns or phonemic feature reception (however, see Giguère *et al.*, 1997). Several recent studies have suggested a dependence of phoneme feature reception on specific acoustic modulation frequency ranges, including cross-channel correlation effects (e.g., van der Horst *et al.*, 1999; Christiansen *et al.*, 2007). For example, Christiansen *et al.* (2007), concluded that for decoding place of articulation, significant cross-channel integration was required particularly for acoustic envelope fluctuations greater than 6 Hz. Voicing was suggested to require modulation rates between 3 and 6 Hz, whereas manner appeared to depend on modulation rates greater than 12 Hz. The neural metrics developed in the present study may help

to expand current approaches for predicting speech intelligibility by allowing the quantification of the relation between neural envelope fluctuations in specific modulation frequency ranges to different speech features that are thought to be critical for phoneme identification.

A significant issue for predicting the effects of SNHL on speech intelligibility and for fitting hearing aids to individual subjects is the large degree of across-subject variability in speech reception even for patients with similar audiograms. Much insight into perceptual effects of SNHL has come from simple and complex models of basilar-membrane responses and the effects of OHC dysfunction on those responses. However, recent studies suggest that IHC dysfunction can also significantly affect perceptually relevant response properties in the AN related to intensity and speech coding (Bruce *et al.*, 2003; Heinz and Young, 2004). Physiological evidence suggests that many common forms of SNHL are likely to involve mixed OHC and IHC damage, including noise-induced and age-related hearing loss (Liberman and Dodds, 1984; Schmiedt *et al.*, 2002). Evidence for mixed OHC and IHC damage has also been observed in psychophysical studies, even for mild-to-moderate degrees of SNHL (Plack *et al.*, 2004). Predictions from the present study suggest that there are differences in cross-fiber envelope coding across different speech modulation frequency ranges with selective hair-cell damage (Fig. 4). For example, predictions for selective IHC damage were comparable to predictions in the NH case for syllabic envelopes, whereas the decay in cross-CF correlation was steeper than NH for phonemic modulations and broader than NH for periodicity modulations. Thus, the quantitative physiological framework in the present study could be extended to predict the effects of varying degrees of mixed OHC and IHC damage on envelope coding of speech (rather than the selective OHC or IHC damage considered so far). Ultimately, applying this approach to individual patients will require improvements in psychophysical and physiological estimates of the degree of OHC and IHC damage as a function of CF in individual patients (Lopez-Poveda *et al.*, 2009; Heinz, 2010).

VI. CONCLUSIONS

1. As expected, OHC damage increased the across-fiber envelope correlation for both broadband noise and speech; however, this degradation was more severe for speech, suggesting that predictions from simple stimuli may underestimate the effects of SNHL for complex stimuli.
2. OHC damage was predicted to increase the number of spectral channels over which correlated activity occurred for all speech modulation frequency ranges in quiet, and for phonemic and periodicity modulations for speech in noise. This reduction in the number of spectral channels providing independent neural information may contribute to reduced speech intelligibility with SNHL, and is consistent with perceptual data from HI listeners and cochlear-implant users.
3. In contrast to phonemic and periodicity modulations, across-channel correlation for syllabic modulations of speech in noise was predicted to decrease with OHC

impairment. This effect may contribute to the reduced ability of HI listeners to segregate speech in complex backgrounds.

4. The predicted effects of IHC damage on across-fiber envelope correlation differed across speech modulation frequency ranges. This result suggests that the effects of SNHL on speech coding may differ across patients with varying degrees of OHC and IHC damage. The present study provides a quantitative physiological framework for evaluating within- and across-fiber/channel neural coding, which may be useful for expanding current approaches to predicting speech intelligibility for individual patients.

ACKNOWLEDGMENTS

The authors would like to thank Kenneth Grant, Brian Moore, and two anonymous reviewers for their invaluable comments on previous versions of this manuscript. We are also grateful to Jonathan Boley, Kenneth Henry, Skyler Jennings, and Elizabeth Strickland for their thorough comments, which improved an earlier version of the manuscript. Thanks are also given to Ananthakrishna Chintanpalli for providing tuning-curve Q_{10} values for NH and HI versions of the AN model. This research was funded by grants from the Purdue Research Foundation and NIH (NIDCD) Grant No. R01-DC009838. Michael Heinz was also supported through a joint appointment held between the Department of Speech, Language, and Hearing Sciences, and the Weldon School of Biomedical Engineering at Purdue University.

- Apoux, F., and Bacon, S. P. (2004). "Relative importance of temporal information in various frequency regions for consonant identification in quiet and in noise," *J. Acoust. Soc. Am.* **116**, 1671–1680.
- Apoux, F., and Bacon, S. P. (2008). "Differential contribution of envelope fluctuations across frequency to consonant identification in quiet," *J. Acoust. Soc. Am.* **123**, 2792–2800.
- Arai, T., and Greenberg, S. (1997). "The temporal properties of spoken Japanese are similar to those of English," in *Proceedings of EUROSPEECH*, Rhodes, Greece, pp. 1011–1014.
- Baskent, D. (2006). "Speech recognition in normal hearing and sensorineural hearing loss as a function of the number of spectral channels," *J. Acoust. Soc. Am.* **120**, 2908–2925.
- Bruce, I. C., Sachs, M. B., and Young, E. D. (2003). "An auditory-periphery model of the effects of acoustic trauma on auditory nerve responses," *J. Acoust. Soc. Am.* **113**, 369–388.
- Carney, L. H. (1993). "A model for the responses of low-frequency auditory-nerve fibers in cat," *J. Acoust. Soc. Am.* **93**, 401–417.
- Carney, L. H., Heinz, M. G., Evilsizer, M. E., Gilkey, R. H., and Colburn, H. S. (2002). "Auditory phase opponency: A temporal model for masked detection at low frequencies," *Acta. Acust.* **88**, 334–347.
- Christiansen, T. U., Dau, T., and Greenberg, S. (2007). "Spectro-temporal processing of speech—An information-theoretic framework," in *Hearing—From Sensory Processing to Perception*, edited by B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, S. Uppenkamp, and J. Verhey (Springer, Berlin), pp. 517–523.
- Crouzet, O., and Ainsworth, W. A. (2001). "On the various influences of envelope information on the perception of speech in adverse conditions: An analysis of between-channel envelope correlation," in *Workshop on Consistent and Reliable Cues for Sound Analysis*, Aalborg, Denmark, September.
- Dinath, F., and Bruce, I. C. (2008). "Hearing aid gain prescriptions balance restoration of auditory nerve mean-rate and spike-timing representations of speech," in *30th International IEEE Engineering in Medicine and Biology Conference*, IEEE, Piscataway, New Jersey, pp. 1793–1796.
- Drullman, R., Festen, J. M., and Plomp, R. (1994a). "Effect of temporal envelope smearing on speech reception," *J. Acoust. Soc. Am.* **95**, 1053–1064.
- Drullman, R., Festen, J. M., and Plomp, R. (1994b). "Effect of reducing slow temporal modulations on speech reception," *J. Acoust. Soc. Am.* **95**, 2670–2680.
- Elhilali, M., Chi, T., and Shamma, S. A. (2003). "A spectro-temporal modulation index (STMI) assessment of speech intelligibility," *Speech Commun.* **41**, 331–348.
- Flanagan, J. L. (1980). "Parametric coding of speech spectra," *J. Acoust. Soc. Am.* **68**, 412–430.
- Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (2001). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," *J. Acoust. Soc. Am.* **110**, 1150–1163.
- Giguère, C. A. J., Bosman, A. J., and Smoorenburg, G. F. (1997). "Automatic speech recognition experiments with a model of normal and impaired peripheral hearing," *Acta. Acust.* **83**, 1065–1076.
- Gilbert, G., and Lorenzi, C. (2006). "The ability of listeners to use recovered envelope cues from speech fine structure," *J. Acoust. Soc. Am.* **119**, 2438–2444.
- Glasberg, B. R., and Moore, B. C. J. (1986). "Auditory filter shapes in subjects with unilateral and bilateral cochlear impairments," *J. Acoust. Soc. Am.* **79**, 1020–1033.
- Greenberg, S. (1999). "Speaking in shorthand - A syllable-centric perspective for understanding pronunciation variation," *Speech Commun.* **29**, 159–176.
- Healy, E. W., and Bacon, S. P. (2002). "Across-frequency comparison of temporal speech information by listeners with normal and impaired hearing," *J. Speech Lang. Hear. Res.* **45**, 1262–1275.
- Healy, E. W., Kannabiran, A., and Bacon, S. P. (2005). "An across-frequency processing deficit in listeners with hearing impairment is supported by acoustic correlation," *J. Speech Lang. Hear. Res.* **48**, 1236–1242.
- Heil, P., Neubauer, H., Irvine, D. R., and Brown, M. (2007). "Spontaneous activity of auditory-nerve fibers: Insights into stochastic processes at ribbon synapses," *J. Neurosci.* **27**, 8457–8474.
- Heinz, M. G. (2007). "Spatiotemporal encoding of vowels in noise studied with the responses of individual auditory nerve fibers," in *Hearing—From Sensory Processing to Perception*, edited by B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp, and J. Verhey (Springer-Verlag, Berlin), pp. 107–115.
- Heinz, M. G. (2010). "Computational modeling of sensorineural hearing loss," in *Computational Models of the Auditory System*, edited by R. Meddis, E. A. Lopez-Poveda, A. N. Popper, and R. R. Fay (Springer, New York), pp. 177–202.
- Heinz, M. G., Colburn, H. S., and Carney, L. H. (2001a). "Auditory nerve model for predicting performance limits of normal and impaired listeners," *ARL* **2**, 91–96.
- Heinz, M. G., Colburn, H. S., and Carney, L. H. (2001b). "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.
- Heinz, M. G., and Young, E. D. (2004). "Response growth with sound level in auditory-nerve fibers after noise-induced hearing loss," *J. Neurophysiol.* **91**, 784–795.
- Heinz, M. G., and Swaminathan, J. (2009). "Quantifying envelope and fine-structure coding in auditory nerve responses to chimaeric speech," *J. Assoc. Res. Otolaryngol.* **10**, 407–423.
- Heinz, M. G., Swaminathan, J., Boley, J. D., and Kale, S. (2010). "Across-fiber coding of temporal fine-structure: Effects of noise-induced hearing loss on auditory-nerve responses," in *The Neurophysiological Bases of Auditory Perception*, edited by E. A. Lopez-Poveda, A. R. Palmer, and R. Meddis (Springer, New York), pp. 621–630.
- Hopkins, K., and Moore, B. C. J. (2007). "Moderate cochlear hearing loss leads to a reduced ability to use temporal fine structure information," *J. Acoust. Soc. Am.* **122**, 1055–1068.
- Hopkins, K., and Moore, B. C. J. (2009). "The contribution of temporal fine structure to the intelligibility of speech in steady and modulated noise," *J. Acoust. Soc. Am.* **125**, 442–446.
- Houtgast, T., and Steeneken, H. J. (1985). "A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria," *J. Acoust. Soc. Am.* **77**, 1069–1077.
- Joris, P. X. (2003). "Interaural time sensitivity dominated by cochlea-induced envelope patterns," *J. Neurosci.* **23**, 6345–6350.

- Joris, P. X., and Yin, T. C. (1992). "Responses to amplitude-modulated tones in the auditory nerve of the cat," *J. Acoust. Soc. Am.* **91**, 215–232.
- Joris, P. X., Van de Sande, B., Louage, D. H., and van der Heijden, M. (2006). "Binaural and cochlear disparities," *Proc. Natl. Acad. Sci. U.S.A.* **103**, 12917–12922.
- Lawson, J. L., and Uhlenbeck, G. E. (1950). "Threshold signals," in *Radiation Laboratory Series*, Vol. 24 (McGraw-Hill, New York), pp. 1–388.
- Lieberman, M. C., and Dodds, L. W. (1984). "Single-neuron labeling and chronic cochlear pathology. III. Stereocilia damage and alterations of threshold tuning curves," *Hear. Res.* **16**, 55–74.
- Lieberman, M. C., and Kiang, N. Y. S. (1984). "Single-neuron labeling and chronic cochlear pathology. IV. Stereocilia damage and alterations in rate- and phase-level functions," *Hear. Res.* **16**, 75–90.
- Loeb, G. E., White, M. W., and Merzenich, M. M. (1983). "Spatial cross-correlation—a proposed mechanism for acoustic pitch perception," *Biol. Cybern.* **47**, 149–163.
- Lopez-Poveda, E. A., Johannesen, P. T., and Merchán, M. A. (2009). "Estimation of the degree of inner and outer hair cell dysfunction from distortion product otoacoustic emission input/output functions," *Audiol. Med.* **7**, 22–28.
- Lorenzi, C., Gilbert, G., Carn, H., Garnier, S., and Moore, B. C. J. (2006). "Speech perception problems of the hearing impaired reflect inability to use temporal fine structure," *Proc. Natl. Acad. Sci. U.S.A.* **103**, 18866–18869.
- Louage, D. H., Van Der Heijden, M., and Joris, P. X. (2004). "Temporal properties of responses to broadband noise in the auditory nerve," *J. Neurophysiol.* **91**, 2051–2065.
- Miller, R. L., Schilling, J. R., Franck, K. R., and Young, E. D. (1997). "Effects of acoustic trauma on the representation of the vowel /e/ in cat auditory nerve fibers," *J. Acoust. Soc. Am.* **101**, 3602–3616.
- Nilsson, M., Soli, S. D., and Sullivan, J. A. (1994). "Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise," *J. Acoust. Soc. Am.* **95**, 1085–1099.
- Patuzzi, R. B., Yates, G. K., and Johnstone, B. M. (1989). "Outer hair cell receptor current and sensorineural hearing loss," *Hear. Res.* **42**, 47–72.
- Payton, K. L., and Braida, L. D. (1999). "A method to determine the speech transmission index from speech waveforms," *J. Acoust. Soc. Am.* **106**, 3637–3648.
- Rice, S. O. (1954). "Mathematical analysis of random noise," in *Selected Papers on Noise and Stochastic Processes*, edited by N. Wax (Dover, New York), pp. 133–162.
- Rosen, S. (1992). "Temporal information in speech: acoustic, auditory, and linguistic aspects," *Philos. Trans. R. Soc. London.* **336**, 367–373.
- Ruggero, M. A., and Rich, N. C. (1991). "Furosemide alters organ of Corti mechanics: Evidence for feedback of outer hair cells upon the basilar membrane," *J. Neurosci.* **11**, 1057–1067.
- Ruggero, M. A., and Temchin, A. N. (2005). "Unexceptional sharpness of frequency tuning in the human cochlea," *Proc. Natl. Acad. Sci. U.S.A.* **102**, 18614–18619.
- Schmiedt, R. A., Lang, H., Okamura, H. O., and Schulte, B. A. (2002). "Effects of furosemide applied chronically to the round window: a model of metabolic presbycusis," *J. Neurosci.* **22**, 9643–9650.
- Sewell, W. F. (1984). "Furosemide selectively reduces one component in rate-level functions from auditory-nerve fibers," *Hear. Res.* **15**, 69–72.
- Shamma, S. A. (1985). "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.
- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). "Speech recognition with primarily temporal cues," *Science (N.Y.)* **270**, 303–304.
- Shera, C. A., Guinan, J. J., Jr., and Oxenham, A. J. (2002). "Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements," *Proc. Natl. Acad. Sci. U.S.A.* **99**, 3318–3323.
- Silipo, R., Greenberg, S., and Arai, T. (1999). "Temporal constraints on speech intelligibility as deduced from exceedingly sparse spectral representation," in *Proceedings of EUROSPEECH*, Budapest, Hungary, pp. 2687–2690.
- Steeneken, H. J., and Houtgast, T. (1980). "A physical method for measuring speech-transmission quality," *J. Acoust. Soc. Am.* **67**, 318–326.
- Steeneken, H. G. M. and Houtgast, T. (1999). "Mutual dependence of the octave-band weights in predicting speech intelligibility," *Speech Commun.* **28**, 109–123.
- Swaminathan, J. (2010). "The role of envelope and temporal fine structure in the perception of noise degraded speech," Ph.D. dissertation, Purdue University, West Lafayette, IN.
- Tan, Q., and Carney, L. H. (2003). "A phenomenological model for the responses of auditory-nerve fibers. II. Nonlinear tuning with a frequency glide," *J. Acoust. Soc. Am.* **114**, 2007–2020.
- Turner, C. W., Chi, S. L., and Flock, S. (1999). "Limiting spectral resolution in speech for listeners with sensorineural hearing loss," *J. Speech Lang. Hear. Res.* **42**, 773–784.
- van der Horst, R., Rens Leeuw, A., and Dreschler, W. A. (1999). "Importance of temporal-envelope cues in consonant recognition," *J. Acoust. Soc. Am.* **105**, 1801–1809.
- Vickers, D. A., Robinson, J., Füllgrabe, C., Baer, T., and Moore, B. C. J. (2009). "Relative importance of different spectral bands to consonant identification: Relevance for frequency transposition in hearing aids," *Int. J. Audiol.* **48**, 334–345.
- Young, E. D., and Barta, P. E. (1986). "Rate responses of auditory nerve fibers to tones in noise near masked threshold," *J. Acoust. Soc. Am.* **79**, 426–442.
- Zhang, X., Heinz, M. G., Bruce, I. C., and Carney, L. H. (2001). "A phenomenological model for the responses of auditory-nerve fibers: I. Nonlinear tuning with compression and suppression," *J. Acoust. Soc. Am.* **109**, 648–670.
- Zilany, M. S. A., and Bruce, I. C. (2006). "Modeling auditory-nerve responses for high sound pressure levels in the normal and impaired auditory periphery," *J. Acoust. Soc. Am.* **120**, 1446–1466.
- Zilany, M. S. A., and Bruce, I. C. (2007a). "Predictions of speech intelligibility with a model of the normal and impaired auditory-periphery," in *3rd International IEEE EMBS Conference on Neural Engineering*, IEEE, Piscataway, New Jersey, pp. 481–485.
- Zilany, M. S. A., and Bruce, I. C. (2007b). "Representation of the vowel /e/ in normal and impaired auditory nerve fibers: model predictions of responses in cats," *J. Acoust. Soc. Am.* **122**, 402–417.