Analysis of Spatiotemporal Pattern Correction Using a Computational Model of the Auditory Periphery

Timothy J. Zeyl and Ian C. Bruce

Objectives: The purpose of this study was to determine the cause of poor experimental performance of a spatiotemporal pattern correction (SPC) scheme that has been proposed as a hearing aid algorithm and to determine contexts in which it may provide benefit. The SPC scheme is intended to compensate for altered phase response and group delay differences in the auditory nerve spiking patterns in impaired ears. Based on theoretical models of loudness and the hypothesized importance of temporal fine structure for intelligibility, the compensations of the SPC scheme are expected to provide benefit; however, preliminary experiments revealed that listeners preferred unprocessed or minimally processed speech as opposed to complete SPC processed speech.

Design: An improved version of the SPC scheme was evaluated with a computational auditory model in response to a synthesized vowel at multiple SPLs. The impaired model auditory nerve response to SPC-aided stimuli was compared to the unaided stimuli for spectrotemporal response similarity to the healthy auditory model. This comparison included analysis of synchronized rate across auditory nerve characteristic frequencies and a measure of relative phase response of auditory nerve fibers to complex stimuli derived from cross-correlations.

Results: Analysis indicates that SPC can improve a metric of relative phase response at low SPLs, but may do so at the cost of decreased spectrotemporal response similarity to the healthy auditory model and degraded synchrony to vowel formants. In-depth analysis identifies several technical and conceptual problems associated with SPC that need to be addressed. These include the following: (1) a nonflat frequency response through the analysis–synthesis filterbank that results from time-varying changes in the relative temporal alignment of filterbank channels, (2) group delay corrections that are based on incorrect frequencies because of spread of synchrony in auditory nerve responses, and (3) frequency modulations in the processed signal created by the insertion of delays.

Conclusions: Despite these issues, SPC provided benefit to an error metric derived from auditory nerve response cross-correlations at low SPLs, which may mean phase adjustment is achieved at the expense of other metrics, but could be beneficial for low-level speech.

Key words: Auditory nerve response, Group delay adjustment, Phase response, Hearing aid algorithm, Computational model.

(Ear & Hearing 2014;35;246–255)

INTRODUCTION

The mammalian cochlea performs a frequency decomposition of incident sound waves, in which high frequencies resonate at the base of the basilar membrane and low frequencies resonate at the apex. Auditory nerve (AN) fibers inherit the tuning properties of the basilar membrane according to the place where they innervate inner hair cells along the organ of Corti (Ruggero et al. 2000), which allows for approximation of AN fiber response as the output of a band-pass filter with a specific center frequency, bandwidth, phase response, and group delay (Glasberg & Moore 1990; Carney 1994). The compressive nonlinearity of the basilar membrane, which has been associated with an active process in the inner ear (Patuzzi et al. 1989; Ruggero & Rich 1991), results in systematic changes in the bandwidth of peripheral auditory filters such that frequency selectivity is diminished as sound level increases (Rhode 1971; Robles et al. 1986). As a consequence of broadening bandwidths with increasing level, the phase response of the auditory filters is flattened and filter group delay is shortened. Similarly, impairment to the cochlear nonlinearity, which results from sensorineural hearing loss, causes auditory filters to have broad bandwidths at low stimulus presentation levels when compared to the healthy case (Florentine et al. 1980; Moore 1985). This results in a reduced dynamic range of the filter's bandwidth, phase response, and group delays, which determine the timing of action potentials in the AN. Modern hearing aids account for reduced audibility with level-dependent gain; however, they do not address the flattened phase and shortened group delay that are associated with sensorineural hearing loss. In this article, we evaluate a phase adjustment scheme designed to address these issues with a computational auditory model.

Correction of flattened phase and shortened group delay may be important for hearing-impaired listeners because the phase and group delays determine the relative spike timing across AN fibers tuned to different frequencies. Precise spike timing may have several perceptual correlates related to intelligibility and loudness. Pichora-Fuller et al. (2007) found that intelligibility decreased significantly when speech was temporally jittered for normal hearing listeners, and Lorenzi et al. (2006) showed that listeners with sensorineural hearing loss had difficulty understanding speech that was processed to isolate the temporal fine structure in the presence of background noise. Temporal fine structure speech cues are encoded in the synchronization of AN spikes to a particular phase of the stimulus, known as phase locking (Young 2008), which suggests that precise spike timing may be important for intelligibility in background noise.

A modeling study by Carney (1994) has also suggested that loudness may be encoded in the relative phase of the response patterns of AN fibers and decoded in coincidence detecting neurons in the cochlear nucleus. When outer hair cell damage is incorporated, the model predicts recruitment-like behavior in the discharge rates of neurons in the cochlear nucleus that resembles perceptual loudness balance curves for individuals with unilateral hearing loss (Hood 1977; Moore et al. 1985; Moore 2004). Cai et al. (2009) have shown that chopper-type neurons in the ventral cochlear nucleus of cats receive AN input across a range of AN fibers and display recruitment-like behavior in their discharge rates in ears suffering from noise-induced hearing loss.

To address the issue of degraded phase response with sensorineural hearing loss, Shi et al. (2006) proposed a spatiotemporal

Department of Electrical and Computer Engineering, McMaster University, Hamilton, Ontario, Canada.

pattern correction (SPC) scheme that attempts to correct for the shortened group delays in impaired auditory filters by inserting the missing delay into the signal, as prescribed by a computational ear model. A filter's group delay is defined by Oppenheim et al. (1999) as the negative derivative of the phase response as a function of frequency. It describes the delay the envelope of a narrowband signal will experience when passing through the filter. Thus, group delay can be used for determining the relative timing of spectral energy across frequency and manipulating that timing accordingly.

The SPC scheme first measures the group delay through a bank of healthy and impaired auditory filters across a range of characteristic frequencies (CFs) and finds the difference between the delays. It then inserts the delay difference as a temporal shift in the corresponding channels of an analysis–synthesis gammatone filterbank. Shi et al. (2006) conducted a preliminary human listening test for each subject with several SPC strengths because the listeners' levels of impairment to the cochlear nonlinearity were not known. They found that the difference in hearing-impaired listeners' preference level for minimally processed versus unprocessed sounds was not significant for the group of subjects, but the two listeners who had the greatest degree of hearing loss preferred some amount of SPC processing. Subsequently, Calandruccio et al. (2007) evaluated a version of the SPC scheme that used rectangular filters in place of the gammatone filters when implementing the analysis–synthesis filterbank. Their study found that listeners preferred unprocessed sounds to SPC processed sounds. As well, they found no systematic improvement in syllable and sentence intelligibility with the SPC-processed speech.

The delays introduced by the SPC scheme were shown by Calandruccio et al. (2007) to be near the range of 0.4 to 0.6 milliseconds for a vowel; these small values, along with the listening results described earlier, might call into question the importance of correcting for group delay because the literature cites larger detection thresholds for across-frequency delay differences. Specifically, Kates and Arehart (2005) found that it is unlikely that listeners with hearing impairment would detect a difference in group delay of around 4 milliseconds between the low and high frequencies. Stone and Moore (2003) also found that, although there was a statistically significant effect on vowel-consonant-vowel recognition rates for increasing across-frequency delay, the effects were quite small; percent correct scores only degraded by around 4.5% on average when a delay difference of 24 milliseconds was inserted between the low and high frequencies. However, the results of Kates and Arehart and Stone and Moore were for static delays, whereas the delays are inserted dynamically by the SPC scheme, which could have a different effect. Additionally, the prescribed delays in the present study are found to be on the order of milliseconds as opposed to tenths of milliseconds reported by Calandruccio et al.

Work by Uppenkamp et al. (2001) and Wojtczak et al. (2012) support the idea of a higher-level mechanism that could compensate for the different cochlear delays that are imposed by the basilar membrane as a function of frequency. However, these studies were performed on normal hearing listeners, and it remains to be shown whether a higher-level delay compensation mechanism would be dependent on the natural delay profiles generated by a healthy basilar membrane or if it would be able to adapt to changes that result from basilar membrane impairment. Wojtczak et al. (2012) found that asynchronous tones were more easily detected when a low-frequency tone was leading a high-frequency tone and discussed the possibility that this could occur because of an adaptation to the natural statistics of sound, in which low frequencies typically lag high frequencies in a sound complex because of longer impulse responses of low-frequency resonators. A higher-level delay compensation mechanism could also be adapted to the delays imposed by the healthy basilar membrane, such that delay mismatch sensitivity may be heightened when the basilar membrane delays are distorted, as in the case of sensorineural hearing loss. Thus, accounting for the distortion in delays imposed by sensorineural hearing loss, as intended by the SPC scheme, may be important for high-level compensation of relative delays across frequency.

In this article, we evaluate an implementation of the SPC scheme of Shi et al. (2006) with a computational model of the auditory periphery to determine possible causes of the poor listening results found by Shi et al. and Calandruccio et al. (2007). We find several implementation and conceptual issues associated with the SPC scheme that cause distortion in the neural firing pattern; these need to be addressed before phase adjustment can be incorporated into hearing aids. First, there is a nonflat frequency response through the gammatone filterbank that changes in time because of time-varying insertion delays that are prescribed by the SPC scheme. Second, applying time-varying delays to individual frequency channels in the filterbank causes frequency modulation in each channel, which results in modulations in the filtered signal. Third, the group delays inserted could be based on incorrect frequencies because of a mismatch in the frequency response of the auditory filters in the computer model and the gammatone filters in the analysis–synthesis processing path. As well, the group delays in the auditory filters are calculated at the filter's center frequency as opposed to the dominant stimulus frequency, resulting in further errors in the group delay prescribed in the modeling path.

These limitations contribute to model results that suggest SPC has increased the absolute difference between healthy and impaired spatiotemporal response patterns and degraded the synchrony to the formants in a synthesized vowel. However, SPC provided modest benefit to an error metric based on cross-correlations between AN firing patterns, which indicates that some phase correction is occurring.

MATERIALS AND METHODS

Auditory Model

The computational model of the cat auditory periphery by Zilany and Bruce (2006, 2007) is used in the present implementation and evaluation of the SPC scheme. Figure 1 shows a schematic of the auditory model, which describes the processing done from the middle ear to the AN where the output is given as spike times for a CF of interest. The error metrics used in this article make use of the synapse model output, which gives the probability of spikes at each instant before refractoriness is incorporated, and each fiber has a spontaneous rate of 50 spikes per second. The model consists of a signal path that is separated into two paths to represent the C1/C2 transition hypothesis (Sewell 1984; Kiang et al. 1986): the C1 and C2 filters represent two modes of vibration on the basilar membrane, and their subsequent transduction functions represent two modes of transduction in the inner hair cells (Zilany & Bruce 2006). In addition to the signal path, there is a control path that represents

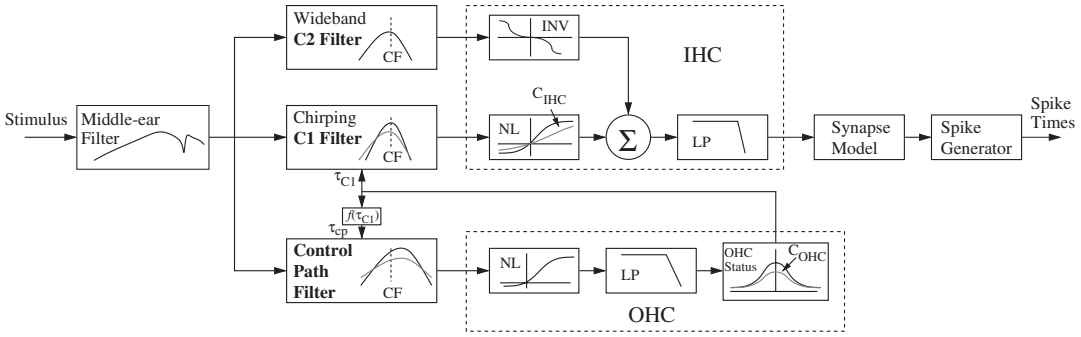

Fig. 1. Block diagram of the Zilany and Bruce (2006) cat auditory nerve model. C_{IHC} and C_{OHC} are scaling constants that control IHC and OHC status, respectively. CF, characteristic frequency; IHC, inner hair cell; OHC, outer hair cell; LP, low-pass filter; NL, static nonlinearity; INV, inverting nonlinearity. Reprinted with permission from the *J Acoust Soc Am* 2006;120:1446–1466.

the active feedback mechanisms of the the outer hair cells and affects the sharpness of tuning in the cochlear filters of the signal path. Impairment to the outer hair cells can be incorporated into the control path. This broadens the bandwidth of the C1 auditory filter in the signal path and shortens its group delay.

Implementation of SPC

We implemented a version of the SPC processing scheme as described by Shi et al. (2006), which is depicted in Figure 2. Their scheme consists of two main paths: (1) a modeling path to determine the group delays through a healthy and an impaired bank of auditory filters, and (2) a processing path that introduces the insertion delays found in the modeling path into the signal using an analysis–synthesis filterbank.

Shi et al. (2006) used the computational model of Heinz et al. (2001) in the modeling path. That auditory model uses gammatone filters to represent the tuning of the basilar membrane. The bandwidth of these filters fluctuates with sound level, resulting in a time-varying filter group delay; the range of bandwidths in which these filters fluctuate is affected by outer hair cell impairment. The difference between the healthy and impaired group delay was found for a range of CFs that correspond to the center frequencies in the SPC signal path. These insertion delays are introduced into the processing path by time-shifting the appropriate channels of the analysis–synthesis filterbank.

The present implementation of the SPC scheme uses the auditory model of Zilany and Bruce (2006) in the modeling path because it provides an improved representation of AN response in cats that is extensively validated on physiological data. The cat model provides insight into the effects of SPC; however, a human version of the auditory model is currently being developed (Ibrahim & Bruce 2010). The cat model includes a middle ear filter, a

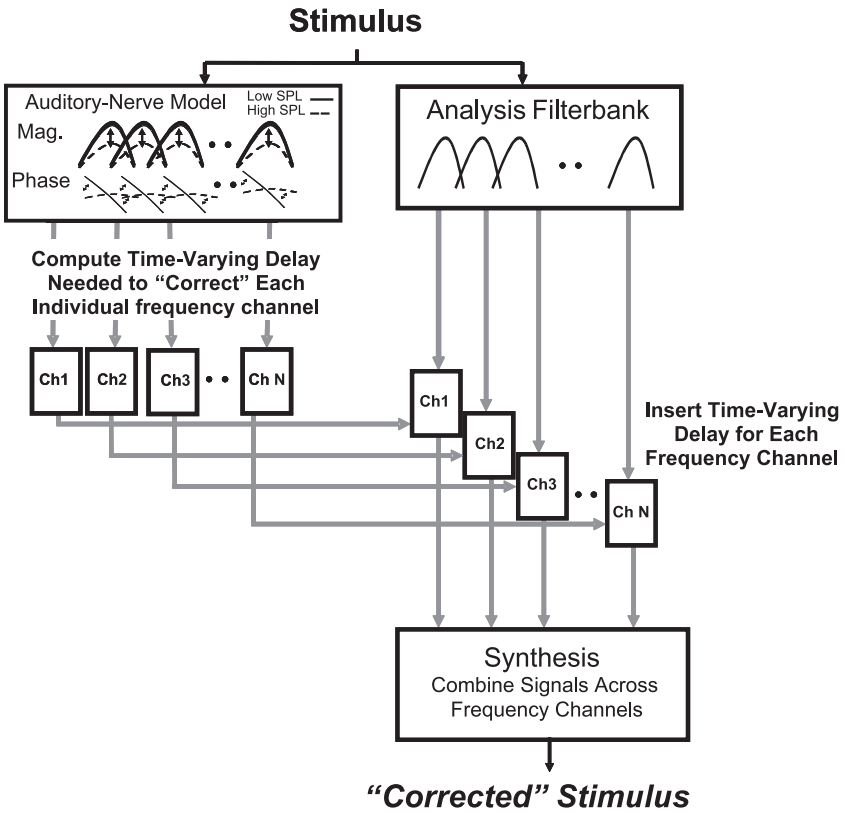

Fig. 2. A block diagram of the processing done in the spatiotemporal pattern correction scheme. The modeling path is on the left, and the processing path is on the right. Reprinted from Shi et al. with permission from the *Speech Lang Hear Res* 2006;49:848–855.

feature not found in the model used by Shi et al. (2006), that provides spectral shaping, which affects the responses to broadband stimuli such as speech. The Zilany and Bruce model also provides improved accuracy in AN response patterns over a wider range of sound levels. The group delay through this model is found as the group delay through the C1 filter, which describes the tuning of the basilar membrane at low to moderate input levels.

C1 Filter Group Delay

The C1 filter in the Zilany and Bruce model is a 10th-order digital filter with two complex second-order poles, one complex first-order pole, their complex conjugates, and a real fifth-order zero. It is implemented as a cascade of five second-order difference equations, each with a single pole and a single zero. The group delay through the C1 filter is found using the analog description of these filters given in Eqs. (9) to (15) of Zilany and Bruce (2006) for computational simplicity. Any change in group delay that would be brought about by analog-to-digital conversion is found to be three orders of magnitude smaller than the group delays inserted by SPC and considered negligible. First, we take an analog transfer function of the complex variable $s=\sigma+j\omega$ given by:

$$H(s)=\frac{(s-q)}{(s-p)(s-p^*)}, \tag{1}$$

where q is a real zero, $p=\sigma_p+j\omega_p$ is a complex pole, and $p^*=\sigma_p-j\omega_p$ is its complex conjugate. Then the group delay is found by evaluating the frequency response and applying the definition of group delay given by Oppenheim et al. (1999) as $-(d/d\omega)\angle H(j\omega)$. This yields the group delay, τ, for this filter as:

$$\tau[H(j\omega)]=\frac{\sigma_q}{\sigma_q^2+\omega^2}-\frac{\sigma_p}{\sigma_p^2+(\omega-\omega_p)^2}-\frac{\sigma_p}{\sigma_p^2+(\omega+\omega_p)^2}. \tag{2}$$

The appropriate poles and zeros derived in Eqs. (9) to (15) of Zilany and Bruce (2006) are applied to (Eq. 2), which is evaluated at the fiber's CF because the precise frequency content of the input is not known. These group delays are added for each of the five second-order filters to determine the C1 filter's total group delay.

Analysis–Synthesis Filterbank

The difference in group delay through the healthy and impaired C1 filters was introduced as a temporal shift in the corresponding frequency channels of the analysis–synthesis gammatone filterbank. The filterbank, as described by Hohmann (2002) and used by Shi et al. (2006), consists of 52 linear fourth-order gammatone filters evenly spaced on the equivalent rectangular bandwidth scale (Glasberg & Moore 1990) with two filters per equivalent rectangular bandwidth from 100 to 5000 Hz. Details of the gammatone filterbank design are thoroughly described by Hohmann (2002). The modeling path evaluates AN fibers with center frequencies matching the filterbank's center frequencies. Delays are only inserted in the frequency channels below 2000 Hz because Shi et al. suggest that there will be no benefit from SPC above this frequency. Temporal shifts are applied in integer sample numbers, where the sampling frequency is Fs = 32 kHz.

Shi et al. (2006) and Calandruccio et al. (2007) inserted delays at various SPC strengths, which scaled the amount of delay inserted based on the assumed percentage of remnant cochlear nonlinearity. Since we used the auditory model to characterize the hearing loss, the retained nonlinearity was known and we inserted the full difference in group delay as prescribed by the healthy and impaired C1 filters at each filterbank channel. In addition to these inserted delays, the Hohmann filterbank applies static temporal shifts to each channel to account for the different delays of the impulse responses in each gammatone filter. Upon successful temporal alignment, the real part of the filterbank's channels are linearly combined into a weighted sum

$$y(n)=\sum_{k=1}^{52} g_k \cdot y_k(n), \tag{3}$$

where the y_k's are the filterbank channels and the g_k's are the channel weights. Hohmann (2002) suggests that the weights be optimized to approximate a flat frequency response of the system, but the optimization method is not reported. The method used in the present study measures the gain of the filterbank's impulse response at the center frequencies and then sets the weights according to

$$g_k=1/|Y_k[l_k']| \tag{4}$$

$$\text{with}\quad Y_k[l]=\sum_{n=0}^{N-1} y_k'[n]e^{\frac{-2\pi i}{N}ln} \tag{5}$$

$$\text{and}\quad l_k'=\arg\min_l\left(\frac{l}{N}\text{Fs}-fc_k\right), \tag{6}$$

where the Y_k's represent the discrete Fourier transform (DFT) of the filterbank channel impulse responses, y_k'; k indexes the filterbank channels; and N=8192. The Fourier coefficient frequencies are indexed by l, and the l_k''s represent the indices of the nearest Fourier frequencies to the true filterbank center frequencies, fc_k. This brings the gain-frequency response close to flat when there are no insertion delays applied; however, the time-varying nature of the insertion delays results in a nonflat frequency response in the SPC implementation. Dynamic reweighting of the filterbank's frequency channels may be required for a flat gain-frequency response but was not attempted in this study.

Latency in SPC's Pathways

We found that the latencies through the modeling path and the processing path were different and may result in application of the insertion delays at the incorrect time. For this reason, our SPC implementation measures the difference in latencies between the middle ear filter in the modeling path and the gammatone filters in the processing path. First, the delay required to account for the difference in group delay is found for each auditory filter CF. Then the delay required to account for the remaining difference in phase delay is found. This information is used to adjust the time at which the prescribed SPC delays are applied. This adjustment was important because it resulted in better formant synchrony and improved peak-lag error, as discussed later. This adjustment may not have been necessary in Shi et al.'s SPC implementation because they used gammatone filters both in the processing and modeling paths.

TABLE 1. Model audiogram

Frequency (kHz)	0.25	0.5	0.75	1	1.5	2	3	4	6	8
Hearing loss (dB)	20	20	25	30	35	40	45	50	50	50

RESULTS

Evaluation of SPC

Our implementation of the SPC processing scheme is evaluated in response to a synthesized vowel, /ε/, of duration 100 milliseconds with a fundamental frequency of 100 Hz; formants at 500, 1700, 2500, and 3300 Hz; and presented at levels between 50 and 80 dB SPL in 10 dB increments. The AN response predicted by the Zilany and Bruce (2006) model is used to evaluate the SPC scheme. The model consisted of 41 AN fibers with a spontaneous rate of 50 spikes per second and CFs logarithmically spaced from 250 Hz to 8 kHz to appropriately include the frequencies of voiced speech. The impairment represented by the model was a sloping mild, mixed hearing loss, with damage to the outer hair cells accounting for two third of the loss and damage to the inner hair cells accounting for the other one third (Plack et al. 2004). The audiogram that appears in Table 1 was linearly interpolated to the model CFs.

We compare the neural firing patterns, or neurograms, generated by the healthy and impaired auditory models at the level of synaptic output to determine SPC effectiveness. A qualitative illustration of the synaptic output for the healthy, impaired, and SPC-aided case is shown in Figure 3 for a brief portion of the vowel across a range of CFs. The healthy neurogram (A) displays visible synchrony to multiple frequency components, including the first and second formants across a range of CFs. The synchronous response to formant frequencies at higher CFs becomes diminished and smeared in the impaired response (B). In the SPC-aided case (C), a disruption of synchrony is evident particularly to the second formant in the 2 kHz CF range, and the higher CFs seem to be synchronizing to additional frequency components not observed in the healthy response.

A number of evaluation metrics are used to quantify the difference between healthy, impaired, and SPC-aided neural firing patterns. First, an absolute difference is taken between the healthy and impaired neurograms

$$E_{\text{abs}} = \sum_{t} \sum_{\text{CF}=1}^{41} | N_{\text{Healthy}}(\text{CF},t) - N_{\text{Impaired}}(\text{CF},t) |, \qquad (7)$$

where N is the neurogram, CF is the fiber characteristic frequency, and t indexes time for the duration of the vowel. We find the absolute error between neurograms at two time resolutions: (1) at 62.5 microseconds to preserve spike timing and phase locking information and (2) at 60 microseconds smoothed with a 128 sample Hamming window to emphasize mean rate information. The Hamming window low-pass filters the neural response with 60-microsecond time resolution to exclude synchronization frequencies above approximately 265 Hz,

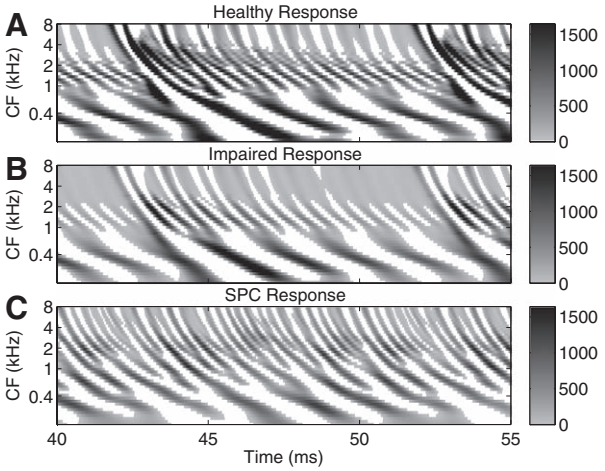

Fig. 3. Synapse model output to a synthesized vowel /ε/ presented at 60 dB SPL shown as a function of time for 41 auditory nerve fiber characteristic frequencies (CFs) logarithmically spaced from 250 to 8000 Hz. The healthy (A) and impaired (B) model responses to the unaided vowel are shown along with the impaired response to the spatiotemporal pattern correction–aided vowel (C). Spiking rates indicated by the gray-scale color bar are given in spikes per second, but these rates are exaggerated because the synapse model does not incorporate refractoriness.

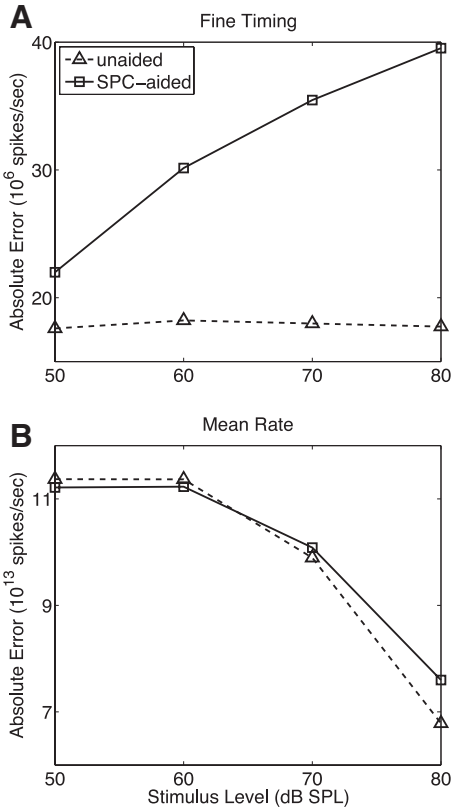

Fig. 4. A, Fine-timing absolute error between the healthy and impaired spatiotemporal response pattern for a synthesized /ε/ vowel. The synapse model is used with a temporal resolution of 62.5 μs. Spiking rates are exaggerated in the synapse model because no refractoriness is incorporated. B, Mean rate absolute error. The synapse model is used with an effective temporal resolution of 3.8 ms. The solid lines with square markers represent the results for the vowel aided with spatiotemporal pattern correction (SPC) and the dashed lines with triangle markers represent the unaided case.

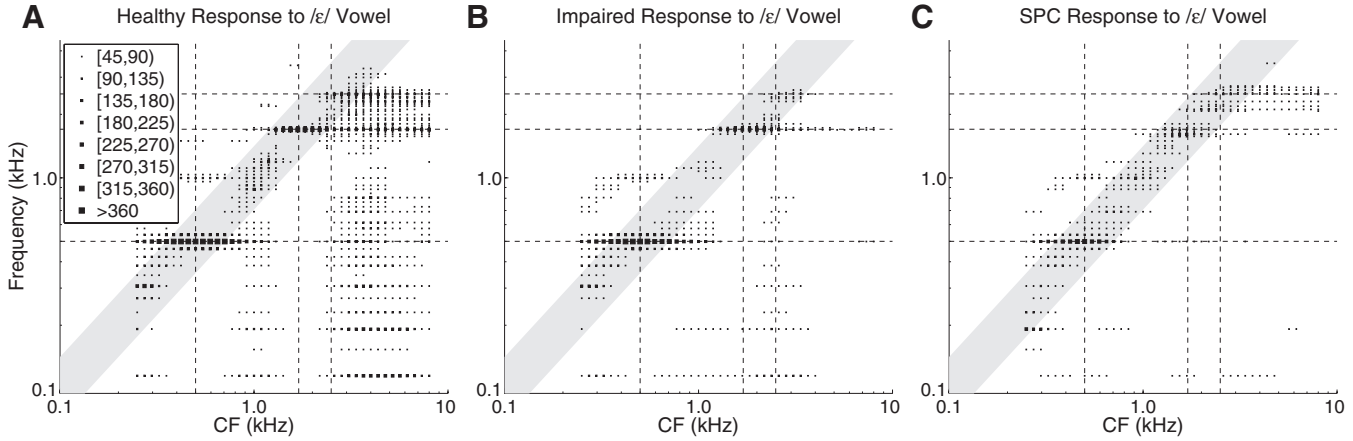

Fig. 5. Box-plots of the synchronized rate of auditory nerve response to a synthesized vowel /ɛ/ presented at 60 dB SPL. A, Healthy model response to the unaided vowel. B, Impaired model response to the unaided vowel. C, Impaired model response to the vowel aided with spatiotemporal pattern correction (SPC). Auditory nerve (AN) fibers are represented along the x axis in order of increasing characteristic frequency (CF). Stimulus frequency components are represented along the y axis. The first three formants of the stimulus are shown as horizontal dashed lines; vertical dashed lines indicate the formant positions relative to AN fiber CF. Marker size is dependent on the value of the synchronized rate (Eq. 8) at each frequency; a large marker size indicates a high degree of response at that frequency.

providing an effective temporal resolution of 3.8 milliseconds. This removes the phase-locked response to the vowel's formant frequencies so that the absolute difference is heavily influenced by the AN fiber's mean firing rate. The results shown in Figure 4 demonstrate that processing with SPC increased the absolute error in the spike timing neural response but had little effect on the mean rate information. This suggests that SPC disrupts certain aspects of the fine temporal structure of the AN response pattern, particularly as sound level increases. Shi et al. (2006) presented their stimuli at 65 dB SPL, which in some cases could be lower than the 30 dB re: speech reception threshold level at which Calandruccio et al. (2007) presented their stimuli. This, together with the fact that fine temporal structure appears disrupted by SPC as sound level increases, may have contributed to relatively better intelligibility results found by Shi et al. (2006).

Second, we compare the synchronized rate of the aided neural response by using box-plots similar to those used by Miller et al. (1999). The synchronized rate, presented by Miller et al. (1997), is a normalized DFT of the instantaneous spiking rate. Here it is found for a 26-millisecond time window $\omega(n)$ as

$$R(kf_t) = \frac{\left|\sum_{n=0}^{N-1} w(n)s(n)e^{-j2\pi kn/N}\right|}{\sqrt{N\sum_{n=0}^{N-1} w(n)^2}}, \qquad (8)$$

where N=416, $s(n)$ is the model synaptic output in spikes per second resampled to 16 kHz, f_t is the frequency spacing of each frequency bin in the DFT, and k indexes the frequency bins. The synchronized rate is found for a population of fibers and discretized in the box-plots of Figure 5, where a marker of corresponding size is plotted.

The effects of processing with SPC on the synchronized rate are demonstrated in Figure 5C. In the impaired response to the unaided vowel in panel B, the response to the third formant is diminished when compared to the healthy response in panel A; however, synchrony to the first and second formants remains evident. In the impaired response to the SPC-aided vowel in panel C, the synchrony to the second and third formants is considerably diminished and synchrony to the first formant is somewhat decreased. As well, there seems to be increased activity to the inter-formant frequencies when compared to the unaided-impaired case. This suggests the representation of the vowel in the SPC-aided neural response is degraded.

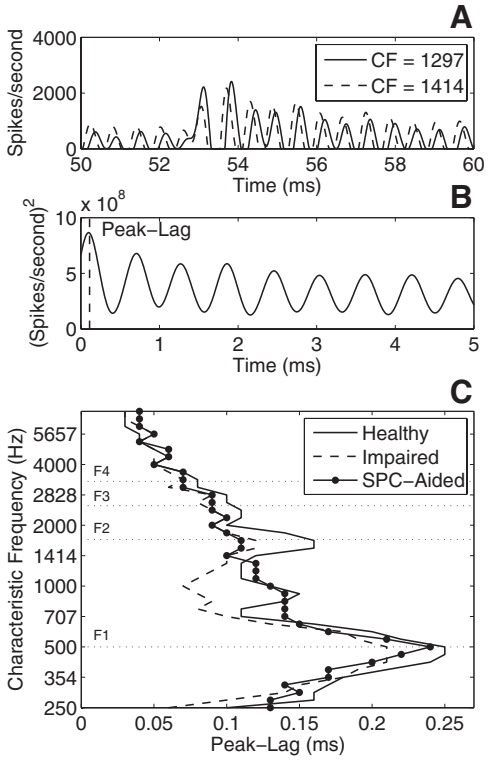

Fig. 6. A, Synaptic output from two healthy model fibers with characteristic frequencies (CFs) 1297 and 1414 Hz. Spiking rates are exaggerated because the synapse model does not include refractoriness. B, Positive half of the cross-correlation between the two neural responses in A. The sampling frequency is 100 kHz. Peak-lag is shown by the vertical dotted line; see text for explanation. C, Peak-lags shown at the lower CF of each neighboring fiber pair for the healthy (solid line), impaired (dashed line), and spatiotemporal pattern correction–aided (circle-marked line) response to the /ɛ/ vowel. The vowel is presented at 60 dB SPL, and formant frequencies are indicated by horizontal dotted lines.

Third, to analyze the phase response, we develop a cross-correlation metric that compares the lags at which neighboring fiber cross-correlations are maximal. This is referred to as the cross-correlation peak-lag and is related to the relative phase of the two relevant fibers. The derivation of the peak-lag is demonstrated in Figure 6, where the cross-correlation is shown for the neural response of two healthy AN fibers. If the two fibers are synchronizing to the same frequency component of the stimulus, and there is a large difference in phase between their responses, then the peak-lag will have a relatively large value. In contrast, if the two responses have similar phase, that is their spike times are more coincident, then the peak-lag of the cross-correlation will have a small value. In this way, the cross-correlation can give an indication of the relative phase between two fibers without prior knowledge of the frequency to which they are synchronized. As well, the cross-correlation peak-lag is affected by neighboring fibers that are synchronized to different stimulus frequency components. This information can be used to derive a neural spiking error metric for two model responses that compares the relative phase of activity across all neighboring CFs.

The difference between the peak-lags of the healthy response and the impaired response is found at each CF and summed across frequencies to obtain a peak-lag error value. Figure 6C shows the peak-lags across fiber CFs for the healthy, impaired, and SPC-aided model response to the /ε/ vowel presented at 60 dB SPL. The peak-lags in the impaired model are lower than the peak-lags in the healthy model because the phase response that is associated with broader cochlear filters is flattened. The SPC-aided vowel seems to provide some correction of peak-lag at the first formant frequency and at frequencies between the first and second formant, but not at the higher formant frequencies. The lack of improvement for the second formant seems to be due to the abnormal frequency modulations caused by the SPC insertion delays, while the third and fourth formants fall above the 2 kHz upper limit of the SPC scheme.

The difference in peak-lag between the healthy and impaired models, or peak-lag error, is shown for a range of stimulus presentation levels in Figure 7. This is compared to the SPC-aided vowel presented to the impaired auditory model, which results in a smaller peak-lag error at low stimulus presentation levels when compared to the unaided case.

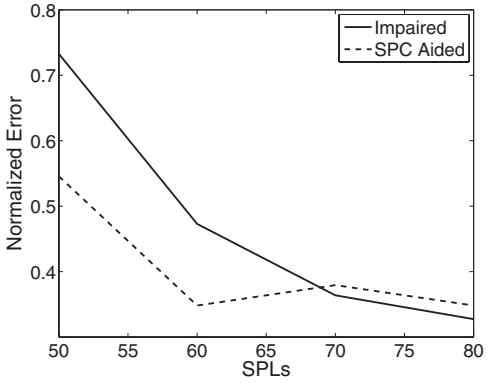

Fig. 7. Error between healthy and impaired model peak-lags summed across characteristic frequencies and plotted as a function of SPL. The response to the vowel processed with spatiotemporal pattern correction (SPC) is represented by the dashed line, and the response to the unaided vowel is represented by the solid line.

Analysis of SPC

Evaluation of the SPC processing scheme with the error metrics described previously suggests that the AN response to a synthesized vowel is distorted as a result of phase adjustment via group delay correction. Closer analysis of the SPC scheme provides insight into the reasons for these distortions and has revealed several technical limitations that were found to be associated with the SPC.

First, the frequency response through the analysis–synthesis filterbank is not flat and its shape fluctuates as a function of time. This fluctuation results from the time-varying delays that are inserted into the different frequency channels as prescribed by the modeling path. The gain-frequency response of the analysis–synthesis filterbank is shown at two arbitrarily chosen time points during the processing of the synthesized /ε/ vowel in Figure 8.

Second, the insertion of time-varying delays results in the appearance of frequency modulations in each frequency channel. Figure 9 shows a portion of the filtered synthesized /ε/ vowel in the 1344 Hz channel, the corresponding insertion delays prescribed for that channel, and the resultant waveform after the insertion delays in panels A, B, and C, respectively. Frequency modulations are visually evident in the resultant waveform of panel C. Figure 9D shows the filtered vowel that appears in panel A after it has been processed by the C1 filter of a healthy ear model. Panel E shows the SPC-aided waveform after it has been processed by the impaired C1 filter. Filtering by the impaired ear partially mitigates the frequency modulations present at the output of the C1 filter; however, the modulations generated by SPC processing could also degrade the response of inner ear filters across a range of fiber CFs.

Frequency modulations distort the SPC-aided vowel such that the representation of the formants is deteriorated. This is evident in the frequency spectrum of the unaided and SPC-aided vowel shown in Figure 10. Here there is substantial spectral smear, such that the second, third, and fourth formants are no longer visibly evident in the SPC-aided vowel spectrum. This formant smearing explains how the frequency modulations cause disrupted synchrony to the vowel as shown in the SPC-aided, impaired response box-plot in Figure 5.

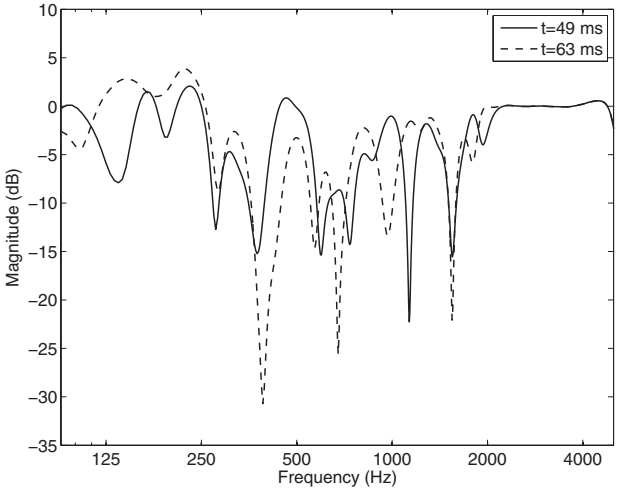

Fig. 8. Gain-frequency response of the spatiotemporal pattern correction filterbank shown at two different time points in the processing of the /ε/ vowel at 60 dB SPL.

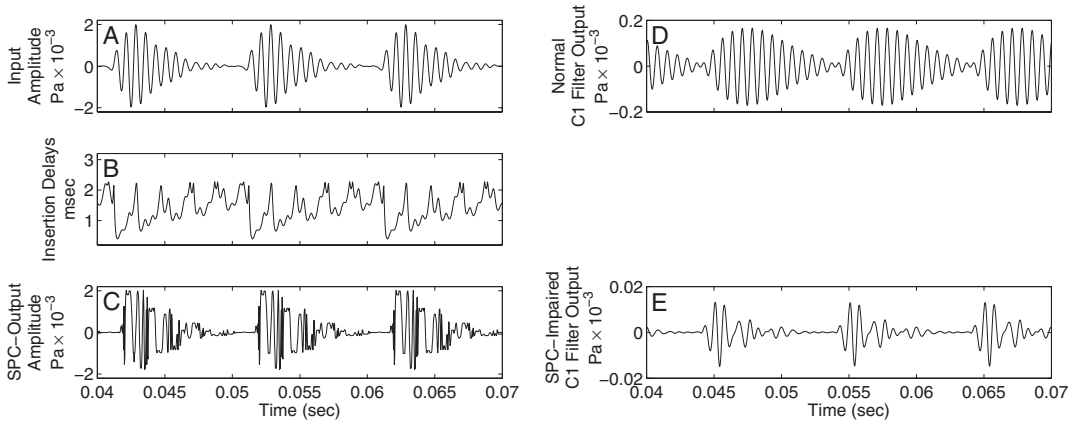

Fig. 9. Demonstration of frequency modulations in individual channels by spatiotemporal pattern correction (SPC). A, Synthesized /ɛ/ vowel at 60 dB SPL filtered through the 1344 Hz SPC filterbank channel. B, Insertion delays prescribed by SPC for this channel. C, Result of applying insertion delays to the filtered vowel. D, Result of passing the synthesized filtered vowel through the healthy C1 filter of the auditory model. E, Result of passing the SPC-aided filtered vowel through the impaired C1 filter.

Third, the SPC processing scheme applies its insertion delays only to approximately correct frequency components. This imprecision has two causes. The first cause results from the group delay calculations in the SPC scheme being evaluated at the AN fiber's CF. Frequency imprecision results when the dominant frequency passing through the filter is not equal to the fiber's CF. This occurs often because the dominant frequency seen in the auditory filters is dependent on the stimulus and the spread of synchrony that occurs in the response of healthy AN. This means that when a prominent off-center frequency component is passed through the filter, the insertion delays prescribed for that channel are based on the filter's center frequency rather than the stimulus frequency. This could result in imprecise insertion delay prescription.

The second cause of imprecision results from a mismatch in the frequency response of SPC's modeling path filters and its processing path filters. The auditory filters in the modeling path are nonlinear and broaden as stimulus level increases, causing spread of synchrony to a particular frequency component across a range of fiber CFs. In contrast, the gammatone filters in the processing path are of fixed bandwidth. This means that the group delay through a wide range of auditory filters can be derived from a single synchronization frequency, whereas the prescribed insertion delays that result are applied to a range of stimulus frequency components, instead of just the single synchronization frequency.

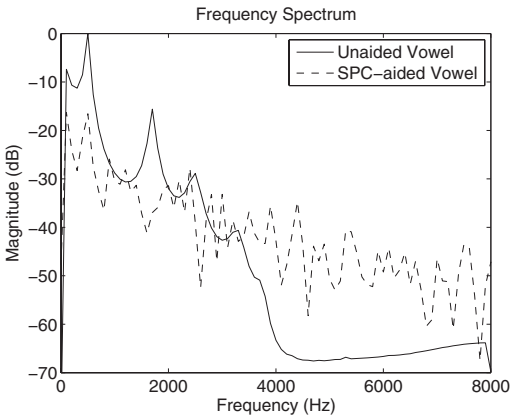

Fig. 10. Relative magnitude-frequency spectra of the unaided synthesized vowel /ɛ/ (solid line) and the corresponding spatiotemporal pattern correction (SPC)-aided vowel (dashed line) when presented to the SPC prescription algorithm at 60 dB SPL.

DISCUSSION

The results demonstrate that processing with SPC increased the absolute difference of the fine-timing neural firing pattern generated by the auditory model and disrupted synchrony to a synthesized vowel. The absolute difference provides a neural metric that includes all information passing through the AN as captured by the auditory model, including spiking rates, spike timing information, and frequency–place information. Processing by SPC degrades some of this information, particularly as stimulus level increases. As well, analysis of the synchronized rate gives a synchrony measure of the response to formants of voiced speech. The frequencies of vowel formants have been suggested to determine the specific vowel type (Peterson & Barney 1952; Kiefte et al. 2010). Synchrony to a formant frequency suggests that an auditory neuron is responding to that formant energy in the stimulus, providing a representation of the formant in neural code (Young 2008). The results indicate that SPC deteriorates the synchrony in response to a vowel's formants and suggest that SPC degrades the representation of the vowel in the AN response. In addition to degrading vowel representation, it is likely that synchrony disruption would affect sound localization as interaural timing differences would be distorted.

Despite this degradation, processing with SPC reduced the near-CF fiber cross-correlation error (peak-lag error) for low stimulus levels. The peak-lag error gives a general indication of the relative phase of AN fibers with neighboring CFs. This could be related to representation of loudness, as proposed by Carney (1994). However, the reduction of peak-lag error provided by SPC seems to come at the cost of disrupted synchrony and increased absolute error in the spatiotemporal response pattern.

Analysis of the SPC scheme with a computational model of the auditory periphery demonstrates several technical limitations associated with the SPC processing scheme, which may explain some of the poor results. These limitations include a

nonflat frequency response and frequency modulations introduced into the signal, which need to be addressed before a phase adjustment scheme like SPC can be implemented successfully.

The Hohmann (2002) filterbank implemented in the SPC processing path addresses the issue of a nonflat frequency response by optimizing the weights in a weighted sum used to combine the gammatone filterbank's channels. However, the time-varying delays of the SPC scheme applied to each channel require a dynamic optimization of the channel weights to approach a flat filterbank frequency response using this method. Dynamic optimization would require evaluation of the filterbank's impulse response at every time the insertion delays are applied, which in this case is every time sample. This would be computationally expensive and is not guaranteed to result in an acceptably flat frequency response.

The frequency modulations introduced by SPC could be reduced by applying a low-pass filter on the insertion delays prescribed for each filterbank channel, or an averaging all-pass filter on the SPC-aided channels to provide the averaged SPC insertion delays. However, both these scenarios conflict conceptually with the operation of the SPC processing scheme, which measures the difference in group delay between the healthy and impaired auditory models in real time to account for them. Low-pass filtering or averaging would alter the insertion delays that were prescribed to correct for group delay. Furthermore, preliminary investigation suggested that applying a 500 Hz cutoff low-pass filter to the SPC delays resulted in further degradation to the synchronized rate representation of the vowel /ɛ/ and did not improve the error in the peak-lag cross correlation error metric at low levels (Zeyl, Reference Note 1). Averaging the SPC-aided channels with a 20-point zero-phase moving-average filter did not result in visibly reduced frequency modulations in the C1 filtered output and did not improve the peak-lag cross correlation error.

Analysis of the SPC scheme also reveals a conceptual problem associated with the imprecision of frequency components used to prescribe the insertion delays. The SPC scheme cannot restore the normal spatiotemporal response to a pure tone. In the impaired AN response, the individual fibers have a shallower phase transfer function at low SPLs, resulting in more coincident spiking across the population that exhibits phase locking. Correcting for impaired group delay in response to a pure tone by inserting a static delay cannot affect the flattened relative phase of AN fibers synchronized to the tone. This is because applying a static delay will simply shift the overall AN response in time. As well, applying a dynamic insertion delay will introduce new frequency components into the signal, which cannot affect the relative AN phase response to the tone of interest but may disrupt the synchrony to that tone.

We presented the SPC response to a vowel, but the identification of consonants might also be improved from the correction of impaired across-frequency delays. Consonants contain transient fluctuations in power across multiple frequencies, which means correct temporal processing of the relative frequencies may be important for consonant recognition. We processed several consonants with SPC and found similar results (not shown) to those of the vowel, namely, decreased similarity in the impaired and healthy neurograms, frequency modulations in each processing channel, and a fluctuating nonflat frequency response. The peak-lag cross correlation metric did not improve for the SPC-aided consonants as it did for vowels; this follows because the peak-lag cross correlation is sensitive to relative phase response of neighboring AN fibers to periodic stimulus components, which are prominent in vowels but not in consonants.

In this study, we have used a phenomenological model of the auditory periphery not only to implement the modeling path of SPC but also to evaluate the SPC scheme; this represents an idealistic scenario. In any real hearing aid application, the usefulness of SPC will be dependent on how well the modeling path represents the group delay changes with sound level and impairment in the real ear. Alternative auditory models, such as the dual-resonance nonlinear filterbank model (Lopez-Poveda & Meddis 2001; Meddis et al. 2001), could be used in SPC's modeling path, but these would suffer from the same issues found in the present study, namely that frequency modulations are introduced by incorporating time-varying delays and that corrections may be implemented at frequencies away from the CF.

The potential benefits of SPC lie in the correction of relative phase response across a population of AN fibers that are responding to a complex stimulus. The present study has assessed this using a near-CF fiber cross-correlation error metric that compares the lag at which the cross-correlation is maximized. The results suggest that SPC may provide some benefit at low stimulus presentation levels. However, this is below conversational speech levels. As well, the perceptual importance of relative phase response across an AN population remains uncertain.

The results of SPC processing on the near-CF fiber cross-correlation metric stimulate future investigation into two avenues for potential perceptual gains by SPC: (1) improvements to intelligibility of speech at low SPLs, and (2) the restoration of loudness representation in listeners with sensorineural hearing loss. One caveat here is that SPC is unable to improve the phase response of an AN population to a pure tone, as discussed previously, which limits the possibility that SPC can restore the loudness representation of a pure tone in accordance with the loudness model proposed by Carney (1994).

We found that SPC increased the error in the absolute spatiotemporal response patterns and synchronized rate patterns, which could explain the poor experimental results obtained by Shi et al. (2006) and Calandruccio et al. (2007). However, these degradations were smallest at low SPLs and SPC provided the best peak-lag cross-correlation improvement at low levels. These results suggest that the benefits of SPC may be optimal at low SPLs and further motivate perceptual experiments that investigate the effects SPC has on speech intelligibility for quiet speech and on loudness growth for broadband sounds.

ACKNOWLEDGMENTS

The authors thank Jeffrey Zeyl and Jason Boulet for proofreading earlier versions of the article prior to submission.

This work was supported by NSERC Discovery Grant 261736 [ICB]. TJZ was supported by an NSERC Postgraduate Scholarship M.

The authors declare no conflict of interest.

Address for correspondence: Ian C. Bruce, Department of Electrical and Computer Engineering, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4K1, Canada. E-mail: ibruce@ieee.org.

Received November 4, 2012; accepted July 11, 2013.

REFERENCES

Cai, S., Ma, W. L., Young, E. D. (2009). Encoding intensity in ventral cochlear nucleus following acoustic trauma: Implications for loudness recruitment. *J Assoc Res Otolaryngol*, *10*, 5–22.

Calandruccio, L., Doherty, K. A., Carney, L. H., et al. (2007). Perception of temporally processed speech by listeners with hearing impairment. *Ear Hear*, *28*, 512–523.

Carney, L. H. (1994). Spatiotemporal encoding of sound level: Models for normal encoding and recruitment of loudness. *Hear Res*, *76*, 31–44.

Florentine, M., Buus, S., Scharf, B., et al. (1980). Frequency selectivity in normally-hearing and hearing-impaired observers. *J Speech Hear Res*, *23*, 646–669.

Glasberg, B. R., & Moore, B. C. (1990). Derivation of auditory filter shapes from notched-noise data. *Hear Res*, *47*, 103–138.

Heinz, M. G., Zhang, X., Bruce, I. C., et al. (2001). Auditory nerve model for predicting performance limits of normal and impaired listeners, *Acoust Res Lett Online*, *2*, 91–96.

Hohmann, V. (2002). Frequency analysis and synthesis using a Gammatone filterbank, *Acta Acust United Ac*, *88*, 433–442.

Hood, J. D. (1977). Loudness balance procedures for the measurement of recruitment. *Audiology*, *16*, 215–228.

Ibrahim, R., & Bruce, I. C. (2010). Effects of peripheral tuning on the auditory nerve's representation of speech envelope and temporal fine structure cues. In E. A. Lopez-Poveda, A. R. Palmer, & R. Meddis (Eds.), *The Neurophysiological Bases of Auditory Perception* (pp. 429–438) New York, NY: Springer.

Kates, J. M., & Arehart, K. H. (2005). Multichannel dynamic-range compression using digital frequency warping. *EURASIP J Appl Sig Proc*. *2005*, 483486.

Kiang, N. Y., Liberman, M. C., Sewell, W. F., et al. (1986). Single unit clues to cochlear mechanisms. *Hear Res*, *22*, 171–182.

Kiefte, M., Enright, T., Marshall, L. (2010). The role of formant amplitude in the perception of /i/ and /u/. *J Acoust Soc Am*, *127*, 2611–2621.

Lopez-Poveda, E. A., & Meddis, R. (2001). A human nonlinear cochlear filterbank. *J Acous Soc Am*, *110*, 3107–3118.

Lorenzi, C., Gilbert, G., Carn, H., et al. (2006). Speech perception problems of the hearing impaired reflect inability to use temporal fine structure. *Proc Natl Acad Sci USA*, *103*, 18866–18869.

Meddis, R., O'Mard, L. P., Lopez-Poveda, E. A. (2001). A computational algorithm for computing nonlinear auditory frequency selectivity. *J Acous Soc Am*, *109*, 2852–2861.

Miller, R., Calhoun, B., Young, E. (1999). Contrast enhancement improves the representation of /ɛ/-like vowels in the hearing-impaired auditory nerve. *J Acoust Soc Am*, *106*, 2693–2708.

Miller, R. L., Schilling, J. R., Franck, K. R., et al. (1997). Effects of acoustic trauma on the representation of the vowel /ɛ/ in cat auditory nerve fibers. *J Acoust Soc Am*, *101*, 3602–3616.

Moore, B. C. (2004). Testing the concept of softness imperception: Loudness near threshold for hearing-impaired ears. *J Acoust Soc Am*, *115*, 3103–3111.

Moore, B. C., Glasberg, B. R., Hess, R. F., et al. (1985). Effects of flanking noise bands on the rate of growth of loudness of tones in normal and recruiting ears. *J Acoust Soc Am*, *77*, 1505–1513.

Moore, B. C. (1985). Frequency selectivity and temporal resolution in normal and hearing-impaired listeners. *Br J Audiol*, *19*, 189–201.

Oppenheim, A. V., Schafer, R. W., Buck, J. R. (1999). *Discrete-Time Signal Processing* (2nd ed.). Upper Saddle River: Prentice Hall.

Patuzzi, R. B., Yates, G. K., Johnstone, B. M. (1989). Changes in cochlear microphonic and neural sensitivity produced by acoustic trauma. *Hear Res*, *39*, 189–202.

Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of the vowels. *J Acoust Soc Am*, *24*, 175–184.

Pichora-Fuller, M. K., Schneider, B. A., Macdonald, E., et al. (2007). Temporal jitter disrupts speech intelligibility: A simulation of auditory aging. *Hear Res*, *223*, 114–121.

Plack, C. J., Drga, V., Lopez-Poveda, E. A. (2004). Inferred basilar-membrane response functions for listeners with mild to moderate sensorineural hearing loss. *J Acoust Soc Am*, *115*, 1684–1695.

Rhode, W. S. (1971). Observations of the vibration of the basilar membrane in squirrel monkeys using the Mössbauer technique. *J Acoust Soc Am*, *49*, 1218–1231.

Robles, L., Ruggero, M. A., Rich, N. C. (1986). Basilar membrane mechanics at the base of the chinchilla cochlea. I. Input-output functions, tuning curves, and response phases. *J Acoust Soc Am*, *80*, 1364–1374.

Ruggero, M. A., Narayan, S. S., Temchin, A. N., et al. (2000). Mechanical bases of frequency tuning and neural excitation at the base of the cochlea: Comparison of basilar-membrane vibrations and auditory-nerve-fiber responses in chinchilla. *Proc Natl Acad Sci USA*, *97*, 11744–11750.

Ruggero, M. A., & 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.

Sewell, W. F. (1984). Furosemide selectively reduces one component in rate-level functions from auditory-nerve fibers. *Hear Res*, *15*, 69–72.

Shi, L. F., Carney, L. H., Doherty, K. A. (2006). Correction of the peripheral spatiotemporal response pattern: A potential new signal-processing strategy. *J Speech Lang Hear Res*, *49*, 848–855.

Stone, M. A., & Moore, B. C. (2003). Tolerable hearing aid delays. III. Effects on speech production and perception of across-frequency variation in delay. *Ear Hear*, *24*, 175–183.

Uppenkamp, S., Fobel, S., Patterson, R. D. (2001). The effects of temporal asymmetry on the detection and perception of short chirps. *Hear Res*, *158*, 71–83.

Wojtczak, M., Beim, J. A., Micheyl, C., et al. (2012). Perception of across-frequency asynchrony and the role of cochlear delays. *J Acoust Soc Am*, *131*, 363–377.

Young, E. D. (2008). Neural representation of spectral and temporal information in speech. *Philos T Roy Soc B*, *363*, 923–945.

Zilany, M. S., & 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., & Bruce, I. C. (2007). 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.

REFERENCE NOTE

1. Zeyl, T. J. (2010). *Evaluation of combined spectral enhancement and phase adjustment in hearing-aids*, Master's thesis, McMaster University, 1280 Main Street West, Hamilton Ontario, Canada, L8S 4L8.