Companding to improve cochlear-implant speech recognition in speech-shaped noise[a)]

Aparajita Bhattacharya[b)]
Hearing and Speech Research Laboratory, Department of Biomedical Engineering, University of California, Irvine, 316 Med Surge II, Irvine, California 92697

Fan-Gang Zeng[c)]
Hearing and Speech Research Laboratory, Departments of Anatomy and Neurobiology, Biomedical Engineering, Cognitive Sciences, and Otolaryngology—Head and Neck Surgery, University of California, Irvine, 364 Med Surge II, Irvine, California 92697

(Received 12 October 2006; revised 25 May 2007; accepted 25 May 2007)

Nonlinear sensory and neural processing mechanisms have been exploited to enhance spectral contrast for improvement of speech understanding in noise. The "companding" algorithm employs both two-tone suppression and adaptive gain mechanisms to achieve spectral enhancement. This study implemented a 50-channel companding strategy and evaluated its efficiency as a front-end noise suppression technique in cochlear implants. The key parameters were identified and evaluated to optimize the companding performance. Both normal-hearing (NH) listeners and cochlear-implant (CI) users performed phoneme and sentence recognition tests in quiet and in steady-state speech-shaped noise. Data from the NH listeners showed that for noise conditions, the implemented strategy improved vowel perception but not consonant and sentence perception. However, the CI users showed significant improvements in both phoneme and sentence perception in noise. Maximum average improvement for vowel recognition was 21.3 percentage points ($p < 0.05$) at 0 dB signal-to-noise ratio (SNR), followed by 17.7 percentage points ($p < 0.05$) at 5 dB SNR for sentence recognition and 12.1 percentage points ($p < 0.05$) at 5 dB SNR for consonant recognition. While the observed results could be attributed to the enhanced spectral contrast, it is likely that the corresponding temporal changes caused by companding also played a significant role and should be addressed by future studies. [DOI: 10.1121/1.2749710]

PACS number(s): 43.66.Ts, 43.71.Ky, 43.71.Gv, 43.71.Es [AJO] Pages: 1079–1089

I. INTRODUCTION

In realistic listening situations, speech signals are often degraded by noise. While normal-hearing (NH) listeners are remarkably adept at extracting the critical information from noisy speech, this ability decreases rapidly with hearing loss. With the latest generation cochlear implants, the majority of implant users derive substantial benefits in quiet listening conditions, including communication over the telephone. However, their performance in noise is severely impeded, thereby significantly affecting the implant efficiency in real-life situations. In general, cochlear-implant (CI) users require much higher signal-to-noise ratios (SNRs) to match the performance of NH listeners on speech recognition tasks in noise (Fu and Nogaki, 2005; Hochberg *et al.*, 1992; Müller-Deiler *et al.*, 1995; Stickney *et al.*, 2004; Zeng and Galvin, 1999, Zeng *et al.*, 2005). Zeng *et al.* (2005) found that the speech reception threshold (SRT) of NH listeners was approximately 14 dB better than that of the CI users in steady-state noise. The difference was more drastic in fluctuating background noise. With a female talker as the masker, the SRT of NH listeners was approximately 32 dB higher than that of the CI users. SRT was defined as the SNR necessary for a listener to produce 50% correct score.

The auditory system uses nonlinear processing to provide the necessary spectral and temporal resolution. Outer hair cells (OHCs) in the cochlea are responsible for the active mechanisms observed in the peripheral auditory system. These OHCs employ nonlinear adaptive gain processing to encode the large dynamic range by a relatively narrow physiological range of the auditory nerve fibers. The same OHCs are also responsible for the sharp frequency selectivity of the auditory system. Another nonlinear phenomenon arising from complex interactions between OHCs and the basilar membrane is two-tone suppression (Rhode, 1974; Ruggero *et al.*, 1992). It is characterized by a decrease in the evoked response to a tone in the presence of a second tone (Sachs and Kiang, 1968). Two-tone suppression is considered to be the primary mechanism underlying spectral enhancement and is thought to improve the SNR of the stronger components (Rhode *et al.*, 1978; Sachs *et al.*, 1983; Stoop and Kern, 2004). Spectral enhancement is defined as an increase in peak-to-valley difference which is also referred to as "spectral sharpening."

[a)]Portions of this work were presented in "Companding to improve cochlear implants' speech processing in noise," *2005 Conference on Implantable Auditory Prosthesis*, Asilomar Conference Center, Pacific Grove, California, August 2005.

[b)]Author to whom correspondence should be addressed. Electronic mail: abhattac@uci.edu

[c)]Electronic mail: fzeng@uci.edu

Spectral sharpening may also result from competitive interactions between the neurons. Early studies on the Limulus retina demonstrated a neural contrast enhancement phenomenon, termed lateral inhibition, in which the output of a neuron is inhibited by the inputs from the adjacent neurons (Hartline, 1969). Lateral inhibition has also been observed in the auditory system. Blackburn and Sachs (1990) studied the discharge patterns of the cat anteroventral cochlear nucleus neurons in response to steady-state vowels. They found that certain neurons receive strong inhibitory inputs from surrounding neurons so that the overall neural response can maintain the valley between peaks (i.e., formants), particularly at moderate to high sound pressure levels and in the presence of background noise. Along with two-tone suppression, lateral inhibition too is likely responsible for preserving the spectral contrast between peaks and the valleys (Rhode and Greenberg, 1994).

Cochlear damage reduces the dynamic range, broadens the tuning curves (Kiang *et al.*, 1976) and eliminates two-tone suppression (Ruggero *et al.*, 1992). The combined effect is poor spectral resolution and reduced spectral contrast. As a result, the neural representation of formants in response to steady-state vowels is degraded, that results in poor vowel recognition(Loizou *et al.*, 2000; Miller *et al.*, 1997). Degraded spectral resolution also causes abnormal susceptibility to noise in hearing-impaired listeners (Horst, 1987).

In CI users, several factors limit the available spectral resolution and spectral contrast. Current cochlear implants that use amplitude compression to map the large acoustic amplitude range into a narrow electrical dynamic range have a side effect of reducing the spectral contrast. Loizou *et al.* (2000) studied the effect of input envelope amplitude compression on the recognition of vowels and consonants. They used stimuli processed by a six-channel cochlear implant simulator and tested NH listeners. They found that compression degraded the perception of vowels and recognition of place of articulation of the consonants. Studies on intensity discrimination with electric stimulation suggest that sometimes the size of just discriminable steps is smaller than NH listeners (Hochmair-Desoyer *et al.*, 1981; Nelson *et al.*, 1996) but it is not enough to compensate for the degrading effects of reduced dynamic range. This is because the number of discriminable steps within the dynamic range is considerably smaller compared to that of the NH listeners. In addition, nonlinear mechanisms such as two-tone suppression have not been implemented in cochlear implants. As a result, current CI users typically need a spectral contrast of 4–6 dB more than the NH listeners to accurately identify vowels in quiet listening condition (Loizou and Poroy, 2001). Spectral resolution in cochlear implants is limited by the number and location of the electrodes along the cochlea, the number of surviving neurons and the amount of current spreading from the stimulating electrodes. To a large extent, impaired spectral resolution and reduced spectral contrast are responsible for the susceptibility of cochlear implants to noise.

To compensate for the impaired ability to understand speech in noise, many noise suppression algorithms have been proposed for cochlear implants (Hamacher *et al.*, 1997; Toledo *et al.*, 2003; Weiss, 1993; Wouters and Vanden Berghe, 2001; van Hoesel and Clark, 1995). These algorithms use either single microphone or dual microphones and can be adaptive or nonadaptive in nature. Weiss (1993) showed that the INTEL method, which is used to suppress the random wideband noise, reduces the noise-induced deviations in the second formant frequency. Adaptive beamforming is a noise reduction technique, which uses signals from two or more microphones, to attenuate signal coming from directions other than the front. The portable beamformer using two microphones implemented by van Hoesel and Clark (1995) showed considerable improvements in speech intelligibility at 0 dB SNR. Wouters and Vanden Berghe (2001) used a two-channel two-stage adaptive filtering beamformer as a pre-processing stage in LAURA cochlear implants and found an improvement of 10 dB in SNR for the perception of consonant-vowel-consonant words. It is important to note that the speech source was angled at 90° with respect to the noise source in azimuth and the improvement may be smaller for angles less than 90°.

Several spectral enhancement techniques have particularly focused on compensating for the degraded spectral resolution of an impaired ear (Baer *et al.*, 1993; Bunnell, 1990; Clarkson and Bahgat, 1991; Franck *et al.*, 1999; Lyzenga *et al.*, 2002). Bunnell (1990) used a contrast enhancement technique in which the envelope amplitude of each fast Fourier transform bin was enhanced proportionately to the difference in the original envelope amplitude and the average spectrum level. He found a small improvement in the identification of stop consonants in quiet. Baer *et al.* (1993) convolved the spectrum with a difference of Gaussian filter to provide spectral enhancement. They showed that their NH subjects preferred speech in noise with moderate enhancement in terms of quality and intelligibility. This technique, combined with phonemic compression, improved the perception of vowels in hearing-impaired listeners but degraded the understanding of consonants (Franck *et al.*, 1999). To provide noise suppression, Clarkson and Bahgat (1991) expanded the temporal envelopes in different frequency bands. They found a small but significant improvement at 0 dB SNR in NH listeners. Lyzenga *et al.* (2002) employed a similar enhancement technique but followed it by an additional "lift" stage to counteract the effect of upward spread of masking. They applied spectral smearing in the end to simulate loss of frequency selectivity and tested NH listeners. They found that enhancement employed separately did not produce any improvement in SRT but enhancement and lift applied together improved the SRT by approximately 1 dB.

To counterbalance the degraded neural representations of the second ($F2$) and third ($F3$) formant frequencies in the impaired ear (Miller *et al.*, 1999a), Miller *et al.* (1999b) proposed a contrast enhancing frequency shaping algorithm that selectively amplifies $F2$ and $F3$ without modifying the spectral valleys. They found considerable improvements in the neural representation of $F2$ in acoustically traumatized cats.

Recently, Turicchia and Sarpeshkar (2005) have proposed a novel spectral enhancement scheme, companding, which combines two-tone suppression and dynamic gain

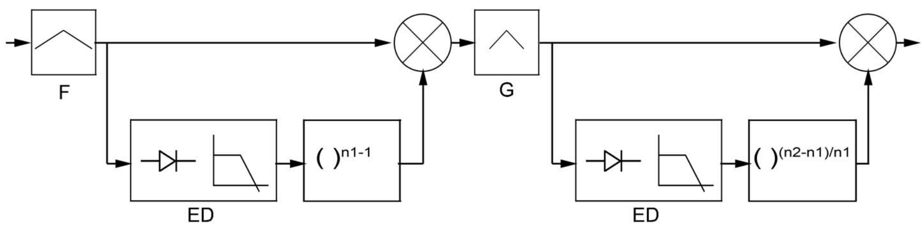

FIG. 1. Detailed architecture of a single channel. ED-envelope detector.

control in order to increase the spectral contrast. One specific goal in their study is to use the scheme to improve speech recognition in noise in cochlear implants. Studies have shown that companding is also present along the auditory pathway. Both cochlea and the cochlear nucleus perform logarithmic compression on the input signals, while the brain performs exponential expansion (Zeng and Shannon, 1994; Zeng and Shannon, 1999).

The overall companding architecture can be found in Turicchia and Sarpeshkar, 2005, and is briefly described here. First, the incoming signal is divided into a number of frequency channels by a bank of relatively broad bandpass filters F. Figure 1 shows the detailed architecture of a single channel companding pathway. The signal within each channel is subjected to amplitude compression. The extent of compression depends on the output of the envelope detector, ED, and the compression index, n_1. The compressed signal is then passed through a relatively narrow bandpass filter G before being expanded. The gain of the expansion block depends on the corresponding ED output and the ratio n_2/n_1. The outputs from all the channels are summed to obtain the processed signal.

The present work implemented and evaluated the companding architecture as a front end for a CI processor. This paper has the following organization. Section II describes implementation of the companding architecture and discusses the rationale behind choosing different parameter values. Section III shows the effect of companding on the acoustic features in both time and frequency domains. Section IV describes the evaluation results of companding in vowel, consonant, and sentence recognition tasks performed by both NH and CI subjects. Section V discusses the importance and possible mechanisms underlying the present findings. Section VI summarizes the present findings and points out future research directions.

II. IMPLEMENTATION

The companding architecture was implemented in MATLAB (The MathWorks, Natick, MA, USA). The precompression filter F and the postcompression filter G are zero-phase, bandpass filters with magnitudes described by the following transfer functions:

$$F'(s) = \left[\frac{2\left(\dfrac{\tau}{q_1}\right)s}{\tau^2 s^2 + 2\left(\dfrac{\tau}{q_1}\right)s + 1} \right]^2$$

$$G'(s) = \left[\frac{2\left(\dfrac{\tau}{q_2}\right)s}{\tau^2 s^2 + 2\left(\dfrac{\tau}{q_2}\right)s + 1} \right]^2.$$

Turicchia and Sarpeshkar used zero-phase filters for the companding-off cases to avoid interference between the channels. However, they did not use zero-phase filters for the companding-on cases. The time constant is given by $\tau = 1/2\pi f_r$, where f_r is the resonant frequency of the filters in the channels. The envelope detector in each channel consists of an ideal full wave rectifier and a first order low-pass filter. The time constant of the low-pass filter was set as $\tau_{\mathrm{ED}} = w\tau$. Turicchia and Sarpeshkar discussed the importance of the relative tuning between pre- and postcompression filters as well as the compression index in the companding strategy.

Here, we found that the degree of spectral enhancement depends on the number of channels. Figure 2 compares the frequency spectra of the steady-state portions of the original vowel /hid/ (lighter trace) and the same vowel processed with companding strategy (darker trace) as a function of the number of channels. The resonant frequencies for each channel were logarithmically spaced between 100 and 8000 Hz. We chose $q_1=2$, $q_2=12$, $w=40$, $n_2=1$ and $n_1=0.3$. The first formant frequency is at 400 ± 50 Hz, the second formant is at 1800 ± 100 Hz and the third formant is at 2570 ± 140 Hz. We compared the peak-to-valley differences of the power spectra with peak corresponding to the first formant and valley corresponding to the dip between the first and second formant frequencies. We found that companding always enhances the highest peak (typically the first formant), but reducing the number of channels to less than 40 produces an undesirable suppression of the relatively weaker second and third formants. On the other hand, increasing the number of channels beyond 50 does not further enhance spectral contrast. Based on these results, the number of channels was chosen to be 50 for all the experiments.

In addition, we found that time constant of the envelope detectors affects the level of spectral enhancement. Figure 3 shows the outputs of a 50 channel companding processing of the vowel /hid/ as a function of the time constant. The lighter traces represent the unprocessed vowel and the darker traces represent the processed vowel. We see that the amount of spectral enhancement increases as w increases from 5 to 30, and plateaus thereafter. For all the experiments from here on, w was chosen as 40. Moreover, values of q_1, q_2, n_1 and n_2 remained the same as in the previous analysis (Fig. 2) for all

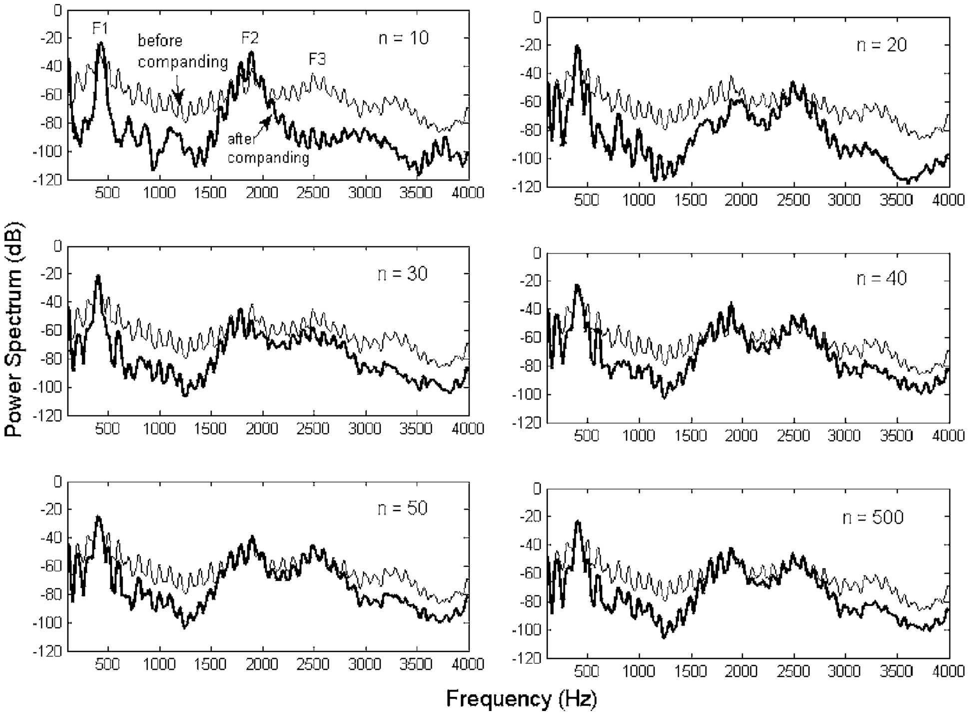

FIG. 2. Spectra of the vowel /hid/ as a function of the number of channels (n). Lighter traces correspond to the original stimuli and the darker traces represent the stimuli after companding.

the experiments reported in this work. All values were chosen based on the simulation results giving the maximum spectral contrast without degrading the signal quality.

Acoustic simulation of the eight-channel cochlear implant consisted of eight fourth-order bandpass Butterworth filters with frequencies between 100 and 5000 Hz (Shannon *et al.*, 1995; Dorman *et al.*, 1997a). Frequency spacing followed the Greenwood model, emulating equal spacing on the basilar membrane. The envelope of the signal in each band was extracted by full-wave rectification and low-pass filtering (eighth-order Butterworth) with a 160 Hz cutoff frequency. The envelope of each band was used to amplitude modulate a sinusoid at the center frequency of the channel. The modulated signals from all the channels were summed to form the acoustic simulation of an eight-channel cochlear implant.

III. ACOUSTIC ANALYSIS

A. Vowels

Figure 4 shows the temporal wave forms of the vowel /hid/ before (panel a) and after (panel b) companding, respectively. It is somewhat surprising to observe that companding also enhances the temporal contrast by sharpening the onset.

Figure 5 shows the spectra of the vowel /hid/ in steady-state speech-shaped noise (SSN) as a function of SNR. The lighter traces represent the inputs (original vowel in noise without companding), whereas the darker traces represent the corresponding outputs after companding. As the SNR decreases, the formant peaks are increasingly lost. We see that

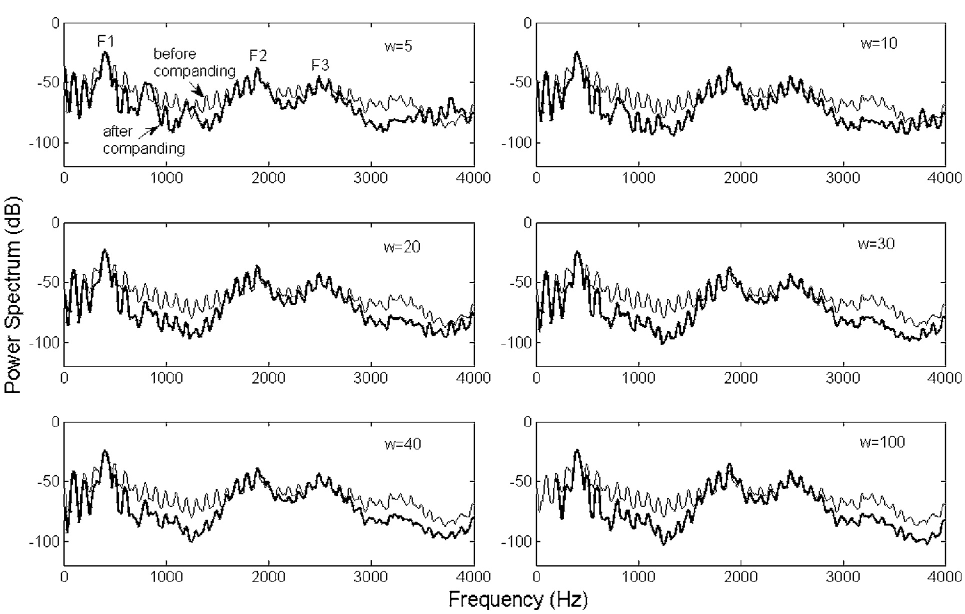

FIG. 3. Spectra of the vowel /hid/ as a function of time constant of the envelope detectors ($w\tau$). Lighter traces correspond to the original stimuli and the darker traces represent the stimuli after companding.

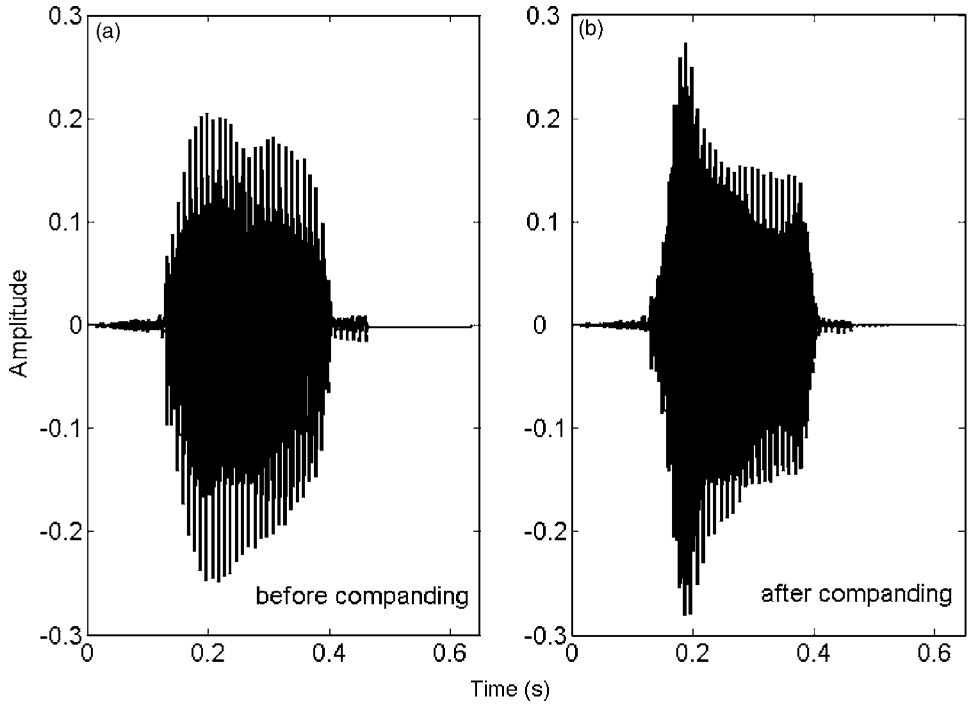

FIG. 4. Temporal wave forms of the vowel /hid/ before and after companding.

companding reduces the background noise while preserving the formant peaks, even at −10 dB SNR, when the decibel difference between the adjacent peaks and valleys is not apparent in the original stimulus.

B. Consonants

Figure 6 shows the spectra and the temporal wave forms of the consonant /aFa/. Panel a shows the spectra of the initial vowel part of /aFa/ and panel b shows the spectra of the consonant part following the initial vowel part. The lighter traces correspond to the input (i.e., the original consonant) and the darker traces correspond to the data after companding. We see that the formant peaks are enhanced during the initial vowel part but the spectral sharpening during the consonant part is relatively weak. This result is understandable, because, unlike vowels, consonants generally have flat spectra and lack prominent spectral peaks. Panels c and d show the temporal wave forms of the same consonant before and after companding, respectively. Similar to the vowel results, companding enhances changes in the temporal wave form envelope of the consonants.

IV. SPEECH RECOGNITION EVALUATION

Speech recognition experiments were conducted in NH and CI subjects. The tests included recognition of vowels,

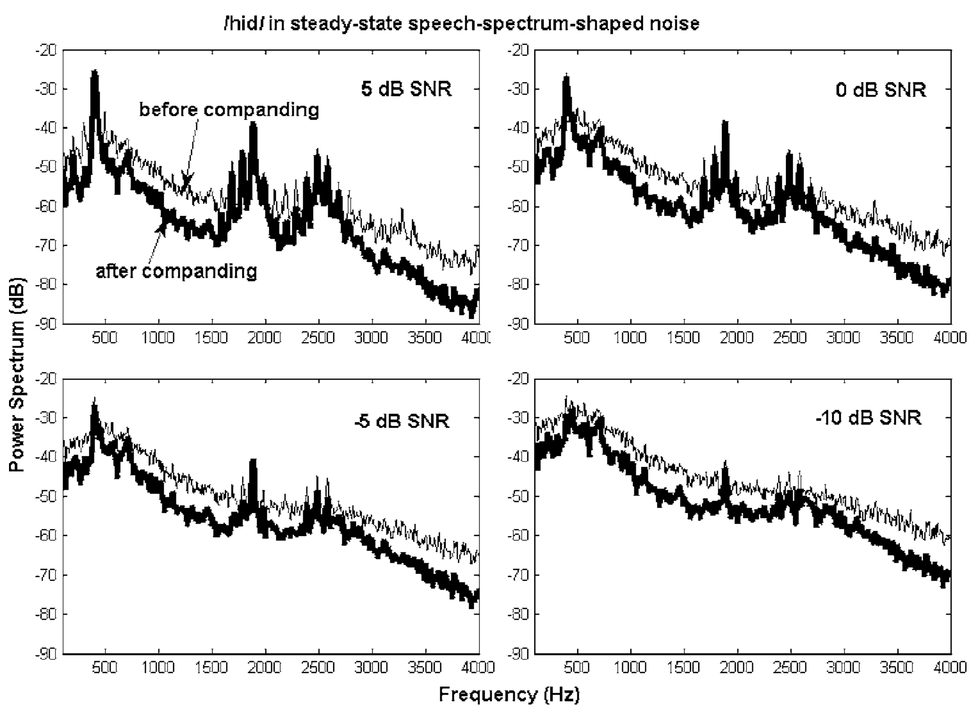

FIG. 5. Spectra of the vowel /hid/ in steady-state speech-spectrum-shaped noise at different SNRs. Lighter traces correspond to the original stimuli and the darker traces represent the stimuli after companding.

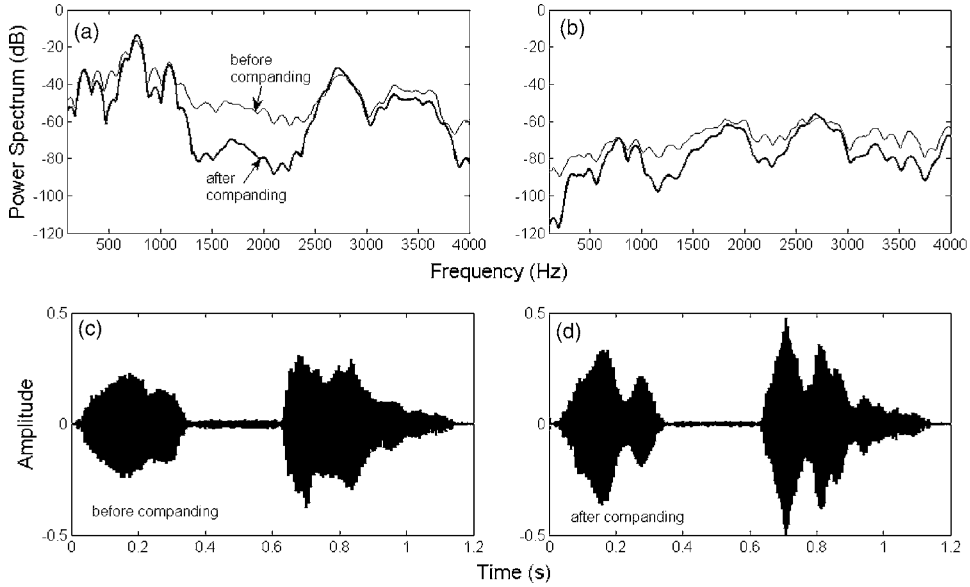

FIG. 6. (a) Spectra of the initial vowel part of the consonant /aFa/. (b) Spectra of the consonant part following the initial vowel part. Lighter traces correspond to the original stimuli and the darker traces represent the stimuli after companding. (c), (d) Temporal wave forms of the same consonant before and after companding, respectively.

consonants and sentences in quiet and in the presence of a steady-state SSN. NH subjects performed these tests using unprocessed and CI simulated stimuli, with and without companding in both cases. The CI subjects used their clinical processors to perform the same tests but with the unprocessed stimuli (with and without companding) only. The root mean square levels of all the stimuli were equalized.

A. Subjects

A total of seven NH subjects participated in the phoneme recognition tests and nine subjects in the sentence recognition tests. Out of these, six subjects participated in both phoneme and sentence recognition tests. The subjects were aged between 19 and 36 years with no reported history of hearing loss. Tests were also conducted on seven implant users between the ages of 56 and 79 years. The implant subjects included 5 Nucleus 24 (C1 to C5) and 2 Clarion II (C6 and C7) users. Subject C6 was pre-lingually deafened. The subjects had at least two years of implant experience at the time of testing. Detailed information on the CI users is presented in Table I. All subjects were native English speakers. They were compensated for their participation.

B. Test material

The phoneme materials included 12 /hvd/ vowels (Hillenbrand *et al.*, 1995) and 20 /aCa/ consonants (Shannon *et al.*, 1999) spoken by a male and a female speaker. The target sentence material consisted of 250 hearing in noise test (HINT) sentences spoken by a male speaker (Nilsson *et al.*, 1994). Both phonemes and the sentences were presented in quiet and in steady-state SSN at different SNRs between −10 dB and +10 dB, spaced at 5 dB intervals. The SSN was constructed by filtering white noise with the talker's long-term speech spectrum envelope derived using a tenth-order LPC analysis. The stimuli were presented via headphones (Sennheiser HDA 200) to the NH subjects while the CI subjects listened to the signal coming out of a speaker (Grason-Stadler 61 Clinical Audiometer). The signals were presented at 70 dB sound pressure level. The noise level was varied to produce different SNRs.

C. Procedure

All experiments were conducted in a double-walled, sound-attenuated booth. For the phoneme recognition tests, a graphical user interface containing 12 vowels or 20 conso-

TABLE I. Detailed information of the cochlear implant users.

Subject	Gender	Age (yrs)	Cause of deafness	Duration of implant use (yrs)	Clinical speech strategy	Vowel	Consonant	HINT
C1	F	77	Blood clot	2	ACE	42%	76%	84%
C2	F	71	Fever	4	ACE	74%	69%	98%
C3	F	62	Unknown	2	ACE	82%	91%	100%
C4	F	70	Virus	7	ACE	64%	84%	94%
C5	M	79	Unknown	2	CIS	58%	49%	92%
C6	F	56	Unknown	4	Clarion CII	21%	25%	15%
C7	F	52	Overnight SNHL	3	Clarion CII	58%	85%	100%

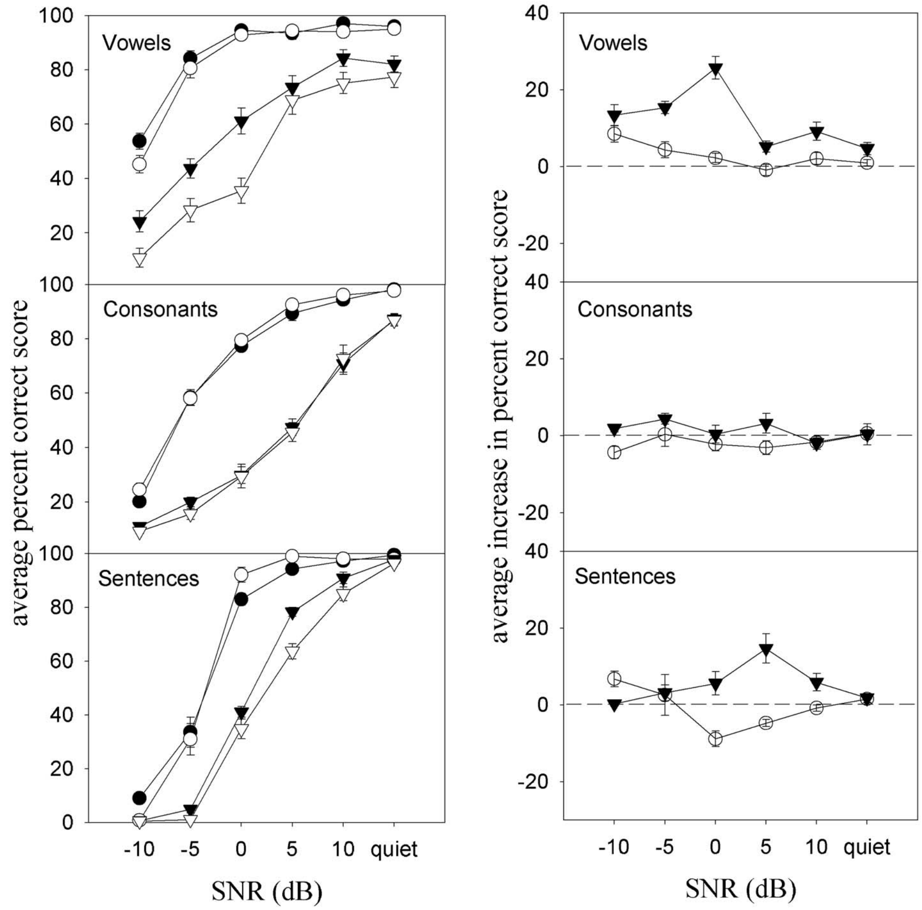

FIG. 7. Phoneme and sentence recognition scores for the NH subjects. Left panels show average percent correct scores and the right panels show average increase in the percent correct scores. Left panel: Circled traces represent the stimuli without CI simulation, whereas the inverted triangular traces correspond to the CI simulated stimuli. Open symbols correspond to the original stimuli and the filled symbols correspond to the stimuli after companding. Right panel: Filled inverted triangular traces correspond to the CI simulated stimuli and the open circular traces correspond to the stimuli without CI simulation. Vertical bars represent the standard errors.

nants displayed as buttons was presented on the computer screen. After a phoneme was presented, the subjects were instructed to click on the button corresponding to the presented phoneme. The subjects were provided with feedback regarding the response after each phoneme presentation and were instructed to guess if they were not sure. All the NH subjects had to take a pretest consisting of unprocessed stimuli in quiet and only those subjects who scored above 90% correct were allowed to participate in the study. The subjects did not receive any training prior to the tests. The noise conditions were presented in the order of increasing level of difficulty, to counterbalance any learning effect. The order of processing conditions for each noise condition was randomized and balanced across subjects. The phonemes were presented randomly for each condition.

In the sentence recognition tests, the subjects were presented with a target sentence. They were asked to type in as many words as possible from the sentence using a computer keyboard. The number of correctly identified words was calculated to give the final percent correct score. No feedback was provided during the test and the subjects were instructed to guess if they were not sure. Similar to the phoneme recognition tests, all NH subjects took a pretest consisting of unprocessed stimuli in quiet and only those subjects who scored above 85% correct were allowed to participate in the study. The noise conditions were presented in the order of increasing level of difficulty, to counterbalance any learning effect. The processed and unprocessed stimuli were presented randomly in each test condition. The sentences were not repeated.

D. Results

Figure 7 shows the phoneme recognition and sentences recognition scores as a function of SNR for the NH subjects. The left panels show the average percent correct score and the right panels show the average increase in percent correct score as a function of SNR. In the left panels, open symbols correspond to stimuli before companding and the filled symbols correspond to stimuli after companding. The circled traces correspond to stimuli without eight-channel CI simulation and the inverted triangular traces correspond to stimuli with eight-channel CI simulation. In the right panels, the open circular traces correspond to stimuli without CI simulation and the filled inverted triangular traces correspond to stimuli with CI simulation.

For the vowel recognition test, analysis of variance (ANOVA) showed that companding improved the performance for both stimuli with $[F(1,6)=98.10, p<0.001]$ and without $[F(1,6)=62.07, p<0.01]$ CI simulation. Increase in percent correct scores varied nonmonotonically with larger improvements seen in the case of stimuli with CI simulation. At -10 dB SNR, difference in the average increases in percent correct scores between stimuli with CI simulation and stimuli without CI simulation were 5%, which increased to 25% at 0 dB SNR and thereafter decreased to 4% in quiet.

For the consonant recognition test, companding did not produce any significant change in the performance for stimuli without CI simulation $[F(1,6)=1.99, p=0.22]$. No significant improvement was seen in the case of consonants with CI simulation either $[F(1,6)=1.51, p=0.27]$. Unlike the

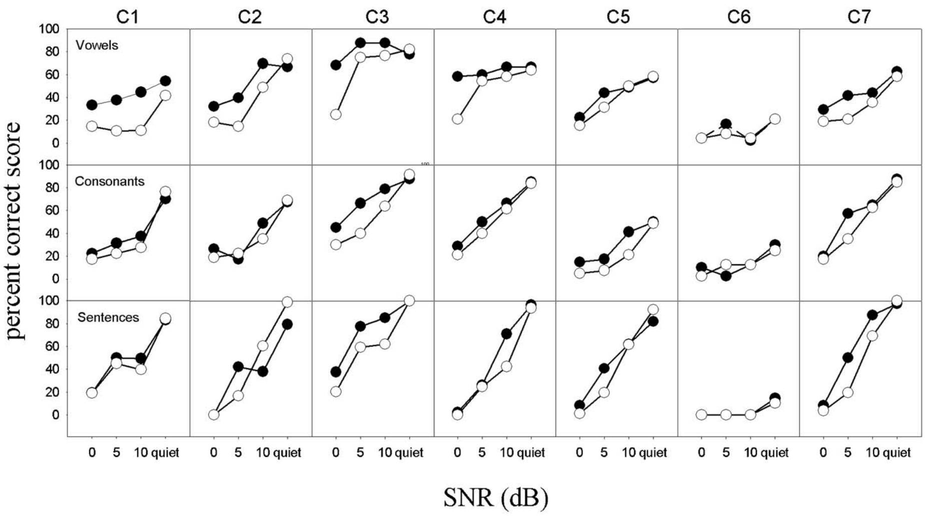

FIG. 8. Phoneme and sentence recognition scores for the CI users. C1 to C5 - Nucleus 24 users, C6 and C7 - Clarion II users. Upper, middle and the bottom panels show scores on vowel, consonant and sentence recognition tests, respectively. Open circles correspond to the original stimuli and the filled circles correspond to stimuli after companding.

vowel recognition results, there was no significant difference in the performances between consonants with CI simulation and the consonants without CI simulation.

For the sentence recognition test, companding seemed to degrade the performance in the case of stimuli without CI simulation at 0, 5, and 10 dB SNRs $[F(1,8)=44.38, p<0.001]$. However, subjects benefited from companding for the stimuli with CI simulation $[F(1,8)=25.53, p<0.05]$. Similar to the vowel recognition results, the improvement varied nonmonotonically with SNR. Larger improvement was seen in the case of stimuli with CI simulation at 0, 5, and 10 dB SNRs ($p<0.05$). At −10 dB SNR, improvement in the case of sentences without CI processing was better than the sentences with CI simulation ($p<0.05$), whereas there was no significant difference in the improvements at −5 dB SNR and in quiet. In general, the performance did not vary significantly with companding in quiet.

Figure 8 shows phoneme recognition and sentence recognition scores as a function of SNR for the CI users. The upper, middle and the bottom panels show the scores on vowel, consonant and sentence recognition tests. Open circles correspond to the original stimuli and the filled circles correspond to stimuli after companding. The improvement in performance for the vowel recognition tests was found to be as high as 43% (C3, 0 dB SNR). Similarly the maximum increases in percent scores were 26% (C3, 5 dB SNR) and 30% (C3, 5 dB SNR) for consonant recognition and sentence recognition tasks, respectively.

Figure 9 shows the average performance of CI users as a function of SNR. For the left panels, open symbols represent stimuli without companding and the filled symbols correspond to stimuli with companding. Because subject C6 showed a floor effect without any notable improvement with companding, her data were excluded from the above average analysis and also from the following statistical analysis.

For the vowel recognition test, subjects performed better with companding $[F(1,5)=35.50, p<0.005]$ at all SNRs except in quiet $[F(1,5)=0.18, p=0.69]$. The maximum average improvement was 21.3% ($p<0.05$) at 0 dB SNR. Similarly, consonant recognition in noise improved with companding $[F(1,5)=26.94, p<0.005]$ with a maximum average improvement of 12.1% ($p<0.05$) at 5 dB SNR. Companding also produced better performance for sentence recognition in noise $[F(1,5)=11.50, p<0.05]$, with a maximum average improvement of 17.7% ($p<0.05$) at 5 dB SNR. No significant improvement was seen in quiet ($p=0.36$). In general, the cochlear implant users benefited from companding in noise conditions.

V. DISCUSSION

A. Comparison with previous studies

Turicchia and Sarpeshkar showed that spectral contrast is an emergent property of the companding strategy and had speculated that the strategy has the potential to improve speech performance in noise. Loizou (2005) implemented a 16 channel companding strategy in actual CI users and found a modest improvement in vowel recognition but no improvement in consonant and sentence recognition. Here we found that a 50 channel companding implementation significantly improved the recognition of both phonemes and sentences in noise. Based on the acoustic analysis presented in Sec. III, we attribute the observed improvement to optimization of the companding parameters in the present implementation (discussed in the next section).

In another study, also investigating the effects of companding, Oxenham *et al.* (2007) tested NH listeners using sentences processed through a noise-excited envelope vocoder. Speech intelligibility was measured in steady-state SSN at 0, 3 and 6 dB SNRs. They varied the number of channels, channel bandwidths and time constants of the envelope detectors. They showed that a 50 channel companding implementation, using parameters similar to that used in the present study, improved the performance by 6 percentage points, averaged across subjects and SNRs. We found an average improvement of 15 and 6 percentage points at 5 and 0 dB SNRs, respectively, which is consistent with their finding. Further, they showed that by reducing the number of analysis channels to 16, the performance was improved by 4 percentage points (2 percentage points less than 50 analysis channels). Thus the degree of benefit from companding drops with reducing the number of channels. They also

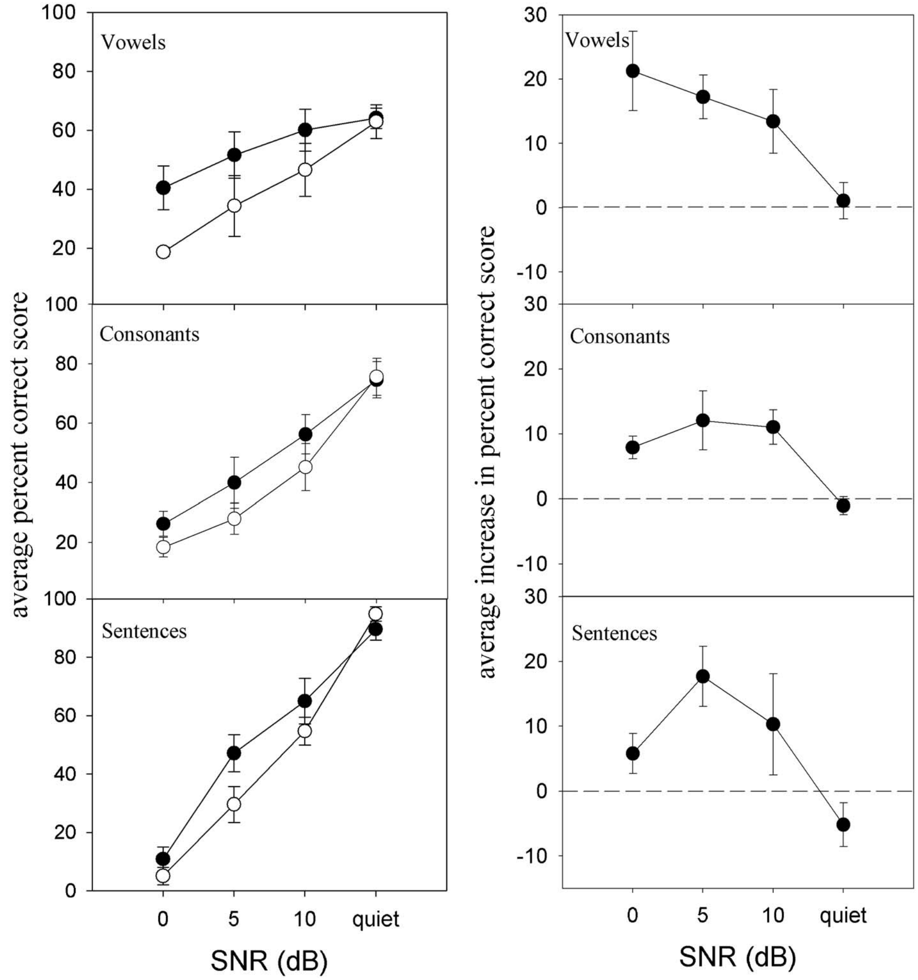

FIG. 9. Average percent correct scores and the average increase in percent correct scores for the CI subjects. Left panel: Open symbols correspond to the original stimuli and the filled symbols correspond to the stimuli after companding. Vertical bars represent the standard errors.

showed that decreasing the sharpness of tuning of the postcompression filter (G) by a factor between 2 and 3 did not decrease the improvement in intelligibility. Finally, as we had predicted from acoustic analysis, they showed that companding with smaller envelope detector time constants produced no benefit.

B. Optimization of the companding parameters

We identified four key parameters that need to be optimized to achieve a desirable spectral enhancement and speech performance in noise. These parameters are (1) number of channels, (2) the relative tuning between the pre- and postcompression filters (F and G), (3) the compression index (n_1), and (4) the time constant of the envelope detector ($w\tau$).

First, the effectiveness of companding depends upon the number of channels. Increasing the number of channels increases the processing time. Too few channels will result in the suppression of local spectral peaks, namely the weaker higher formants (Fig. 2). An optimal number of channels should achieve maximal spectral enhancement with sufficient frequency representation while minimizing the suppression of useful peaks. We found the optimal number of channels to be about 50 for the values tested.

In addition, the relative tuning of the two filters creates a difference in the levels of compression and expansion. A tone at the resonant frequency f_r of a channel is suppressed when the narrow filter G blocks a stronger tone entering the same channel. The tuning of F determines the range of frequencies above and below f_r of the channel that can suppress f_r. The tuning of G determines the range of frequencies around f_r that will not suppress f_r (Turicchia and Sarpeshkar, 2005). Hence the relative tuning of the two filters along with the number of channels determines the localness of spectral enhancement. We used a $q_1=2$ for the pre-compression filter F and $q_2=12$ for the postcompression filter G.

Furthermore, the degree of spectral enhancement depends on the value of the compression index, n_1. The smaller the value of n_1, the greater the difference between the degree of compression and the degree of expansion, resulting in greater enhancement of spectral contrast. We chose n_1 to be 0.3 and n_2 to be 1 in the present implementation.

Finally, the time constant of the envelope detectors can also affect the performance of the strategy. A smaller value of w (higher envelope cutoff frequency) generates a large number of unwanted frequencies within the channel thereby reducing the spectral contrast. We chose w to be 40 in the present implementation.

C. Factors affecting companding performance

The present results show that the effectiveness of companding depends on speech materials, listeners, and listening conditions. We found that companding enhances the spectral peaks, but mainly for vowels and not consonants. Since the companding strategy in this study has higher channel density at lower frequencies than at higher frequencies, and vowels generally contain stronger spectral peaks at lower frequencies than consonants, greater spectral enhancement is produced for vowels compared to consonants. As a result, more improvement was seen in vowel perception scores for both NH listeners and CI users. This interpretation is consistent with previous studies showing that spectral smearing affected vowel perception more than consonant perception (Boothroyd *et al.*, 1996).

NH subjects listening to CI simulations showed better performance than CI users for both phoneme and sentence recognition tests. The results suggest that implant users are unable to fully utilize the temporal and the spectral cues. Potential reasons behind this limitation are electrode interaction (Fu *et al.*, 1998), mismatch between acoustic frequency and electric pitch (Dorman *et al.*, 1997b) and the patient's surviving neural population. The present eight-channel CI simulation does not simulate these limitations.

Furthermore, the degree of improvement varied greatly between individual CI users. It appears that subjects with better performance in quiet showed larger improvements (e.g., C3 vs. C6). As a matter of fact, subject C6 did not benefit from companding at all. This subject has been deaf for 50 years, with a severe degree of hearing loss compared to the other subjects. The SNR corresponding to maximum improvement was subject dependent (see Fig. 8). Most implant users did not show any improvement in quiet, likely due to the ceiling effect.

Companding improved vowel recognition in both NH listeners and CI users but improved consonant recognition only in CI users. One particular finding of this study was that apart from improving the spectral contrast, companding also enhances the temporal contrast. This along with the fact that CI users can detect smaller modulation amplitudes than the NH listeners (Shannon, 1992), suggests that the improvement was a result of better detection of the enhanced temporal contrast in consonants. Sequential information analysis (Wang and Bilger, 1973) was performed on the pooled confusion matrix obtained from four NH listeners tested using stimuli with CI simulation. Unfortunately, the confusions matrices from CI users and the rest of NH subjects were not saved. Results indicated that transmission of information about plosive and nasal was higher for stimuli with companding compared to stimuli without companding. Similar effects were observed with the temporal enhancement technique known as transient emphasis spectral maxima (Vandali, 2001) which amplifies the low-intensity short-duration rapid changes in temporal envelope. It is also likely that the implant subjects were able to use the temporal envelope information more efficiently when the background noise was suppressed with companding. The effect of enhanced temporal contrast needs to be investigated by future studies.

D. Applications

The present study showed that companding improved CI speech performance in steady-state noise. It remains to be seen whether companding will work in realistic conditions including multiple-talker babbles as noise. Because of the dynamic changes in both signal and noise in these conditions, it is unlikely that companding will improve the actual SNR to the same degree as in the steady-state noise condition. However, the enhanced spectral and temporal contrasts between signal and the noise (another talker or masker) may help in perceptual segregation of the signal from noise.

Companding can be incorporated into cochlear implants by using either the present front-end approach or an integrated signal processing strategy. In order for it to act as the latter, simultaneous control of compression, noise suppression and channel gain should be feasible. The processing architecture and parameter optimization can be significantly different depending on the approach taken. Different designs may have to be considered in real-time implementations of companding with these different approaches.

Finally, companding may be used to enhance hearing aid performance. It has been shown that multi-channel compression in hearing aids causes spectral smearing (Plomp, 1988). Spectral enhancement using companding has the potential to remedy this problem. Future studies are needed to examine the effectiveness of companding in realistic listening conditions and its utility in assistive listening devices.

VI. CONCLUSIONS

This study optimized implementation of the companding strategy and evaluated its effectiveness in enhancing speech performance in noise for CI users. We found that choice of parameters was critical to produce optimal spectral enhancement and adequate noise suppression. We also found that companding enhanced the temporal contrast. Once the parameters were optimized, the implemented strategy significantly improved the CI speech perception in noise. Future studies are needed to evaluate the relative contribution of spectral and temporal contrast enhancement, the effectiveness of companding under realistic listening conditions (e.g., multiple talkers), and implementation and optimization of companding in cochlear implants and hearing aids.

ACKNOWLEDGMENTS

The authors would like to thank all the subjects who participated in the experiments. This work was supported by NIH (2RO1 DC002267).

Baer, T., Moore, B. C., and Gatehouse, S. (**1993**). "Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment: Effects on intelligibility, quality, and response times," J. Rehabil. Res. Dev. **30**, 49–72.

Blackburn, C. C., and Sachs, M. B. (**1990**). "The representation of the steady-state vowel /eh/ in the discharge patterns of cat anteroventral cochlear nucleus neurons," J. Neurophysiol. **63**, 1191–1212.

Boothroyd, A., Mulhearn, B., Gong, J., and Ostroff, J. (**1996**). "Effects of spectral smearing on phoneme and word recognition," J. Acoust. Soc. Am. **100**, 1807–1818.

Bunnell, H. T. (**1990**). "On enhancement of spectral contrast in speech for hearing-impaired listeners," J. Acoust. Soc. Am. **88**, 2546–2556.

Clarkson, P., and Bahgat, Sayed F. (**1991**). "Envelope expansion methods for speech enhancement," J. Acoust. Soc. Am. **89**, 1378–1382.

Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997a**). "Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs," J. Acoust. Soc. Am. **102**,

2403–2411.
Dorman, M. F., Loizou, P. C., and Rainey, D. (**1997b**). “Simulating the effect of cochlear-implant electrode insertion depth on speech understanding,” J. Acoust. Soc. Am. **102**, 2993–2996.
Franck, B. A. M., van Kreveld-Bos, C. S. G. M., Dreschler, W. A., and Verschuure, H. (**1999**). “Evaluation of spectral enhancement in hearing aids, combined with phonemic compression,” J. Acoust. Soc. Am. **106**, 1452–1468.
Fu, Q. J., and Nogaki, G. (**2005**). “Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing,” Trans.- Geotherm. Resour. Counc. **6**, 19–27.
Fu, Q. J., Shannon, R. V., and Wang, X. (**1998**). “Effects of noise and spectral resolution on vowel and consonant recognition: Acoustic and electric hearing,” J. Acoust. Soc. Am. **104**, 3586–3596.
Hamacher, V., Doering, W. H., Mauer, G., Fleischmann, H., and Hennecke, J. (**1997**). “Evaluation of noise reduction systems for cochlear implant users in different acoustic environment,” Am. J. Otol. **18**, S46–49.
Hartline, H. K. (**1969**). “Visual receptors and retinal interaction,” Science **164**, 270–278.
Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). “Acoustic characteristics of American English vowels,” J. Acoust. Soc. Am. **97**, 3099–3111.
Hochberg, I., Boothroyd, A., Weiss, M., and Hellman, S. (**1992**). “Effects of noise and noise suppression on speech perception by cochlear implant users,” Ear Hear. **13**, 263–271.
Hochmair-Desoyer, I. J., Hochmair, E. S., Burian, K., and Fischer, R. E. (**1981**). “Four years of experience with cochlear prosthesis,” Med. Prog. Technol. **8**, 107–119.
Horst, J. W. (**1987**). “Frequency discrimination of complex signals, frequency selectivity, and speech perception in hearing-impaired subjects,” J. Acoust. Soc. Am. **82**, 874–885.
Kiang, N. Y., Liberman, M. C., and Levine, R. A. (**1976**). “Auditory-nerve activity in cats exposed to ototoxic drugs and high-intensity sounds,” Ann. Otol. Rhinol. Laryngol. **85**, 752–768.
Loizou, P. (**2005**). “Evaluation of the companding and other strategies for noise reduction in cochlear implants,” *2005 Conference on Implantable Auditory Prosthesis*, Asilomar, Monterey, California.
Loizou, P. C., Dorman, M., and Fitzke, J. (**2000**). “The effect of reduced dynamic range on speech understanding: Implications for patients with cochlear implants,” Ear Hear. **21**, 25–31.
Loizou, P. C., and Poroy, O. (**2001**). “Minimum spectral contrast needed for vowel identification by normal hearing and cochlear implant listeners,” J. Acoust. Soc. Am. **110**, 1619–1627.
Lyzenga, J., Festen, J. M., and Houtgast, T. (**2002**). “A speech enhancement scheme incorporating spectral expansion evaluated with simulated loss of frequency selectivity,” J. Acoust. Soc. Am. **112**, 1145–1157.
Miller, R. L., Calhoun, B. M., and Young, E. D. (**1999a**). “Discriminability of vowel representations in cat auditory-nerve fibers after acoustic trauma,” J. Acoust. Soc. Am. **105**, 311–325.
Miller, R. L., Calhoun, B. M., and Young, E. D. (**1999b**). “Contrast enhancement improves the representation of /ε/ like vowels in the hearing-impaired auditory nerve,” J. Acoust. Soc. Am. **106**, 2693–2708.
Miller, R. L., Schilling, J. R., Franck, K. R., and Young, E. D. (**1997**). “Effects of acoustic trauma on the representation of the vowel “eh” in cat auditory nerve fibers,” J. Acoust. Soc. Am. **101**, 3602–3616.
Müller-Deiler, J., Schmidt, B. J., and Rudert, H. (**1995**). “Effects of noise on speech descrimination in cochlear implant patients,” Ann. Otol. Rhinol. Laryngol. **166**, 303–306.
Nelson, D. A., Schmitz, J. L., Donaldson, G. S., Viemeister, N. F., and Javel, E. (**1996**). “Intensity discrimination as a function of stimulus level with electric stimulation,” J. Acoust. Soc. Am. **100**, 2393–2414.
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.
Oxenham, A. J., Simonson, A. M., Turicchia, L., and Sarpeshkar, R. (**2007**). “Evaluation of companding-based spectral enhancement using simulated cochlear-implant processing,” J. Acoust. Soc. Am. **121**, 1709–1716.
Plomp, R. (**1988**). “The negative effect of amplitude compression in multichannel hearing aids in the light of the modulation-transfer function,” J. Acoust. Soc. Am. **83**, 2322–2327.
Rhode, W. S. (**1974**). “Evidence from Mässbauer experiments for nonlinear vibrations in the cochlea,” J. Acoust. Soc. Am. **55**, 588–597.
Rhode, W. S., Geisler, C. D., and Kennedy, D. T. (**1978**). “Auditory nerve fiber responses to wide-band noise and tone combinations,” J. Neurophysiol. **41**, 692–704.
Rhode, W. S., and Greenberg, S. (**1994**). “Lateral suppression and inhibition in the cochlear nucleus of the cat,” J. Neurophysiol. **71**, 493–514.
Ruggero, M. A., Robles, L., and Rich, N. C. (**1992**). “Two-tone suppression in the basilar membrane of the cochlea: Mechanical basis of auditory-nerve rate suppression,” J. Neurophysiol. **68**, 1087–1099.
Sachs, M. B., and Kiang, N. Y. (**1968**). “Two-tone inhibition in auditory-nerve fibers,” J. Acoust. Soc. Am. **43**, 1120–1128.
Sachs, M. B., Voigt, H. F., and Young, E. D. (**1983**). “Auditory nerve representation of vowels in background noise,” J. Neurophysiol. **50**, 27–45.
Shannon, R. V. (**1992**). “Temporal modulation transfer functions in patients with cochlear implants,” J. Acoust. Soc. Am. **91**, 2156–2164.
Shannon, R. V., Jensvold, A., Padilla, M., Robert, M. E., and Wang, X. (**1999**). “Consonant recordings for speech testing,” J. Acoust. Soc. Am. **106**, L71–L74.
Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). “Speech recognition with primarily temporal cues,” Science **270**, 303–304.
Stickney, G. S., Zeng, F. G., Litovsky, R., and Assmann, P. (**2004**). “Cochlear implant speech recognition with speech maskers,” J. Acoust. Soc. Am. **116**, 1081–1091.
Stoop, R., and Kern, A. (**2004**). “Essential auditory contrast-sharpening is preneuronal,” Proc. Natl. Acad. Sci. U.S.A. **101**, 9179–9181.
Toledo, F., Loizou, P., and Lobo, A. (**2003**). “Subspace and envelope subtraction algorithms for noise reduction in cochlear implants,” *Proceedings of 25th Annual International Conference of IEEE-EMBC*, Cancun, Mexico, pp. 2002–2005.
Turicchia, L., and Sarpeshkar, R. (**2005**). “A bio-inspired companding strategy for spectral enhancement,” IEEE Trans. Acoust., Speech, Signal Process. **13**, 243–253.
Vandali, A. E. (**2001**). “Emphasis of short-duration acoustic speech cues for cochlear implant users,” J. Acoust. Soc. Am. **109**, 2049–2061.
van Hoesel, R. J., and Clark, G. M. (**1995**). “Evaluation of a portable two-microphone adaptive beamformer speech processor with cochlear implants patients,” J. Acoust. Soc. Am. **97**, 2498–2503.
Wang, M. D., and Bilger, R. C. (**1973**). “Consonant confusions in noise: A study of perceptual features,” J. Acoust. Soc. Am. **54**, 1248–1266.
Weiss, M. R. (**1993**). “Effects of noise and noise reduction processing on the operation of the Nucleus-22 cochlear implant processor,” J. Rehabil. Res. Dev. **30**, 117–128.
Wouters, J., and Vanden Berghe, J. (**2001**). “Speech recognition in noise for cochlear implantees with a two-microphone monaural adaptive noise reduction system,” Ear Hear. **22**, 420–430.
Zeng, F. G., and Galvin, J. J. (**1999**). “Amplitude mapping and phoneme recognition in cochlear implant listeners,” Ear Hear. **20**, 60–74.
Zeng, F. G., Nie, K., Stickney, G. S., Kong, Y. Y., Vongphoe, M., Bhargave, A., Wei, C., and Cao, K. (**2005**). “Speech recognition with amplitude and frequency modulations,” Proc. Natl. Acad. Sci. U.S.A. **102**, 2293–2298.
Zeng, F. G., and Shannon, R. V. (**1994**). “Loudness-coding mechanisms inferred from electric stimulation of the human auditory system,” Science **264**, 564–566.
Zeng, F. G., and Shannon, R. V. (**1999**). “Psychophysical laws revealed by electric hearing,” NeuroReport **10**, 193l–1935.