Evaluation of companding-based spectral enhancement using simulated cochlear-implant processing

Andrew J. Oxenham,[a] Andrea M. Simonson, Lorenzo Turicchia, and Rahul Sarpeshkar
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 4 January 2006; revised 14 December 2006; accepted 21 December 2006)

This study tested a time-domain spectral enhancement algorithm that was recently proposed by Turicchia and Sarpeshkar [IEEE Trans. Speech Audio Proc. **13**, 243–253 (2005)]. The algorithm uses a filter bank, with each filter channel comprising broadly tuned amplitude compression, followed by more narrowly tuned expansion (companding). Normal-hearing listeners were tested in their ability to recognize sentences processed through a noise-excited envelope vocoder that simulates aspects of cochlear-implant processing. The sentences were presented in a steady background noise at signal-to-noise ratios of 0, 3, and 6 dB and were either passed directly through an envelope vocoder, or were first processed by the companding algorithm. Using an eight-channel envelope vocoder, companding produced small but significant improvements in speech reception. Parametric variations of the companding algorithm showed that the improvement in intelligibility was robust to changes in filter tuning, whereas decreases in the time constants resulted in a decrease in intelligibility. Companding continued to provide a benefit when the number of vocoder frequency channels was increased to sixteen. When integrated within a sixteen-channel cochlear-implant simulator, companding also led to significant improvements in sentence recognition. Thus, companding may represent a readily implementable way to provide some speech recognition benefits to current cochlear-implant users.
[DOI: 10.1121/1.2434757]

PACS number(s): 43.71.Ky, 43.66.Ts [DOS] Pages: 1709–1716

I. INTRODUCTION

In a recent paper, Turicchia and Sarpeshkar (2005) proposed a new strategy for time-domain spectral enhancement, based on relatively broadband compression followed by more frequency-selective expansion. This compressing-and-expanding (companding) approach leads to certain properties shared by the peripheral auditory system. In particular, it can produce a suppression of the response to one tone by the presence of another, more intense, tone at a nearby frequency—an effect known as two-tone suppression. At a more global level, the companding scheme can lead to the enhancement of spectral peaks in a stimulus, relative to nearby spectral valleys. Spectral enhancement techniques have often been used in an attempt to improve speech intelligibility in noise for listeners with hearing loss. Despite many different approaches and solutions (e.g., Bunnell, 1990; Simpson *et al.*, 1990; Stone and Moore, 1992; Baer *et al.*, 1993; Franck *et al.*, 1999; Lyzenga *et al.*, 2002; Yang *et al.*, 2003), the results so far, in terms of improved speech intelligibility, have been not been particularly promising (for a review, see Moore, 2003).

In contrast to earlier studies using spectral enhancement techniques, an initial evaluation of Turicchia and Sarpheshkar's (2005) algorithm has shown that it can provide some improvement in speech intelligibility for cochlear-implant users and for normal-hearing subjects listening to acoustic simulations of cochlear-implant processing (Bhattacharya and Zeng, 2007): In both groups, recognition of words in sentences increased on average by between 10 and 20 percentage points. These new findings are encouraging and stimulate a number of questions. In particular, it is not clear what properties of the new algorithm (Turicchia and Sarpeshkar, 2005) or the evaluation method (Bhattacharya and Zeng, 2007) allow it to provide some benefit, where others have shown little or no improvement. A number of possible explanations exist. One reason may be that, in contrast to most other schemes, this was tested with cochlear-implant users, rather than hearing-impaired listeners. Frequency selectivity is generally poorer in cochlear-implant users, who typically only have access to about eight broadly tuned channels of spectral information (Friesen *et al.*, 2001), than in listeners with moderate-to-severe hearing loss. It is thus possible that cochlear-implant users might benefit more than hearing-impaired listeners from spectral enhancement schemes. For instance, it has been shown that the benefits of enhanced spectral contrast can increase with decreasing initial spectral resolution (Loizou and Poroy, 2001).

Another potential reason is that the scheme is implemented in the time domain, using logarithmically scaled and spaced filters. This is closer to the processing carried out by the normal auditory system than the more typical linear-frequency FFT processing, in which the windowing can lead to temporal distortions that may be more audible than those produced by a logarithmically scaled system. One other potential reason is the number of analysis channels used. One

[a]Author to whom correspondence should be addressed. Current address: Department of Psychology, University of Minnesota, Minneapolis, MN 55455. Electronic mail: oxenham@umn.edu

other earlier study used a filterbank approach, which was implemented in real time with 16 contiguous frequency channels, where the outputs from adjacent channels were used to adjust the gain of each channel, thereby producing some enhancement of spectral peaks (Stone and Moore, 1992). The authors suggested that the lack of improvement they found might be due to the small number of frequency analysis bands (16) used in their implementation. In contrast, Bhattacharya and Zeng's (2007) implementation of the companding algorithm had 50 analysis channels.

The present study had three main aims. The first aim was to test the robustness of the improvement provided by the companding algorithm. This was done by using very similar parameters to those used by Bhattacharya and Zeng (2007). The second aim was to identify the parameters that were critical in providing the improvement in speech intelligibility. Specifically, changes to the companding time constants and filter bandwidths were made and their effects on performance were measured. The third aim was to test whether a "pared-down" version of the companding algorithm, which might be suitable for implementation in cochlear-implants in the near future, might also provide improved speech intelligibility. This aim was also able to test the effect of reducing the number of analysis channels from 50 to 16, in order to test whether this was a critical difference between the evaluation of the companding algorithm (Bhattacharya and Zeng, 2007) and the earlier filter-bank study of Stone and Moore (1992) in hearing-impaired listeners.

The aims were achieved by measuring speech intelligibility in normal-hearing listeners using a steady-state background noise, after the stimuli (speech and noise) had been processed through a noise-excited envelope vocoder, which is designed to simulate certain aspects of cochlear-implant processing, such as the poor frequency selectivity and the lack of temporal fine-structure information (e.g., Shannon *et al.*, 1995; Dorman *et al.*, 1998). The study is divided into five separate experiments, each involving eight different normal-hearing listeners. The first three experiments test a 50-channel version of the companding algorithm by passing its output through an 8-channel cochlear-implant simulator and varying the filter bandwidths and time constants of the algorithm. The next experiment tests the same 50-channel algorithm after passing it through a 16-channel cochlear-implant simulator. The final experiment presents a modified version of the algorithm that uses only 16 analysis channels, in a way that could be readily implemented with current implant technology.

II. GENERAL METHODS

A. Stimuli

Listeners were tested on their ability to correctly report words from the Hearing in Noise Test (HINT) sentence lists (Nilsson *et al.*, 1994), spoken by a male talker. The background noise was steady Gaussian noise that was spectrally shaped to approximate the long-term spectrum of speech, and was generated by filtering white noise with a filter composed of two poles at 200 Hz, two poles at 8 kHz, and a zero in the origin (see Fig. 1). The speech was presented at average signal-to-noise ratios of 0, 3, and 6 dB. The noise and speech were always added prior to any processing.

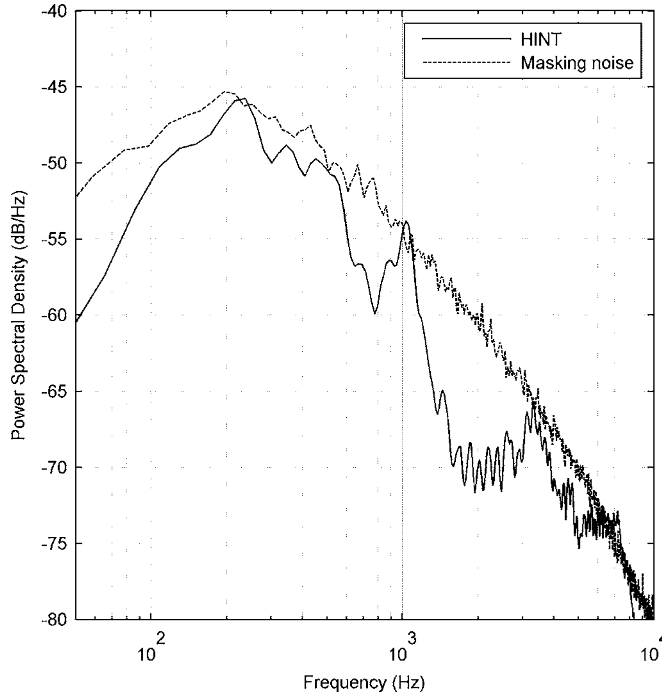

FIG. 1. Long-term power spectrum of the speech used in the HINT data base (solid line) and the steady-state masking noise (dashed line).

1. Companding processing

We implemented the companding algorithm presented in Turicchia and Sarpeshkar (2005). The strategy uses a non-coupled filter bank and compression-expansion blocks as shown in Fig. 2. Every channel in the companding architecture has a relatively broad prefilter, a compression block, a relatively narrowband postfilter, and an expansion block. The prefilter and postfilter in each channel have the same center frequency. The pre- and postfilter banks have logarithmically spaced center frequencies that span the desired spectral range. Finally, the channel outputs of this nonlinear filter bank are summed to generate an output with enhanced spectral peaks.

The broadband filter determines the frequency range that can affect the gain of the compressor in a given channel. The narrowband filter determines a narrower subset of these fre-

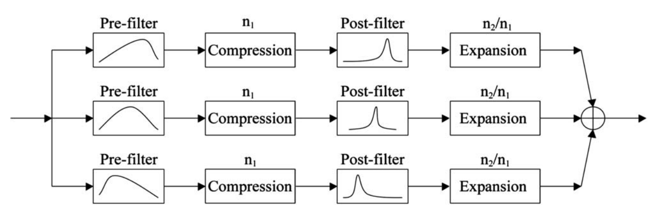

FIG. 2. Block diagram of the companding architecture, showing the stimulus analyzed in a bank of broad prefilters. The output of each prefilter is subjected to compression, is filtered again using sharper postfilters, and then expanded. The outputs from each channel are then summed to produce the processed broadband stimulus. For more details, see Turicchia and Sarpeshkar (2005).

quencies that can affect the gain of the expander. The compressor and expander are normally complementary, such that a tone at the center frequency of the filters will undergo sequential compression and expansion and emerge unchanged in level. However, sufficiently intense frequencies outside the narrowband filter passband but within the broadband filter passband can determine the gain of the compressor only to be filtered out by the narrowband filter. This can result in the suppression of weak frequency components at or near the filters' center frequency by stronger more remote frequency components. Consequently, the output spectrum of the filter bank will have a local winner-take-all-like characteristic with strong spectral peaks in the input suppressing weaker neighboring ones, and high signal-to-noise-ratio channels being emphasized over weaker ones. There are certain similarities between the properties exhibited by the companding algorithm and phenomena such as two-tone suppression that are found at early stages of the auditory periphery (e.g., Sachs and Kiang, 1968; Ruggero *et al.*, 1992). However, there are also differences, such as the differential growth of suppression for low- and high-side suppressors that is found in the auditory periphery (e.g., Delgutte, 1990), but not in the companding algorithm.

The compression and expansion blocks can be instantaneous or use feedforward automatic gain control (AGC), which sets the gain based on the input level of the signal. In this paper, we only use AGCs as in the original description of Turicchia and Sarpeshkar (2005). The time constants of the compression and expansion AGCs are set at $w/(2\pi f_c)$ s with $w=40$ unless otherwise mentioned, where f_c is the center frequency of the channel filters in Hertz. The compression and expansion functions are compressive and expansive power laws of $n_1=0.3$ and $1/n_1=1/0.3$, respectively. The n_2 parameter of the algorithm, which sets the overall compression in the signal after companding, had a value of 1 throughout this paper such that there was no overall compression. Further parameters are listed in Sec. III A. A more detailed analysis of the potential benefits and operation of the algorithm may be found in Turicchia and Sarpeshkar (2005).

2. Noise vocoding

All stimuli (whether enhanced by companding or not) were processed by a noise-excited envelope vocoder in the following manner: The stimuli were first bandpass filtered (with sixth-order Butterworth filters) into 8 or 16 contiguous frequency bands between 250 and 6000 Hz. The frequency range was divided equally in terms of the ERB_N scale (Glasberg and Moore, 1990). For the 8-channel conditions, the Q values of the resulting filters ranged from 1.8 for the lowest-frequency filter to around 3.0 for the mid- and high-frequency filters (above 1 kHz, where the ERB_N is roughly constant as a proportion of the center frequency); for the 16-channel conditions, the Q values were higher, ranging from around 3.5 at the lowest center frequency to between 5 and 6 for the mid- to high-frequency filters. The envelopes of the signals were extracted by half-wave rectification and lowpass filtering using a second-order Butterworth filter with a cutoff frequency of 200 Hz. The envelopes were then used to modulate narrowband noises, filtered by the same bandpass filters that were used to filter the original stimuli. The modulated narrowband noise waveforms were summed together. This simulation preserves the temporal envelope cues in speech, but not the fine structure (e.g., Shannon *et al.*, 1995).

B. Procedure

Listeners were instructed that they would hear distorted sentences in a noisy background. The listeners' task was to type what they heard via a computer keyboard. Listeners were told that some of the utterances would be very hard to understand, and that they should make their best guess at as many words as possible.

Approximately 1 h of practice was provided before testing in order to acclimatize listeners to the sound of the noise-excited envelope vocoder. For this, listeners were presented with eighteen lists, each comprising seven sentences from the IEEE/TIMIT corpus (IEEE, 1969), spoken by a male talker. The sentences were initially presented without noise and then at signal-to-noise ratios of 10 and then 5 dB. At each signal-to-noise ratio a list of seven sentences was presented without companding, followed by a list with companding. After each practice sentence, listeners were required to type what they had heard via a computer keyboard. Feedback was then provided by displaying the sentence on the computer monitor, and listeners had the option to repeat the sentence as often as they wished before moving on to the next sentence.

The actual experiment comprised the three different signal-to-noise ratios (0, 3, and 6 dB SNR), each presented with and without companding, resulting in a total of six conditions. A list of 10 HINT sentences was presented in a block for each condition. The order of presentation of the conditions was randomized across listeners, and each condition was presented once before any was repeated. A total of 24 HINT sentence lists were presented in each experiment, resulting in a total of four lists (i.e., 40 sentences) per condition per listener. No feedback was given during the actual experiment, which took approximately 1 h and immediately followed the practice session.

For both the practice and actual experiments, the stimuli were presented diotically at an overall level of 75 dB SPL (after processing) in each ear over Sennheiser 580 headphones. All processing (companding and cochlear-implant simulation) was carried out in advance and the sentences were stored as 16-bit sound files on the computer hard drive and were played out via a LynxOne (LynxStudio) sound card. After the session, the listeners' responses were scored off-line. All words were counted, and the percent of correct words in each condition was calculated. Obvious misspellings were counted as correct.

C. Listeners

Eight listeners participated in each of the experiments. No listener participated in more than one experiment, and no listener had taken part in any previous study involving HINT sentences. All listeners were paid for their participation, and

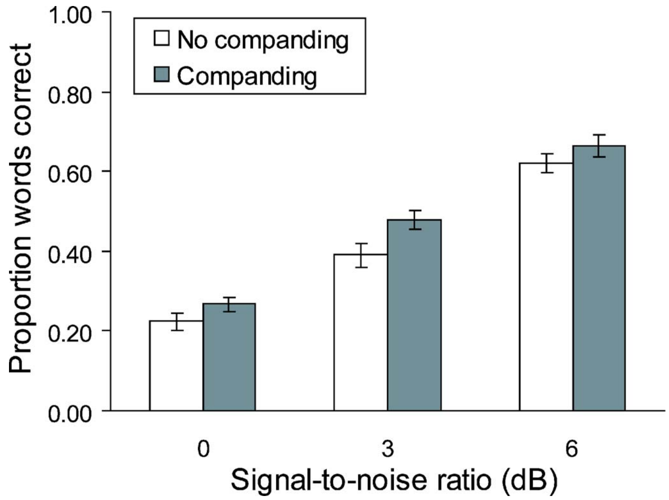

FIG. 3. Initial speech intelligibility results with and without companding. Proportion of words correctly reported are plotted for the three signal-to-noise ratios tested. Error bars represent ±1 standard error of the mean across the eight listeners.

had audiometric thresholds of 25 dB HL or less at octave frequencies between 250 and 8000 Hz. The ages of the listeners ranged from 18 to 60, with a mean age of 29.

III. EXPERIMENT 1: PERFORMANCE USING AN 8-CHANNEL IMPLANT SIMULATOR WITH AND WITHOUT COMPANDING

A. Companding parameters

The first experiment involved testing the companding algorithm with parameters essentially identical to those used by Bhattacharya and Zeng (2007). The Q values of the prefilters and postfilters were set to 2.3 and 13.8, respectively (corresponding to Q_{10} values of 1.1 and 6.8, respectively).[1] The time-constant parameter for the AGCs, w, was set to 40, which corresponds to an overall time constant of about 13 ms at 500 Hz and 1.3 ms at 5000 Hz. Stimuli were processed using a 50-channel implementation of the companding algorithm, with the lowest and highest filter center frequencies set to 100 and 8000 Hz, respectively. The filters were spaced at equal distances along a log scale.

B. Results and discussion

The results, averaged across the eight listeners, are shown in Fig. 3. The percentage of words correctly reported is shown with and without companding at the three signal-to-noise ratios tested. The companding provides a small but reliable improvement in performance. This was confirmed using a repeated-measures analysis of variance (ANOVA), with companding (on or off) and signal-to-noise ratio as fixed factors, which showed that both factors were statistically significant [companding: $F(1,7)=13.65$, $p<0.01$; signal-to-noise ratio: $F(2,7)=436.252$, $p<0.001$], with no significant interaction. The improvement in performance, averaged across subjects and signal-to-noise ratios, was 6 percentage points. This corresponds roughly to a 1-dB improvement in signal-to-noise ratio at a given level of performance. These results are broadly consistent with those reported by Bhattacharya and Zeng (2007). They showed an average improvement of around 5 and 15 percentage points in sentence word recognition at speech-to-noise ratios of 0 and +5 dB, respectively, with normal-hearing subjects listening through a 16-channel implant simulator. More importantly, they also showed an average improvement of between 5 and 20 percentage points at speech-to-noise ratios of between 0 and +10 dB for cochlear-implant users. Thus, the present experiment, together with the study of Bhattacharya and Zheng (2007) confirms the small, but robust, improvements produced by the algorithm in both normal-hearing subjects (through a noise-excited vocoder simulation of implant processing) and cochlear-implant users. The remaining experiments probe the effects of various parameters within the algorithm, such as filter bandwidth, compression time constants, and the number of analysis and presentation filter channels.

IV. EXPERIMENT 2: EFFECTS OF ALTERING ANALYSIS FILTER BANDWIDTH AND COMPRESSION AND EXPANSION TIME CONSTANTS

A. Rationale

One difference between the parameters used by us and Bhattacharya and Zeng (2007), and those proposed in the original paper (Turicchia and Sarpeshkar, 2005) is the bandwidth of the filters. The original paper proposed using relatively broad filters, with overall Q values of 2.3 and 4.6 (corresponding to Q_{10} values of 1.1 and 2.3) for the pre- and postfilter, respectively. The narrower postfilters used in experiment 1 are closer to those found in normal hearing, where estimates of Q range from between about 9 and 12 at 1 kHz, depending on how the bandwidths are estimated (Oxenham and Shera, 2003). On the other hand, broader (lower-order) filters are less computationally expensive, have less inherent group delay, and are amenable to very low-power analog implementation, which may make them of practical value in cochlear implants. Experiment 2A investigated the effect of decreasing the Q value of the postfilters by a factor of about 3.

Another potentially important set of parameters within the companding algorithm are the time constants associated with compression and expansion. The question of what time constants are most appropriate for compression in hearing aids has been studied extensively. On one hand, very short attack time constants can be deleterious to intelligibility (e.g., Stone and Moore, 2004). On the other hand, the compression found in the normal auditory system at the level of the basilar membrane tends to be very fast acting, and can be thought of as acting quasi-instantaneously on the temporal envelope (e.g., Ruggero *et al.*, 1997). In addition, studies using companding in automatic speech recognition, suggested potential benefits of shorter time constants (Guinness *et al.*, 2005). The time constant of $40/(2\pi f_c)$ s, used in experiment 1 corresponds to around 6 cycles of the center frequency, which is longer than that associated with cochlear compression and suppression effects. Experiment 2B tested the effect of decreasing the time constants.

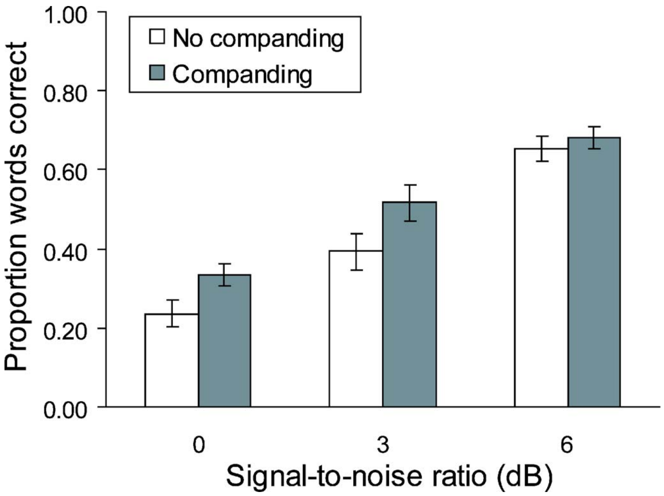

FIG. 4. Speech intelligibility with broader filter tuning in the postfilter stage of the compander. Bars show the mean and standard error from eight new subjects. Other details are as in Fig. 3.

B. Companding parameters

In experiment 2A, both compression and expansion AGC time constants were $40/(2\pi f_c)$ s, as in experiment 1, but the postfilter Q value was changed. Specifically, in experiment 2A, the prefilter and postfilter Q values were 2.3 and 4.6. In experiment 2B, the Q values were the same as in experiment 1, but the compression time constant was reduced to $5/(2\pi f_c)$ s and the expansion time constant was reduced to $20/(2\pi f_c)$ s. All other parameters were left unchanged from experiment 1.

C. Results and discussion

The average results from experiments 2A and 2B are shown in Figs. 4 and 5, respectively. Consider first the results using the broader filters (Fig. 4). As in experiment 1, a repeated-measures ANOVA showed significant effects of both companding $[F(1,7)=15.883, p=0.005]$ and signal-to-noise ratio $[F(2,7)=168.491, p<0.001]$, with no interaction.

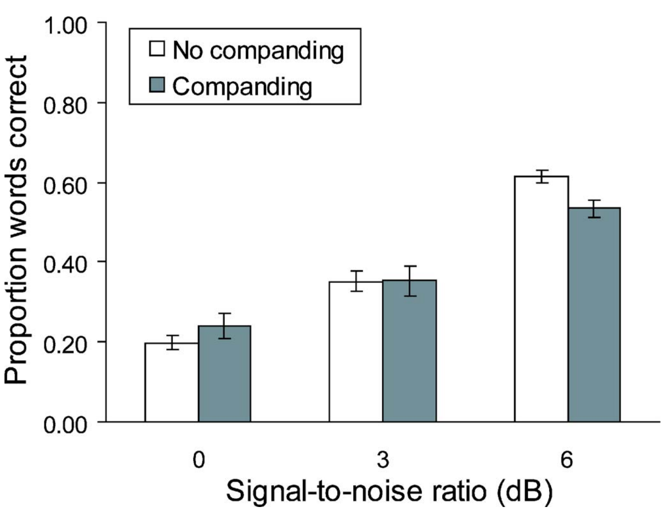

FIG. 5. Speech intelligibility with reduced time constants for both the compression and expansion stages of the compander. Bars show the mean and standard error from eight new subjects. Other details are as in Fig. 3.

A comparison of these results with those from experiment 1 (Fig. 3) suggests that decreasing the sharpness of tuning of the companding filters by a factor of between 2 and 3 produced no decrease in the improvement in intelligibility over the original parameters $[F(2,14)<1, \mathrm{n.s.}]$.

Consider next the results using the shorter time constants. Here, an ANOVA revealed no significant main effect of companding $[F(1,7)=0.359, p=0.568]$, with a significant interaction $[F(2,7)=9.333, p=0.003]$, suggesting that decreasing the time constants actually worsened performance, especially at the higher signal-to-noise ratio, relative to the original parameter choices. A between-subjects comparison between the companding conditions of experiment 1 and experiment 2B indicated a significant effect $[F(2,14)=5.323, p=0.012]$, further confirming the significantly poorer performance with the shorter time constants.

Overall, the results confirm that companding can improve speech intelligibility somewhat and show that the improvement is not critically dependent on the filter bandwidths of the companding scheme. This is encouraging in terms of the prospects for implementing the algorithm in actual cochlear implants, because broader filters are less computationally expensive. Furthermore, the architecture is amenable to very low-power analog VLSI implementations, which may result in a much longer battery life than is currently available. The results do, however, seem to depend on the compression and expansion time constants, with time constants of $40/(2\pi f_c)$ s producing a significant improvement, and shorter time constants of $5/(2\pi f_c)$ s and $20/(2\pi f_c)$ s for compression and expansion, respectively, producing no significant benefit relative to no processing. Thus, in contrast to what was found in automatic speech recognition (Guinness *et al.*, 2005), shorter, perhaps more physiologically realistic, time constants do not appear to improve intelligibility, at least for the conditions tested here. It remains to be seen what effect would result from time constants longer than those used in experiment 1.

V. EXPERIMENT 3: COMPANDING FOLLOWED BY 16-CHANNEL VOCODER PROCESSING

A. Rationale

All the results so far were obtained using an 8-channel noise-excited envelope vocoder. In an earlier study using vowel stimuli, Loizou and Poroy (2001) found that the minimum spectral contrast necessary for vowel recognition increased with decreasing number of spectral channels. Thus, it is possible that if our simulations had used a larger number of more narrowly tuned channels, no further benefit of companding would have been observed. Given that current cochlear implants are rarely able to deliver more than eight independent channels of information, the caveat associated with numbers of channels may not be an important one for the present. However, it is possible that technical improvements will lead to a larger number of channels becoming available for implant patients in the future. It is therefore of interest to test the extent to which the improvement in performance found for companding will generalize to cochlear-

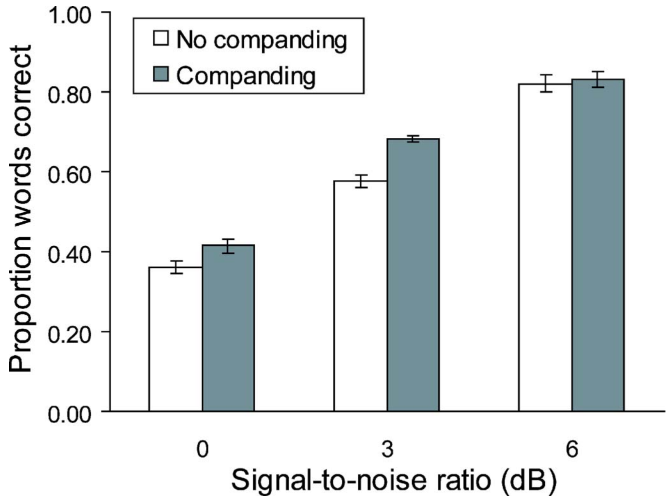

FIG. 6. Speech intelligibility with a 16-band vocoder. Frequency resolution was improved by presenting subjects with 16 bands of spectral information. The companding algorithm itself was the same as used for experiment 1 (see Fig. 3). Again, eight new subjects participated.

implant simulations with larger numbers of spectral channels. This was tested in experiment 3 using a 16-channel implant simulator.

B. Companding and vocoder parameters

The parameters used in the companding algorithm were the same used in experiment 1. The only difference is that the sounds were then processed by a cochlear-implant simulator implemented with 16 bandpass filter (sixth-order Butterworth filters) between 250 and 6000 Hz rather than 8, as in experiments 1 and 2. The frequency range was again divided equally in terms of the ERB_N scale (Glasberg and Moore, 1990).

C. Results

The results using a 16-channel implant simulator are shown in Fig. 6. A repeated-measures ANOVA showed significant effects of both companding $[F(1,7)=18.611, p<0.005]$ and signal-to-noise ratio $[F(2,7)=308.605, p<0.001]$. This time, the interaction also reached significance $[F(2,7)=7.883, p=0.005]$, reflecting the fact that little or no improvement was observed at the highest signal-to-noise ratio. The improvement in the conditions with companding, averaged across listeners and signal-to-noise ratios, was 5 percentage points. Thus, despite the improved spectral resolution of the cochlear-implant simulations (16 instead of 8 channels), some benefit was still provided by the companding algorithm.

VI. EXPERIMENT 4: SPECTRAL ENHANCEMENT WITH ONLY 16 ANALYSIS CHANNELS

A. Rationale

The experiments so far have demonstrated the ability of the companding system described by Turicchia and Sarpeshkar (2005) to provide small but significant benefits in speech intelligibility when the speech is passed through 8- and 16-channel cochlear-implant simulators. However, the algorithm as tested so far still comprises 50 independent frequency analysis channels, the outputs of which are combined before being fed to the cochlear-implant simulator. It would not be possible to implement such processing within current implants without increasing their complexity. Another possibility is to simply modify existing filters within cochlear implants to incorporate the benefits of companding in each of the channels of a cochlear implant. The final experiment tests whether it is possible to reduce the number of analysis channels within the companding algorithm from 50 to 16 while still retaining the beneficial effects of companding that were observed in the original experiments.

B. Algorithm

Stimuli were processed by first implementing companding using only a 16-channel filter bank, as shown in Fig. 2 but without the final summation. The outputs of the resulting companding channels were combined as in a cochlear-implant simulator by extracting the final envelope outputs of each companding channel, using these envelopes to modulate bandpass noise corresponding to that channel, and then finally summing these outputs. In effect, we directly use companded spectral-analysis outputs, rather than speech enhanced by companding, as inputs to the noise vocoder.

The parameters used for companding were as in experiment 1 except that the prefilter was a second-order Butterworth filter and the postfilter was a fourth-order Butterworth filter, so that the overall filter order was the same as the sixth-order Butterworth filter used in the original vocoder. To provide a quantitative comparison of the filters used here and in Experiment 1, the Q_{10} values were derived. In experiment 1, the Q_{10} values for the pre- and postfilter were 1.1 and 6.8, respectively. Here, for the Butterworth filters, the Q_{10} values were between 1.2 and 2.1 for the prefilters and between 2 and 3.5 for the postfilters, depending on the center frequency. All envelope detectors for compression and expansion were implemented as second-order lowpass Butterworth filters, with cut-off frequencies of 70 and 12 Hz, respectively.

C. Results

The results for the 16-channel companding system are shown in Fig. 7. A repeated-measures ANOVA showed significant effects of both companding $[F(1,7)=13.825, p<0.01]$ and signal-to-noise ratio $[F(2,7)=190.8, p<0.001]$, with no significant interaction. The mean increase in intelligibility, pooled across subjects and signal-to-noise ratios, was 4 percentage points. Thus, even with a reduction in the number of analysis frequency channels in the companding algorithm from 50 to 16, a small benefit of companding was still observed.

VII. DISCUSSION AND SUMMARY

The results in many of the conditions show that companding can result in a small but significant improvement in the intelligibility of speech in a background of steady-state noise. The improvements proved robust to some, but not other, changes in the companding parameters. In particular, it was possible to reduce the sharpness of tuning of the post-

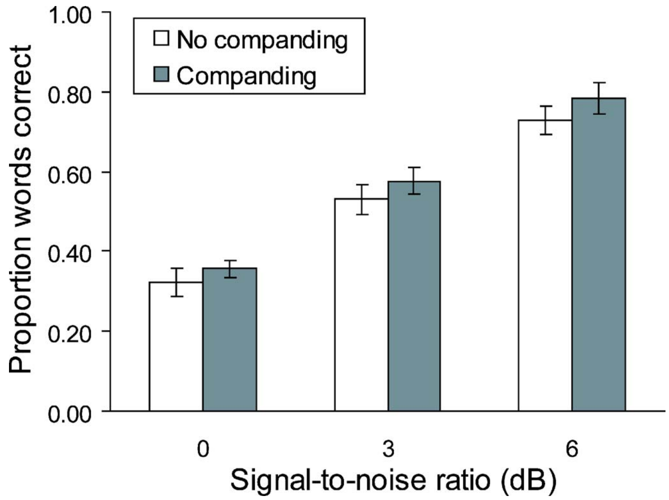

FIG. 7. Speech intelligibility with 16-channel companding, followed by a 16-band vocoder. Instead of the 50 channels used in all previous experiments, this graph shows results using only 16 analysis channels, the outputs of which are fed directly to 16 presentation channels within the implant simulator. Other graph details are as in Fig. 3.

filter by a factor of 3 without altering the benefit provided by the companding algorithm. On the other hand, shortening the time constants to more closely resemble the instantaneous compression and suppression provided by the basilar membrane resulted in the loss of any benefit provided by the algorithm. The benefit of companding was observed for both 8-channel and 16-channel simulations, suggesting that the spectral enhancement continues to be of benefit even with a spectral resolution that exceeds what is currently available to most cochlear-implant users. Perhaps most encouragingly, a simplified version of the companding algorithm, using only 16 analysis channels in a way that could be implemented into current implants, still resulted in a small improvement in speech intelligibility.

The present study tested speech intelligibility only in a steady-state noise background. Some of the most challenging acoustic environments for cochlear implants and implant simulations are those that involve maskers with spectro-temporal fluctuations, such as a competing talker (Nelson *et al.*, 2003; Qin and Oxenham, 2003; Stickney *et al.*, 2004). Further study will be required to determine whether the companding algorithm described here continues to provide an advantage in more complex masking situations.

There are obvious limits to the applicability of the current research using acoustic stimulation to actual cochlear implants. Nevertheless, as shown by a number of studies (e.g., Shannon *et al.*, 1995; Friesen *et al.*, 2001; Turner *et al.*, 2004) such simulations can capture many important aspects of cochlear implant processing, without some of the confounding individual differences often observed in patients, and at much lower effort and cost. In further support of our approach, the results of experiment 1 are in good agreement with those of Bhattacharya and Zeng (2007) in actual cochlear-implant users, which provides some optimism that the new results from this study—in particular the finding that a readily implementable system can provide some benefit—may generalize to cochlear-implant users.

ACKNOWLEDGMENTS

This work was supported primarily by the Advanced Bionics Corporation. A.J.O. and A.M.S. are also supported by the National Institutes of Health [No. R01 DC 03909 and R01 DC 05216 (awarded to A.J.O.)]. We thank Johannes Lyzenga, Michael Stone, and one anonymous reviewer for their helpful comments on an earlier version of this manuscript.

[1]Turicchia and Sarpeshkar (2005) used the parameters $q1$ and $q2$ to describe the pre- and postfilter properties. Here for simplicity we use the overall Q of the filters. The original filters were built with 4 stages and therefore the overall Q value is 1.15 times the value of $q1$ or $q2$. As an alternative measure, Q_{10} values refer to the center frequency of the filter, divided by the bandwidth of the filter at its 10-dB (rather than 3-dB) down points.

Baer, T., Moore, B. C. J., 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.

Bhattacharya, A., and Zeng, F. G. (**2007**). "Companding to improve cochlear-implant speech recognition in noise," J. Acoust. Soc. Am. (submitted).

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

Delgutte, B. (**1990**). "Two-tone suppression in auditory-nerve fibers: Dependence on suppressor frequency and level," Hear. Res. **49**, 225–246.

Dorman, M. F., Loizou, P. C., Fitzke, J., and Tu, Z. (**1998**). "The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6–20 channels," J. Acoust. Soc. Am. **104**, 3583–3585.

Franck, B. A., van Kreveld-Bos, C. S., Dreschler, W. A., and Verschuure, H. (**1999**). "Evaluation of spectral enhancement in hearing aids, combined with phonemic compression," J. Acoust. Soc. Am. **106**, 1452–1464.

Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (**2001**). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," J. Acoust. Soc. Am. **110**, 1150–1163.

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

Guinness, J., Raj, B., Schmidt-Nielsen, B., Turicchia, L., and Sarpeshkar, R. (**2005**). "A companding front end for noise-robust automatic speech recognition," in International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 249–252.

IEEE. (**1969**). "IEEE recommended practice for speech quality measurements," IEEE Trans. Audio Electroacoust. **17**, 225–246.

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.

Moore, B. C. J. (**2003**). "Speech processing for the hearing-impaired: Successes, failures, and implications for speech mechanisms," Speech Commun. **41**, 81–91.

Nelson, P. B., Jin, S. H., Carney, A. E., and Nelson, D. A. (**2003**). "Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners," J. Acoust. Soc. Am. **113**, 961–968.

Nilsson, M., Soli, S., 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., and Shera, C. A. (**2003**). "Estimates of human cochlear tuning at low levels using forward and simultaneous masking," J. Assoc. Res. Otolaryngol. **4**, 541–554.

Qin, M. K., and Oxenham, A. J. (**2003**). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," J. Acoust. Soc. Am. **114**, 446–454.

Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). "Basilar-membrane responses to tones at the base of the chinchilla cochlea," J. Acoust. Soc. Am. **101**, 2151–2163.

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. S. (**1968**). "Two-tone inhibition in auditory nerve fibers," J. Acoust. Soc. Am. **43**, 1120–1128.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science **270**, 303–304.

Simpson, A. M., Moore, B. C. J., and Glasberg, B. R. (**1990**). "Spectral enhancement to improve the intelligibility of speech in noise for hearing-impaired listeners," Acta Oto-Laryngol., Suppl. **469**, 101–107.

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.

Stone, M. A., and Moore, B. C. J. (**1992**). "Spectral feature enhancement for people with sensorineural hearing impairment: Effects on speech intelligibility and quality," J. Rehabil. Res. Dev. **29**, 39–56.

Stone, M. A., and Moore, B. C. J. (**2004**). "Side effects of fast-acting dynamic range compression that affect intelligibility in a competing speech task," J. Acoust. Soc. Am. **116**, 2311–2323.

Turicchia, L., and Sarpeshkar, R. (**2005**). "A bio-inspired companding strategy for spectral enhancement," IEEE Trans. Speech Audio Process. **13**, 243–253.

Turner, C. W., Gantz, B. J., Vidal, C., Behrens, A., and Henry, B. A. (**2004**). "Speech recognition in noise for cochlear implant listeners: Benefits of residual acoustic hearing," J. Acoust. Soc. Am. **115**, 1729–1735.

Yang, J., Luo, F., and Nehorai, A. (**2003**). "Spectral contrast enhancement: Algorithms and comparisons," Speech Commun. **39**, 33–46.