

A Bio-Inspired Companding Strategy for Spectral Enhancement

L. Turicchia and R. Sarpeshkar, *Member, IEEE*

Abstract—This work presents a compressing-and-expanding, i.e., *companding*, strategy for spectral enhancement inspired by the operation of the auditory system. The companding strategy simulates the two-tone suppression phenomena of the auditory system and implements a soft local winner-take-all-like enhancement of the input spectrum. It performs multichannel syllabic compression without degrading spectral contrast. The companding strategy works in an analog fashion without explicit decision making, without the use of the fast Fourier transform, and without any cross-coupling between spectral channels. The strategy is useful in cochlear-implant processors, hearing aids, and speech recognition for enhancing spectral contrast. It is well suited to low-power analog circuit implementations.

Index Terms—Auditory filtering, auditory masking, auditory model, cochlear implant, companding, hearing aids, spectral enhancement, two-tone suppression.

I. INTRODUCTION

IN SOME OF our prior work ([1] and [2]), we have shown that a cochlear model with traveling-wave amplification and distributed gain control can exhibit two-tone suppression, achieve a desired input–output compression curve, and show the level-dependent frequency–response curves seen in the biological cochlea. We argued that such a model is useful for advanced cochlear-implant processors that seek to improve patient performance in noisy environments and is amenable to low-power analog very large-scale integration (VLSI) implementations. However, such a model is complex to program and test on patients because of the intimate dependence of all of the feedback and feedforward parameters on one another in a traveling-wave architecture. Thus, a better first step is to ask if we can reproduce the two-tone suppression and compressibility effects of the biological cochlea in a set of independent filters with independent and easily programmable parameters. In this paper, we describe an architecture that achieves this goal: The gain control or compression in the architecture improves audibility; nevertheless, the presence of two-tone suppression prevents the stimulus spectral contrast from being degraded while audibility is improved. Our architecture is not intended to explicitly enhance spectral contrast. The contrast is enhanced as an emergent property, and has the benefits of improving performance in noise. It is possible that this is one of the reasons for

why the human auditory system with a biological-cochlea front end is known to have superb performance in noise even when compared with state-of-the-art speech-recognition front ends.

A series of recent articles [3]–[5] show that a healthy cochlea exhibits good phase-locking with the vowel formants and that this behavior is not present in an unhealthy cochlea. In particular this work shows the strong benefits of using vowels with well-contrasted formants in the auditory nerves of acoustically traumatized cats. These results suggest that using a cochlear model with compression and two-tone suppression to increase formant contrast might be beneficial.

Spectral contrast is often formally defined as the decibel difference between adjacent peaks and valleys in the spectrum. The problem of increasing the spectral contrast of a signal is a well-known problem in the literature [6]–[10]. Recently, Yang *et al.* [11] proposed new fast algorithms for improving spectral contrast. Architectures that try to mimic masking are usually used in perceptual coding [12]. Our work proposes a new architecture for increasing spectral contrast without the explicit goal of trying to do so. Spectral enhancement is a natural consequence of our trying to mimic the operation of the biological and silicon cochleas. Our architecture is explicitly designed to be amenable to future low-power analog VLSI implementations, extremely important for energy-efficient operation of cochlear-implant processors, hearing aids, or low-power speech-recognition front ends [13].

There is strong evidence that our work is likely to be beneficial to cochlear-implant patients: recently, Loizou *et al.* [14] showed that cochlear-implant listeners required 4–6 dB of spectral contrast to identify vowels with relatively high accuracy and that some of them obtained significantly higher scores with vowels enhanced to 6 dB of contrast by a contrast-enhancing algorithm. Consonant intelligibility can be improved by enhancing differences in the pattern of the signal-processor channel outputs as well [15].

In hearing-aid patients, increasing spectral contrast and simultaneously performing compression yield a modest but significant improvement for speech perception in noise [9]. However, it is worth noting that the latter work was done with broadband compression rather than with the frequency-dependent compression of this work. Lyzenga *et al.* [16] have suggested how further improvements may be possible if the signal processing is done differently. Moore [17] have pointed out that spectral enhancement has not been evaluated in real-time wearable devices and that the benefits of prolonged experience have yet to be tested. A low-power version of our architecture should be of great benefit in such chronic tests.

Manuscript received January 17, 2003; revised November 24, 2003. This work was supported by grants from the Packard Foundation, the Swartz Foundation, and the Office of Naval Research under Award N00014-02-1-0434. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Gerald Schuller.

The authors are with the Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: rahuls@mit.edu; turic@mit.edu).

Digital Object Identifier 10.1109/TSA.2004.841044

Our proposed strategy performs simultaneous multichannel syllabic compression and spectral-contrast enhancement via two-tone suppression. Multichannel compression by itself improves audibility but degrades spectral contrast: A weak tone at one frequency is strongly amplified so that it is concurrently audible with a weakly amplified strong tone at another frequency; the asymmetric amplification due to compression degrades the spectral contrast that was present in the uncompressed stimulus. When two-tone suppression strategies that enhance contrast are also simultaneously present, the compression is prevented from degrading spectral contrast in regions close to a strong spectral peak while allowing the benefits of improved audibility in regions distant from the peak.

Our strategy uses a noncoupled filter bank and compression-expansion blocks, as shown in Fig. 1(b). One crucial and novel element in our architecture that is essential for obtaining the two-tone suppression is the presence of a second filter bank between the compression and expansion blocks. Programmability in the two-tone suppression and compression characteristics may be attained through simple parametric changes in the compression, expansion, or filter blocks.

The idea of using companding for improving the signal-to-noise ratio in systems where the dominant noise occurs after the compression block is well known [18] and is currently being explored for use in novel analog filtering circuits [19], [20]. Although the compression benefits for improving audibility are similar in our architecture to other companding architectures, the two-tone suppression benefits for enhancing spectral contrast are achieved because of the nonlinear nature of the interaction between signals in the first filter bank, the compressor, and the second filter bank.

The outline of this paper is as follows. In Section II, we describe the essential ideas behind the architecture and show how various parametric changes affect its operation. In Section III, we present MATLAB simulation data for a system with various inputs that illustrate the workings and benefits of our architecture. In Section IV, we discuss the use of our architecture in cochlear-implant processors and speech-recognition front ends. In Section V, we conclude by summarizing our contributions.

II. ESSENTIAL IDEAS BEHIND THE ARCHITECTURE

Fig. 1(a) shows a common multichannel syllabic compression architecture. Fig. 1(b) shows our companding architecture. Every channel in the companding architecture has a pre-filter, a compression block, a post-filter, and an expansion block. The pre-filter and post-filter in every channel have the same resonant frequency. The pre-filter and post-filter banks have logarithmically-spaced resonant frequencies that span the desired spectral range.

Fig. 2 shows the details of a single channel of the architecture. The pre-filter is labeled as F and the post-filter is labeled as G . The compression and expansion are implemented with an envelope detector (ED) block, a nonlinear block, and a multiplier in a feed-forward fashion. The time constant of the envelope detector governs the dynamics of the compression or expansion and is typically scaled with the resonant frequency of each

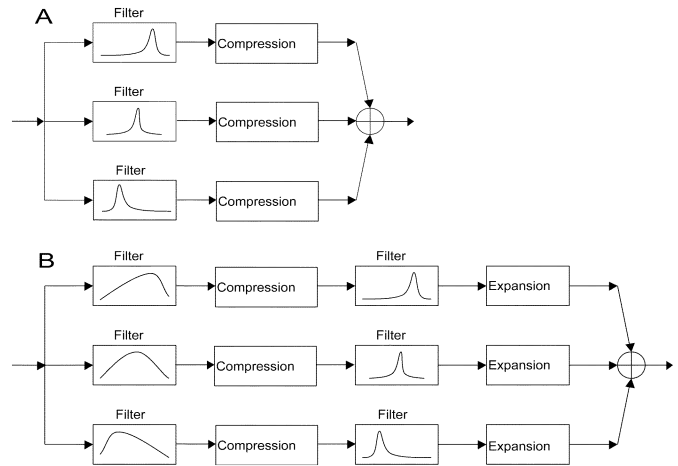


Fig. 1. (a) Block diagram of a common multichannel syllabic compression strategy. (b) Block diagram of our companding strategy.

channel. In general, compression or expansion schemes can involve sophisticated dynamics and energy extraction strategies (peak versus root-mean-square (rms), etc).

In our scheme, n_1 represents the compression index of the compression block, i.e., $n_1 = 0.3$ would yield third-root compression on the input in the compression block. If $n_2 = 1$, then the expansion block simply undoes the effect of the compression block and the channel is input-output linear on the time-scale of the envelope-detector dynamics. If $0 < n_2 < 1$, then the effect of the channel is to implement syllabic compression with an overall channel compression index of n_2 . The expansion block implements a n_2/n_1 power law and is thus really an expansion block only if $n_2 > n_1$, which we shall assume, without loss of generality, throughout this article. In all cases, setting $n_1 = 1$ will shut off the companding strategy and create a multichannel syllabic compression system like that of Fig. 1(a) with a compression index of n_2 .

To start with, let us suppose that n_2 is 1 so that the overall effect of a channel is input-output linear. If we input a sinusoid at the resonant frequency of the channel, the compression stage compresses the signal and the expansion stage undoes the effects of the compression. Fig. 3 illustrates how this works by plotting the effects of the compression and expansion on a decibel or logarithmic scale: The compression line has a slope less than 1 on this plot and the expansion line has a slope greater than 1 on this plot. A sinusoid with amplitude A_1 is transformed to a sinusoid with amplitude B_1 after the compression block. The sinusoid with amplitude B_1 is transformed back to a sinusoid of amplitude A_1 after expansion, i.e., we traverse the square with corners at A_1 and B_1 as we compress and expand the signal and return to the A_1 starting point. Note that we use the 1:1 line in Fig. 3 to map the output of one stage of processing to the input of the next stage of processing.

A. Intuition Behind the Two-Tone Suppression

We shall now intuitively explain why this architecture also allows us to implement tone-to-tone suppression through the use of the post-filter. To start with, let us assume that the pre-filter F is a broad almost perfectly flat filter and that post-filter G is very narrowly tuned. If, in addition to A_1 at the resonant frequency

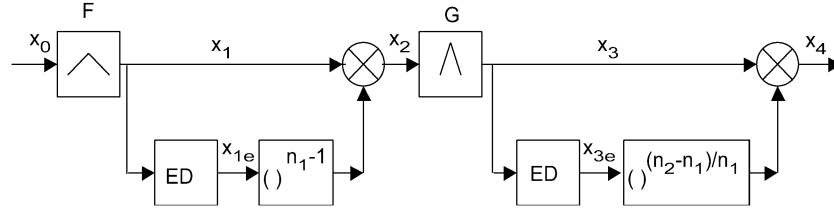


Fig. 2. Detailed view of a single channel of processing in Fig. 1.

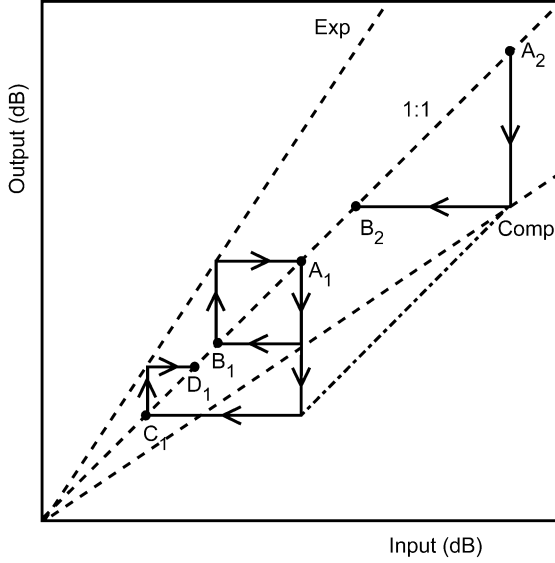


Fig. 3. Intuitive view of the companding strategy.

of the channel, we also have a sinusoid of stronger amplitude A_2 at a different frequency in the input, then, after filtering by F , we obtain two sinusoids represented as A_1 (the weaker) and A_2 (the stronger) in Fig. 3. Since the envelope detector sets the gain of the compression block based primarily on the stronger tone, A_2 is transformed to B_2 and A_1 is transformed to C_1 after compression. If the post-filter G is sharply tuned to suppress the louder tone A_2 , the expansion stage will only see a weak tone of amplitude C_1 at its input and expand that tone to a tone of amplitude D_1 at its output. Since D_1 is clearly less than A_1 in Fig. 3, we observe that an out-of-band strong tone A_2 has effectively suppressed an in-band weak tone A_1 to an output of amplitude D_1 . If A_2 were not simultaneously present, the A_1 tone would have had its amplitude unchanged by the overall channel. The suppression arises because the decibel reduction in gain caused by the compression is large because of the strong out-of-band tone A_2 but the decibel increase in gain caused by the expansion is small because of the weak in-band tone C_1 . The decibel suppression of the input A_1 by A_2 is given by the difference in decibels between the asymmetric compression and expansion. Note that if A_1 were much stronger than A_2 , then the G filter would simply attenuate A_2 and leave A_1 almost unchanged. Thus, in all cases, the stronger tone has the effect of suppressing the weaker tone.

From the above arguments, it is easy to imagine how relaxing our various assumptions would affect the overall architecture: if F is not perfectly flat, but has a finite bandwidth, then the suppressive effect of A_2 on A_1 will be reduced as the frequencies

of the tones get more distant from each other. If G is not perfectly narrow and relatively flat, then the compression and expansion gains in decibels will be determined by the strong A_2 and B_2 tones, respectively, be nearly equal, and result in little suppression of A_1 by A_2 ; A_2 will dominate the response of the channel. Thus, if F is broad, distant tones cause stronger suppression of A_1 , while if G is broad, tones for a broad range of frequencies near A_1 are ineffective in causing suppression of A_1 . Together, the shapes of F and G determine the two-tone suppression frequency profile. The smaller the value of n_1 , the more flat is the compression curve and the more steep is the expansion curve. Thus, the difference in compression and expansion gains in decibels is larger for smaller n_1 , and the suppressive effects of the two-tone suppression are stronger for smaller n_1 . The value of n_2 affects the overall compression characteristics of the channel but does not change the two-tone suppression properties as we have discussed.

B. Analysis of the Two-Tone Suppression

We may express this intuition more formally by computing the value of the signal at various stages of processing in Fig. 2. Suppose, that at the input, we have

$$x_0 = a_1 \sin(w_1 t) + a_2 \sin(w_2 t + \varphi_0). \quad (1)$$

If the gain and phase of the filter F at frequencies w_1 and w_2 are given by

$$\begin{aligned} f_1 &= |F(jw_1)|, & f_2 &= |F(jw_2)| \\ \varphi_1 &= \text{ang}(F(jw_1)), & \varphi_2 &= \text{ang}(F(jw_2)) \end{aligned} \quad (2)$$

then

$$x_1 = f_1 a_1 \sin(w_1 t + \varphi_1) + f_2 a_2 \sin(w_2 t + \varphi_0 + \varphi_2). \quad (3)$$

Suppose we have nearly ideal peak detection in the envelope detector, and that the frequency ratio w_1/w_2 is not a small rational number. Then the envelope of x_1 may be approximated by

$$x_{1e} = f_1 a_1 + f_2 a_2. \quad (4)$$

Thus, after compression

$$x_2 = x_{1e}^{(n_1-1)}. \quad (5)$$

If

$$\begin{aligned} g_1 &= |G(jw_1)|, & g_2 &= |G(jw_2)| \\ \vartheta_1 &= \text{ang}(G(jw_1)), & \vartheta_2 &= \text{ang}(G(jw_2)) \end{aligned} \quad (6)$$

then

$$x_3 = \left[g_1 f_1 a_1 \sin(w_1 t + \varphi_1 + \vartheta_1) + g_2 f_2 a_2 \sin(w_2 t + \varphi_0 + \varphi_2 + \vartheta_2) \right] x_{1e}^{(n_1-1)} \quad (7)$$

and the envelope of x_3 may be approximated by

$$x_{3e} = (g_1 f_1 a_1 + g_2 f_2 a_2) x_{1e}^{(n_1-1)} \quad (8)$$

where x_{3e} is the output of the envelope detector

$$\begin{aligned} x_4 &= x_3 x_{3e}^{((n_2-n_1)/n_1)} \\ &= \left[g_1 f_1 a_1 \sin(w_1 t + \varphi_1 + \vartheta_1) + g_2 f_2 a_2 \sin(w_2 t + \varphi_0 + \varphi_2 + \vartheta_2) \right] x_{1e}^{(n_1-1)} \\ &\quad \cdot \left((g_1 f_1 a_1 + g_2 f_2 a_2) x_{1e}^{(n_1-1)} \right)^{((n_2-n_1)/n_1)} \\ &= \left[g_1 f_1 a_1 \sin(w_1 t + \varphi_1 + \vartheta_1) + g_2 f_2 a_2 \sin(w_2 t + \varphi_0 + \varphi_2 + \vartheta_2) \right] \\ &\quad \cdot (g_1 f_1 a_1 + g_2 f_2 a_2)^{((n_2-n_1)/n_1)} x_{1e}^{n_2((n_1-1)/n_1)}. \quad (9) \end{aligned}$$

If $g_1 = f_1 = 1$ (the pre and post filters have a resonance frequency of w_1) and $g_2 = 0$ (G is sharply tuned and w_2 is distant from w_1), then

$$\begin{aligned} x_4 &= \left[a_1^{n_2/n_1} (a_1 + f_2 a_2)^{n_2((n_1-1)/n_1)} \right] \sin(w_1 t + \varphi_1 + \vartheta_1) \\ &= \left[a_1 \left(\frac{a_1 + f_2 a_2}{a_1} \right)^{((n_1-1)/n_1)} \right]^{n_2} \sin(w_1 t + \varphi_1 + \vartheta_1). \quad (10) \end{aligned}$$

Thus, the presence of a second tone with amplitude a_2 suppresses the tone with amplitude a_1 . If there is only one tone ($a_2 = 0$), then

$$x_4 = \sin(w_1 t + \varphi_1 + \vartheta_1) a_1^{n_2} \quad (11)$$

such that, if $n_2 = 1$, the output has amplitude a_1 .

C. Parametric Dependence of Comping in a Single Channel

Figs. 4–6 show the amplitude of x_4 in (10) versus the amplitude ratio of the two tones a_2 and a_1 expressed in decibels. We use $n_2 = 1$ in all figures. The amplitude of the suppressed tone a_1 is fixed while the amplitude of the suppressor tone a_2 varies. Fig. 4 shows that, with a small suppressor amplitude a_2 , the output is equal to the amplitude of the suppressed tone a_1 . As a_2 becomes large, the output becomes very small due to suppression. Fig. 5 shows that smaller values of n_1 result in greater suppression. Fig. 6 shows that narrow filters that result in small values of f_2 in (10) cause less suppression than broad filters with larger values of f_2 .

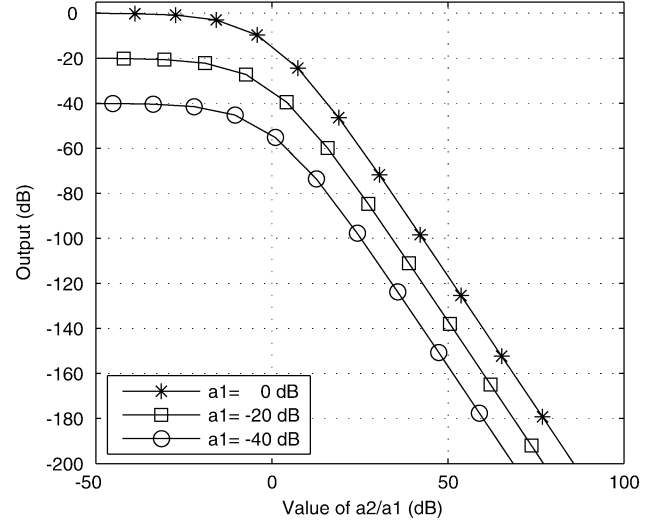


Fig. 4. Tone-to-tone suppression in one channel as suppressor tone's amplitude a_2 varies with respect to the fixed suppressed tone's amplitude, a_1 equal to 0, -20, and -40 dB in the three plots, respectively. The amplitude of a_2/a_1 is plotted in decibels on the X-axis while the output amplitude of the suppressed tone is plotted on the Y-axis. The filter parameters in (10) are $f_2 = 1$ (F is broad), and $n_1 = 0.3$. With a small suppressor amplitude a_2 , the output is equal to the amplitude of the suppressed tone a_1 . As a_2 becomes large, the output becomes very small due to suppression.

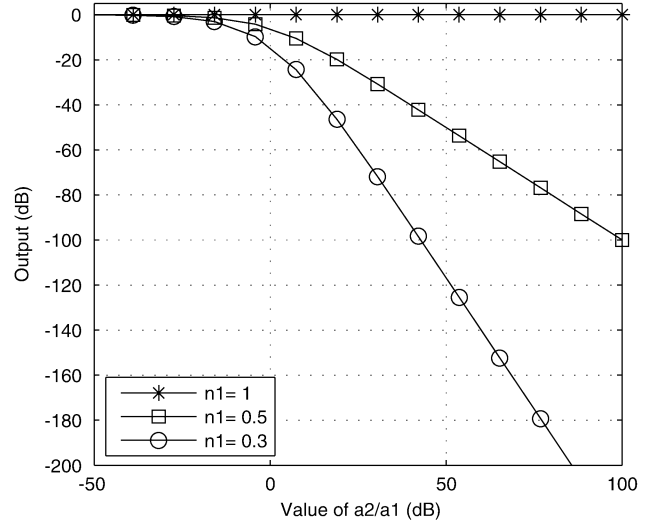


Fig. 5. Tone-to-tone suppression in one channel plotted as in Fig. 4 but the three plots are for different values of n_1 . The suppressed tone's amplitude a_1 is fixed at 0 dB while the amplitude a_2 varies. When $n_1 = 1$ the companding strategy is off and there is no suppression. All plots have $f_2 = 1$ (F is broad). We observe that smaller values of n_1 result in greater suppression.

Thus, we can achieve the two-tone suppression profile that we want by varying the filter, compression, and expansion parameters: an asymmetric profile in F will result in asymmetric two-tone suppression and a broader profile in F will result in broader band two-tone suppression. Small values of n_1 yield stronger two-tone suppression while the value of n_2 affects the overall compression characteristics of the system. The sharpness in tuning of the G filter determines the frequency region around the suppressed tone where the two-tone suppression is ineffective. The dynamics of the envelope detectors determine the attack and release time constants of the compression and

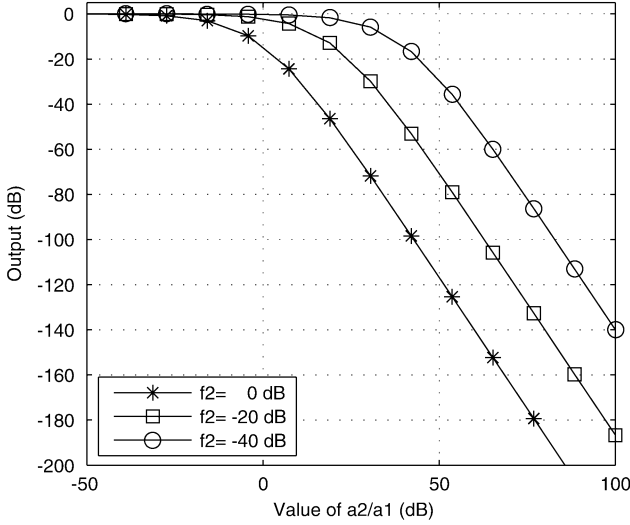


Fig. 6. Tone to tone suppression in one channel plotted as in Fig. 4 but with different values of f_2 corresponding to different F filters. The plot with $f_2 = 0$ dB corresponds to a broad F filter and results in more suppression, while that with $f_2 = -40$ dB is sharp and results in less suppression. The suppressed tone's amplitude a_1 is fixed at 0 dB while the amplitude a_2 varies; $n_1 = 0.3$.

thus the time course of overshoots and undershoots in transient responses. Nonlinear gain control due to saturation in the envelope detectors is important in determining the transient distortion of the system.

In this paper, because of their simplicity and ease of construction in analog VLSI implementations, we use low-order band-pass filters to implement F and G [21]. In general, more sophisticated filters may be also be used. For certain experiments, we use zero-phase versions of these filters to illustrate the ideal case.

III. SIMULATION EXPERIMENTS AND RESULTS

We implemented the companding architecture shown in Figs. 1(b) and 2 with 50 channels in MATLAB. The number of channels was chosen to reflect numbers that could soon be seen in advanced cochlear-implant processors. The architecture does not necessarily need so many channels. We chose band-pass filters for F and G with transfer functions described by $F_i(s) = F_i'^2(s)$ and $G_i(s) = G_i'^2(s)$, where $F_i'(s)$ and $G_i'(s)$ are given by (12) and (13) below¹

$$F_i'(s) = \left(\frac{2 \left(\frac{\tau_i}{q_1} \right) s}{\tau_i^2 s^2 + 2 \left(\frac{\tau_i}{q_1} \right) s + 1} \right)^2 \quad (12)$$

$$G_i'(s) = \left(\frac{2 \left(\frac{\tau_i}{q_2} \right) s}{\tau_i^2 s^2 + 2 \left(\frac{\tau_i}{q_2} \right) s + 1} \right)^2. \quad (13)$$

¹These filters can be implemented in the digital domain with second-order sections in a straightforward fashion: $H(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2}}{1 + a_1 z^{-1} + a_2 z^{-2}}$. The parameters a_i and b_i can be computed with ordinary techniques for discretizing continuous-time systems, e.g., zero-order hold, Bilinear (Tustin) approximation or the matched pole-zero method. At 44.1-kHz sampling rate, all methods yielded very similar performance.

Effectively, to create $F_i(s)$ and $G_i(s)$ we apply $F_i'(s)$ and $G_i'(s)$ twice, respectively. As we later explain, if we need zero-phase versions of $F_i(s)$ and $G_i(s)$, then we apply $F_i'(s)$ or $G_i'(s)$ once in the forward time direction and once in the reverse time direction. Each channel has a resonance frequency given by $f_r = 1/(2\pi\tau_i)$. The filters have resonance frequencies that are logarithmically spaced between 250 and 4000 Hz across the 50 channels. For most experiments, we used $q_1 = 2.8$ and $q_2 = 4.5$.

The envelope detector in each channel was built with an ideal rectifier and a first-order low-pass filter that is applied twice. For our zero-phase experiments, the low-pass filter is applied once in the forward time direction and once in the reverse time direction. The poles of the low-pass filter were chosen to scale with the resonant frequency of the channel, i.e., $(\tau_{ED_i} = w\tau_i)$. We chose $w = 40$ for all experiments except for the cochlear-implant simulations of Section IV, where we chose $w = 10$.

The properties of the entire architecture are similar to the properties of a single channel except for the final summation at the output. The sum of a bunch of filtered outputs can cause interference effects due to phase differences across channels. The interference effects can be severe if the filters are not sharply tuned because the same sinusoidal component is present in several channel outputs with different phases. The companding architecture alleviates interference effects because its local winner-take-all behavior suppresses the outputs of interfering channels.

When companding is turned off in our architecture, i.e., $n_1 = 1$, interference across channels due to phase differences results in severe attenuation of the output. However, in some of our experiments, we wanted to compare the effects of using companding versus not using companding. To permit such comparisons, we used zero-phase versions of the F and G filters to avoid interference problems. This kind of simulation is only for permit a comparison with a perfect-ideal structure. For companding architectures where interference across channels is not a big problem, the use of zero-phase filters appears to make little difference. However, for architectures where the companding is turned off, the use of zero-phase filters appears to be essential. To create zero-phase versions of the $F_i(s)$ or $G_i(s)$ we time reverse the filtered outputs of $F_i'(s)$ or $G_i'(s)$, respectively, filter with the same $F_i'(s)$ or $G_i'(s)$ filter again, and time reverse the final output. The zero-phase version of $F_i(s)$ then has the same magnitude transfer function as $F_i(s)$ but an identically zero phase transfer function. The zero-phase version of the low-pass filter in the envelope detector is created in a similar fashion.

Fig. 7 shows the magnitude transfer function of the overall architecture of Fig. 1(b) for different values of n_1 . With $n_1 = 1$, there is no companding, and we observe a huge attenuation for frequencies in the central portions of the spectrum due to interference effects. At the borders of the spectrum, there is less attenuation because of a reduction in the amount of interference caused by edge effects. As the value of n_1 falls, the effects of companding grow stronger, the spectrum is sharpened and there is less interaction and interference amongst channels. Thus, the central portions of the spectrum suffer increasingly smaller amounts of attenuation. The results of Fig. 7 were obtained with $q_1 = 2.8$ and $q_2 = 4.5$. Of course, the interference

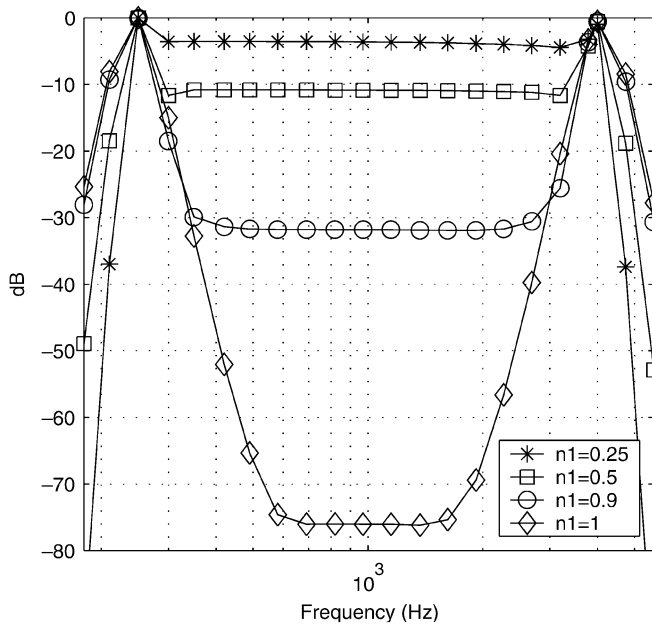


Fig. 7. Magnitude transfer function of the overall architecture of Fig. 1(b) for different values of n_1 . The companding strategy is off for $n_1 = 1$. Higher amounts of compression (smaller values of n_1) flatten the transfer function's profile because they result in less interference amongst channels. Small ripples in the transfer function, not visible in the figure, are caused by the resonances of the 50 channel filters.

effects are less pronounced when higher-Q filters or fewer filters/octave are used (e.g., for obtain a similar profile obtained with $n_1 = 0.25$, $q_1 = 2.8$ and $q_2 = 4.5$ in the companding-on architecture we need a $q_1 = q_2 = 9$ with companding-off). With zero-phase filters there is no interference and the magnitude transfer function of Fig. 7 with companding and without companding is almost identical and flat for all values of n_1 . Our use of the zero-phase filter is simply because we want an ideal and simple noncompanding case to compare with.

A. Two Tone Suppression Experiments

Figs. 8–10 reveal tone-to-tone suppression data for different values of n_1 , q_1 , and q_2 respectively. All experiments were performed by inputting a fixed 970 Hz sinusoid of amplitude $a_2 = 0$ dB (the suppressor tone) and varying the frequency of a second sinusoid with fixed amplitude $a_1 = -20$ dB (the suppressed tone). The output plots the two-tone output spectrum after companding, which was extracted by performing a fast Fourier transform (FFT) on the final output of Fig. 1(b). The suppressor tone is invariant in all output spectra and results in a big spectral peak at 970 Hz in all plots. The suppressed tone strength varies in the output depending on how close in frequency it is to the suppressor and what the parameter settings of the companding architecture are.

Fig. 8 shows that far from 970 Hz, the output amplitude of a_1 is unchanged at -20 dB because the finite bandwidth of the F filter prevents suppression from happening at frequencies distant from 970 Hz. As the a_1 tone frequency approaches 970 Hz, it is suppressed by the strong a_2 tone and its output amplitude falls below -45 dB. When the a_1 tone frequency is very close in frequency to the a_2 tone, however, the G filter has similar gains to both tones and there is again no suppression. As n_1 is

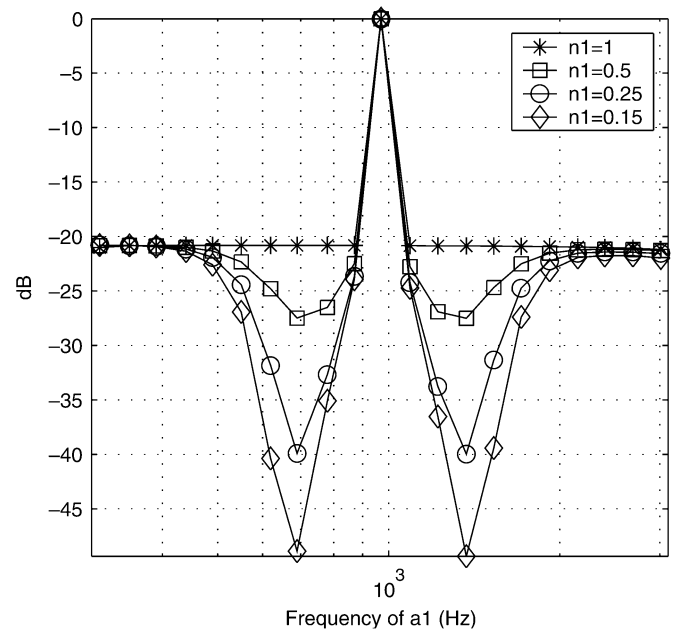


Fig. 8. Tone-to-tone suppression in the entire system as the frequency of the suppressed a_1 tone is varied for different values of n_1 . The suppressor tone is fixed at 970 Hz with an amplitude $a_2 = 0$ dB. The suppressed tone has an amplitude $a_1 = -20$ dB. The value of n_2 is 1 in all curves. The case $n_1 = 1$ corresponds to turning off the companding. The filters are created with $q_1 = 2.8$; $q_2 = 4.5$. The two-tone FFT of the companding architecture's output is plotted as the frequency of the suppressed tone varies.

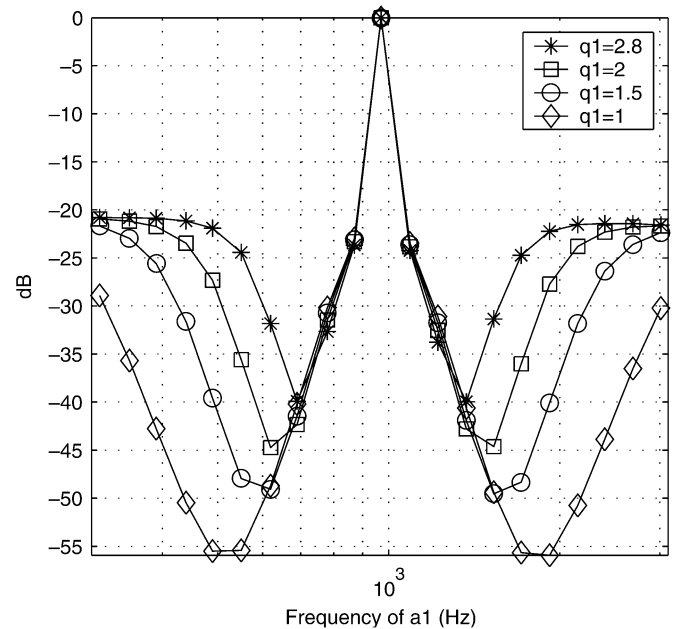


Fig. 9. Tone-to-tone suppression in the entire system as the frequency of the suppressed a_1 tone is varied for different parameters of the F filter. The data are plotted as in Fig. 8 with $n_1 = 0.25$, $n_2 = 1$, $q_2 = 4.5$, $a_1 = -20$ dB, $a_2 = 0$ dB, and the fixed a_2 tone = 970 Hz. As q_1 is decreased, broadening the F filter, the spatial extent and magnitude of the suppression are increased.

reduced, the suppression increases, consistent with the theory of Section II and Fig. 5. At $n_1 = 1$, there is no companding or suppression.

Fig. 9 shows that if the Q of the F filter as parametrized by q_1 is lowered, the extent of the suppression is more widespread in frequency; the suppression is also larger at any given frequency

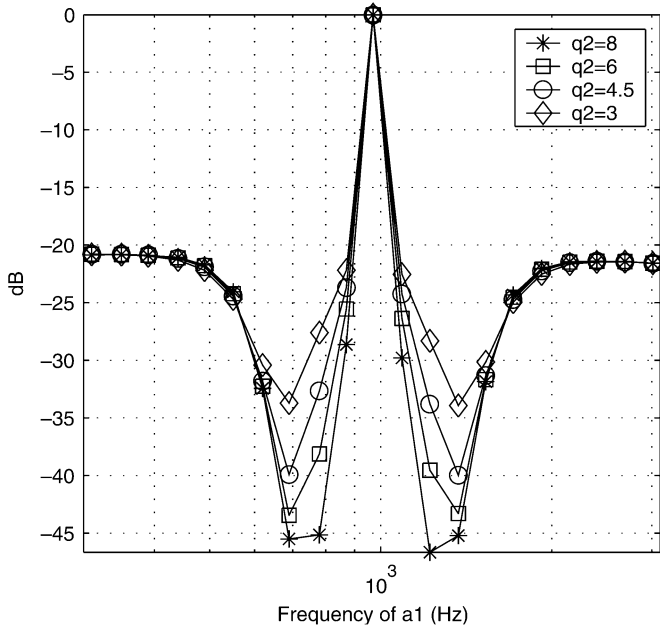


Fig. 10. Tone-to-tone suppression in the entire system as the frequency of the suppressed a_1 tone is varied for different parameters of the G filter. The data are plotted as in Fig. 8 with $n_1 = 0.25$, $n_2 = 1$, $q_1 = 2.8$, $a_1 = -20$ dB, $a_2 = 0$ dB, and the fixed a_2 tone at 970 Hz. As q_2 is decreased, broadening the G filter, the spatial region where suppression is ineffective is broadened, and the magnitude of the suppression decreases in these regions as well.

because the pre-filtered value of the suppressor tone (value after filtering by F) is larger and therefore more effective in causing suppression.

Fig. 10 shows that if the Q of the G filter as parametrized by q_2 is lowered, then the frequency region where the suppression is not effective is broadened; the suppression is also smaller at any given frequency because the G filter is less effective at removing the strong a_2 tone, a necessary condition for having a small expansion gain and large suppression.

The two-tone suppression curves are reminiscent of the consequences of lateral inhibition used in speech enhancement [7]. It is interesting to note that the two-tone suppression is achieved without any lateral coupling between channels and without the use of inhibition.

B. Vowel Experiments

Figs. 11–14 illustrate data obtained from a companding architecture with a synthetic vowel $/u/$ input. The asterisked trace of Fig. 11 shows that the pitch of the vowel input is at 100 Hz, the first formant is at 300 Hz, the second formant is at 900 Hz, and the third formant is at 2200 Hz. The spectral output of the companding architecture was extracted by performing an FFT. For clarity, the harmonics in the spectrum are joined with lines in the figures.

Fig. 11 compares output spectra with the companding strategy on ($n_1 = 0.25$) and with the companding strategy off ($n_1 = 1$) for a zero-phase filter bank. The filter banks span a 300–3500 Hz range and therefore attenuate some of the input energy at very low frequencies. Apart from this low-frequency filtering, however, we observe that the no-companding strategy yields a faithful replica of the input and the companding

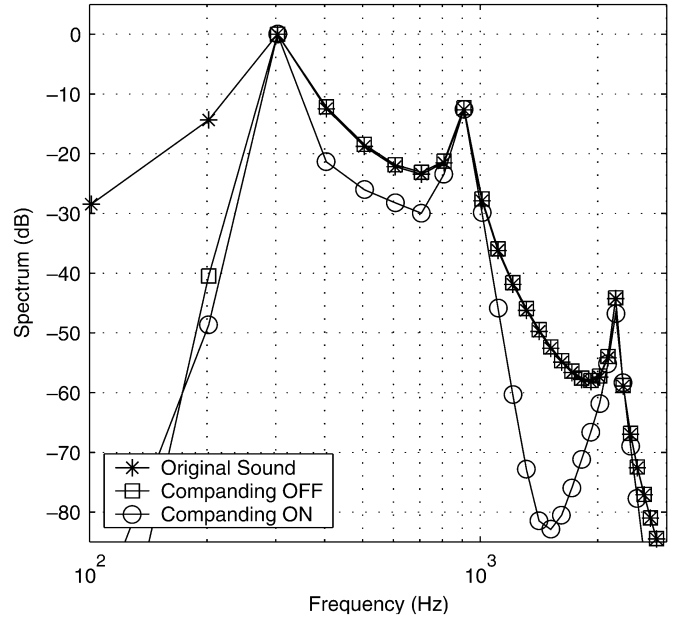


Fig. 11. Spectrum of the output of the vowel $/u/$. The companding-off case corresponds to $n_1 = 1$ and $n_2 = 1$. The companding-on case corresponds to $n_1 = 0.25$, and $n_2 = 1$. Zero-phase filters were used in both cases.

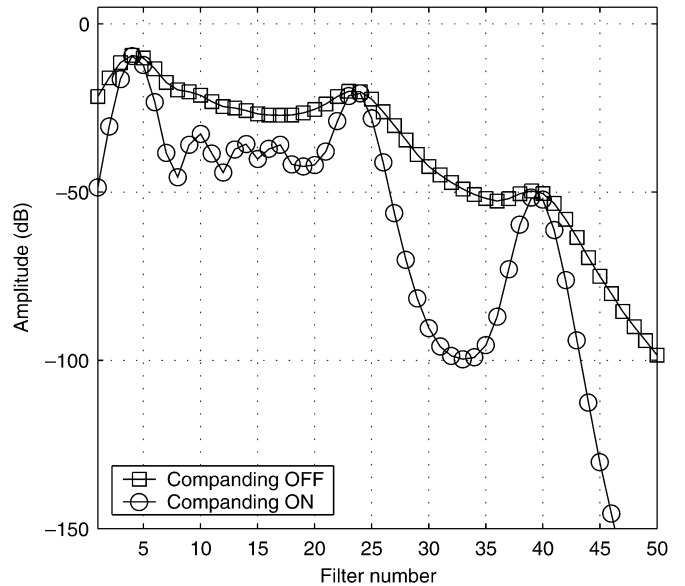


Fig. 12. Maximum output of every channel versus filter number for the vowel input $/u/$. The companding-off case corresponds to $n_1 = 1$ and $n_2 = 1$. The companding-on case corresponds to $n_1 = 0.25$ and $n_2 = 1$.

strategy enhances the spectrum by suppressing harmonics near the formants.

Fig. 12 plots the maximum output of every channel (we do not perform summation) for the companding and no-companding strategies with zero-phase filter banks. We see that the companding strategy sharpens the spectrum and enhances the formant structure. Using nonzero-phase filters made little difference to the output of Fig. 12.

Fig. 13 shows that if zero-phase filter banks are not used, the companding-off strategy results in a strong attenuation of the vowel spectrum due to interference amongst channels. There is less attenuation at the borders of the spectrum due to reduced

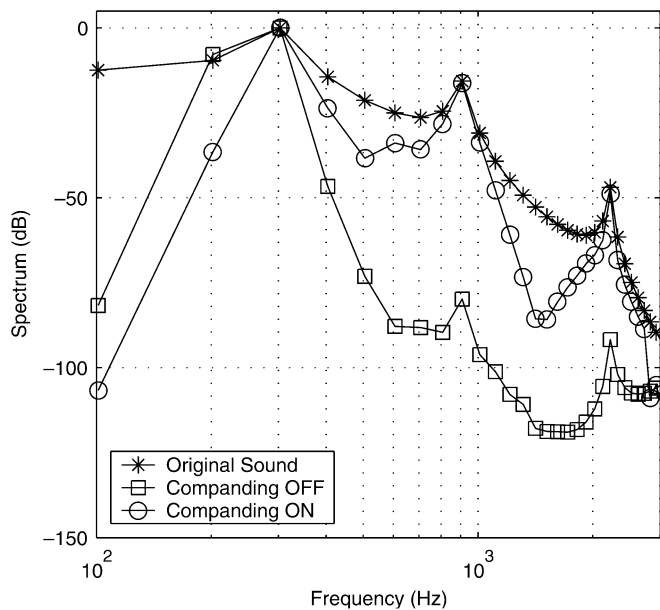


Fig. 13. Spectrum of the output of a vowel /u/. The companding-off case corresponds to $n_1 = 1$ and $n_2 = 1$. The companding-on case corresponds to $n_1 = 0.25$ and $n_2 = 1$. No zero-phase filters were used in either case.

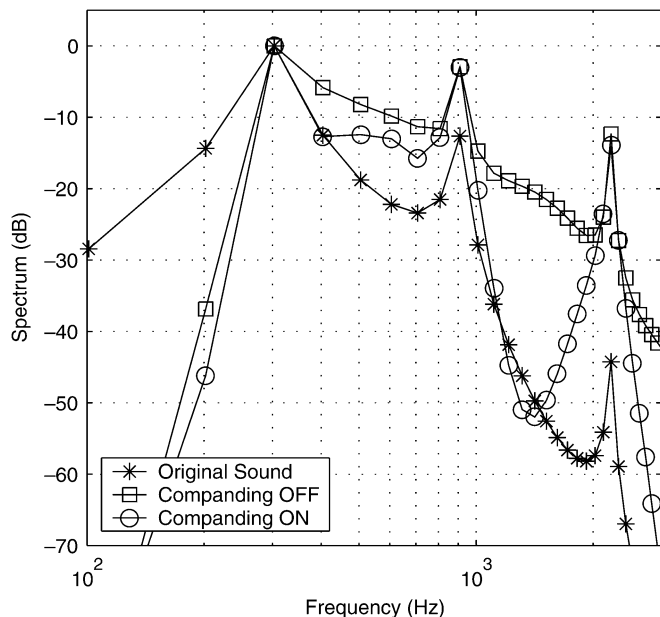


Fig. 14. Spectrum of the output of a vowel /u/. The companding-off case corresponds to $n_1 = 1$ and $n_2 = 0.3$. The companding-on case corresponds to $n_1 = 0.08$ and $n_2 = 0.25$. Zero-phase filters were used in both cases.

interference at the edges of the filter bank. In contrast, the companding-on strategy yields an output spectrum that is almost identical to that obtained with zero-phase filters (Fig. 11) because of its immunity to interference amongst channels.

Fig. 14 shows that the companding architecture performs multichannel syllabic compression of the sound without flattening the spectrum and reducing spectral contrast: In the figure, we compare the results of compression without companding ($n_1 = 1$, $n_2 = 0.3$) with the results of companding ($n_1 = 0.08$, $n_2 = 0.25$). The numbers were chosen to have formant peaks with the same amplitude in both cases. We see that compression alone degrades spectral contrast but

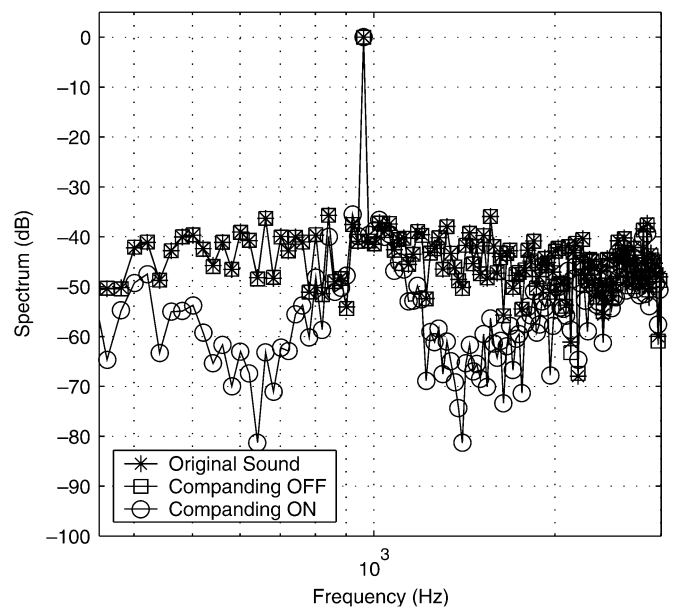


Fig. 15. Output spectrum of a 970-Hz sinusoid amidst Gaussian white noise. The suppression of the noise around the tone is evident. The original sound's spectrum is identical to the spectrum observed in the companding-off case.

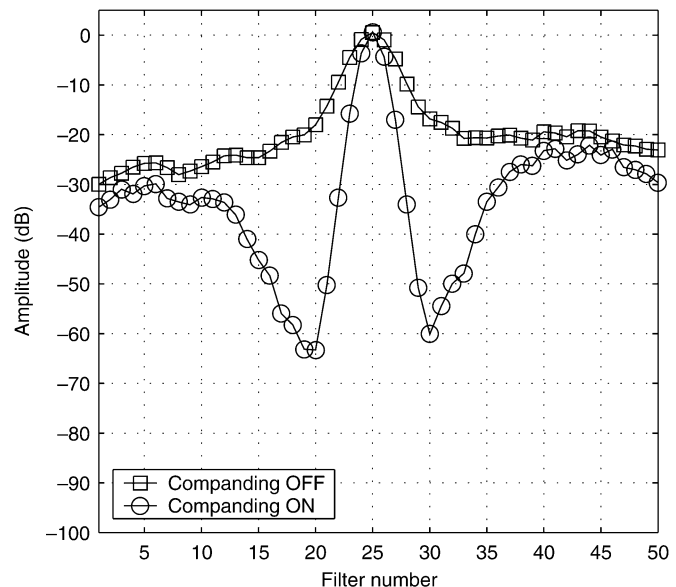


Fig. 16. Maximum output of the various channels (in 250 ms) versus channel number for the input of Fig. 15, i.e., a sine tone in noise.

companding is capable of compression while preserving good contrast in the spectrum.

It is possible to architect filter shapes and choose parameters to mimic auditory system or auditory nerve behavior [3]–[5]. We could customize the two-tone suppression extent for each channel by having different F filters for each channel. It may be advantageous to have more two-tone suppression of low-frequency tones by high-frequency tones such that the low-frequency formant does not create excessive suppression of higher frequencies in the damage-impaired cochlea.

C. Tone-in-Noise Experiments

Figs. 15 and 16 illustrate the effects of noise suppression in the companding architecture: the input to the architecture is a

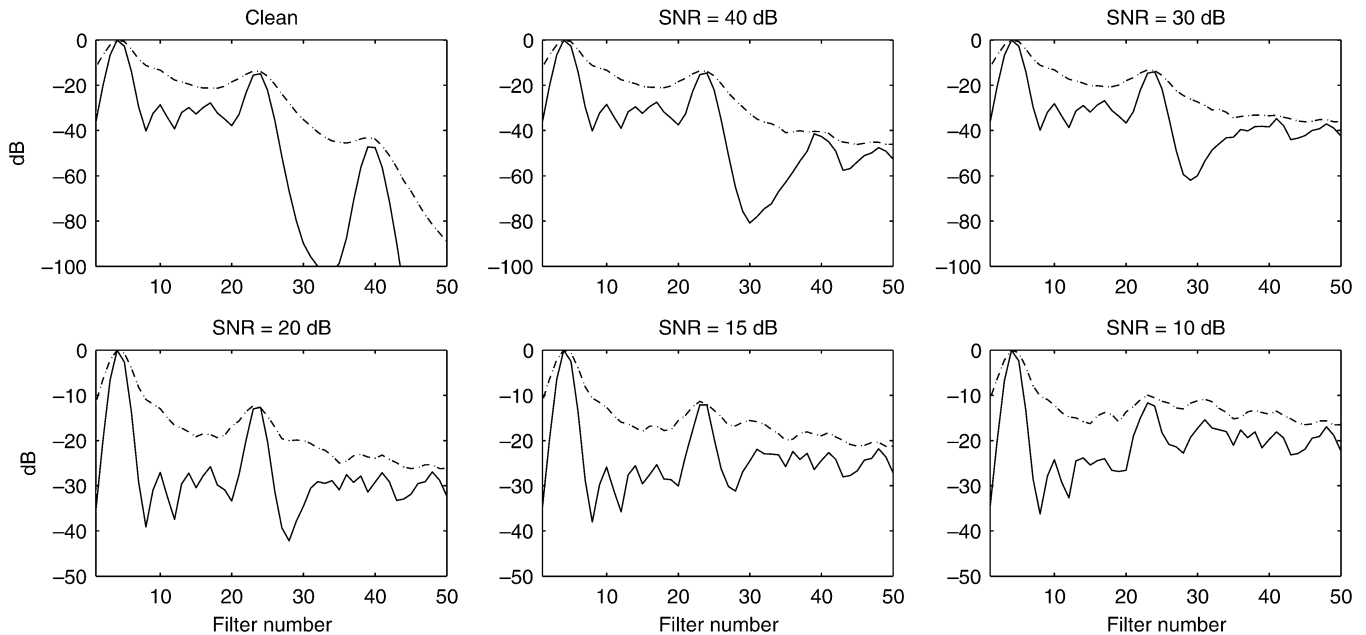


Fig. 17. Maximum output of each channel versus filter number for the vowel /u/ in Gaussian noise for different SNR values. The continuous line represents the companding-on architecture and the dashed line represents the companding-off architecture.

970-Hz sinusoid amidst Gaussian white noise. The output and input spectra extracted via FFT operations are shown in Fig. 15. We see that the tone suppresses the noise in regions of the spectrum near it. Fig. 16 plots the maximum output of every channel (in 250 ms) versus channel number. We see that companding suppresses the effects of channels near the strongest channel.

D. Vowel-in-Noise Experiments

We performed experiments with the vowel /u/ in Gaussian noise. The noise was filtered with a second-order low- and high-pass filter. The cutoff frequencies were 100 Hz and 3 kHz, respectively. Fig. 17 illustrates the maximum output of every channel versus filter number with different signal-to-noise ratio (SNR). The continuous line represents the companding-on architecture and the dashed line represents the companding-off architecture (parameters are set as in Fig. 11). The top three plots illustrate the progressive loss of the third formant and the bottom three plots illustrate the progressive loss of the second formant. We see that the companding architecture improves the recognition of all formants.

IV. LOMPANDED SPECTRA: COMPANDING WITHOUT SUMMATION

We can use our companding architecture to perform non-linear spectral analysis if we omit the final summation operation at the end of Fig. 1(b). The local winner-take-all properties of the architecture then enhance the peaks in the spectrum just like tone-to-tone suppression and lateral inhibition in the auditory system do. We shall now discuss potential uses of such *companded spectra* for cochlear-implant processing and speech-recognition front ends.

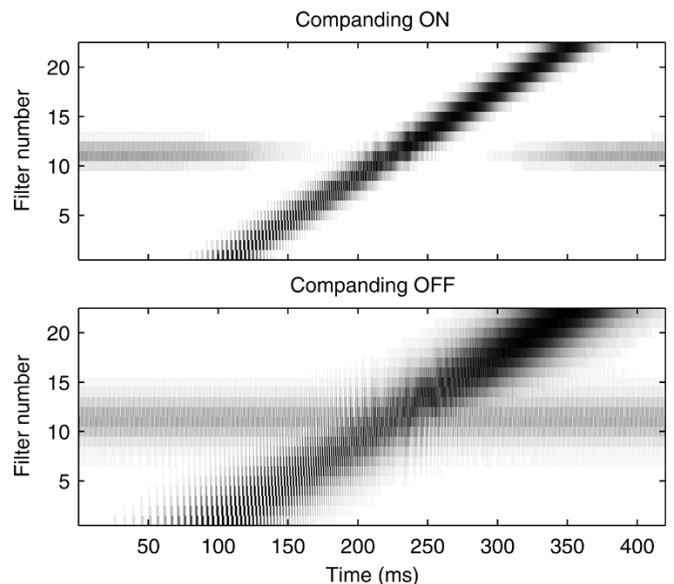


Fig. 18. Spectrogram-like plots illustrating tone-to-tone suppression: The suppressed input is a sinusoid fixed at 1000 Hz and the suppressor is a logarithmic chirp with an amplitude five times that of the tone. In the top figure, the companding strategy is active ($n_1 = 0.3$) and in the lower figure the companding strategy is disabled ($n_1 = 1$).

A. Cochlear-Implant Processing

Strategies called *N-of-M strategies* in cochlear-implant processing pick only those M channels with the largest spectral energies amongst a set of N channels for electrode stimulation [22]. Our companding architecture naturally enhances channels with spectral energies significantly above their surround and suppresses weak channels. Effectively we can create an analog N-of-M-like strategy without making any explicit decisions or completely shutting off weak channels. The companding strategy could thus preserve more information and degrade

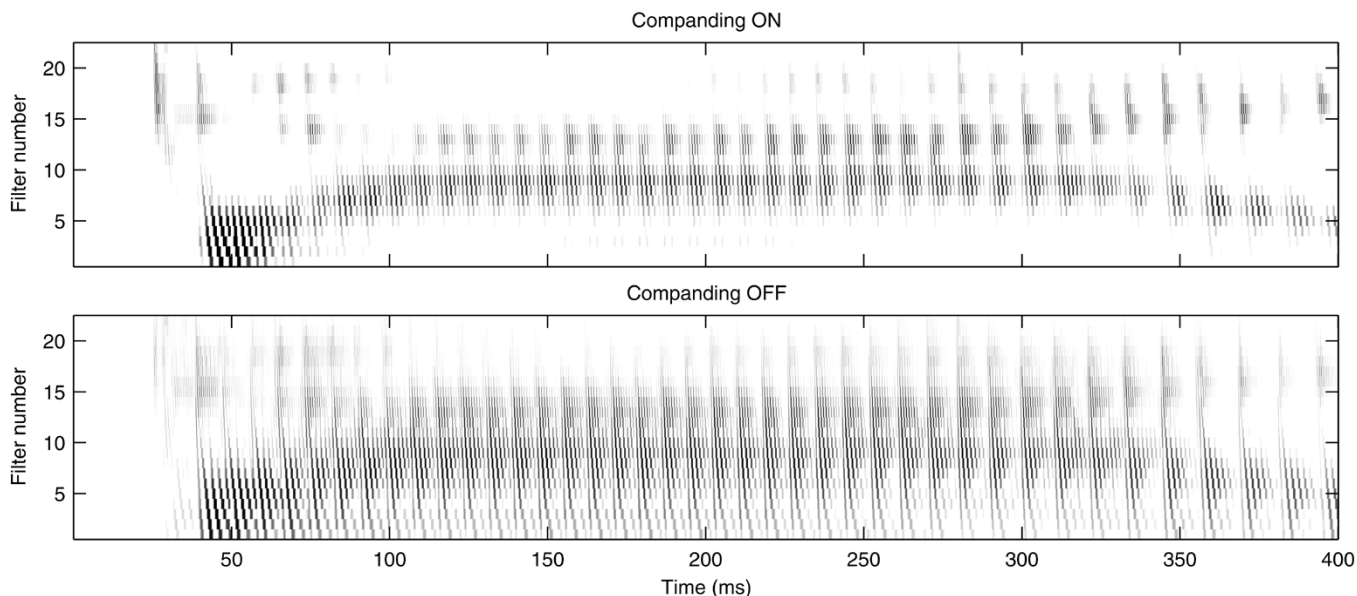


Fig. 19. Spectrogram-like plots for the word “die” illustrating the clarifying effect of companding. In the top figure, the companding strategy is active ($n_1 = 0.3$) and in the lower figure the companding strategy is disabled ($n_1 = 1$).

more gracefully in low signal-to-noise environments than the N-of-M strategy.

Given that improving patient performance in noise is one of the key unsolved problems in cochlear implants, companding spectra could yield a useful spectral representation for implant processing. The effects of compression and the two-tone suppression can be modeled in an intertwined fashion as in the biological cochlea and customized to each patient. The parameter n_2 will always be between 0 and 1 in this application because we need to compress the wide dynamic range of input sounds to the limited electrode dynamic range of the patient. The architecture requires filters of modest Q and relatively low order and is amenable to very low-power analog VLSI implementations such as those described in [23] and [24].

Recently, Loizou [14] showed that cochlear-implant listeners required 4–6 dB contrast to identify vowels with relatively high accuracy and that some of them obtained significantly higher scores with vowels enhanced to 6 dB with a spectral-contrast enhancement algorithm. Previously, he showed [15] that consonant intelligibility can be improved by enhancing differences in the pattern of the signal-processor channel outputs. These results show that it is likely that our architecture will improve cochlear-implant patient performance due to the presence of spectral-contrast enhancement.

Figs. 18 and 19 show the evolution in time of the channel outputs of Fig. 1(b) right before the final summation point for two inputs. A dark black color encodes a large positive signal. We used 22 channels logarithmically spaced between 300 and 3500 Hz with $q_1 = 1.5$, $q_2 = 4.5$, $n_1 = 0.3$, $n_2 = 1$, and $w = 10$. Effectively, Figs. 18 and 19 are spectrogram-like plots for companding spectra. In these plots, we used $F_i(s) = F_i'(s)$, $G_i(s) = G_i'(s)$, and first-order low-pass filter in the envelope detector.

In the experiment illustrated by Fig. 18, the input (not shown) consists of a fixed tone at 1000 Hz with an amplitude that is 1/5 the amplitude of a logarithmically chirped tone. The top part of

Fig. 18 shows that the chirp suppresses the background weak tone when its frequency is near that of the tone and companding is on ($n_1 = 0.3$). The bottom part of Fig. 18 shows that the background weak tone is confounded with the chirp when there is no companding ($n_1 = 1$). As we have previously discussed in Section III, we may vary the amount and extent of suppression by altering compression or filter parameters. Note also that when companding is on, the overall response is sharper due to fewer channels being active.

In the experiment illustrated by Fig. 19, the input is an intentionally low-quality rendition of the word “die” with two visible formant transitions. The top part of Fig. 19 shows that the companding architecture is able to follow the formant transitions with clarity and suppress the surrounding clutter. In contrast, the bottom part of Fig. 19 shows that, in the absence of companding, the formant transitions lie buried in an environment with lots of active channels and lack clarity.

B. Speech-Recognition Front-Ends

The use of Automatic Gain Control strategies for modeling forward masking in filter-bank front ends for automatic speech recognition (ASR) has been described in [25] and shown to be important in noisy environments. Our companding architecture adds simultaneous masking through nonlinear interactions to achieve compression without degrading spectral contrast. Thus, it offers promise for speech-recognition front ends in noisy environments. The architecture is also very amenable to low-power analog VLSI implementations, which are important for portable speech recognizers of the future.

V. CONCLUSIONS

Our companding architecture performs multichannel syllabic compression without degrading local spectral contrast due to the presence of the two-tone suppression. The two-tone suppression arises from implicit nonlinear interactions in the architec-

ture and is not explicitly due to any interactions between channels. The architecture is inspired by suggestions from recent biophysical experiments that suggest that local winner-take-all behavior and formant capture, which are seen in healthy auditory nerves, are not present in impaired cochleae. The compression and the two-tone suppression properties of the architecture may easily be altered by changing filter shapes and compression and expansion parameters. Due to its simplicity, its ease of programmability, its modest requirements on filter Q 's and filter order, its ability to suppress interference effects when channels are combined, and its ability to enhance noisy spectra, the architecture is useful for hearing aids, cochlear-implant processing, and speech-recognition front ends. In effect, a nonlinear spectral analysis may be performed generating a "companding spectrum". The architecture is suited to low-power analog VLSI implementations.

REFERENCES

- [1] L. Turicchia and R. Sarpeshkar, "The silicon cochlea: From biology to bionics," in *Biophysics of the Cochlea: From Molecules to Models*, A. W. Gummer, Ed. Singapore: World Scientific, 2003, pp. 417–423.
- [2] R. Sarpeshkar, R. F. Lyon, and C. A. Mead, "A low-power wide-dynamic-range analog VLSI cochlea," *Analog Integr. Circuits Signal Process.*, vol. 16, no. 3, pp. 245–274, Aug. 1998.
- [3] J. R. Schilling, R. L. Miller, M. B. Sachs, and E. D. Young, "Frequency-shaped amplification changes the neural representation of speech with noise-induced hearing loss," *Hearing Res.*, vol. 117, pp. 57–70, Mar. 1998.
- [4] R. L. Miller, B. M. Calhoun, and E. D. Young, "Contrast enhancement improves the representation of ϵ -like vowels in the hearing-impaired auditory nerve," *J. Acoust. Soc. Amer.*, vol. 106, no. 5, pp. 2693–708, Nov. 1999.
- [5] M. B. Sachs, I. C. Bruce, R. L. Miller, and E. D. Young, "Biological basis of hearing-aid design," *Ann. Biomed. Eng.*, vol. 30, no. 2, pp. 157–68, 2002.
- [6] H. T. Bunnell, "On enhancement of spectral contrast in speech for hearing-impaired listeners," *J. Acoust. Soc. Amer.*, vol. 88, no. 6, pp. 2546–2556, Dec. 1990.
- [7] Y. M. Cheng and D. O'Shaughnessy, "Speech enhancement based conceptually on auditory evidence," *IEEE Trans. Signal Process.*, vol. 39, no. 1, pp. 1943–54, Jan. 1991.
- [8] M. A. Stone and B. C. J. Moore, "Spectral feature enhancement for people with sensorineural hearing impairment: Effects on speech intelligibility and quality," *J. Rehabil Res. Dev.*, vol. 29, no. 2, pp. 39–56, 1992.
- [9] T. Baer, B. C. J. Moore, and S. Gatehouse, "Spectral contrast enhancement of speech in noise for listeners with sensorineural hearing impairment: Effects on intelligibility, quality, and response times," *J. Rehabil Res. Dev.*, vol. 30, no. 1, pp. 49–72, 1993.
- [10] Z. Ribic, Y. Jun, and M. Latzel, "Adaptive spectral contrast enhancement based on masking effect for the hearing impaired," in *Proc. 1996 IEEE Int. Conf. Acoustics, Speech, and Signal Processing Conf.*, vol. 2, May 2, 1996, pp. 937–940.
- [11] J. Yang, F. Luo, and A. Nehorai, "Spectral contrast enhancement: Algorithms and comparisons," *Speech Commun.*, vol. 39, Jan. 2003.
- [12] T. Painter and A. Spanias, "Perceptual coding of digital audio," *Proc. IEEE*, vol. 88, no. 4, pp. 451–515, Apr. 2000.
- [13] M. W. Baker, T. K.-T. Lu, C. Salthouse, J. J. Sit, S. Zhak, and R. Sarpeshkar, "A 16-channel analog VLSI processor for bionic ears and speech-recognition front ends," in *Proc. 2003 IEEE Custom Integrated Circuits Conf.*, Orlando, FL, Oct. 2003.
- [14] P. C. Loizou and O. Poroy, "Minimum spectral contrast needed for vowel identification by normal hearing and cochlear implant listeners," *J. Acoust. Soc. Amer.*, vol. 110, no. 3, pp. 1619–1627, Sep. 2001.
- [15] M. Dorman and P. Loizou, "Improving consonant intelligibility for ineraid patients fit with CIS processors by enhancing contrast among channel outputs," *Ear Hearing*, vol. 17, no. 4, pp. 308–313, Aug. 1996.
- [16] J. Lyzenga, J. M. Festen, and T. Houtgast, "A speech enhancement scheme incorporating spectral expansion evaluated with simulated loss of frequency selectivity," *J. Acoust. Soc. Amer.*, vol. 112, no. 3, pp. 1145–1157, Sep. 2002.
- [17] B. C. J. Moore, "Speech processing for the hearing-impaired: Successes, failures, and implications for speech mechanisms," *Speech Commun.*, vol. 41, no. 1, pp. 81–91, Aug. 2003.
- [18] R. Dolby, "An audio noise reduction system," *J. Audio Eng. Soc.*, vol. 15, no. 4, Oct. 1967.
- [19] Y. Tsvividis, "Externally linear, time-invariant systems and their application to companding signal processors," *IEEE Trans. Circuits Syst. II*, vol. 44, no. 2, pp. 65–85, Feb. 1997.
- [20] D. Frey, Y. Tsvividis, G. Efthivoulidis, and N. Krishnapura, "Syllabic-Companding log domain filters," *IEEE Trans. Circuits Syst. II*, vol. 48, no. 4, pp. 329–339, Apr. 2001.
- [21] C. D. Salthouse and R. Sarpeshkar, "A practical micropower programmable bandpass filter for use in bionic ears," *IEEE J. Solid-State Circuits*, vol. 38, no. 1, pp. 63–70, Jan. 2003.
- [22] P. C. Loizou, "Mimicking the human ear," *IEEE Signal Process. Mag.*, vol. 15, no. 5, pp. 101–30, May 1998.
- [23] R. Sarpeshkar, R. F. Lyon, and C. A. Mead, "A low power wide dynamic range analog VLSI cochlea," *Analog Integr. Circuits Signal Process.*, vol. 16, no. 3, pp. 245–274, Aug. 1998.
- [24] T. K.-T. Lu, M. Baker, C. Salthouse, J. J. Sit, S. Zhak, and R. Sarpeshkar, "A micropower analog VLSI processing channel for bionic ears and speech recognition front ends," in *Proc. IEEE Symp. Circuits and Systems*, May 2003.
- [25] B. Strobe and A. Alwan, "A model of dynamic auditory perception and its application to robust word recognition," *IEEE Trans. Speech Audio Process.*, vol. 5, no. 5, pp. 451–64, Sep. 1997.



L. Turicchia was born in Bologna, Italy. He received the Laurea degree in electrical engineering from the University of Padova, Padova, Italy. He is currently pursuing the Ph.D. degree in computer science in the Department of Mathematics and Computer Science, University of Udine, Udine, Italy.

Since 2002, he has been with the Analog VLSI and Biological Systems Group, Massachusetts Institute of Technology, Cambridge. His research involves speech recognition based on properties of the human auditory system, and nonlinear signal processing.



R. Sarpeshkar (M'98) received the B.S. degree in electrical engineering and physics from Massachusetts Institute of Technology (MIT), Cambridge, in 1990 and the Ph.D. degree from California Institute of Technology, Pasadena, CA, in 1997.

He then joined Bell Labs as a Member of Technical Staff. Since 1999, he has been on the faculty at MIT, where he is currently the Robert J. Shillman Associate Professor of electrical engineering and computer science. He heads a group on analog VLSI and biological systems at MIT. He holds over a dozen patents and has authored several publications, including one that was featured on the cover of *Nature*. His research interests include analog and mixed-signal VLSI, ultra-low-power circuits and systems, biologically inspired circuits and systems, and control theory.

Dr. Sarpeshkar has received several awards, including the Packard Fellow Award given to outstanding young faculty, the Office of Naval Research Young Investigator Award, and the National Science Foundation Career Award.