A new sound coding strategy for suppressing noise in cochlear implants

Yi Hu and Philipos C. Loizou[a)]
Department of Electrical Engineering, University of Texas at Dallas, Richardson, Texas 75083-0688

(Received 15 August 2007; revised 10 April 2008; accepted 16 April 2008)

In the *n-of-m* strategy, the signal is processed through *m* bandpass filters from which only the *n* maximum envelope amplitudes are selected for stimulation. While this maximum selection criterion, adopted in the advanced combination encoder strategy, works well in quiet, it can be problematic in noise as it is sensitive to the spectral composition of the input signal and does not account for situations in which the masker completely dominates the target. A new selection criterion is proposed based on the signal-to-noise ratio (SNR) of individual channels. The new criterion selects target-dominated (SNR $\geq$ 0 dB) channels and discards masker-dominated (SNR $<$ 0 dB) channels. Experiment 1 assessed cochlear implant users' performance with the proposed strategy assuming that the channel SNRs are known. Results indicated that the proposed strategy can restore speech intelligibility to the level attained in quiet independent of the type of masker (babble or continuous noise) and SNR level (0–10 dB) used. Results from experiment 2 showed that a 25% error rate can be tolerated in channel selection without compromising speech intelligibility. Overall, the findings from the present study suggest that the SNR criterion is an effective selection criterion for *n-of-m* strategies with the potential of restoring speech intelligibility.
© *2008 Acoustical Society of America.* [DOI: 10.1121/1.2924131]

PACS number(s): 43.66.Ts, 43.66.Sr [RYL] Pages: 498–509

I. INTRODUCTION

Current cochlear implant manufacturers offer several speech coding strategies to users (see review by Loizou, 2006). The Cochlear Corporation, for instance, offers the advanced combination encoder (ACE) strategy and the continuous interleaved sampling (CIS) strategy (Vandali *et al.*, 2000). Both ACE and CIS strategies are based on channel vocoder principles dating back to Dudley's VODER in the 1940s (Dudley, 1939; Peterson and Cooper, 1957). Signal is decomposed into a small number of bands (16–22) via the fast Fourier transform or a bank of bandpass filters, and the envelopes are extracted from each band. The envelopes are used to modulate biphasic pulses which are in turn sent to the electrodes for stimulation. The number of envelopes (and number of electrode sites) selected for stimulation at each cycle differs between the CIS and ACE strategies. In the ACE strategy, only a subset *n* (n=8–10) out of 22 envelopes is selected and used for stimulation at each cycle and all 22 electrode sites are utilized for stimulation. In the CIS strategy, a fixed number (8–10) of envelopes are computed, and only the corresponding electrode sites (8–10) are used for stimulation. Several studies (Kim *et al.*, 2000; Kiefer *et al.*, 2001; Skinner *et al.*, 2002a, 2002b) have shown that most Nucleus-24 users prefer the ACE over the CIS strategy[1] and in most conditions perform as well or slightly better on speech recognition tasks (Kiefer *et al.*, 2001; Skinner *et al.*, 2002b). The ACE strategy belongs to the general category of *n-of-m* strategies, which select (based on an appropriate criterion) *n* envelopes out of a total of *m* ($n<m$) envelopes for stimulation, where *m* is typically set to the number of electrodes available.

The selection criterion used in the ACE strategy is the maximum amplitude. More specifically, 8–12 maximum envelope amplitudes are typically selected out of 22 envelopes for stimulation in each cycle.[2] Provided the signal is preemphasized for proper spectral equalization (needed to compensate for the inherent low-pass nature of the speech spectrum), the maximum selection works well as it captures the perceptually relevant features of speech such as the formant peaks. In most cases, the maximum selection criterion performs spectral peak selection. Alternative selection criteria were proposed by Noguiera *et al.* (2005) based on a psychoacoustic model currently adopted in audio compression standards (MP3). In their proposed scheme, the amplitudes which are farthest away from the estimated masking thresholds are retained. The idea is that amplitudes falling below the masking threshold would not be audible and should therefore be discarded. The new strategy was tested on sentence recognition tasks in speech-shaped noise (SSN) at 15 dB signal-to-noise ratio (SNR) and compared to ACE. A large improvement over ACE was noted when four channels were retained in each cycle, but no significant difference was found when eight channels were retained.

The maximum selection criterion adopted in the ACE strategy works well in quiet as cochlear implant (CI) users fitted with the ACE strategy have been found to perform as well or slightly better than when fitted with the CIS strategy (Kiefer *et al.*, 2001; Skinner *et al.*, 2002b). In the study by Skinner *et al.* (2002b), 6 of the 12 subjects tested had significantly higher CUNY sentence scores with the ACE strategy than with the CIS strategy. Group mean scores on CUNY sentence recognition were 62.4% with the ACE strat-

[a)]Author to whom correspondence should be addressed. Tel.: (972)883-4617. FAX: (972) 883-2710. Electronic mail: loizou@utdallas.edu.

egy and 56.8% with the CIS strategy. The ACE strategy offers the added advantage of prolonged battery life since not all electrodes need to be stimulated at a given instant. In noise, however, this criterion could be problematic for several reasons. First, the selected amplitudes could include information from the masker-dominated channels, thereby confusing the listeners as to which is the target and which is the masker. Second, the selection is done all the time for all segments of speech, including the low-energy segments where noise will most likely dominate and mask the target signal. Third, the maximum criterion may be influenced by the spectral distribution (e.g., spectral tilt) of the target and/or masker. If, for instance, the masker has high-frequency dominance, then the selection will be biased toward the high-frequency channels in that the high-frequency channels will be selected more often than the low-frequency channels. Clearly, a better selection criterion needs to be used to compensate for the above shortcomings of ACE in noise.

In the present study, we propose the use of channel-specific SNR as the criterion for selecting envelope amplitudes. More specifically, we propose to select a channel if its corresponding SNR is larger than or equal to 0 dB and discard channels whose SNR is smaller than 0 dB. The idea is that channels with low SNR, i.e., SNR <0 dB, are heavily masked by noise and therefore contribute little, if any, information about the speech signal. As such, those channels should be discarded. On the other hand, target-dominated channels (i.e., SNR $\geqslant 0$ dB) should be retained as they contain reliable information about the target. The proposed approach is partly motivated by the articulation index (AI) theory (French and Steinberg, 1947) and partly by intelligibility studies utilizing the ideal binary mask (IdBM) (e.g., Roman *et al.*, 2003; Brungart *et al.*, 2006; Li and Loizou, 2008). The AI model predicts speech intelligibility based on the proportion of time the speech signal exceeds the masked threshold (Kryter, 1962; ANSI, 1997). Hence, just like the AI model, the new SNR selection criterion assumes that the contribution of each channel to speech intelligibility depends on the SNR of that channel. As such, it is hypothesized that the SNR-based selection criterion will improve speech intelligibility.

A number of studies with normal-hearing listeners recently demonstrated high gains in intelligibility in noise with the IdBM technique (e.g., Roman *et al.*, 2003; Brungart *et al.*, 2006; Anzalone *et al.*, 2006; Li and Loizou, 2007, 2008). The IdBM takes values of 0 and 1, and is constructed by comparing the local SNR in each time-frequency (T-F) unit against a threshold (e.g., 0 dB). It is commonly applied to the T-F representation of a mixture signal and eliminates portions of a signal (those assigned to a "0" value) while allowing others (those assigned to a "1" value) to pass through intact. When the IdBM is applied to a finite number of channels, as in cochlear implants, it would retain the channels with a mask value of 1 (i.e., SNR $\geqslant 0$ dB) and discard the channels with a mask value of 0 (i.e., SNR <0 dB). Hence, the SNR selection criterion proposed in the present study is similar to the IdBM technique in many respects.

In the first experiment, we make the assumption that the true SNR of each channel is known at any given instance and assess performance of the proposed SNR selection criterion under ideal conditions. The results from this study will tell us about the full potential of using SNR as the new selection criterion and whether efforts need to be invested in finding ways to estimate the SNR accurately. It is not the intention of this study to compare the performance of ACE against CIS, as this has been done by others (Kiefer *et al.*, 2001; Skinner *et al.*, 2002b). Rather, the objective is to assess whether the new criterion, based on SNR, can restore speech intelligibility to the level attained in quiet as predicted by IdBM studies (Brungart *et al.*, 2006). One of the primary differences between prior IdBM studies and the present study (aside from the subjects used, normal-hearing versus cochlear implant users) is the number of channels used to process the stimuli. A total of 128 channels were used to synthesize the stimuli by Brungart *et al.*, (2006), while in the present study, only 16 channels of stimulation are available. Hence, it is not clear whether the intelligibility benefit seen in noise with the IdBM technique by normal-hearing listeners will carry through to cochlear implant users who only receive a limited amount of spectral information. The first experiment investigates the latter question. In a real system, signal processing techniques can be used to estimate the SNR (e.g., Ephraim and Malah, 1984; Hu *et al.*, 2007; Loizou, 2007, Chap. 7.3.3). Hence, in the second experiment, we assess the impact on intelligibility of the errors that can potentially be introduced when the SNR is estimated via an algorithm. The latter experiment addresses the real-world implementation of the proposed technique and will inform us about the required accuracy of SNR estimation algorithms.

II. EXPERIMENT 1: EVALUATION OF SNR CHANNEL SELECTION CRITERION

A. Subjects and material

A total of six postlingually deafened Clarion CII implant users participated in this experiment. All subjects had at least four years of experience with their implant device. Biographical data for all subjects are presented in Table I. IEEE sentences (IEEE subcommittee, 1969) corrupted in multitalker babble (MB) (ten female and ten male talkers) and continuous speech-shaped noise (SSN) were used in the test. The IEEE sentences were produced by a male speaker and were recorded in our laboratory in a double-walled sound-attenuating booth. These recordings are available from Loizou (2007). The babble recording was taken from the AUDITEC CD (St. Louis, MO). The continuous (steady-state) noise had the same long-term spectrum as the test sentences in the IEEE corpus.

B. Signal processing

The block diagram of the proposed speech coding algorithm is shown in Fig. 1. The mixture signal is first bandpass filtered into 16 channels and the envelopes are extracted in each channel using full-wave rectification and low-pass filtering (200 Hz, sixth-order Butterworth). The frequency spacing of the 16 channels is distributed logarithmically

TABLE I. Biographical data for the subjects tested.

Subject	Gender	Age (yr)	Duration of deafness prior to implantation (yr)	CI use (yr)	Number of active electrodes	Stimulation rate (pulses/s)	Etiology
S1	Female	60	2	4	15	2841	Medication
S2	Male	42	1	4	15	1420	Hydrops/Menier's syndrome
S3	Female	47	>10	5	16	2841	Unknown
S4	Male	70	3	5	16	2841	Unknown
S5	Female	62	<1	4	16	1420	Medication
S6	Female	53	2	4	16	2841	Unknown

across a 300 Hz–5.5 kHz bandwidth. In parallel, the true SNR values of the envelopes in each channel are determined by processing independently the masker and target signals via the same 16 bandpass filters and extracting the corresponding envelopes. The SNR computation process (shown at the bottom of Fig. 1) yields a total of 16 SNR values (1 for each channel) in each stimulation cycle (the SNR of channel i at time instant t is defined as $\mathrm{SNR}_i(t) = 10\log_{10}[x_i^2(t)/n_i^2(t)]$, where $x_i(t)$ is the envelope of the target signal and $n_i(t)$ is the envelope of the masker signal. Of the 16 mixture envelopes, only the mixture envelopes with SNR $\geqslant 0$ dB are retained while the envelopes with SNR <0 dB are discarded. The number of channels selected in each stimulation cycle (corresponding to a stimulation rate of 2841 pulses/s for most of our subjects) varies from 0 (i.e., none are selected) to 16 (i.e., all are selected). The selected mixture envelopes are finally smoothed with a low-pass filter (200 Hz) and log compressed to the subject's electrical dynamic range. The latter low-pass filter is used to ensure that the envelopes are smoothed and are free of any abrupt amplitude changes that may be introduced by the dynamic selection process.[3]

The SNR threshold used in the present study in the amplitude selection was 0 dB. This was a reasonable and intuitive criterion, as the objective was to retain the target-dominated channels and discard the masker-dominated channels. This threshold (0 dB) has been found to work well in prior studies utilizing the IdBM (Wang, 2005; Brungart *et al.*, 2006; Li and Loizou, 2008). The intelligibility study by Brungart *et al.* (2006) with normal-hearing listeners, for instance, showed that near perfect word identification scores can be achieved not only with a SNR threshold of 0 dB but with other SNR thresholds between −12 and 0 dB. Thus, we cannot exclude the possibility that other SNR thresholds can be used for cochlear implant users (and perhaps work equally well) and these thresholds might even vary across different subjects.

The above algorithm was implemented off-line in MATLAB and the stimuli were presented directly (via the auxiliary input jack) to CI users via the Clarion research interface platform. As the above algorithm was motivated by IdBM studies, we will be referring to it as the IdBM strategy.

C. Procedure

The listening task involved sentence recognition in noise. Subjects were tested in four different noise conditions: 5 and 10 dB SNRs in babble and 0 and 5 dB SNRs in SSN. Lower SNR levels were chosen for the SSN conditions to avoid ceiling effects as the pilot data showed that most subjects performed very well at 10 dB SNR. Two sentence lists (ten sentences/list) were used for each condition. The sen-

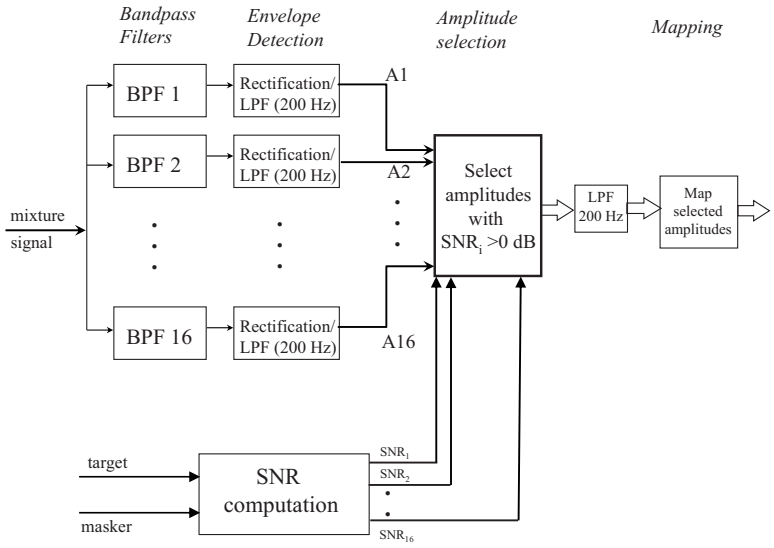
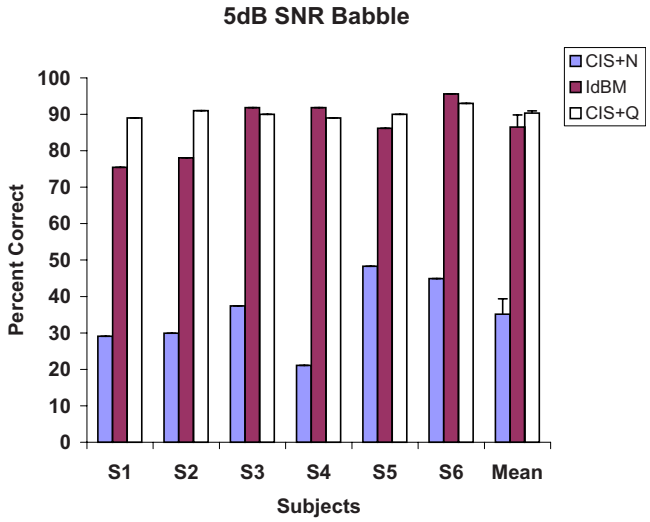
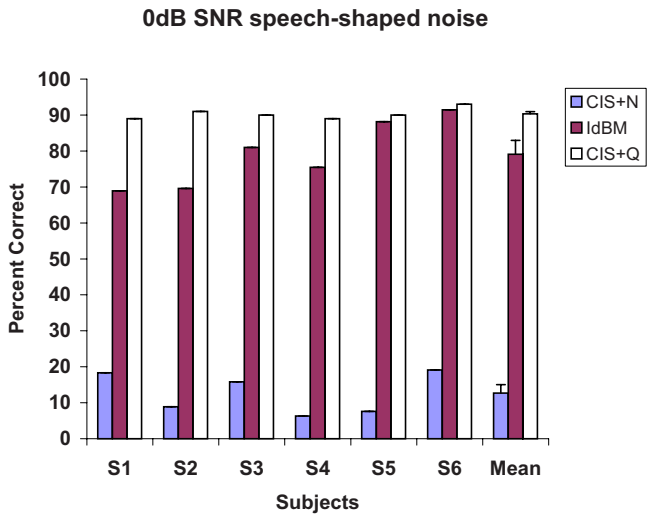


FIG. 1. Block diagram of the proposed coding strategy (IdBM).





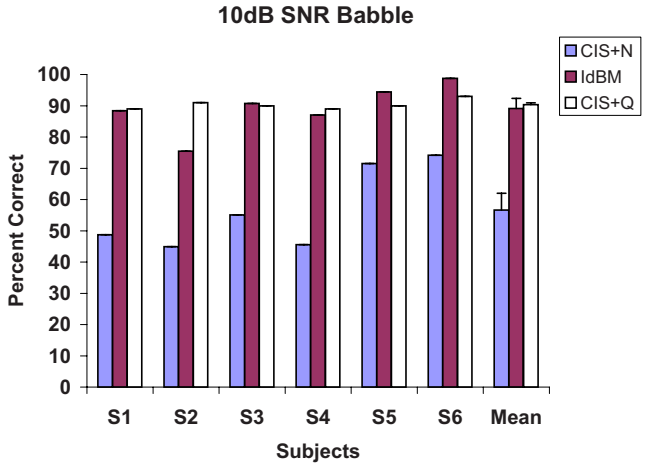


FIG. 2. (Color online) Percentage of correct scores of individual subjects, obtained with IdBM for recognition of sentences presented with MB at 5 and 10 dB SNRs. Scores obtained with the subject's everyday processor in quiet (CIS+Q) and in babble (CIS+N) are also shown for comparative purposes. The error bars indicate standard errors of the mean.

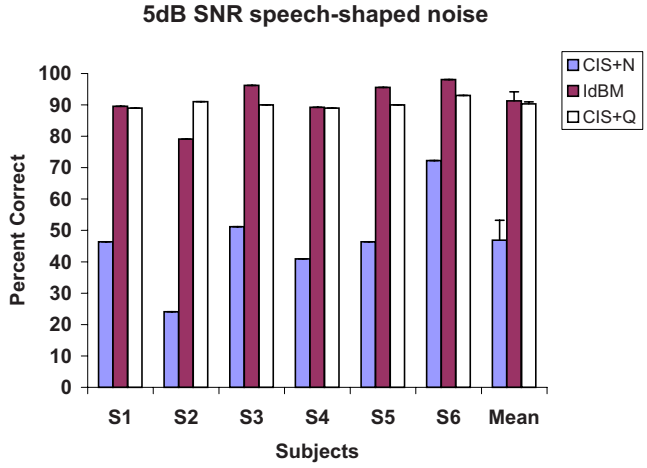


FIG. 3. (Color online) Percentage of correct scores of individual subjects, obtained with IdBM for recognition of sentences presented with SSN at 0 and 5 dB SNRs. Scores obtained with the subject's everyday processor in quiet (CIS+Q) and in noise (CIS+N) are also shown for comparative purposes. The error bars indicate standard errors of the mean.

tences were processed off-line in MATLAB by the proposed algorithm and presented directly (via the auxiliary input jack) to the subjects using the Clarion CII research platform at a comfortable level. For comparative purposes, subjects were also presented with unprocessed noisy sentences using the experimental processor. More specifically, the noisy sentences were processed via our own CIS implementation that utilized the same filters, same stimulation parameters (e.g., pulse width, stimulation rate, etc.), and same compression functions used in the IdBM strategy. Subjects were also presented with sentences in quiet. Sentences were presented to the listeners in blocks, with 20 sentences/block per condition. Different sets of sentences were used in each condition. Subjects were instructed to write down the words they heard, and no feedback was given to them during testing. The presentation order of the processed and control (unprocessed sentences in quiet and in noise) conditions was randomized for each subject.

D. Results and discussions

The sentences were scored by the percentage of the words identified correctly, where all words in a sentence were scored. Figure 2 shows the individual scores for all subjects for the multitalker babble (5 and 10 dB SNR) conditions and Fig. 3 shows the individual subject scores for the SSN (0 and 5 dB SNR) conditions. The scores obtained in quiet are also shown for comparison.

A separate statistical analysis was run for each masker condition. Two-way analysis of variance (ANOVA) (with repeated measures) was run to assess the effect of the noise level (quiet, 5 dB SNR, 10 dB SNR), effect of the processing (CIS versus IdBM), and possible interaction between the two. For the babble conditions, ANOVA indicated a highly significant effect of processing ($F[1,5]=142.5$, $p<0.0005$), significant effect of the noise level ($F[2,10]=51.5$, $p<0.0005$), and significant interaction ($F[2,10]=99.1$, $p<0.0005$). For the SSN conditions, ANOVA indicated a highly significant effect of processing ($F[1,5]=419.4$, $p<0.0005$), significant effect of noise level ($F[2,10]=105.7$, $p<0.0005$), and significant interaction ($F[2,10]=93.6$, $p<0.0005$).

Post hoc tests were run, according to Fisher's least significant difference (LSD) test, to assess differences between

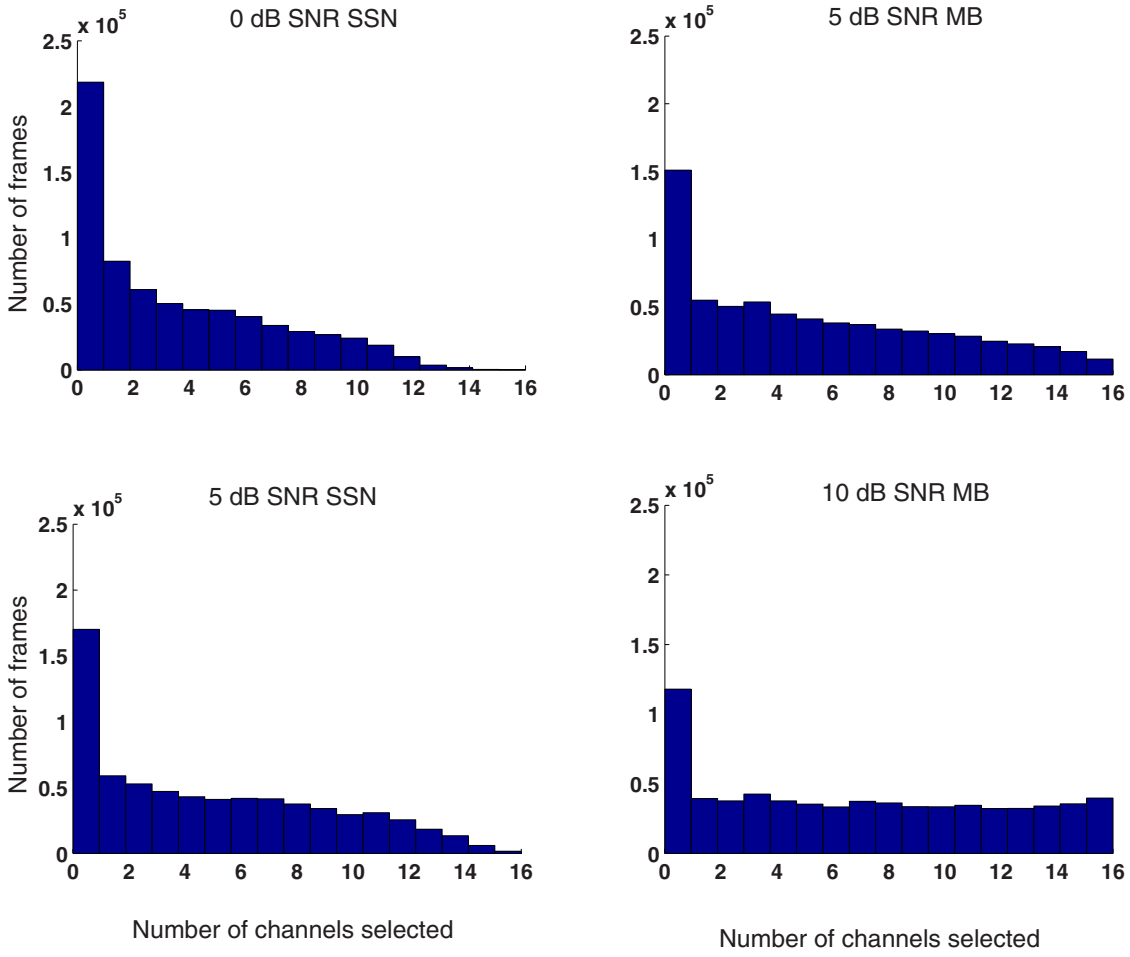


FIG. 4. (Color online) Histograms of the number of channels selected in each cycle by the IdBM strategy. The histograms were computed using a total of 20 IEEE sentences (~1 min of data) processed in the various conditions using MB and SSN as maskers.

scores obtained in noise with the proposed algorithm (IdBM) and scores obtained in quiet with the subject's daily strategy (CIS). Results indicated nonsignificant differences ($p > 0.3$) between scores obtained in noise with IdBM and scores obtained in quiet in nearly all conditions. The scores obtained with IdBM in 0 dB SNR SSN were significantly ($p = 0.009$) lower than the scores obtained in quiet. Nevertheless, the improvement over the unprocessed condition was quite dramatic, nearly 70 percentage points. The difference between scores obtained with IdBM and the scores obtained in noise with the subject's daily strategy (CIS) was highly significant ($p < 0.005$) in all conditions. Previous studies (Kiefer *et al.*, 2001; Skinner *et al.*, 2002b) have shown that ACE performs as well or better (by at most 10 percentage points) than CIS on various speech recognition tasks (some variability in the subject's scores and ACE versus CIS preferences was noted). Pilot data [4] collected with one subject indicated a similar outcome. Hence, we speculate that IdBM will perform significantly better than ACE in noise.

As shown in Figs. 2 and 3, the improvement obtained with IdBM over the subject's daily strategy was quite substantial and highly significant. The improvement was largest (nearly 70 percentage points) in 0 dB SSN as it improved consistently the subjects' scores from 10%–20% correct (base line noise condition) to 70%–90% correct. In nearly all conditions, the IdBM strategy restored speech intelligibility to the level obtained in quiet independent of the type of masker used (babble or steady noise) or input SNR level. The large improvements in intelligibility are consistent with those reported in IdBM studies (e.g., Brungart *et al.*, 2006), although in those studies, the signal was decomposed into 128 channels using fourth-order gammatone filters. The binary mask was applied in those studies to a fine T-F representation of the signal, whereas in the present study, it was applied to a rather coarse time-frequency representation (16 channels). Yet, the intelligibility gain was equally large.

Unlike the ACE strategy which selects the same number of channels (8–12 maximum) in each stimulation cycle based on the maximum criterion, the proposed IdBM strategy selects a different number of channels in each cycle depending on the SNR of each channel. In fact, IdBM may select as few as 0 or as many as 16 channels in each cycle for stimulation. To gain a better understanding of how many channels, on the average, are selected by IdBM or, equivalently, how many electrodes (on the average) are stimulated, we computed histograms of the number of channels selected in each cycle. The histograms were computed by using a total of 20 IEEE sentences processed in four noise conditions (two in MB and two in SSN). The four histograms are shown in Fig. 4 for the various SNR levels tested. As shown in Fig. 4, the most frequent number of channels selected was zero. In SSN, no channel was selected 25%–31% of the time, and in MB, no

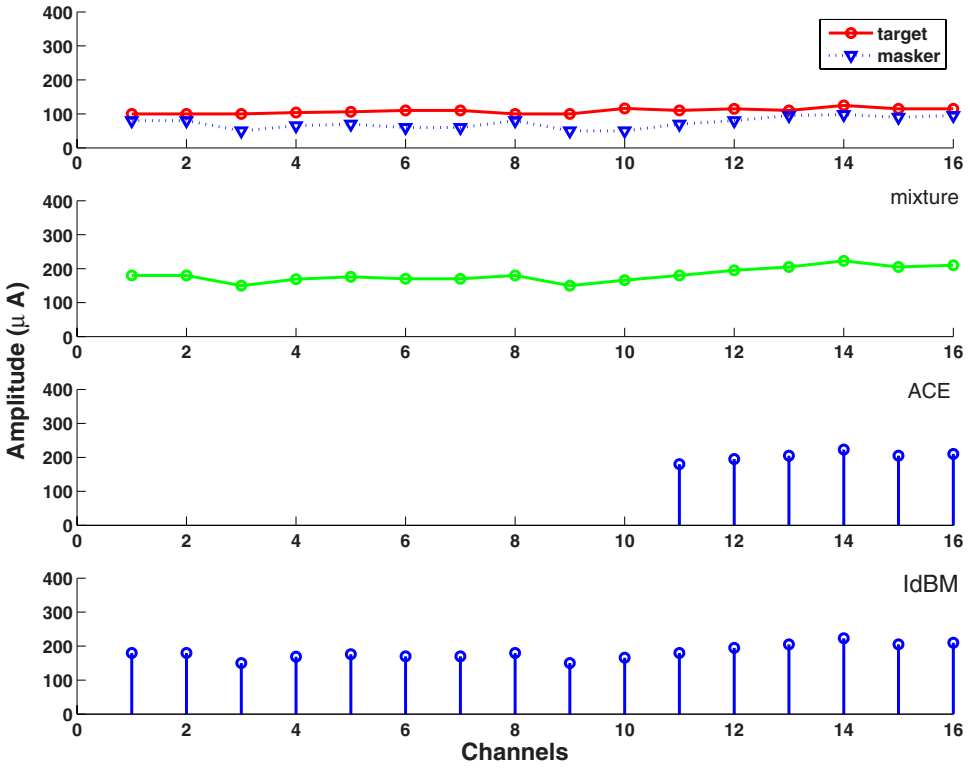


FIG. 5. (Color online) Example illustrating the selection process by ACE and IdBM strategies for a frame in which the target and mixture spectra are flat. The top panel shows the target and masker envelope amplitudes (in μAs) and the second panel from the top shows the mixture envelopes. The bottom two panels show the amplitudes selected by ACE and IdBM, respectively.

channel was selected 17%–21% of the time. This reflects the fact that low-energy speech segments (e.g., fricatives, stops, stop closures) occur quite often in fluent speech. These low-energy segments are easily and more frequently masked by background interference (compared to the high-energy voiced segments) yielding in turn a large number of occurrences of channels with SNR$<$0 dB. The distribution of the number of channels selected was skewed toward the low numbers for low SNR levels and became uniform for higher SNR levels. This reflects perhaps the fact that as the input global SNR level decreases, fewer channels with SNR $>$0 dB are available. The average number of channels selected (excluding zero) was five to six for the SSN conditions (0 and 5 dB SNRs) and seven to eight for the MB conditions (5 and 10 dB SNRs). The probability, however, of selecting a specific number of channels was roughly equal, indicating the flexibility of the SNR selection criterion in accommodating different target/masker scenarios and different spectral distributions of the input signal.

Two major factors influence the channel selection process and those include the spectral distribution of the target and the underlying SNR in each channel. Both factors are accommodated by the SNR selection criterion but not by the maximum selection criterion. Figures 5 and 6 show two examples in which the SNR criterion offers an advantage over the maximum criterion in selecting channels in the presence of background interference. Consider the example in Fig. 5 wherein the target (and mixture) spectrum is flat (e.g., fricative /f/) and the channel SNRs are positive. The IdBM strategy will select all channels, while the ACE strategy will only select a subset of the channels, i.e., the largest in amplitude. In this example, the ACE-selected channels might be perceived by listeners as belonging to a consonant with a rising-tilt spectrum or a spectrum with high-frequency dominance (e.g., /sh/, /s/, /t/). Hence, the maximum selection approach (ACE) might potentially create perceptual confusion between flat-spectra consonants (e.g., /f/, /th/, /v/) and rising-tilt or high-frequency spectra consonants (e.g., /s/, /t/, /d/). Consider a different scenario in Fig. 6, in which the target is completely masked by background interference, as it often occurs, for instance, during stop closures or weak speech segments. The IdBM strategy will not select any channel (i.e., no electrical stimulation will be provided) due to the negative SNR of all channels, whereas the ACE strategy will select a subset (the largest) of the channels independent of the underlying SNR. Providing no stimulation during stop closures or during low-energy segments in which the masker dominates is important for two reasons. First, it can, at least in principle, reduce masker-target confusions, particularly when the masker(s) is a competing voice(s) and happens to be present during speech-absent regions. In practice, an accurate algorithm would be required that would signify when the target is stronger than the masker (more on this in Sec. III D). Second, it can enhance access to voicing cues and reduce voicing and/or manner errors. As demonstrated in Fig. 4, the latter scenario happens quite often and the IdBM strategy can offer a significant advantage over the ACE strategy in target segregation. In brief, the IdBM strategy is more robust than ACE in terms of accommodating the spectral composition of the target and the underlying SNR. It is interesting to note that the SPEAK strategy (the predecessor of the ACE strategy), which was used in the Spectra 22 processor (Seligman and McDermott, 1995), selected five to ten channels depending on the spectral composition of the input signal, with an average number of six maxima. The SPEAK strategy, however, made no consideration for the underlying SNR of each channel and is no longer used in the latest Nucleus-24 speech processor (Freedom).

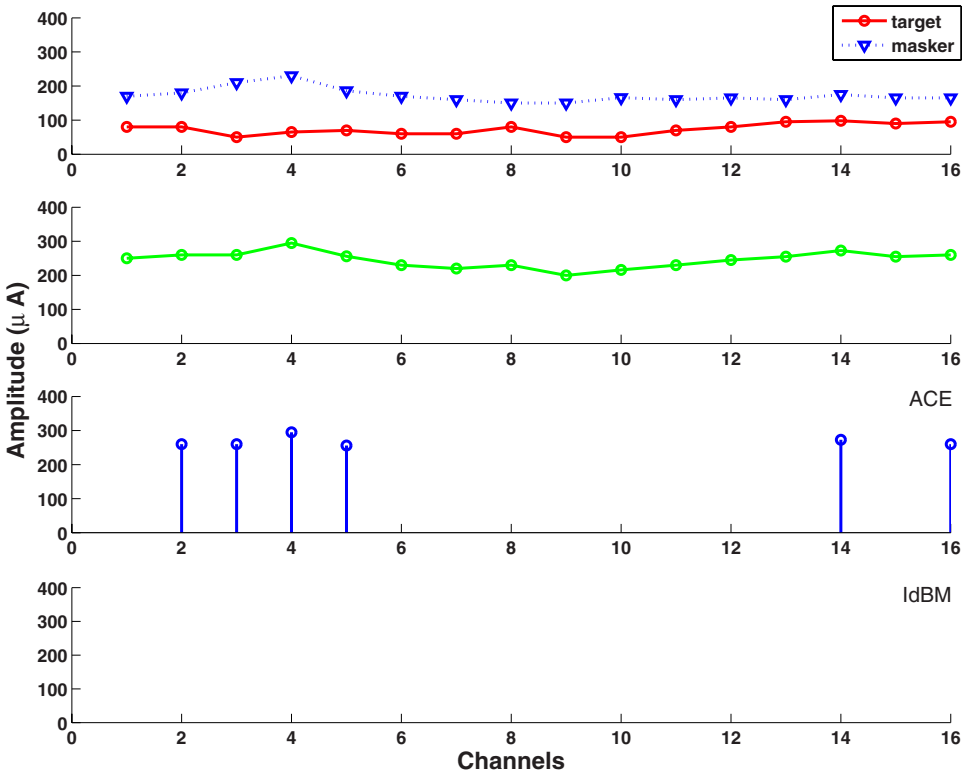


FIG. 6. (Color online) Example illustrating the selection process by ACE and IdBM strategies for a frame in which the masker dominates the target. The top panel shows the target and masker envelope amplitudes (in μAs) and the second panel from the top shows the mixture envelopes. The bottom two panels show the amplitudes selected by ACE and IdBM, respectively.

In fairness, it should be pointed out that there exist scenarios in which the maximum and SNR selection criteria select roughly the same channels (see example in Fig. 7). In voiced segments, for instance, where spectral peaks (e.g., formants) are often present, the maximum and SNR criteria will select roughly the same channels. Channels near the spectral peaks will likely have a high SNR (relative to the channels near the valleys) and will therefore be selected by both ACE and IdBM strategies. We therefore suspect that the partial agreement in channel selection between ACE and IdBM (more on this in experiment 2) occurs during voiced speech segments.

The SNR threshold used in the present study in the amplitude selection was 0 dB. Negative SNR thresholds might be used as well, as we acknowledge the possibility that masker-dominated channels could also contribute, to some

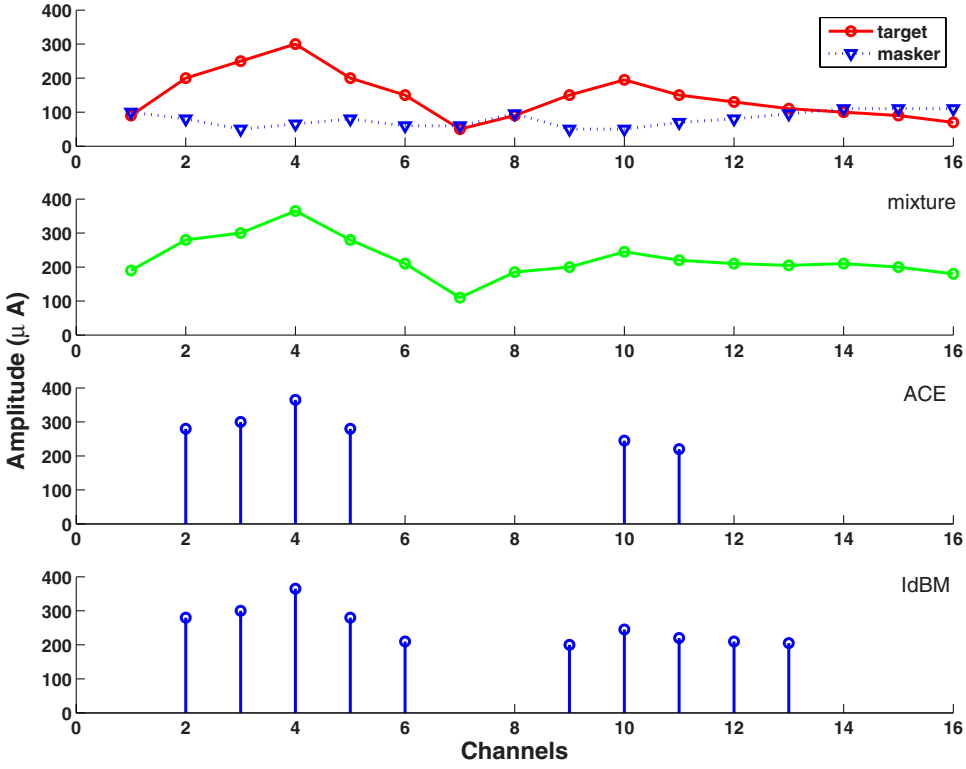


FIG. 7. (Color online) Example illustrating the selection process by ACE and IdBM strategies for a segment extracted from a vowel. The top panel shows the target and masker envelope amplitudes and the second panel from the top shows the mixture envelopes. The bottom two panels show the amplitudes selected by ACE and IdBM, respectively.

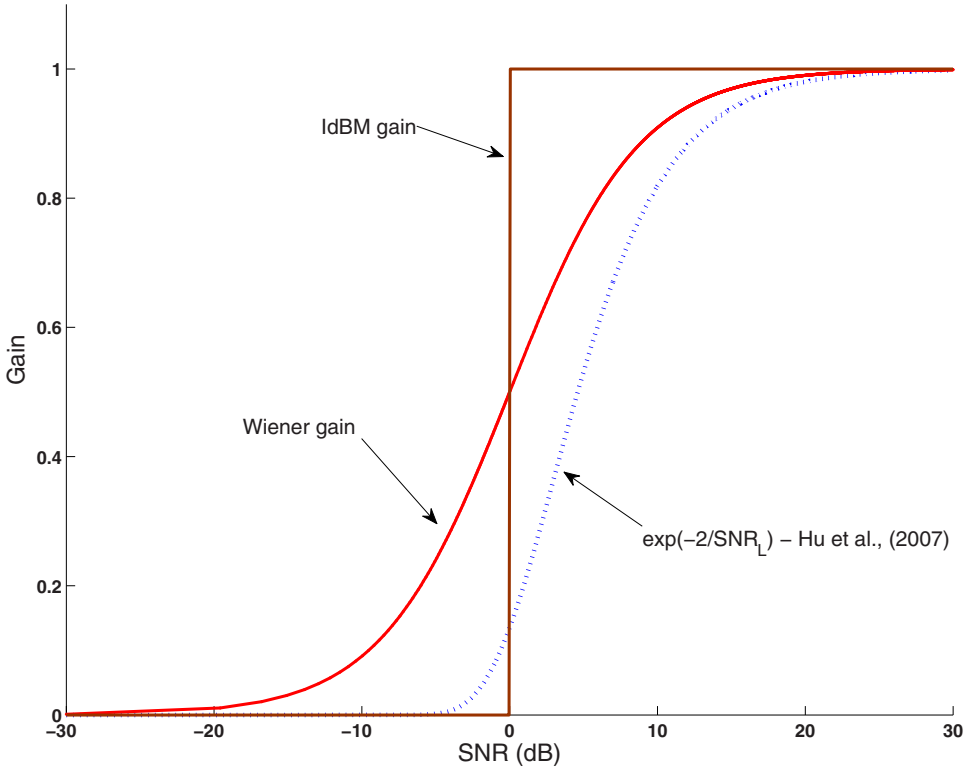


FIG. 8. (Color online) Plots of various gain functions that can be applied to mixture envelopes for noise suppression. The proposed IdBM strategy uses a binary function. The gain function used to Hu *et al.* (2007) was of the form $g(\mathrm{SNR}_L)=\exp(-2/\mathrm{SNR}_L)$, where SNR_L is the estimated SNR expressed in linear units. The Wiener gain function is superimposed for comparison and is given by the expression $g(\mathrm{SNR}_L)=\mathrm{SNR}_L/(\mathrm{SNR}_L+1)$.

extent, to intelligibility. In fact, Brungart *et al.* (2006) observed a plateau in performance (near 100% correct) for a range of SNR thresholds (−12 to 0 dB) smaller than 0 dB. Hence, we cannot exclude the possibility that other values (smaller than 0 dB) of SNR threshold might prove to be as effective as the 0 dB threshold.

The proposed *n-of-m* algorithm (IdBM) based on the SNR selection criterion can be viewed as a general algorithm that encompasses characteristics from both the ACE and CIS algorithms. When the SNR is sufficiently high (as, for instance, in quiet environments), $n=m$ (i.e., all channels will be selected) most of the time and the IdBM algorithm will operate like the CIS strategy. When n is fixed (for all cycles) to, say, $n=8$, then IdBM will operate similar to the ACE algorithm. In normal operation, the IdBM algorithm will be operating somewhere between the CIS and ACE algorithms. More precisely, in noisy environments, the value of n will not remain fixed but will change dynamically in each cycle depending on the number of channels that have positive SNR values.

The IdBM algorithm belongs to the general class of noise-reduction algorithms which apply a weight or a gain (typically in the range of 0–1) to the mixture envelopes (e.g., James *et al.*, 2002; Loizou, 2006; Hu *et al.*, 2007). The gain function of the IdBM algorithm is binary and takes the value of 0 if the channel SNR is negative and the value of 1 otherwise (see Fig. 8). Most noise-reduction algorithms utilize gain functions which provide a smooth transition from gain values near 0 (applied at extremely low SNR levels) to values of 1 (applied at high SNR values). Figure 8 provides two such examples. The Wiener gain function (known to be the optimal gain function in the mean-square error sense, see Loizou, 2007, Chap. 6) is plotted in Fig. 8 along with the sigmoidal-shaped function used by Hu *et al.* (2007). The implication of using sigmoidal-shaped functions, such as those shown in Figure 8, is that within a narrow range of SNR levels (which in turn depend on the steepness of the sigmoidal function), the envelopes (presumed to be masker dominant) will be heavily attenuated rather than zeroed out, as done in the IdBM algorithm when the SNR is negative. It remains to be seen whether such attenuation if applied to target-dominant envelopes will introduce any type of noise/speech distortion that is perceptible by the CI users. The findings by Hu *et al.* (2007) seem to suggest otherwise, but further experiments are warranted to investigate this possibility.

The binary function (see Fig. 8) used in the IdBM algorithm suggests turning off channels with SNR below threshold (0 dB, in this study) while keeping channels with SNR above threshold. In a realistic scenario, this might not be desirable as that will completely eliminate all environmental sounds, some of which (e.g., sirens, fire alarms, etc.) may be vitally important to the listener. One way to rectify this is to make the transition in the weighting function from 0 to 1 smooth rather than abrupt. This can be achieved by using a sigmoidal-shaped weighting function, such as the Wiener gain function shown in Fig. 8. Such a weighting function would provide environmental awareness, since the envelopes with SNR <0 dB would be attenuated rather than set to zero.

III. EXPERIMENT 2: EFFECT OF SNR ESTIMATION ERRORS ON SPEECH INTELLIGIBILITY

In the previous experiment, we assumed access to the true SNR value of each channel. In practice, however, the SNR of each channel needs to be estimated from the mixture

envelopes. Algorithms (e.g., Hu and Wang, 2004; Hu *et al.*, 2007) can be used in a practical system to estimate the SNR in each channel. Such algorithms will likely result in errors in estimating the SNR, as we lack access to the masker signal and, consequently, will make errors in selecting the right channels. In the present experiment, we assess the perceptual effect of SNR estimation errors on speech intelligibility. At issue is how accurate do SNR estimation algorithms need to be without compromising the intelligibility gain observed in experiment 1.

A. Subjects and material

Five of the six CI users who participated in experiment 1 also participated in the present experiment (subject S1 was not available for testing). The same speech material (IEEE Subcommittee, 1969) was used as in experiment 1. Different sentence lists were used for the new conditions.

B. Signal Processing

The stimuli were processed with the same method as described in experiment 1. We randomly selected a fixed number of channels in each cycle and reversed the decisions made using the true SNR values so as to model the errors that might be introduced when the channel SNRs are computed via an algorithm. That is, channels that were originally selected according to the ideal SNR criterion (i.e., SNR ≥ 0 dB) were now discarded. Similarly, channels that were originally discarded (i.e., SNR < 0 dB) were now retained. We varied the number of channels with erroneous decision from 2 to 12 (2, 4, 8, and 12) channels. In the 4-channel error condition, for instance, a total of 4 (out of 16) channels were wrongly discarded or selected in each cycle.

C. Procedure

The procedure was identical to that used in experiment 1. Subjects were tested with a total of 16 conditions (=4 channel errors $\times$ 2 maskers $\times$ 2 SNR levels). Two lists of sentences (i.e., 20 sentences) were used per condition, and none of the lists was repeated across conditions. The order of the test conditions was randomized for each subject.

The errors in channel selection were introduced off-line in MATLAB and presented directly (via the auxiliary input jack) to the CI users via the Clarion research interface platform.

D. Results and discussions

The sentences were scored in terms of percentage of words identified correctly (all words were scored). The top panel in Fig. 9 shows the mean percentage correct scores obtained in MB and the bottom panel of Fig. 9 shows the mean scores obtained in SSN, both as a function of the number of channels with errors. The mean scores obtained in experiment 1 for the five subjects tested are also shown and indicated as "0 number of channels with errors" for comparative purposes. A repeated-measure ANOVA with the main factors of SNR and number of channels with error was applied to the babble conditions. A significant effect of the

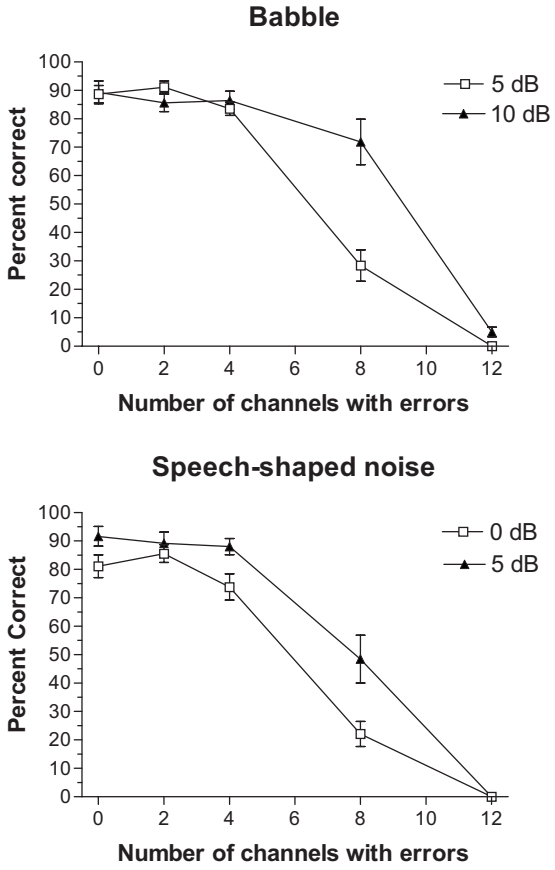


FIG. 9. Percentage of correct scores, averaged across subjects and shown as a function of the number of channels (out of 16) introduced with errors. The top panel shows the scores obtained in multitalker babble and the bottom panel shows the scores obtained in SSN. The error bars indicate standard errors of the mean.

SNR level ($F[1,4]=41.6$, $p=0.003$), significant effect of the number of channels with errors ($F[3,12]=230.7$, $p<0.0005$), and significant interaction ($F[3,12]=170.3$, $p<0.0005$) were observed. Two-way ANOVA (with repeated measures) applied to the speech-shaped conditions indicated significant effect of the SNR level ($F[1,4]=49.8$, $p=0.002$), significant effect of the number of channels with errors ($F[3,12]=222.2$, $p<0.0005$), and significant interaction ($F[3,12]=11.8$, $p=0.001$).

As shown in Fig. 9, performance remained high even when four channels were wrongly selected (or discarded). *Post hoc* tests (Fisher's LSD) confirmed that performance obtained with four wrongly selected (or discarded) channels was not statistically different ($p>0.05$) from the ideal performance obtained when no errors were introduced in the channel selection (Figs. 2 and 3). This was found to be true for both maskers and all SNR levels. In brief, the SNR selection algorithm (IdBM) presented in experiment 1 can tolerate up to a 25% (4 channels in error out of a total of 16 channels) error rate without compromising performance. With the exception of one condition (10 dB SNR babble), performance drops substantially (Fig. 9) for error rates higher than 25%.

The above findings raised the following question: How close is the maximum selection criterion used in ACE to the SNR criterion used in IdBM? This question led us to com-

TABLE II. Percentage of correct agreement between the channels selected by ACE and the channels selected by IdBM in different background conditions.

	SNR			
Masker type	0 dB	5 dB	5 dB	10 dB
SSN	60.7%	60.0%		
MB			57.4%	55.0%

pare the set of channels selected by ACE to those selected by IdBM. To that end, we processed 20 sentences through our implementation of the ACE and examined the agreement between the number of channels selected by ACE with those obtained by IdBM. To keep the proportion of channels (8/22=36%) selected by the commercial ACE strategy the same, we implemented a 6-of-16 strategy. For each stimulation cycle, we compared the six maximum channels selected by ACE with those selected by IdBM, and considered the selection decision correct if both ACE and IdBM selected or discarded the same channels. The results are tabulated in Table II for both masker types and all SNR levels tested. As shown in Table II, ACE makes the same decisions as IdBM (with regards to channel selection) 55%–60% of the time. The corresponding error rate is 40%, which falls short of the error rate needed to restore speech intelligibility (Fig. 9).

In the present experiment, we made no distinction between the two types of error that can potentially be introduced due to inaccuracies in SNR estimation. The first type of error occurs when a channel that should be discarded (because SNR$<$0 dB) is retained, and the second type of error occurs when a channel that should be retained (because SNR$>$0 dB) is discarded. From signal detection theory, we can say that the first type of error is similar to type I error (false alarm) and the second type of error is similar to type II error [5] (miss). The type I error will likely introduce more noise distortion or more target-masker confusion, as channels that would otherwise be discarded (presumably belonging to the masker or dominated by the masker) would now be retained. The type II error will likely introduce target speech distortion, as it will discard channels that are dominated by the target signal and should therefore be retained. The perceptual effect of these two types of errors introduced is likely different (e.g., Li and Loizou, 2008). Further experiments are thus needed to assess the effect of these two types of errors on speech intelligibility by CI users.

The present study, as well as others with normal-hearing listeners (e.g., Brungart *et al.*, 2006; Li and Loizou, 2007, 2008), have demonstrated the full potential of using the SNR selection criterion to improve (and in some cases restore) intelligibility of speech in multitalker or other noisy environments. Algorithms capable of estimating the SNR accurately can therefore yield significant gains in intelligibility. A number of techniques have been proposed in the computational auditory scene analysis (CASA) literature (see review by Wang and Brown, 2006) for estimating the IdBM and include methods based on pitch continuity information (Hu and Wang, 2004; Roman and Wang, 2006) and sound-localization cues (Roman *et al.*, 2003). Most of the CASA techniques proposed thus far are based on elaborate auditory models and make extensive use of grouping principles (e.g., pitch continuity, onset detection) to segregate the target from the mixture. Alternatively, the IdBM, or equivalently the SNR, can be estimated using simpler signal processing algorithms that compute the SNR in each channel from the mixture envelopes based on estimates of the masker spectrum and past estimates of the enhanced (noise-suppressed) spectrum (e.g., Hu *et al.*, 2007, Loizou, 2007, Chap. 7.3.3). Several such algorithms do exist and are commonly used in speech enhancement applications to improve the quality of degraded speech (see review by Loizou, 2007). To assess how accurate such algorithms are, we processed 20 IEEE sentences embedded in 5 and 10 dB SNR babbles (20 talkers) via two conventional noise-reduction algorithms, which we found in a previous study to preserve intelligibility (Hu and Loizou, 2007a), and computed the hits and false-alarm rates [6] of the SNR estimation algorithms (see Table III). We also processed the mixtures via the SNR estimation algorithm that was used by Hu *et al.* (2007) and tested with cochlear implant users. Overall, the percentage of errors (type I and II) made by the two algorithms, namely, the Wiener (Scalart and Filho, 1996) and the minimum mean-square error (MMSE) algorithms (Ephraim and Malah, 1984), were quite high (see Table III), thus providing a plausible explanation as to why current noise-reduction algorithms do not improve speech intelligibility for normal-hearing listeners (Hu and Loizou, 2007a), although they improve speech quality (Hu and Loizou, 2007b). In contrast, the SNR estimation algorithm used by Hu *et al.* (2007) was relatively more accurate (smaller percentage of type II errors) than the other two algorithms (MMSE and Wiener) accounting for the moderate intelligibility improvement reported by Hu *et al.* (2007) by CI users. The data shown in Table III were computed using sentences corrupted in MB, and required an algorithm for tracking the background noise (needed for the estimation of the SNR). While several noise-estimation algorithms exist (see Loizou, 2007, Chap. 9) that perform reasonably well for stationary (and continuous) noise, no algorithms currently exist that would track accurately a single competing talker. Better noise-tracking algorithms are thus needed for tackling the situation in which the target speech signal is embedded in single competing talker(s). Estimates of the masker (competing talker) spectra would be needed for accurate estimation of the instantaneous SNR in such listening situations. Hence, further research is warranted in developing algorithms capable of estimating more accurately the IdBM in various noise background conditions.

IV. CONCLUSIONS

A new channel selection criterion was proposed for *n-of-m* type of coding strategies based on the SNR values of individual channels. The new SNR criterion can be used in lieu of the maximum selection criterion presently used by the commercially available ACE strategy in the Nucleus-24 cochlear implant system. The new strategy (IdBM) requires access to accurate values of the SNR in each channel. Re-

TABLE III. Average performance, in terms of hits and false-alarm rates (Ref. 6), of three SNR estimation algorithms that were used to compute the binary mask.

Global SNR	Noise-reduction algorithm	Hits (%)	False alarm (%)
5 dB	Wiener (Scalart and Filho, 1996)	26.72	20.67
	MMSE (Ephraim and Malah, 1984)	23.25	15.53
	Hu *et al.* (2007)	53.19	17.41
10 dB	Wiener (Scalart and Filho, 1996)	28.66	18.46
	MMSE (Ephraim and Malah, 1984)	25.55	14.01
	Hu *et al.* (2007)	58.76	18.38

sults from experiment 1 indicated that if such SNR values are available, then the proposed strategy (IdBM) can restore speech intelligibility to the level attained in quiet independent of the type of masker or SNR level (0–10 dB) used. Results in experiment 2 showed that IdBM can tolerate up to a 25% error rate in channel selection without compromising speech intelligibility. Overall, the outcomes from the present study suggest that the SNR criterion has proven to be a good and effective channel selection criterion with the potential of restoring speech intelligibility. Thus, much effort needs to be invested in developing signal processing algorithms capable of estimating accurately the SNR of individual channels from the mixture envelopes.

ACKNOWLEDGMENTS

This research was supported by Grant Nos. R01 DC007527 and R03 DC008887 from the National Institute of Deafness and other Communication Disorders, NIH. The authors would like to thank the three anonymous reviewers for the valuable suggestions and comments they provided.

[1]Aside from the method used to select the envelopes, the ACE and CIS strategies implemented on the Nucleus-24 device differ in the number of electrodes stimulated. In the study by Skinner *et al.* (2002b), for instance, only 12 electrodes were stimulated in the CIS strategy, and 8 (out of 20) electrodes were stimulated in the ACE strategy. The selected (and activated) electrodes in the ACE strategy vary from cycle to cycle depending on the location of the eight maximum amplitudes, whereas in the CIS strategy, the same set of electrodes is activated for all cycles.

[2]The duration of each cycle depends largely on the stimulation rate, which might in turn vary depending on the device. The ACE strategy, for instance, operates at a higher rate compared to the SPEAK strategy.

[3]Anecdotally, subjects did not report any quality degradation in the processed speech stimuli due to the dynamic selection process of the IdBM strategy.

[4]Pilot data were collected with one subject (S2) to assess whether ACE performs better than CIS in noise. More specifically, we assessed the performance of our own implementation of a 6-of-15 strategy (ACE) on speech recognition in noise. The subject was tested on a different day with a different set of IEEE sentences following the same experimental protocol described in experiment 1. Mean percentage correct scores in the 5 and 10 dB SNR babble conditions were 21.2%. Mean percentage correct scores in the 0 and 5 dB SNR SSN were 14.3%. Comparing these scores with the scores obtained with the CIS strategy (see Figs. 2 and 3), we note that the difference in scores is small (six to eight percentage points). While we cannot assess statistical significance, it is noteworthy to mention that the small differences (six to eight percentage points) in score between CIS and ACE are consistent with those reported by Skinner *et al.* (2002b).

[5]Type I error (also called *false alarm*) is produced when deciding hypothesis H_1 (signal is present) when H_0 is true (signal is absent). Type II error (also called *miss*) is produced when deciding H_0 when H_1 is true (Kay, 1998).

[6]The estimated SNR of each T-F unit was compared against a threshold (0 dB), and T-F units with positive SNR were classified as target-dominated T-F units and units with negative SNR were classified as masker-dominated units. The binary mask pattern estimated using the MMSE and Wiener algorithms was compared against the (true) IdBM pattern. The noise power spectrum, needed in the computation of the SNR, was computed using the algorithm proposed by Rangachari and Loizou (2006). Errors were computed in each frame by comparing the true decision made by the idBM with the decision made by the SNR estimation algorithm for each T-F unit. The hits (=1-type II errors) and false-alarm (type I error) rates were averaged across 20 IEEE sentences and are reported in Table III. It should be noted that the data in Table III were computed using a SNR threshold of 0 dB in order to be consistent with the data collected with cochlear implant users in experiment 1. Use of a smaller SNR threshold (−5 dB) yielded higher hit rates (~40%), however, at the expense of increasing the false-alarm rates to near 30%. Similarly, increasing the SNR threshold to +5 dB yielded lower false-alarm rates (<10%) but decreased the hit rate to 17%.

ANSI (**1997**). "Methods for calculation of the speech intelligibility index," ANSI S3.5-1997, American National Standards Institute, New York.

Anzalone, M., Calandruccio, L., Doherty, K., and Carney, L. (**2006**). "Determination of the potential benefit of time-frequency gain manipulation," Ear Hear. **27**, 480–492.

Brungart, D., Chang, P., Simpson, B., and Wang, D. (**2006**). "Isolating the energetic component of speech-on-speech masking with ideal time-frequency segregation," J. Acoust. Soc. Am. **120**, 4007–4018.

Dudley, H. (**1939**). "Remaking speech," J. Acoust. Soc. Am. **11**, 1969–1977.

Ephraim, Y., and Malah, D. (**1984**). "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," IEEE Trans. Acoust., Speech, Signal Process. **32**, 1109–1121.

French, N. R., and Steinberg, J. D. (**1947**). "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Am. **19**, 90–119.

Hu, G., and Wang, D. (**2004**). "Monaural speech segregation based on pitch tracking and amplitude modulation," IEEE Trans. Neural Netw. **15**, 1135–1150.

Hu, Y., and Loizou, P. (**2007a**). "A comparative intelligibility study of single-microphone noise reduction algorithms," J. Acoust. Soc. Am. **122**, 1777–1786.

Hu, Y., and Loizou, P. (**2007b**). "Subjective comparison and evaluation of speech enhancement algorithms," Speech Commun. **49**, 588–601.

Hu, Y., Loizou, P., Li, N., and Kasturi, K. (**2007**). "Use of a sigmoidal-shaped function for noise attenuation in cochlear implants," J. Acoust. Soc. Am. **122**, EL128–EL134.

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

James, C., Blamey, P., Martin, L., Swanson, B., Just, Y., and Macfarlane, D. (**2002**). "Adaptive dynamic range optimization for cochlear implants: A preliminary study," Ear Hear. **23**, 49S–58S.

Kay, S. (**1998**). *Fundamentals of Statistical Signal Processing: Detection Theory* (Prentice-Hall, Upper Saddle River, NJ).

Kiefer, J., Hohl, S., Sturzebecher, E., Pfennigdorff, T., and Gstoettner, W. (**2001**). "Comparison of speech recognition with different speech coding strategies (SPEAK, CIS, and ACE) and their relationship to telemetric measures of compound action potentials in the nucleus CI 24M cochlear implant system," Audiology **40**, 32–42.

Kim, H., Shim, Y. J., Chung, M. H., and Lee, Y. H. (**2000**). "Benefit of ACE compared to CIS and SPEAK coding strategies," Adv. Oto-Rhino-Laryngol. **57**, 408–411.

Kryter, K. D. (**1962**). "Validation of the articulation index," J. Acoust. Soc. Am. **34**, 1698–1702.

Li, N., and Loizou, P. (**2007**). "Factors influencing glimpsing of speech in noise," J. Acoust. Soc. Am. **122**, 1165–1172.

Li, N., and Loizou, P. (**2008**). "Factors influencing intelligibility of ideal binary-masked speech: Implications for noise reduction," J. Acoust. Soc. Am. **123**, 2287–2294.

Loizou, P. (**2006**). "Speech processing in vocoder-centric cochlear implants," Adv. Oto-Rhino-Laryngol. **64**, 109–143.

Loizou, P. (**2007**). *Speech Enhancement: Theory and Practice* (CRC, Boca Raton, FL).

Noguiera, W., Buchner, A., Lenarz, T., and Edler, B. (**2005**). "A psychoacoustic 'nofm'-type speech coding strategy for cochlear implants," EUR-

ASIP J. Appl. Signal Process. **18**, 3044–3059.
Peterson, G., and Cooper, F. (**1957**). "Peakpicker: A bandwidth compression device," J. Acoust. Soc. Am. **29**, 777.
Rangachari, S., and Loizou, P. (**2006**). "A noise estimation algorithm for highly nonstationary environments," Speech Commun. **28**, 220–231.
Roman, N., and Wang, D. (**2006**). "Pitch-based monaural segregation of reverberant speech," J. Acoust. Soc. Am. **120**, 458–469.
Roman, N., Wang, D., and Brown, G. (**2003**). "Speech segregation based on sound localization," J. Acoust. Soc. Am. **114**, 2236–2252.
Scalart, P., and Filho, J. (**1996**). "Speech enhancement based on a priori signal to noise estimation," Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 629–632.
Seligman, P., and McDermott, H. (**1995**). "Architecture of the Spectra 22 speech processor," Ann. Otol. Rhinol. Laryngol. **104**, 139–141.
Skinner, M., Arndt, P., and Staller, S. (**2002a**). "Nucleus 24 advanced encoder conversion study: Performance versus preference," Ear Hear. **23**, 2S–17S.
Skinner, M., Holden, L. K., Whitford, L. A., Plant, K. L., Psarros, C., and Holden, T. A. (**2002b**). "Speech recognition with the nucleus 24 SPEAK, ACE, and CIS speech coding strategies in newly implanted adults," Ear Hear. **23**, 207–223.
Vandali, A. E., Whitford, L. A., Plant, K. L., and Clark, G. M. (**2000**). "Speech perception as a function of electrical stimulation rate using the Nucleus 24 cochlear implant system," Ear Hear. **21**, 608–624.
Wang, D. (**2005**). "On ideal binary mask as the computational goal of auditory scene analysis," *Speech Separation by Humans and Machines*, edited by P. Divenyi (Kluwer Academic, Dordrecht), pp. 181–187.
Wang, D., and Brown, G. (**2006**). *Computational Auditory Analysis* (Wiley, New York).