A channel-selection criterion for suppressing reverberation in cochlear implants

Kostas Kokkinakis, Oldooz Hazrati, and Philipos C. Loizou[a)]
Department of Electrical Engineering, The University of Texas at Dallas, Richardson, Texas 75080

(Received 18 September 2010; revised 2 February 2011; accepted 2 February 2011)

Little is known about the extent to which reverberation affects speech intelligibility by cochlear implant (CI) listeners. Experiment 1 assessed CI users' performance using Institute of Electrical and Electronics Engineers (IEEE) sentences corrupted with varying degrees of reverberation. Reverberation times of 0.30, 0.60, 0.80, and 1.0 s were used. Results indicated that for all subjects tested, speech intelligibility decreased exponentially with an increase in reverberation time. A decaying-exponential model provided an excellent fit to the data. Experiment 2 evaluated (off-line) a speech coding strategy for reverberation suppression using a channel-selection criterion based on the signal-to-reverberant ratio (SRR) of individual frequency channels. The SRR reflects implicitly the ratio of the energies of the signal originating from the early (and direct) reflections and the signal originating from the late reflections. Channels with SRR larger than a preset threshold were selected, while channels with SRR smaller than the threshold were zeroed out. Results in a highly reverberant scenario indicated that the proposed strategy led to substantial gains (over 60 percentage points) in speech intelligibility over the subjects' daily strategy. Further analysis indicated that the proposed channel-selection criterion reduces the temporal envelope smearing effects introduced by reverberation and also diminishes the self-masking effects responsible for flattened formants. © *2011 Acoustical Society of America*. [DOI: 10.1121/1.3559683]

PACS number(s): 43.66.Ts, 43.71.Ky, 43.71.Gv [EB] Pages: 3221–3232

I. INTRODUCTION

Acoustic reverberation is primarily caused by multiple reflections and diffractions of sounds on the walls and objects in enclosed spaces. Reverberation is present in everyday situations: at home, at work, in public spaces, or, in other words, in all enclosed rooms. Reverberation can cause significant changes in speech quality and can also have a very negative impact on speech intelligibility, since it blurs temporal and spectral cues, flattens formant transitions, reduces amplitude modulations associated with the fundamental frequency of speech, and increases low-frequency energy, which in turn results in masking of higher speech frequencies (e.g., see Bolt and MacDonald, 1949; Nabelek and Picket, 1974; Nabelek and Letowski, 1988; Nabelek *et al.*, 1989; Assmann and Summerfield, 2004).

Although, in general, overall speech identification by normal-hearing listeners may not be compromised until the reverberation time (RT_{60}) exceeds approximately 1.0 s (e.g., see Nabelek and Letowski, 1988; Kjellberg, 2004), speech intelligibility measured in listeners with sensorineural hearing loss has been shown to deteriorate considerably even in situations where the reverberation time exceeds 0.5 s. Nabelek and Letowski (1985) assessed vowel recognition performance of ten elderly adults with binaural sensorineural hearing loss and concluded that the mean vowel recognition score in reverberation time of 1.2 s was approximately 12 percentage points lower than the mean score obtained in a non-reverberant (anechoic) listening condition. A more recent study with normal-hearing children and adults evaluated speech intelligibility in terms of the speech reception threshold (SRT) and concluded that speech recognition performance can drop substantially in reverberation even in quiet (e.g., see Neuman *et al.*, 2010). More precisely, it was shown that when the initial reverberation time was increased by almost a factor of 3, such that $RT_{60} = 0.8$ s, the SRT increased by approximately 2–3 dB for all the normal-hearing participants (Neuman *et al.*, 2010).

The literature on the effects of reverberation on speech recognition by cochlear implant (CI) users is sparse. Only a few studies were reported, and those studies involved primarily vocoder simulations conducted with normal-hearing listeners. For instance, in the studies by Poissant *et al.* (2006) and Whitmal and Poissant (2009), normal-hearing adult listeners were presented with reverberant stimuli processed into 6–24 channels using tone-excited vocoders. Percent correct recognition scores were found to be significantly worse when speech inside a reverberant field with $RT_{60} = 0.7$ s was vocoded using a small number of channels (Whitmal and Poissant, 2009). The authors concluded that in all conditions tested, the reverberation time and the direct-to-reverberant ratio (DRR)[1] had a negative impact on the speech identification performance of the listeners.

Tackling speech degradation due to reverberation has recently become an area of intense research activity and has given rise to several dereverberation[2] algorithms for speech enhancement (e.g., see Furuya and Kataoka, 2007; Kokkinakis and Loizou, 2009; Krishnamoorthy and Prasanna, 2009). Reducing distortion due to additive reverberation through

a)Author to whom correspondence should be addressed. Electronic mail: loizou@utdallas.edu

inverse filtering is one of the first and remains one of the most commonly used methods today (Miyoshi and Kaneda, 1988). In most methods, the main idea of reverberation cancellation or speech dereverberation is to pass the reverberant (corrupted) signal through a finite impulse response (FIR) filter that inverts the reverberation process, thus recovering the original signal (e.g., see Haykin, 2000; Huang *et al*., 2007). However, the main drawback of inverse filtering approaches is that the acoustic impulse response must be known in advance or alternatively needs to be "blindly" estimated for successful dereverberation. This is known to be a fairly difficult and computationally expensive task (e.g., see Kokkinakis and Loizou, 2009).

From the literature summarized above, it quickly becomes evident that the effects of additive reverberant energy on speech identification by both normal-hearing and hearing-impaired listeners can be quite detrimental. However, to-date, very little is known about the extent to which reverberation can affect the sentence recognition abilities of CI recipients. Nowadays, most CI processors utilize the advanced combination encoder (ACE) speech coding strategy, which operates by selecting only a subset of envelopes (typically around 8–12) for stimulation at each cycle (e.g., see McDermott *et al*., 1992; McKay and McDermott, 1993; Vandali *et al*., 2000; Kiefer *et al*., 2001; Loizou, 2006). The principle underlying the use of the ACE strategy is that speech can be well understood even if only the peaks in the short-term spectrum are transmitted. Although, CI devices perform well in quiet listening conditions and many CI users can now achieve open-set speech recognition scores of 80% or higher regardless of the device or speech coding strategy used (e.g., see Skinner *et al*., 2002; Spahr and Dorman, 2004), in the presence of reverberation, the maximum amplitude selection criterion used in the ACE coding strategy can become problematic. This is so because during the unvoiced segments (e.g., stops) of the utterance, where the overlap-masking effects dominate, the ACE strategy will mistakenly select the channels containing reverberant energy, since those channels have the highest energy.

This study describes two experiments. Experiment 1 investigates the impact of reverberation on speech identification by CI listeners. It is hypothesized that although the maximum selection criterion (ACE) works well in quiet and can very efficiently capture all the perceptually relevant features of speech, it is quite vulnerable to the effects of reverberation. The effects of temporal envelope smearing, in particular, are examined. Experiment 2 proposes a solution to the challenging problem of reverberation. More precisely, in an attempt to alleviate the negative effects of reverberation on speech perception by CI listeners, we propose a new speech coding strategy for dereverberation based on a new channel-selection criterion. In place of the maximum selection criterion currently implemented in the ACE strategy, we propose the use of a new selection criterion that is based on the signal-to-reverberant ratio (SRR) of the individual channels. In the proposed strategy, amplitudes with SRR greater than a preset threshold are selected, while amplitudes with SRR values smaller than the threshold are zeroed out (eliminated). The SRR reflects implicitly the ratio of the energies of the signal originating from the early (and direct) reflections and the signal originating from the late reflections. Note here that the resulting reverberant signal is composed of the superposition of these two aforementioned signals. Hence, the underlying motivation in using the proposed SRR criterion is to retain the signal components arising from the early reflections, while discarding the signal components generated from the late reflections. Early reflections are known to be beneficial (at least for normal-hearing listeners) due to the precedence effect (e.g., see Litovsky *et al*., 1999), whereas late reflections are known to be detrimental to speech intelligibility as they are responsible predominantly for the smearing of the temporal envelopes and filling of the gaps (e.g., closures) in unvoiced segments (e.g., stops) of the utterance. The extent to which CI users exhibit precedence effect is unclear (Seeber and Hafter, 2007). The construction of the SRR criterion assumes *a priori* knowledge of the input target signal, and the aim of experiment 2 is to assess the potential of the proposed SRR channel-selection criterion in suppressing reverberation in highly reverberant conditions ($RT_{60} = 1.0$ s).

II. EXPERIMENT 1. EFFECTS OF REVERBERATION ON SPEECH INTELLIGIBILITY BY CI LISTENERS

To assess the effects of reverberation on speech intelligibility by CI listeners, sentence recognition tests were conducted in different reverberant conditions with RT_{60} values ranging from 0.3 up to 1.0 s.

A. Methods

1. Subjects

A total of six CI listeners participated in this study. All participants were native speakers of American English with postlingual deafness, who received no benefit from hearing aids preoperatively. Their ages ranged from 47 to 76 yr ($M = 65.5$ yr) and they were all paid to participate in this research study. All subjects were fitted with the Nucleus 24 multi-channel implant device (CI24M, Cochlear Corp., Sydney, Australia). The participants used their devices routinely and had a minimum of 5 yr experience with their CIs. Detailed biographical data for the subjects are given in Table I.

2. Research processor

Three of the subjects tested were using the Cochlear ESPrit 3G and three were using the Nucleus Freedom speech processor on a daily basis. During their visit, all the participants were temporarily fitted with the SPEAR3 wearable research processor. SPEAR3 was developed by the Cooperative Research Center (CRC) for Cochlear Implant and Hearing Aid Innovation. The SPEAR3 has been used in a number of investigations to-date as a way of controlling inputs to the CI system (e.g., see Kokkinakis and Loizou, 2010). Prior to the scheduled visit of the subjects, the Seed-Speak GUI application was used to program the SPEAR3 processor with the individual threshold and comfortable loudness levels for each user. All CI listeners used the SPEAR3 device programmed with the ACE speech coding strategy (e.g., see Vandali *et al*., 2000). The ACE strategy implemented in the SPEAR3 processor is very similar to that implemented in the

TABLE I. CI patient description and history.

	S1	S2	S3	S4	S5	S6
Age	47	71	76	69	55	62
Gender	F	F	M	M	F	F
Years implanted (L/R)	7/7	8/8	5/5	7/7	10/10	7/7
Years of deafness	46	60	52	34	24	48
CI processor	ESPrit 3G	Freedom	Freedom	ESPrit 3G	Freedom	ESPrit 3G
Etiology of hearing loss	Unknown	Unknown	Hereditary	Noise	Unknown	Rubella

Nucleus 24 system and most coding parameters of the SPEAR3 ACE strategy matched those of the Nucleus 24 CI system. In addition, all parameters used (e.g., stimulation rate, number of maxima, frequency allocation table) were matched to each patient's clinical settings. The volume of the speech processor was also adjusted to a comfortable loudness prior to initial testing. Institutional review board approval was obtained and informed consent was obtained from all participants before testing commenced.

3. Stimuli

The speech stimuli used for testing were sentences from the Institute of Electrical and Electronics Engineers (IEEE) database (IEEE, 1969). Each sentence is composed of approximately 7–12 words and in total there are 72 lists of 10 sentences each produced by a single talker. The root-mean-square value of all sentences was equalized to the same value corresponding to approximately 65 dBA. All stimuli were recorded at the sampling frequency of 16 000 Hz.

4. Simulated reverberant conditions

Head-related transfer functions (HRTFs) recorded by Van den Bogaert *et al*. (2009) were used to simulate reverberant conditions. To obtain measurements of HRTFs, Van den Bogaert *et al*. (2009) used a CORTEX MKII manikin artificial head placed inside a rectangular reverberant room with dimensions $5.50\,\mathrm{m} \times 4.50\,\mathrm{m} \times 3.10\,\mathrm{m}$ (length $\times$ width $\times$ height) and a total volume of 76.80 m^3. The sound pressure level (SPL) measured at the center of the artificial head was fixed at 70 dB SPL. The overall reverberant characteristics of the experimental room were altered by adding floor carpeting and absorptive panels on the walls and the ceiling, as described in more detail in the Van den Bogaert *et al*. (2009) study.

The average reverberation time of the experimental room (average in one-third-octave bands with center frequencies between 125 and 4000 Hz) before any modification was equal to $\mathrm{RT}_{60} = 1.0$ s. When just two absorptive panels were hung from hooks mounted on the walls close to the ceiling, the average reverberation time of the room was reduced to $\mathrm{RT}_{60} = 0.8$ s. By increasing the number of acoustic panels and by adding floor carpeting to the room the average reverberation time was reduced even further to around $\mathrm{RT}_{60} = 0.6$ s. Finally, a reverberation time equal to $\mathrm{RT}_{60} = 0.3$ s was obtained by partitioning the room with a custom partitioning wall system composed of several highly absorbent rectangular acoustic boards (RESOPAL). This latter RT_{60} value corresponds to a well-dampened room and is typical of the lowest reverberation time that might be found in a small office room.

To obtain HRTF recordings for each reverberation condition, the artificial head was placed in the middle of a ring of 1.25 m inner diameter. A single-cone loudspeaker (FOSTEX 6301 B) with a 10 cm diameter was placed at a 0° azimuth in the frontal plane. A two-channel sound card (VX POCKET 440 DIGIGRAM) and DIRAC 3.1 software type 7841 (Bruel and Kjaer Sound and Vibration Measurement Systems) were used to generate the stimuli. All recordings were facilitated using identical microphones to those used in modern BTE speech processors.

To generate the stimuli used in our study, the HRTFs obtained for each reverberation condition were convolved with the speech files from the IEEE test materials using standardized linear convolution algorithms in MATLAB. All stimuli were presented to the listener through the auxiliary input jack of the SPEAR3 processor in a double-walled sound attenuated booth (Acoustic Systems, Inc.). Prior to testing, each subject participated in a short practice session to gain familiarity with the listening task. During the practice session, the subjects were allowed to adjust the volume to reach a comfortable level.

5. Procedure

Subjects participated in a total of four conditions, each corresponding to a different RT_{60} condition. Two IEEE lists (20 sentences) were used per condition. Unprocessed IEEE sentences in quiet were also used as the control or anechoic condition ($\mathrm{RT}_{60} = 0$ s). Each participant completed all conditions in a single session. Subjects were given a 15 min break every 90 min during the test session. Following initial instructions, each user participated in a brief practice session to gain familiarity with the listening task and to also get acclimatized to the SPEAR3 processor settings. No score was calculated for this practice set. None of the lists used were repeated across different conditions. To minimize any order effects, the order of the test conditions was randomized across subjects. During testing, each sentence was presented once and the participants were instructed to type as many of the words as they could identify via a computer keyboard. The responses of each individual were collected, stored in a written sentence transcript, and scored off-line based on the number of words correctly identified. All words were scored. The percent correct scores for each condition were calculated by dividing the number of words correctly identified by the total number of words in the particular sentence list.

B. Results and discussion

The individual speech intelligibility scores are displayed in Fig. 1(a) as a function of the room reverberation time (RT_{60}). For comparative purposes, the scores obtained in the anechoic condition corresponding to $RT_{60} = 0$ s are also shown. For all subjects tested, speech intelligibility decreased with an increase in the reverberation time of the room. An analysis of variance (ANOVA) (with repeated measures) confirmed a significant effect ($F[3,15] = 61.1$, $p < 0.0005$) of reverberation time (RT_{60}) on speech intelligibility. The average speech intelligibility scores for all listeners dropped from 90% to around 60% for $RT_{60} = 0.3$ s, while for $RT_{60} = 1.0$ s, the intelligibility scores were on average 70 percentage points lower when compared to the anechoic listening condition ($RT_{60} = 0$ s).

The intelligibility scores obtained from experiment 1 suggest a very strong, and negative, relationship between speech perception and the amount of additive acoustical reverberation. According to Fig. 1(b), there is a decaying-exponential relationship between the intelligibility scores obtained and the reverberation time of the room. Accordingly, to model the effects of reverberation on speech

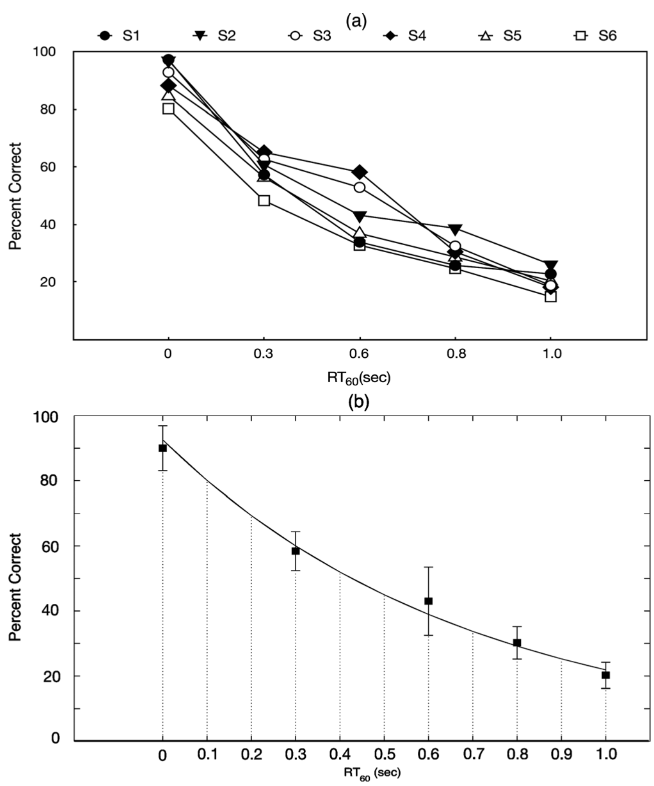

FIG. 1. (a) Percent word recognition scores obtained by six Nucleus 24 CI users tested on IEEE sentences corrupted with varying degrees of reverberation. Reverberation times of $RT_{60} = 0$, 0.30, 0.60, 0.80, and 1.0 s are shown along the abscissa. (b) Mean percent word recognition scores for the same six Nucleus 24 CI users. The filled squares represent experimental data obtained in different reverberant conditions and the error bars indicate standard deviations. The solid line depicts the predicted speech intelligibility as modeled by the exponential function described by Eq. (1).

identification, percent correct recognition scores were fit (in the least-squares sense) by an exponential function of the form

$$y = \exp(C_1 x + C_2), \tag{1}$$

where variable y corresponds to the (predicted) speech intelligibility scores, variable x denotes reverberation time (seconds), and C_1 and C_2 represent the fitting constants. For our data, these constants were found to be equal to

$$C_1 = -0.0014 \quad \text{and} \quad C_2 = 4.528. \tag{2}$$

The solid line depicted in Fig. 1(b) plots the predicted speech intelligibility performance modeled using the above exponential function. This exponential fit was found to exhibit a fairly high correlation (Pearson's correlation, $\rho = 0.996$) with the experimental data. Based on the good agreement between the experimental data and the exponential function given in Eq. (1), we can conclude that speech intelligibility performance for CI users is expected to decline rapidly as the reverberation time of the room increases.

An exponential effect of reverberation (RT_{60}) on speech intelligibility was also evident in the data reported by Poissant *et al.* (2006) with normal-hearing listeners presented with reverberant speech that has been vocoded to 6–24 channels. In the study by Poissant *et al.* (2006), however, only RT_{60} values in the range of 0–0.425 s were investigated. Performance was found to be the lowest (20% correct) when reverberant speech was vocoded using 6 channels ($RT_{60} = 0.425$ s), while performance was only mildly affected by RT_{60} when the reverberant stimuli were vocoded into 12 or more channels. In our study, performance at $RT_{60} = 0.425$ s was found to be much higher (almost 50% correct) than that reported by Poissant *et al.* (2006) and Whitmal and Poissant (2009).

The vocoder simulation studies by Poissant *et al.* (2006) implied that spectral resolution was one of the main factors influencing performance of CI users in reverberation. Other factors, however, contributed to the low intelligibility scores. For one, reverberation causes temporal envelope smearing due to overlap-masking effects. Temporal smearing is caused by the overlapping of succeeding segments of speech by preceding ones (i.e., by overlap-masking), particularly, when a low-energy consonant follows a high-energy voiced segment (e.g., vowel). The additive reverberant energy fills the gaps and silent intervals associated with the vocal tract closures occurring in the low-energy speech segments (e.g., stop consonants). A secondary effect of reverberation is self-masking, which results in flattened F1 and F2 formants, essentially causing diphthongs and glides to be confused with monophthongs (e.g., see Nabelek and Letowski, 1985; Nabelek *et al.*, 1989). Such flattened (or disrupted) formant transitions can severely hinder speech identification by CI listeners, who already have great difficulty perceiving naturally occurring F1 and F2 movements.

Figure 2 illustrates example stimulus output patterns (electrodograms) of an IEEE sentence processed with the ACE speech coding strategy in the Nucleus 24 device. In all panels shown, the vertical axes represent the electrode position corresponding to a specific frequency, while the

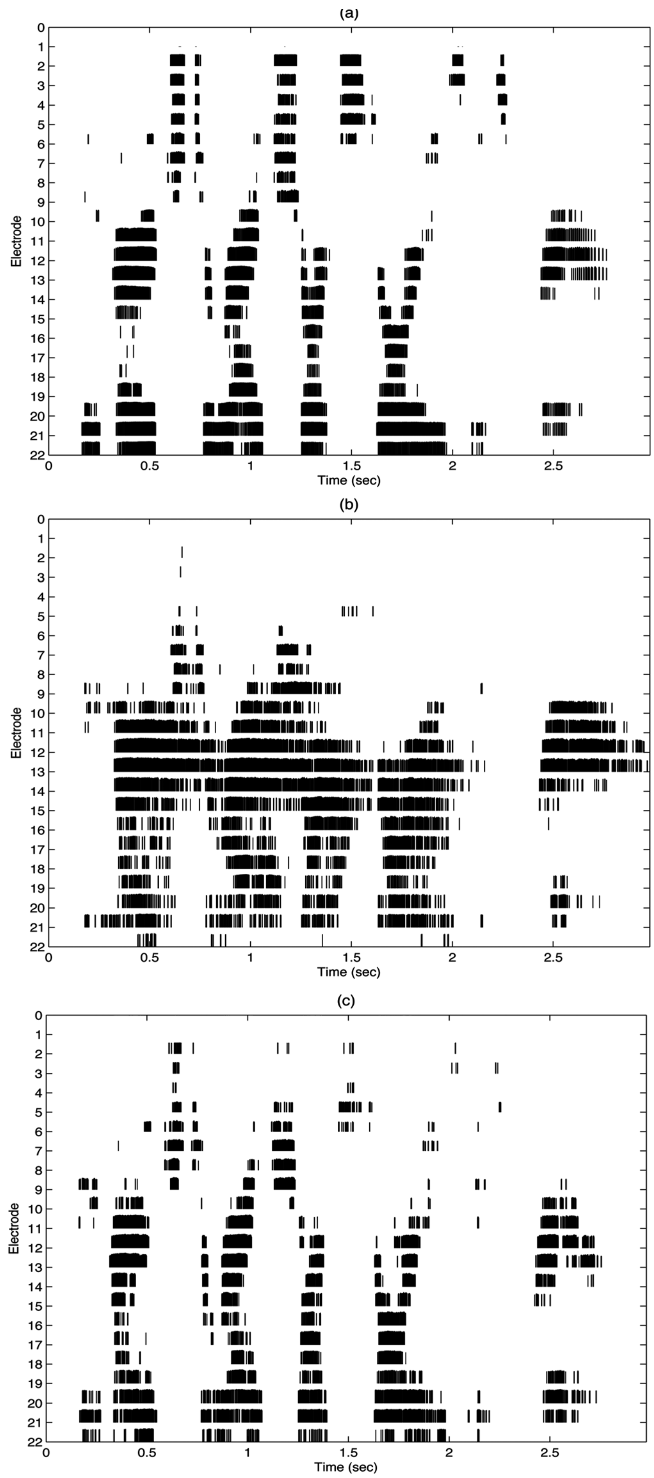

FIG. 2. Stimulus output patterns (electrodograms) of the IEEE sentence "The urge to write short stories is rare" uttered by a male speaker. (a) Electrodogram of unmodified (uncorrupted) sentence processed by the ACE strategy, (b) electrodogram of the same sentence when corrupted by reverberation equal to $\mathrm{RT}_{60} = 1.0$ s and processed by the ACE strategy, and (c) electrodogram of the reverberant sentence when processed by the IRM speech coding strategy with the threshold set to -5 dB. In each electrodogram, time is shown along the abscissa and the electrode number is shown along the ordinate.

horizontal axes show time progression. For this example, speech was corrupted by additive reverberation in a room with $\mathrm{RT}_{60} = 1.0$ s. Temporal envelope smearing is evident in Fig. 2(b). As shown in Fig. 2(b), temporal smearing blurs the vowel and consonant boundaries which are normally present in the anechoic stimuli plotted in Fig. 2(a). The flattened formant transitions caused by self-masking are also evident in Fig. 2(b) (e.g., see envelopes at electrodes 10–14). The associated implications of temporal smearing are discussed in more detail later. The fact that ACE selects in each cycle the channels with the highest amplitude exacerbates further the negative effects of temporal envelope smearing. During the unvoiced segments (e.g., stops) of the utterance, where the overlap-masking effects dominate, the ACE strategy mistakenly selects the channels containing reverberant energy since those channels have the highest energy. During the voiced segments of the utterance (e.g., vowels), ACE correctly selects the amplitudes surrounding the formant regions, however, it fails to adequately capture the information contained in the formant transitions as those are flattened due to self-masking. In fact, given the low spectral resolution, there is seemingly little that can be done about the flattened formant transitions. However, the negative effects of temporal envelope smearing can potentially be diminished (or eliminated) with the use of a better channel-selection criterion designed specifically to suppress reverberation. Such a channel-selection criterion is investigated next.

III. EXPERIMENT 2. EVALUATION OF A NEW CHANNEL-SELECTION CRITERION FOR REVERBERATION SUPPRESSION

A. Methods

1. Subjects and material

The same six postlingually deafened cochlear implantees tested in experiment 1 were asked back on a different day to participate in experiment 2. The same speech material (IEEE, 1969) was used as in experiment 1. None of the sentence lists previously used in experiment 1 was reused in an effort to avoid potential learning effects.

2. Channel-selection criterion and speech coding strategy

a. Motivation. As indicated earlier, the maximum selection criterion adopted in ACE erroneously picks amplitudes during the gaps (e.g., closures) present in most unvoiced segments of the utterance. As a result, the vowel and consonant boundaries are smeared making it difficult for the listeners to use effectively lexical segmentation cues needed for word retrieval. Ideally, we would like to select the amplitudes corresponding only to the direct sound and early reflections, and at the same time discard the amplitudes corresponding to the late reflections. Such a criterion, however, would require access to the acoustic impulse responses, which in practical scenarios may not be available. Instead, we advocate here the use of a new selection criterion based on the SRR, which for each channel is computed as follows:

$$\mathrm{SRR}(t,k) = 10\log_{10}\frac{|X(t,k)|^2}{|Y(t,k)|^2}, \qquad (3)$$

where $X(t, k)$ and $Y(t, k)$ denote the clean and reverberant signals, respectively, t corresponds to the time-frame index, and k defines the frequency or channel index. A large SRR value would suggest that the energy from the direct signal (and early reflections) dominates, as is often the case during the voiced segments (e.g., vowels) of the utterance. In contrast, a small SRR value would suggest that the reverberant energy, composed of the sum of the energies from the early and late reflections, dominates. This happens primarily during the gaps and is caused primarily by overlap-masking. Hence, we could potentially minimize the overlap-masking effects by removing the reverberant energy residing in the gaps. This can be done by comparing the individual channel-specific SRR values against an empirically determined threshold value, T.

To illustrate this concept, we draw attention to Fig. 3(c), which plots the instantaneous SRR values, as well as the clean [see Fig. 3(a)] and reverberant [see Fig. 3(b)] signals bandpass filtered at a center frequency of $f = 500$ Hz. Figure 3(d) plots the synthesized time-domain waveforms of the same IEEE sentence processed with threshold $T = -5$ dB. Figure 3(e) shows the same output processed with the threshold set to $T = +5$ dB. The waveforms depicted in Figs. 3(d) and 3(e) were obtained by retaining the reverberant signal corresponding to $\text{SRR} > T$, while discarding (zeroing out) the reverberant signal when $\text{SRR} < T$. For the example shown, when $T = -5$ dB, we observe that the reverberant energy residing in the gaps is eliminated [compare Fig. 3(b) against Fig. 3(d)]. As shown in Figs. 3(a) and 3(b), during the segments in which SRR is less than -5 dB, the energy of the reverberant signal is more dominant than the energy of the clean (anechoic) speech signal. Thus, a negative threshold value (e.g., $T = -5$ dB) seems to be appropriate for suppressing the reverberation present in the gaps. In contrast, as shown in Fig. 3(e), when the threshold value is set to $T = +5$ dB, the selection process seems to be too aggressive, since apart from discarding the corrupted unvoiced segments and associated gaps, it also zeroes out (eliminates) useful speech information present in the high-energy voiced frames. Since the threshold value will likely influence performance, it is systematically varied in the present experiment from -15 dB to $+5$ dB in steps of 5 dB.

The SRR selection criterion is implemented by multiplying the reverberant signal by a binary time–frequency (T-F) mask or equivalently a binary gain function. This mask (or gain) takes the value of 1 when $\text{SRR} > T$ and is zero otherwise. The dereverberated signal at T-F (t, k) is obtained as follows:

$$\hat{X}_{DE}(t,k) = Y(t,k) \cdot \text{IRM}(t,k), \tag{4}$$

where $\hat{X}_{DE}(t,k)$ denotes the dereverberated signal and $Y(t, k)$ is the reverberant signal. The ideal reverberant mask (IRM) (t, k) is given by

$$\text{IRM}(t,k) = \begin{cases} 1, & \text{SRR}(t,k) > T \\ 0, & \text{otherwise} \end{cases}, \tag{5}$$

where T represents the threshold value, expressed in dB. We refer to IRM as the *ideal reverberant mask* because its

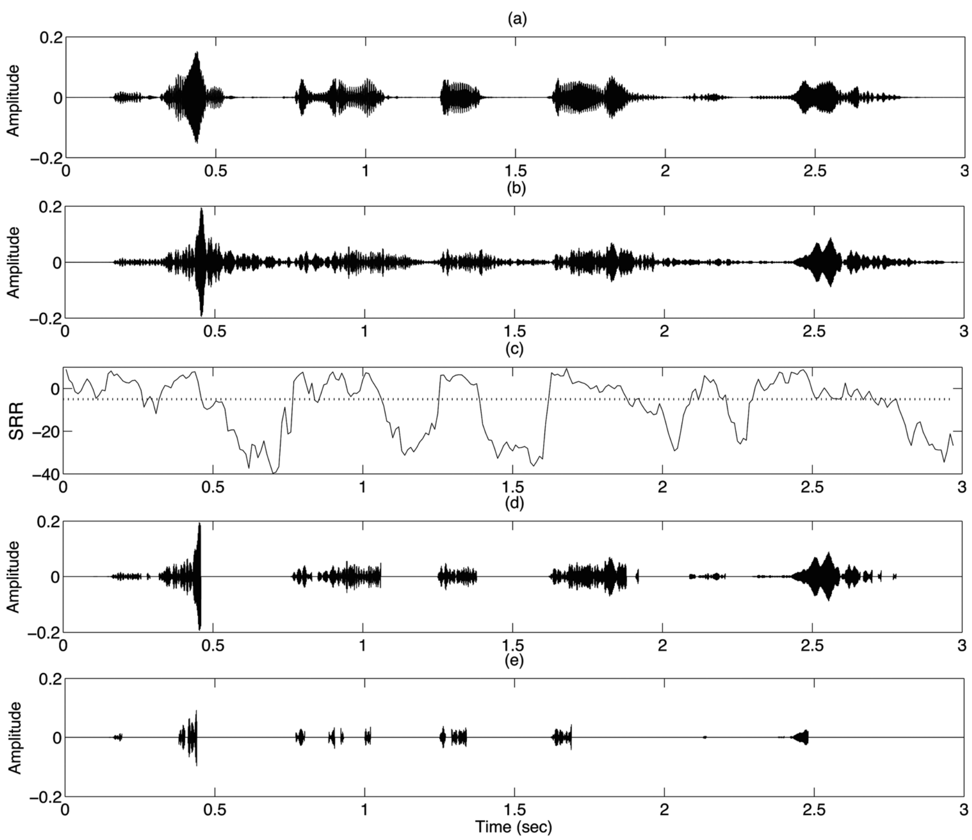

FIG. 3. Example illustrating the bandpass filtered IEEE sentence "The urge to write short stories is rare" extracted at the center frequency of 500 Hz. (a) Unmodified (uncorrupted) sentence, (b) sentence corrupted with reverberation equal to $\text{RT}_{60} = 1.0$ s, (c) instantaneous SRR values (dB) along with threshold value fixed at $T = -5$ dB (dashed line), (d) reverberant sentence processed by the IRM speech coding strategy with the threshold set to -5 dB, and (e) reverberant sentence processed by the IRM speech coding strategy with the threshold set to $+5$ dB.

construction requires prior knowledge of the original (uncorrupted) acoustic information. It is worth mentioning that different forms of binary T-F masks have been previously used in other applications to suppress noise. These T-F masks were based on the local SNR criterion rather than the SRR criterion. Substantial gains in speech intelligibility were observed by both normal-hearing (Brungart *et al.*, 2006; Li and Loizou, 2008; Wang *et al.*, 2009) and CI listeners (Hu and Loizou, 2008) when such masks were applied to speech corrupted by noise even at extremely low input SNR levels (e.g., −10 dB).

b. Speech coding strategy. The block diagram of the proposed speech coding strategy is depicted in Fig. 4. In order to assess the full potential of the proposed channel-selection criterion described in Eqs. (3) and (5) and the associated speech coding strategy without being constrained by implementation or memory issues, we evaluated the proposed strategy as a pre-processor to the SPEAR3 device. That is, speech was first synthesized using the proposed channel-selection criterion and then fed as input to the SPEAR3 speech processor. This way, the number of channels selected in each cycle remained the same as that used in the clinical speech processor. Consequently, the stimulation rate of the proposed speech coding strategy remained the same (as used clinically by the CI users), thus preventing the number of channels and the stimulation rate from confounding the outcomes of our study.

To derive the T-F representation of the clean speech (prior to reverberation) and the reverberant (corrupted) inputs, we use a 128-channel ($N = 128$) fourth-order gammatone filter-bank, with center frequencies equally spaced on the equivalent rectangular bandwidth (ERB) scale covering a frequency range between 50 and 8 kHz. The filtered waveforms are then divided into 20 ms frames with 50% overlap between successive frames, and the short-time energies of the filtered waveforms are computed. In the next stage, a comparison is made between the energy of the clean (or uncorrupted) signal and that of the reverberant (or corrupted) signal. As shown at the bottom of Fig. 4, this comparison is

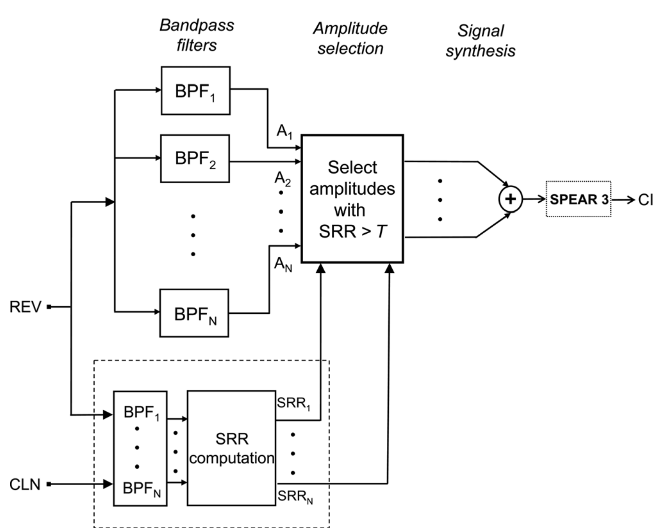

FIG. 4. Block diagram of the IRM strategy.

carried out by calculating the SRR independently in each individual T-F channel [e.g., see Eq. (3)]. The resulting SRR for each T-F unit is then compared against a preset threshold value T to determine whether to retain a specific T-F unit or to discard it. Therefore, out of the 128 initial filtered waveforms, only the T-F units where the energy of the clean signal exceeds that of the reverberant signal by the specified threshold value, such that $\mathrm{SRR}(t, k) > T$, are retained.

In contrast, the T-F regions or channels in which the reverberant signal is dominant, such that $\mathrm{SRR}(t, k) < T$, are eliminated by the proposed strategy. In principle, the number of channels selected can vary from 0 (none selected) to 128 (all selected) in each frame. However, since the dereverberated signal is first synthesized before being fed as input to the SPEAR3 processor, the total number of channels stimulated in each cycle (as well as the stimulation rate) remains the same as that used by the SPEAR3 (programmed using the CI user's clinical parameter settings). At the synthesis stage, phase shifts are corrected by time-reversing the envelope in each channel, passing it through the gammatone filter-bank, time-reversing again, and then summing across all the different channels that were selected. The synthesized stimuli are presented monaurally to the CI listeners via the auxiliary input jack of the SPEAR3 processor.

We refer to the above speech coding strategy as the IRM strategy, since it is based on the IRM described in Eq. (5). For the purpose of this study (and for the reasons stated above), the IRM was implemented in a preprocessing stage (see Fig. 4) and used in conjunction with the ACE strategy. It should be pointed out, however, that in a real-time implementation of IRM, the preprocessing stage would not be required since the IRM would be implemented exactly as ACE, with the exception of the channel-selection stage. That is, the gammatone filter-bank would be replaced by the fast Fourier transform (FFT), and channels with SRR values larger than a prescribed threshold T would be selected for stimulation.

3. Procedure

To generate the reverberant stimuli for the $\mathrm{RT}_{60} = 1.0$ s condition, we convolved the HRTFs (see Sec. II A 4) with the IEEE sentences using standardized linear convolution algorithms in MATLAB. To assess the impact of the SRR threshold T on sentence recognition, we varied its value across five different levels ranging from −15 to +5 dB in increments of 5 dB. The stimuli were presented in the following conditions: (1) unprocessed (reverberant) stimuli with $\mathrm{RT}_{60} = 1.0$ s and (2) reverberant stimuli processed with the IRM strategy for $T = -15$, −10, −5, 0, and +5 dB. The subjects participated in a total of six different test conditions (IRM strategy × five threshold values + one condition involving the unprocessed stimuli). As before, each participant completed all conditions in a single test session. Two IEEE lists (20 sentences) were used for each condition. A total of 120 IEEE sentences were used in this experiment. Each sentence was presented once. The stimuli were presented to each CI user via the auxiliary input jack of the

SPEAR3 processor. The order of the test conditions was randomized across subjects.

B. Results and discussion

The individual speech intelligibility scores for all the aforementioned experimental conditions are shown in Fig. 5(a). The performance in all test conditions was measured in terms of percent of words identified correctly. All words were scored. As can be seen from Fig. 5(a), significant gains in speech intelligibility were obtained for all CI subjects tested with the IRM speech coding strategy for a wide range of threshold values, particularly, when T assumed values in the range of $-15 \leq T \leq -5$ dB. Within this range, the proposed IRM strategy led to substantial gains in speech intelligibility over the subjects' daily strategy. Speech intelligibility scores improved from an average of 20% correct (reverberant baseline) to around 60% correct and almost 70% correct when the subjects utilized the IRM processing strategy with the threshold value set to -15 and -10 dB, respectively. Speech intelligibility scores further increased an additional 15 percentage points reaching peak performance, when the threshold was set to -5 dB. Overall, baseline performance with reverberant stimuli improved from 20% correct to 85% correct when $T = -5$ dB was used. An ANOVA (with repeated measures) indicated a significant effect ($F[5,25] = 256.9$, $p < 0.0005$) of the threshold value T on speech intelligibility. *Post-hoc* comparisons (according to Scheffe's test) were run to assess significant differences in scores obtained between different threshold conditions. Results indicated that performance improved significantly ($p < 0.0005$) relative to the reverberant (unprocessed) scores for all T values except $T = +5$ dB. Performance with $T = -5$ dB was found to be significantly higher ($p < 0.0005$) than all other conditions.

Performance was influenced by the choice of threshold value, T. As shown in Fig. 5(b), negative values of T produced a significant improvement in performance, while non-negative values ($T \geq 0$ dB) did not improve performance or improved it only slightly (e.g., by 19 percentage points when $T = 0$ dB was used). Based on Fig. 5(b), we can conclude that as the T value becomes more negative, i.e., as T approaches $-\infty$, more and more reverberant energy is retained. Although not tested here, we would expect the performance with $T \leq -40$ dB to be comparable to that obtained with the reverberant (baseline) stimuli. We base this on the fact that the lowest SRR value attained in some channels (see Fig. 3) is approximately -40 dB. On the other end, as the value of T increases beyond 0 dB (i.e., approaches $+\infty$), the SRR criterion becomes aggressive and only a small number of channels are selected [see Fig. 3(e)], rendering the synthesized signal quite sparse. This is so

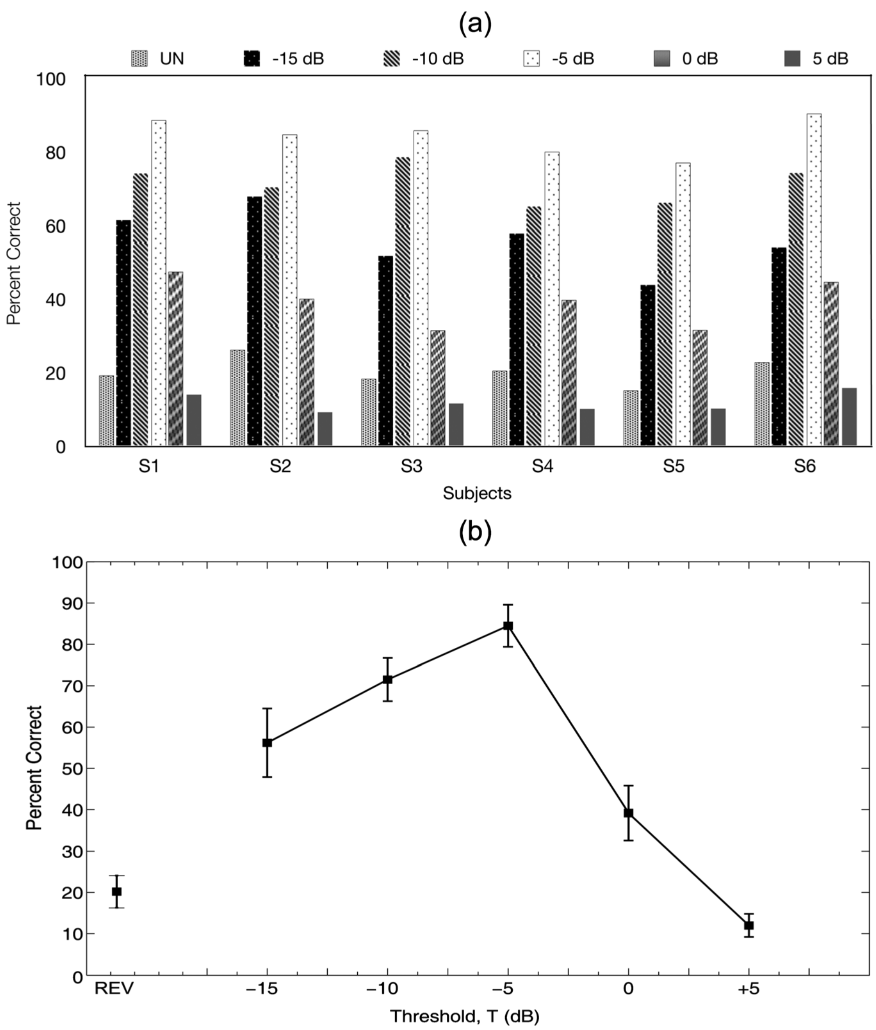

FIG. 5. (a) Individual percent correct scores of six Nucleus 24 CI users tested on IEEE sentences using unprocessed (corrupted) acoustic inputs recorded in $RT_{60} = 1.0$ s and acoustic inputs processed with the IRM coding strategy for different threshold values. (b) Mean percent correct scores for the same users plotted as a function of threshold values (dB). Error bars indicate standard errors of the mean.

because apart from discarding the corrupted unvoiced frames, the SRR criterion also zeroes out (eliminates) useful speech information contained in the high-energy voiced frames. Performance with $T = -5$ dB seems to be the optimal in our study, and that was found to be consistent for all subjects. Based on the example shown in Fig. 3, the choice of $T = -5$ dB seems to detect accurately the vowel and consonant boundaries, making it the ideal threshold value. It is not clear whether the optimal threshold depends on the direct-to-reverberant energy of the room impulse response, which in our case was -5.37 dB, nor is it clear whether it depends on the RT_{60} value or the source-to-listener distance, which was equal to 1 m in our case. Further experiments are needed to assess the dependency (if any) of the optimal threshold value to different room configurations. Based on the findings of the present experiment as well as other pilot experiments using different RT_{60} values, a good choice to consider for T is the range of $-10 < T < -5$ dB.

IV. DISCUSSSION

A. Factors influencing performance in reverberant environments

Based on the data from experiment 1, the performance of CI users in reverberation was greatly affected even with $RT_{60} = 0.3$ s. In fact, an exponential model provided a good fit to the data [see Fig. 1(b)] suggesting that CI users' performance in reverberation degrades exponentially as RT_{60} increases. A number of factors contributed to the low performance observed. These include the low spectral resolution, self-masking effects (causing flat formant transitions), and the detrimental effects of temporal envelope smearing (overlap-masking). Of the three factors, we believe that the negative effects of temporal envelope smearing, introduced predominantly by overlap-masking, contributed the most. As shown in Figs. 2 and 3, reverberation smears the vowel and consonant boundaries which are critical for lexical segmentation and word retrieval (Stevens, 2002). More precisely, the vowel and consonant boundaries serve as acoustic landmarks which are evident as abrupt spectral discontinuities in the signal. These landmarks were posited to be crucial in lexical-access models (Stevens, 2002) and have also been found to be critical in the perception of speech in steady-noise conditions (Li and Loizou, 2008). In brief, the temporal envelope smearing caused by reverberation (particularly by the late reflections) makes the detection of acoustic landmarks extremely difficult, and that in turn disrupts the syllable structure, which is known to be important for determining word boundaries in running speech. As demonstrated in experiment 2, one efficient method for reducing or eliminating the temporal envelope smearing effect is to use the SRR-based channel-selection criterion in place of the maximum criterion adopted in ACE. In doing so, the vowel and consonant boundaries become more evident and intelligibility improves substantially. To some extent, reducing the temporal envelope smearing effect also diminishes the self-masking effects, which are responsible for the flattened formant transitions (Nabelek and Letowski, 1985, 1988). This is evident in the electrodogram shown in Fig. 2(c), where the speech stimuli were processed with the IRM strategy ($T = -5$ dB).

B. Maximum selection criterion vs SRR criterion

The SRR criterion is ideally suited for reverberant conditions and offers a number of advantages over the traditional maximum selection criterion used in ACE. Figures 6 and 7 describe two typical scenarios, in which the proposed channel-selection criterion based on the input SRR, may offer an advantage over the maximum amplitude criterion (e.g., ACE strategy) when selecting the stimulation channels

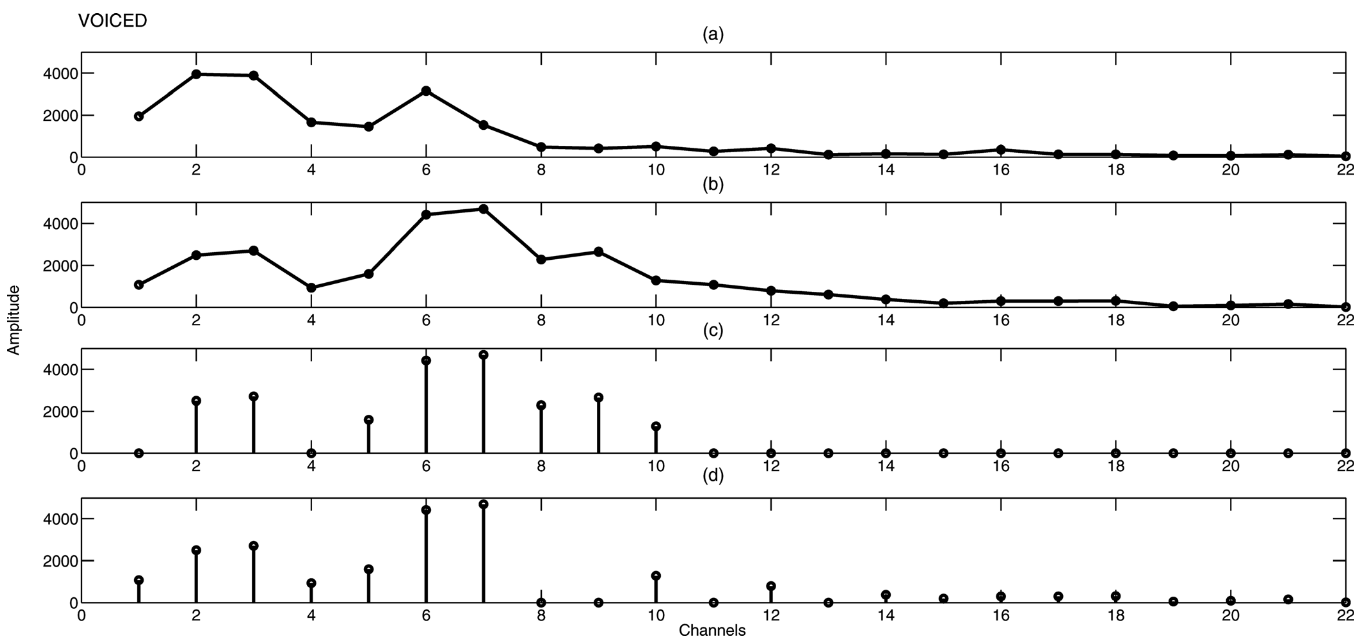

FIG. 6. Example illustrating the selection process by the ACE and IRM speech coding strategies for a *voiced* speech frame. (a) Envelope amplitudes of unmodified (uncorrupted) sentence, (b) envelope amplitudes of the same sentence when corrupted by reverberation equal to $RT_{60} = 1.0$ s, envelope amplitudes of the reverberant sentence selected by ACE, and envelope amplitudes of the reverberant sentence selected by IRM with the threshold set to -5 dB.

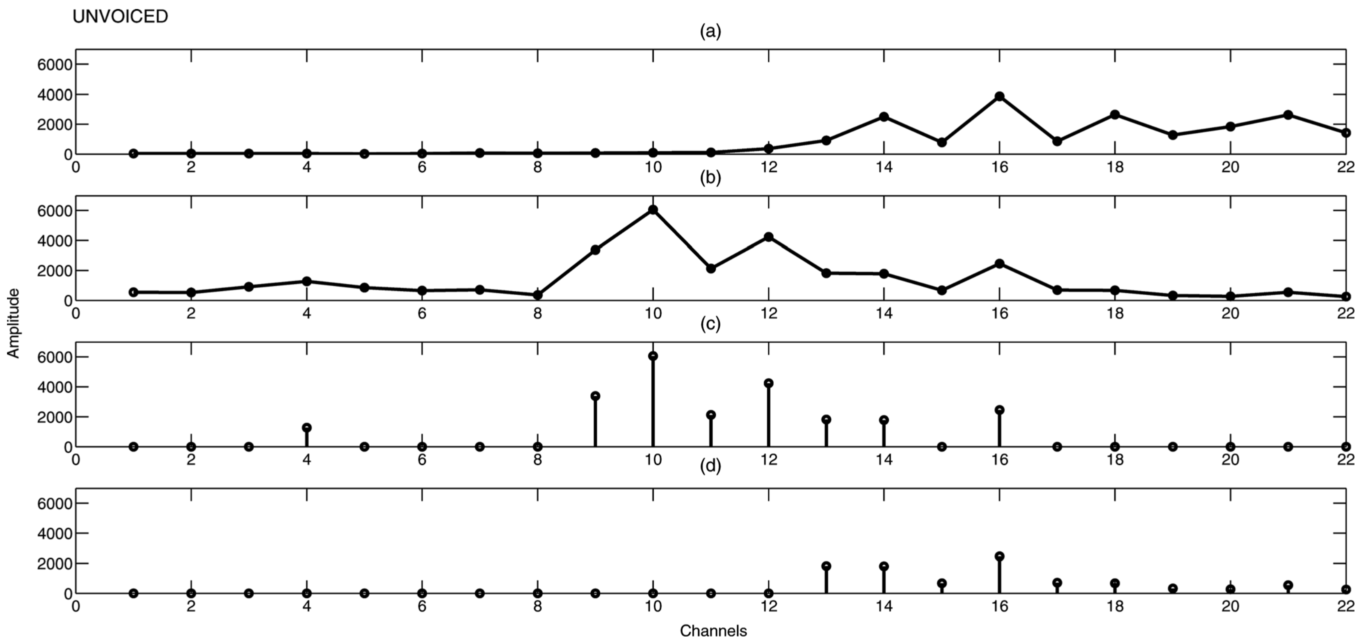

FIG. 7. Example illustrating the selection process by the ACE and IRM speech coding strategies for an *unvoiced* speech frame. (a) Envelope amplitudes of unmodified (uncorrupted) sentence, (b) envelope amplitudes of the same sentence when corrupted by reverberation equal to $RT_{60} = 1.0$ s, (c) envelope amplitudes of the reverberant sentence selected by ACE and envelope amplitudes of the reverberant sentence selected by IRM with the threshold set to −5 dB.

in the presence of additive reverberant energy. First, consider the example shown in Fig. 6, which plots the short-time spectrum (as computed in the ACE strategy and displayed prior to amplitude selection and compression) of a voiced speech segment. In this scenario, as illustrated in Figs. 6(c) and 6(d), despite the presence of reverberation ($RT_{60} = 1.0$ s), both the ACE and IRM strategies are able to correctly select the high-energy spectral regions (e.g., formants) of the voiced speech frame and therefore both stimulate roughly the same number of channels, i.e., the largest in amplitude. Hence, when vowel sounds and other voiced segments are presented in reverberation, the IRM strategy will operate similar to the ACE strategy.

Figure 7 examines a different scenario, where a low-energy unvoiced speech frame is plotted instead. Unvoiced frames are essentially low-energy speech segments (e.g., fricatives, stops, and stop closures) and occur quite frequently in continuous speech. These frames are particularly susceptible to overlap-masking effects, as energy from preceding (and higher energy) phonemes leaks into these segments and fills in the gaps. In this example, as shown in Figs. 7(b) and 7(c), the ACE strategy will mistakenly select the channels that contain mostly reverberant energy. It is clear from this example that the maximum selection criterion is ineffective in terms of capturing relevant spectral information, particularly, in segments where overlap-masking effects are dominant. In contrast, as illustrated in Fig. 7(d), the IRM strategy which relies on the local SRR criterion will correctly discard the reverberant channels corrupted by late reflections and will select only those channels containing primarily the signal from the (direct and) early reflections. This is also evident in the electrodograms shown in Fig. 2. The IRM strategy [see Fig. 2(c)] correctly selected the high-frequency channels (electrodes 1–8) corresponding to the unvoiced consonant segments (e.g., see the high-frequency consonants near $t = 0.6$ s and $t = 1.5$ s). In contrast, the ACE strategy [see Fig. 2(b)] erroneously selected the mid-frequency channels (electrodes 12–15) instead of the high-frequency channels for the consonant segments near $t = 0.6$, 1.2, 1.5, and 2.1 s. It is clear from Fig. 2 that the IRM strategy was quite effective in capturing high-frequency consonant information.

Overall, the IRM strategy, when compared to ACE, appears to be more robust amidst reverberation, since it does not select any of the channels corrupted by reverberation. The channel-selection criterion used in ACE is critical and seems to depend on the application at hand. In situations where additive noise is present, for instance, the optimal channel-selection criterion is the local SNR of each T-F unit. Hu and Loizou (2008) have demonstrated that when the SNR-based selection criterion is used, substantial gains in intelligibility can be achieved in various masker conditions. In fact, performance obtained by CI listeners using the SNR-based selection criterion in low SNR conditions approached that attained in quiet. Similar outcomes were observed with normal-hearing listeners (Brungart *et al.*, 2006; Li and Loizou, 2008). Based on the findings of experiment 2, the SRR criterion seems to be the optimal channel-selection criterion for reverberant conditions. It is optimal in the sense that this criterion enabled CI users to recognize words in extremely reverberant rooms ($RT_{60} = 1.0$ s) at a level near that attained in quiet listening conditions. Although not examined in the present study, the SRR criterion can potentially be applied in situations wherein both reverberation and additive noise are present. Further experiments, however, are needed to assess the performance of the SRR criterion in such paradigms.

C. Practical implications

As demonstrated above, the proposed selection criterion used in the IRM strategy offers numerous advantages over

ACE. However, in a practical system the SRR values need to be estimated directly from the reverberant envelopes. Since the SRR criterion implicitly identifies the presence and absence of unvoiced segments, segmentation techniques can alternatively be used to estimate the SRR criterion from the reverberant envelopes. Signal processing algorithms that detect voiced and unvoiced boundaries and further emphasize spectro-temporal regions, where the direct energy dominates the energy from the late reflections, can be developed. Such an approach was recently proposed by Palomäki *et al.* (2004) who suggested to first filter the low-frequency modulations of the reverberant envelopes, and then to apply a threshold to the filtered envelopes.

In a similar vein, implementing the SNR-selection criterion proposed for noise-reduction applications (Hu and Loizou, 2008) is admittedly more challenging, since in principle it requires estimation of the masker envelopes. Despite the challenge, signal processing techniques have been developed and recently applied to CI users (Hu and Loizou, 2010) and normal-hearing listeners (Kim *et al.*, 2009) with some success. In brief, although challenging, the task of estimating the SRR criterion directly from the reverberant envelopes, either explicitly as per Eq. (3) or implicitly as per Palomäki *et al.* (2004), is feasible. The task of real-time processing for reverberation suppression is technically very difficult, and hence, further research is warranted to pursue such a task. The present study demonstrated the potential, in terms of intelligibility benefits, of using the SRR selection criterion for reverberation suppression.

V. CONCLUSIONS

Based on the outcomes of experiments 1 and 2, the following conclusions can be drawn:

(1) Reverberation adversely affects sentence recognition by CI users. In fact, a decaying-exponential model provided an excellent fit ($r = 0.99$) to the data [see Fig. 1(b)]. Based on this model, speech intelligibility by CI users in reverberation degrades exponentially as RT_{60} increases.

(2) The existing channel-selection criterion used in the ACE strategy is problematic when reverberation is present, especially in unvoiced or low-energy speech segments (e.g., fricatives, stops, and stop closures). During the unvoiced segments (e.g., stops) of the utterance, where the overlap-masking effects dominate, the ACE strategy mistakenly selects the channels containing reverberant energy, since those channels have the highest energy. Thus, the maximum selection criterion used in ACE is not appropriate for reverberant environments.

(3) A new channel-selection criterion based on the SRR of the individual frequency channels was proposed. Unlike ACE, the proposed speech coding strategy (IRM), which incorporates the SRR criterion, only retains channels with SRR values larger than a threshold, while eliminating channels with SRR values lower than the threshold. In highly reverberant conditions ($RT_{60} = 1.0$ s), the IRM strategy was found to yield substantial gains in intelligibility (30–55 percentage points) when the threshold T assumed values in the range of $-15 \leq T \leq -5$ dB. The baseline performance with reverberant stimuli improved from 20% correct to 85% correct when $T = -5$ dB.

(4) The proposed channel-selection criterion, based on SRR, is efficient in suppressing reverberation and in eliminating temporal envelope smearing effects caused by overlap-masking. Following the use of the proposed channel-selection criterion, self-masking effects, which are known to produce flattened formant transitions, were also diminished. As examined in the present study, the construction of the SRR criterion assumed a priori knowledge of the clean target envelopes. Although technically challenging, signal processing techniques can be developed that can estimate the SRR directly from reverberant envelopes. Given the large gains in intelligibility demonstrated with the use of the SRR-based channel-selection criterion (see Fig. 5), further research is warranted to pursue the development of such algorithms.

The outcomes in experiments 1 and 2 have important implications for understanding the overall effects of reverberation and also for designing strategies capable of estimating the SRR criterion with the intent of suppressing reverberation for improved speech intelligibility in CI devices. The present study provided a thorough assessment of the overall effects of reverberation on speech intelligibility by CI listeners and shed new light on the limitations that CI users face in challenging acoustic environments.

ACKNOWLEDGMENTS

This work was supported by Grants R03 DC 008882 (Kostas Kokkinakis) and R01 DC 007527 (Philipos C. Loizou) awarded from the National Institute of Deafness and other Communication Disorders (NIDCD) of the National Institutes of Health (NIH). The authors would like to thank the CI patients for their time and dedication during their participation in this study. The authors would also like to acknowledge the help of Cochlear Limited. We also thank Dr. Jan Wouters of ESAT of K. U. Leuven and Dr. Arlene C. Neuman of the New York University Langone Medical Center.

[1]The DRR is defined as the log energy ratio of the direct and reverberant portions of an impulse response and essentially measures how much of the energy arriving is due to the direct source sound and how much is due to late echoes. Since the DRR depends on the source-to-listener distance, it decreases as perceived reverberation increases.

[2]The term *dereverberation* is often used to define the broad category of signal processing strategies, which attempt to undo (partially or completely) the perceptual artifacts associated with speech reproduction inside reverberant spaces (Huang *et al.*, 2007).

Assmann, P. F., and Summerfield, Q. (**2004**). "The perception of speech under adverse acoustic conditions," in *Speech Processing in the Auditory System*, edited by S. Greenberg, W. A. Ainsworth, A. N. Popper, and R. R. Fay (Springer, New York), pp. 231–308.

Bolt, R. H., and MacDonald, A. D. (**1949**). "Theory of speech masking by reverberation," J. Acoust. Soc. Am. **21**, 577–580.

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.

Furuya, K., and Kataoka, A. (**2007**). "Robust speech dereverberation using multichannel blind deconvolution with spectral subtraction," IEEE Trans. Audio, Speech, Lang. Process. **15**, 1579–1591.

Haykin, S. (**2000**). *Unsupervised Adaptive Filtering, Volume 2: Blind Deconvolution* (Wiley, New York), pp. 1–12.

Hu, Y., and Loizou, P. C. (**2008**). "A new sound coding strategy for suppressing noise in cochlear implants," J. Acoust. Soc. Am. **124**, 498–509.

Hu, Y., and Loizou, P. C. (**2010**). "Environment-specific noise suppression for improved speech intelligibility by cochlear implant users," J. Acoust. Soc. Am. **127**, 3689–3695.

Huang, Y., Benesty, J., and Chen J. (**2007**). "Dereverberation," in *Springer Handbook of Speech Processing*, edited by J. Benesty, M. M. Sondhi, and Y. Huang (Springer, New York), pp. 929–943.

IEEE (**1969**). "IEEE recommended practice speech quality measurements," IEEE Trans. Audio Electroacoust. **AU17**, 225–246.

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, G., Lu, Y., Hi, Y., and Loizou, P. C. (**2009**). "An algorithm that improves speech intelligibility in noise for normal-hearing listeners," J. Acoust Soc. Am. **126**, 1486–1494.

Kjellberg, A. (**2004**). "Effects of reverberation time on the cognitive load in speech communication: Theoretical considerations," Noise Health **7**, 11–21.

Kokkinakis, K., and Loizou, P. C. (**2009**). "Selective-tap blind dereverberation for two-microphone enhancement of reverberant speech," IEEE Signal Process. Lett. **16**, 961–964.

Kokkinakis, K., and Loizou, P. C. (**2010**). "Multi-microphone adaptive noise reduction strategies for coordinated stimulation in bilateral cochlear implant devices," J. Acoust. Soc. Am. **127**, 3136–3144.

Krishnamoorthy, P., and Prasanna, S. R. (**2009**). "Reverberant speech enhancement by temporal and spectral processing," IEEE Trans. Audio, Speech, Lang. Process. **17**, 253–266.

Li, N., and Loizou, P. C. (**2008**). "Factors influencing intelligibility of ideal binary-masked speech: Implications for noise reduction," J. Acoust. Soc. Am. **123**, 1673–1682.

Litovsky, R., Colburn, S., Yost, W., and Guzman, S. (**1999**). "The precedence effect," J. Acoust. Soc. Am. **106**, 1633–1654.

Loizou, P. C. (**2006**). "Speech processing in vocoder-centric cochlear implants," in *Cochlear and Brainstem Implants*, edited by A. Moller (Karger, Basel, Switzerland), Vol. 64, pp. 109–143.

McDermott, H. J., McKay, C. M., and Vandali, A. E. (**1992**). "A new portable sound processor for the University of Melbourne/Nucleus Limited multichannel cochlear implant," J. Acoust. Soc. Am. **91**, 3367–3371.

McKay, C. M., and McDermott, H. J. (**1993**). "Perceptual performance of subjects with cochlear implants using the Spectral Maxima Sound Processor (SMSP) and the Mini Speech Processor (MSP)," Ear Hear. **14**, 350–367.

Miyoshi, M., and Kaneda, Y. (**1988**). "Inverse filtering of room acoustics," IEEE Trans. Speech Audio Process. **36**, 145–152.

Nabelek, A. K., and Letowski, T. R. (**1985**). "Vowel confusions of hearing-impaired listeners under reverberant and non-reverberant conditions," J. Speech Hear. Disord. **50**, 126–131.

Nabelek, A. K., and Letowski, T. R. (**1988**). "Similarities of vowels in non-reverberant and reverberant fields," J. Acoust. Soc. Am. **83**, 1891–1899.

Nabelek, A. K., Letowski, T. R., and Tucker, F. M. (**1989**). "Reverberant overlap- and self-masking in consonant identification," J. Acoust. Soc. Am. **86**, 1259–1265.

Nabelek, A. K., and Picket, J. M. (**1974**). "Monaural and binaural speech perception through hearing aids under noise and reverberation with normal and hearing-impaired listeners," J. Speech Hear. Res. **17**, 724–739.

Neuman, A. C., Wroblewski, M., Hajicek, J., and Rubinstein, A. (**2010**). "Combined effects of noise and reverberation on speech recognition performance of normal-hearing children and adults," Ear Hear. **31**, 336–344.

Palomäki, K. J., Brown, G. J., and Barker, J. P. (**2004**). "Techniques for handling convolutional distortion with 'missing data' automatic speech recognition," Speech Commun. **43**, 123–142.

Poissant, S. F., Whitmal, N. A., and Freyman, R. L. (**2006**). "Effects of reverberation and masking on speech intelligibility in cochlear implant simulations," J. Acoust. Soc. Am. **119**, 1606–1615.

Seeber, B., and Hafter, E. (**2007**). "Precedence-effect with cochlear implant simulation," in *Hearing—from Sensory Processing to Perception*, edited by B. Kollmeier, G. Klump, V. Hohmann, U. Langemann, M. Mauermann, S. Uppenkamp, and J. Verhey (Springer, Berlin, Germany), Chap. 51, pp. 475–484.

Skinner, M., Holden, L. K., Whitford, L. A., Plant, K. L., Psarros, C., and Holden, T. A. (**2002**). "Speech recognition with the Nucleus 24 SPEAK, ACE, and CIS speech coding strategies in newly implanted adults," Ear Hear. **23**, 207–223.

Spahr, A., and Dorman, M. (**2004**). "Performance of subjects fit with the Advanced Bionics CII and Nucleus 3G cochlear implant devices," Arch. Otolaryngol. Head Neck Surg. **130**, 624–628.

Stevens, K. N. (**2002**). "Toward a model for lexical access based on acoustic landmarks and distinctive features," J. Acoust. Soc. Am. **111**, 1872–1891.

Van den Bogaert, T., Doclo, S., Wouters, J., and Moonen, M. (**2009**). "Speech enhancement with multichannel Wiener filter techniques in multi-microphone binaural hearing aids," J. Acoust. Soc. Am. **125**, 360–371.

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. L., Kjems, U., Pedersen, M. S., Boldt, J. B., and Lunner, T. (**2009**). "Speech intelligibility in background noise with ideal binary time-frequency masking," J. Acoust. Soc. Am. **125**, 2336–2347.

Whitmal, N. A, and Poissant, S. F. (**2009**). "Effects of source-to-listener distance and masking on perception of cochlear implant processed speech in reverberant rooms," J. Acoust. Soc. Am. **126**, 2556–2569.