Reverberation suppression in cochlear implants using a blind channel-selection strategy

Oldooz Hazrati^{a)} and Philipos C. Loizou

Department of Electrical Engineering, The University of Texas at Dallas, Richardson, Texas 75080

(Received 10 December 2012; revised 17 April 2013; accepted 22 April 2013)

Reverberation severely degrades speech intelligibility for cochlear implant (CI) users. The ideal reverberant mask (IRM), a binary mask for reverberation suppression which is computed using signal-to-reverberant ratio, was found to yield substantial intelligibility gains for CI users even in highly reverberant environments (e.g., $T_{60} = 1.0$ s). Motivated by the intelligibility improvements obtained from IRM, a monaural blind channel-selection criterion for reverberation suppression is proposed. The proposed channel-selection strategy is blind, meaning that prior knowledge of neither the room impulse response nor the anechoic signal is required. By the use of a residual signal obtained from linear prediction analysis of the reverberant signal, the residual-to-reverberant ratio (RRR) of individual frequency channels was employed as the channel-selection criterion. In each frame, the channels with RRR less than an adaptive threshold were retained while the rest were zeroed out. Performance of the proposed strategy was evaluated via intelligibility listening tests conducted with CI users in simulated rooms with two reverberation times of 0.6 and 0.8 s. The results indicate significant intelligibility improvements in both reverberant conditions (over 30 and 40 percentage points in $T_{60} = 0.6$ and 0.8 s, respectively). The improvement is comparable to that obtained with the IRM strategy. © 2013 Acoustical Society of America.

[http://dx.doi.org/10.1121/1.4804313]

PACS number(s): 43.66.Ts, 43.71.Ky, 43.71.Gv [JFC]

Pages: 4188–4196

I. INTRODUCTION

In addition to ambient noise, room reverberation is one of the main factors affecting the quality and intelligibility of speech for both human and machine listeners (Brown and Pålomaki, 2006; Sadjadi and Hansen, 2011). Room reverberation is a result of multiple reflections and diffractions of sounds from the walls and objects in enclosed spaces. Reverberation does not degrade overall speech intelligibility in normal-hearing (NH) listeners when the reverberation time (T_{60}) is less than 1.0 s (e.g., see Nabelek and Letowski, 1988; Kjellberg, 2004). However, in listeners with sensorineural hearing loss, intelligibility deteriorates considerably even in environments where the reverberation time exceeds 0.5 s. This is due to the fact that reverberation blurs temporal and spectral cues, flattens formant transitions, reduces envelope 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 Pickett, 1974; Nabelek and Letowski, 1988; Nabelek *et al.*, 1989; Assmann and Summerfield, 2004). The intelligibility degradation due to reverberation is attributed to two masking effects: (1) overlap-masking which is caused by temporal smearing of high energy phonemes (e.g., vowels) that mask the succeeding sounds and (2) self-masking, which is due to the internal temporal smearing of energy within each phoneme.

Vowel recognition scores of elderly adults with binaural sensorineural hearing loss were shown to be

12 percentage points lower than the mean score obtained in a non-reverberant (anechoic) listening environment when subjects were tested in a room with reverberation time of 1.2 s (Nabelek and Letowski, 1985). Neuman *et al.* (2010) studied the speech recognition performance of NH children and adults in reverberation and showed that when the initial reverberation time was increased by almost a factor of 3, such that $T_{60} = 0.8$ s, the speech reception threshold (SRT) increased by approximately 2–3 dB.¹ Although there is a high correlation between the results obtained from testing cochlear implant (CI) users and those obtained with presenting simulated vocoder stimuli to NH listeners, all studies on the effects of reverberation on speech intelligibility in CI users, except for the work by Kokkinakis *et al.* (2011), have been conducted only using vocoder stimulations presented to NH listeners (Poissant *et al.*, 2006; Whitmal and Poissant, 2009). In our previous study we evaluated the effects of reverberation on speech intelligibility for six CI users at four different reverberation times ($T_{60} = 0.3, 0.6, 0.8$, and 1.0 s). We found that reverberation adversely affects speech recognition in CI users and their performance degrades exponentially as T_{60} increases (Kokkinakis *et al.*, 2011).

Several speech dereverberation algorithms have been proposed in order to improve the quality or intelligibility of reverberant speech (e.g., see Huang *et al.*, 2007; Naylor and Gaubitch, 2010). However, little is known about the effectiveness of such algorithms in improving speech intelligibility for CI users. In addition, existing dereverberation algorithms are computationally expensive, which makes their integration into CIs a formidable task.

^{a)}Author to whom correspondence should be addressed. Electronic mail: hazrati@utdallas.edu

Regardless of the speech coding strategy used in CI devices, most CI users are able to achieve open-set speech recognition scores of 80% or higher in quiet anechoic conditions (e.g., see Skinner *et al.*, 2002; Spahr and Dorman, 2004). However, current speech coding strategies perform poorly in the presence of noise or reverberation. For example, advanced combination encoder (ACE) which is one of the most commonly used speech coding strategies in CI processors, selects only a subset of channels (8–12) for stimulation at each analysis window. It operates based on the principle that only peaks of speech in the short-term spectrum are sufficient for speech identification (e.g., see McDermott *et al.*, 1992; McKay and McDermott, 1993; Vandali *et al.*, 2000; Kiefer *et al.*, 2001). Therefore, during the unvoiced segments (e.g., stops) of the reverberant utterance, where the reverberation overlap-masking effect dominates, the ACE strategy mistakenly selects the channels containing reverberant energy, since those channels have the highest energy (Kokkinakis *et al.*, 2011).

In our previous study (Kokkinakis *et al.*, 2011) we proposed a channel-selection criterion, ideal reverberant mask (IRM), based on the signal-to-reverberant ratio (SRR) of the individual frequency channels. The proposed speech coding strategy retained channels with SRR values larger than a fixed threshold and eliminated other channels. This technique was found to result in substantial gains in intelligibility in a highly reverberant condition ($T_{60} = 1.0$ s). However, the IRM strategy assumed *a priori* knowledge of the clean target speech.

As shown by Gaubitch *et al.* (2006), and also assumed in most linear prediction (LP) residual-based dereverberation algorithms, reverberation primarily affects the speech excitation (e.g., see Yegnarayana and Murthy, 2000; Krishnamoorthy and Prasanna, 2009). Accordingly, the LP residual obtained from the reverberant speech signal can be approximately expressed as the LP residual of the anechoic target speech convolved with the room impulse response (RIR). The present study was motivated by the success in employing the LP residual signal in processing reverberant speech signals, and also intelligibility improvements obtained using time-frequency (T-F) masking in noisy and reverberant environments (e.g., see Wang *et al.*, 2009; Kokkinakis *et al.*, 2011). A new channel-selection criterion based on the residual-to-reverberant ratio (RRR) is proposed. The RRR is the ratio of the residual signal energy to reverberant signal energy which roughly models the inverse of vocal tract filter response (i.e., the inverse of formant resonances). The residual signal is obtained from the LP analysis of the reverberant signal. By comparing the energy ratio of the two signals at each frequency bin (at each short time-frame) with an adaptive threshold, a decision is made whether to retain or discard that specific frequency bin information. The proposed channel-selection criterion is blind in the sense that it does not require any *a priori* information of the RIR or the anechoic signal. In order to assess the potential of the proposed criterion in suppressing reverberation, listening experiments were conducted with eleven CI users in two reverberant conditions ($T_{60} = 0.6$ and 0.8 s).

II. NEW CHANNEL-SELECTION CRITERION FOR SUPPRESSING REVERBERATION

A. Methods

1. Subjects

Eleven adult post-lingually deafened native speakers of American English CI users with ages ranging from 48 to 77 yrs (with an average age of 64 yrs) participated in this study.

All participants were paid for their participation. All eleven subjects were using a Nucleus (Cochlear, Ltd) device and used their devices routinely with a minimum of 1 yr experience with their device. Detailed biographical data for the subjects are given in Table I.

2. Research processor

Three subjects tested were using the Cochlear ESprit 3 G device, six were using the Nucleus Freedom device, and the other two were using the Nucleus 5 speech processor. The 11 Nucleus users were temporarily fitted with the SPEAR3 research interface² programmed with the ACE speech coding strategy (Vandali *et al.*, 2000). The Seed-Speak GUI application was used to program the SPEAR3 wearable research processor with the threshold and comfortable levels of each individual user.³ In order to assess the full potential of the proposed channel-selection criterion in reverberation suppression, and to prevent the number of channels and the stimulation rate (clinically used by the CI users) from affecting performance, the proposed method was evaluated as a pre-processor to the SPEAR3 device used for testing CI subjects. As a result of this implementation, the number of selected channels in each cycle and the stimulation rate remained the same as that used in the clinical speech processor.

3. Stimuli

The IEEE sentence corpus (IEEE, 1969), taken from a CD ROM available in Loizou (2007), was used for the listening tests. The IEEE corpus includes 72 lists each containing 10 sentences (10 sentences/list) with 7–12 words produced by a male speaker. The root-mean-square energy of all sentences is equalized to the same value corresponding to approximately 65 dBA. All sentence stimuli were recorded

TABLE I. Biographical data of the CI users tested.

Subjects	Age	Gender	Years implanted (L/R)	Number of active electrodes	CI processor	Etiology of hearing loss
S1	48	F	8/8	22	ESpirit 3 G	Unknown
S2	72	F	9/9	22	Freedom	Unknown
S3	77	M	6/6	22	Freedom	Hereditary
S4	68	M	9/9	20	ESpirit 3 G	Noise
S5	63	F	8/8	20	ESpirit 3 G	Rubella
S6	56	F	4/7	22	Freedom	Unknown
S7	70	F	6/-	22	Freedom	Unknown
S8	52	F	2/1	22	Nucleus 5	Unknown
S9	65	F	15/8	22	Freedom	Unknown
S10	60	F	6/9	22	Freedom	Hereditary
S11	70	F	3/1	20	Nucleus 5	Unknown

at a sampling frequency of 25 kHz and down-sampled to 16 kHz.

4. Simulated reverberant conditions

In order to simulate the reverberant conditions, RIRs recorded by Neuman *et al.* (2010) were used. They used a Tannoy CPA5 loudspeaker inside a rectangular reverberant room with dimensions of 10.06 m × 6.65 m × 3.4 m (length × width × height) and a source-to-microphone distance of 5.5 m (beyond the critical distance) to measure the RIRs. The original RIRs were obtained at 48 kHz and down-sampled to 16 kHz for this study. The overall reverberant characteristics of the experimental room were altered by hanging absorptive panels from hooks mounted on the walls close to the ceiling. The average reverberation time (averaged at frequencies of 0.5, 1, and 2 kHz) of the room before modification was 0.8 s with a direct-to-reverberant ratio (DRR) of −3.00 dB. With nine panels hung, the average reverberation time was reduced to approximately 0.6 s with a DRR of −1.83 dB.

To generate the reverberant (Rev) stimuli, the RIRs obtained for each reverberation condition were convolved with the IEEE sentence stimuli (recorded in anechoic conditions) using a standardized linear convolution algorithm in MATLAB (MATLAB, 2011).

5. Proposed channel-selection criterion

As shown in our previous study (Kokkinakis *et al.*, 2011), the overlap masking effect of reverberation

causes temporal smearing especially when a high-energy voiced segment is followed by a low energy consonant. Consequently, the vowel and consonant boundaries become obscured, thus making the use of the lexical segmentation cues for word retrieval challenging. Moreover, this temporal smearing effect causes the maximum selection criterion used in the ACE strategy to mistakenly select channels during the gaps present in most unvoiced segments of the utterance.

In order to overcome the limitations of the ACE strategy in channel-selection in reverberant environments, here we propose the LP channel-selection criterion for reverberation suppression which only uses the information from the reverberant signal.⁴

Gaubitch *et al.* (2006) have shown that LP coefficients of reverberant speech averaged over multiple spatial locations are approximately equal to those of anechoic speech. The LP spectrograms of a sentence in anechoic and reverberant conditions ($T_{60} = 0.6$ s) are shown in Fig. 1. The LP spectrum of a 20 ms frame of the same sentence (starting from the dashed line in Fig. 1, vowel /ah/ in “pup”⁵) is plotted in Fig. 2 for two reverberation times ($T_{60} = 0.6$ and 0.8 s). As seen in this figure, the overall pattern of LP spectra for reverberant signals (for both reverberation times of $T_{60} = 0.6$ and 0.8 s) are close to that of the anechoic signal. As shown in equations (1)–(8), this similarity is effectively exploited to remove the reverberation energy from the gaps, hence recovering the silence regions and consonant onsets of the speech signal. Therefore, one can conclude that the LP residual of a reverberant signal is “approximately” equal to that of the

FIG. 1. LP spectrogram of the IEEE sentence “The small pup gnawed a hole in the sock” in (a) anechoic, and (b) $T_{60} = 0.6$ s conditions. The arrow and dashed line mark the time at which the plots in Fig. 2 are computed.

FIG. 2. (Color online) LP spectrum of a short frame (vowel /ah/ in word “pup”) of the IEEE sentence “The small pup gnawed a hole in the sock” in (a) $T_{60}=0.6$ s, and (b) $T_{60}=0.8$ s. The solid line spectrum in each panel is the LP spectrum of the same anechoic frame.

anechoic signal convolved with the RIR. Motivated by this fact, the residual signal of the reverberant signal is used to make decisions regarding retaining or discarding a channel at each T-F unit. As depicted in Fig. 3, the reverberant signal and its residual are passed through a 128-channel fourth-order gammatone filterbank, with center frequencies equally spaced on the equivalent rectangular bandwidth scale covering a frequency range of 50 and 8 kHz (half of the sampling frequency) to obtain the T-F representation of the signals. The short-time energy ratio (of 20 ms frames with 50% overlap between successive frames) is computed from the band-pass filtered waveforms as

$$RRR(f, t) = 10 \log_{10} \frac{|X_r(f, t)|^2}{|R(f, t)|^2}, \quad (1)$$

where the $RRR(f, t)$ stands for residual-to-reverberant ratio, and $|R(f, t)|^2$ and $|X_r(f, t)|^2$ denote the short-time energy of the reverberant and residual signals, respectively. Time frame and frequency indices are represented by t and f , respectively. Based on the linear source-filter model of speech production

$$S(f, t) = X(f, t)GV(f, t), \quad (2)$$

where $S(f, t)$ is the clean speech spectrum at frame t , $X(f, t)$ is the source excitation, $V(f, t)$ is the vocal tract filter, and G is a gain factor. Similarly, for reverberant speech we have

$$R(f, t) = X_r(f, t)GV(f, t), \quad (3)$$

where $X_r(f, t)$ can be written as

$$X_r(f, t) = X(f, t)H(f, t), \quad (4)$$

with $H(f, t)$ representing the RIR at time frame t and frequency bin f . Accordingly, the $RRR(f, t)$ can be re-written as

$$RRR(f, t) = 10 \log_{10} \frac{1}{|GV(f, t)|^2}, \quad (5)$$

which implicitly represents the inverse of the vocal tract filter at each short time frame. In the next stage, the $RRR(f, t)$ is compared against an adaptive threshold, computed as

$$T(f, t) = \frac{\alpha}{I+1} \sum_{i=0}^I RRR(f, t-i), \quad (6)$$

where α is a constant and I is the number of previous frames contributing to the computation of adaptive threshold which are selected experimentally. At each T-F unit, if the RRR value is less than the threshold (implying that there are strong formant peaks in the spectrum), the channel is retained, otherwise it is “zeroed out” using a binary gain function as

$$\hat{S}(f, t) = R(f, t)BG(f, t), \quad (7)$$

FIG. 3. Block diagram of the proposed channel-selection algorithm.

$$BG(f, t) = \begin{cases} 1, & RRR(f, t) < T(f, t), \\ 0, & \text{otherwise.} \end{cases} \quad (8)$$

In T-F units where the direct speech is dominant (T-F units with higher SRR values), the LP spectrum shows larger peaks. Accordingly, by comparing the RRR of each T-F unit with the weighted average of a few previous T-F units (at the same frequency bin), the higher SRR regions can be determined. In this manner, the number of retained channels at each time-frame can vary from 0 to 128 (total number of channels). At the last stage (synthesis), phase shifts are corrected by time-reversing the envelope in each channel, passing it through the gammatone filterbank, time-reversing again, and then summing across all selected channels. The synthesized stimuli are presented monaurally to the CI users via the auxiliary input jack of the SPEAR3 processor (Nucleus CI users) through the best ear. It is worth emphasizing that the proposed binary gain function applied in Eq. (7) suppresses reverberation by selecting the channels satisfying the RRR criterion in Eq. (8) rather than specifically performing speech dereverberation. By the use of a low order LP analysis (for example, 10th order), the LP spectra only represent the general shape of the vocal tract filter (represented by formant peaks) which is preserved in the presence of reverberation. If the order of the LP analysis is increased, the model will capture the fine structure of the vocal tract response as well. The fine structure is more prone to be affected by reverberation and/or noise, which makes the RRR and threshold estimation erroneous.

The proposed channel-selection criterion is called LP because it is based on the use of the residual signal of the reverberant signal obtained from the LP analysis. As mentioned earlier, this channel-selection criterion, which is evaluated offline, can be implemented in place of the maximum

channel-selection in ACE for real-time applications, by replacing the gammatone filterbank with the fast Fourier transform (FFT). It is worth remarking here that using the SPEAR3 processor allowed the proposed channel-selection strategy to be implemented as a pre-processor, so the number of selected channels in each cycle and the stimulation rate remained the same as that used in the clinical speech processor of each patient.

B. Procedure

Each subject participated in a short practice session to become familiar with the listening task, prior to testing. The stimuli were presented to the 11 Nucleus users unilaterally through the auxiliary input jack of the SPEAR3 processor in a double-wall sound proof booth (Acoustic Systems, Inc.). For the bilateral users, the ear with the highest sentence score in quiet was chosen for testing. During the practice session, the subjects adjusted the volume level to a comfortable level, and the volume level was fixed throughout the tests. To avoid listener fatigue, subjects were given a 15 min break every 60 min during the test session. Subjects participated in a total of seven conditions corresponding to (a) anechoic ($T_{60} \approx 0.0$ s) quiet, (b) two different reverberant ($T_{60}=0.6$ and 0.8 s), (c) two IRM processed reverberant ($T_{60}=0.6$ and 0.8 s), and (d) two LP estimated mask processed reverberant ($T_{60}=0.6$ and 0.8 s) conditions. The unprocessed anechoic condition was used as a control condition. Twenty IEEE sentences (two lists) were used per condition. None of the lists was repeated across conditions. The order of the test conditions was randomized across subjects to minimize the order effects. During testing, the participants were instructed to repeat as many words as they could identify. The responses of each individual were collected and

FIG. 4. Individual percent correct scores of eleven CI users tested on IEEE sentences using anechoic (unprocessed) acoustic inputs recorded in reverberation (black bars) and reverberant acoustic inputs processed with the LP channel-selection criterion (white bars) and IRM strategy (gray bars), (a) $T_{60}=0.6$ s and (b) $T_{60}=0.8$ s. Error bars indicate standard deviations.

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 recognized by the total number of words in the sentence lists tested.

C. Results

The proposed channel-selection criterion has been evaluated in two reverberant conditions ($T_{60} = 0.6$ and 0.8 s). The individual speech intelligibility scores for both experimental conditions are shown in Fig. 4 (the individual as well as mean intelligibility scores of un-processed reverberant and IRM-processed signals are also presented in this figure). The performance was measured in terms of percentage of the words distinguished correctly by each CI user. As can be seen from Fig. 4, significant gains in word recognition scores were obtained in both reverberant conditions using the proposed method ($F^{1,10} = 340.667$, $p < 0.001$). The speech intelligibility scores improved from an average of 44.16 and 32.94%, to 73.87 and 67.65% at $T_{60} = 0.6$ and 0.8 s conditions, respectively. The average score in the anechoic quiet condition was 87.36%.

Post hoc comparisons were performed to assess differences in scores obtained with un-processed reverberant, proposed LP-based masking method, and the IRM strategy. The results indicated that in both reverberant conditions ($T_{60} = 0.6$ and 0.8 s), intelligibility of the processed stimuli improved significantly ($p < 0.0001$, paired samples t -tests, Bonferroni protected) compared to the unprocessed reverberant stimuli. Moreover, paired samples t -tests with Bonferroni protection indicated that the performance of IRM strategy was significantly greater than that of the LP strategy for $T_{60} = 0.8$ s ($p < 0.0001$), but no significant difference for $T_{60} = 0.6$ s condition was observed.

III. DISCUSSION

The intelligibility scores obtained from the LP channel-selection criterion surpass that of traditional maximum channel-selection criterion used in ACE (Fig. 4). This is because of the accurate T-F unit classification as reverberation-dominant or speech-dominant. An example of this fact is demonstrated in Fig. 5 for the stimulus output patterns (electrograms) of an IEEE sentence processed with the ACE speech coding strategy in the Nucleus 24 device. In all panels of Fig. 5, the vertical axes represent the electrode position corresponding to a specific frequency, where the highest electrode number stands for the lowest frequency and vice versa. All vowel and consonant boundaries are present in the anechoic stimuli shown in Fig. 5(a). As seen in Fig. 5(b), the reverberation self- and overlap-masking effects are evident in the channels selected by the ACE strategy in reverberation ($T_{60} = 0.8$ s). The vowel-consonant boundaries are filled with the reverberant energy and therefore the cues used for word recognition by CI users are masked. As shown in Kokkinakis *et al.* (2011), ACE selects the correct channels during the voiced segments of the reverberant utterance, but it mistakenly selects the channels with the highest reverberant energy (instead of low-energy unvoiced information

FIG. 5. (Color online) Stimulus output patterns (electrograms) of the IEEE sentence “the beauty of the view stunned the young boy” uttered by a male speaker. (a) Electrodiagram of the anechoic (uncorrupted) sentence processed by the ACE strategy, (b) electrodiagram of the same reverberant sentence ($T_{60} = 0.8$ s) processed by the ACE strategy, and (c) electrodiagram of the reverberant sentence processed by the LP channel-selection criterion as a pre-processing before ACE strategy. In each electrodiagram, time is shown along the abscissa and the electrode number is shown along the ordinate. The ARPABET phonetic transcription of the sentence is displayed on top of the first electrodiagram in red.

carrying channels) during the unvoiced segments of the speech signal. The electrodiagram of the same reverberant sentence processed with the LP channel-selection criterion as a pre-processing stage before the maximum amplitude selection criterion in ACE, is shown in panel (c) of Fig. 5. As shown in this figure, the vowel-consonant boundaries are recovered to a great extent, which will provide useful acoustic cues for speech recognition by CI users. There are still some effects of reverberant energy present in the channels

FIG. 6. (Color online) Example illustrating the bandpass filtered IEEE sentence “the beauty of the view stunned the young boy” extracted at the center frequency of 1.5 kHz. (a) Anechoic (unprocessed) sentence, (b) reverberant signal ($T_{60}=0.8$ s), (c) instantaneous RRR values (dashed line) along with the adaptive threshold (dB), and (d) reverberant sentence ($T_{60}=0.8$ s) processed by the LP channel-selection strategy.

selected in Fig. 5(c), as well as some missed low-frequency channels, but the effect of reconstructing true boundaries and removing the reverberant energy filling in the gaps seems to be more dominant in the intelligibility scores.

The use of the RRR and adaptive threshold in removing the reverberation energy from the gaps is shown in Fig. 6. Panel (a) of this figure shows an anechoic bandpass filtered IEEE sentence at center frequency of 1.5 kHz. The gaps are clearly visible in this figure (e.g., $t=0.75\text{--}0.8$ s or $t=0.89\text{--}0.99$ s). The reverberant ($T_{60}=0.8$ s) bandpass filtered signal at the same channel is plotted in Fig. 6(b). It is clear that the overlap-masking effect caused by reverberation has filled the gaps. At each time-frame, by comparing the RRR values with the adaptive threshold of that frame, a decision as to whether retain or discard different frequency channels is made [as shown for one example channel in Fig. 6(c)]. As seen in panel (d) of the same figure, the gaps of the utterance previously filled with reverberant energy are recovered. Although some low-level information (e.g., $t=0.85\text{--}1.05$ s) is lost at this example frequency band, they are retained in other frequency bands resulting in recovery of an attenuated version of that segment that still provides useful intelligibility cues to the listeners. If this restoration fails in the majority of the frequency bands then the segment is entirely lost which results in a drop in speech intelligibility of the enhanced signal. Here, the constant value, α , in Eq. (6) is set to 1.05, and 15 previous frames are used for computing the adaptive threshold. The values for different algorithm parameters (e.g., α and D), although effective for RIRs and reverberation times considered in this study, need to be

tuned for different reverberant conditions for optimal performance. As perceived from Figs. 5 and 6, the maximum selection criterion is ineffective in capturing useful spectral information especially in the low-energy regions following high-energy segments. On the contrary, the LP method, which relies only on the information obtained from the reverberant signal (and its residual signal), helps remove the reverberation from these segments, thereby resulting in more accurate channel-selection by the ACE strategy (Figs. 5 and 6). The proposed algorithm has the potential to be implemented for real-time processing in CIs. However, depending on the digital signal processor (DSP) used in CI devices, the computation delay may vary and might be noticeable to the CI users.⁶

As presented in Table II, the results obtained after applying the proposed LP channel-selection criterion to the reverberant speech are comparable to those obtained with IRM channel-selection strategy which assumed *a priori* knowledge of the anechoic quiet target signal. Also evident from the analysis of variance tests, performance obtained

TABLE II. Average word recognition scores obtained by eleven CI users tested on IEEE reverberant sentences processed using LP and IRM strategies. The numbers in parentheses indicate standard deviations.

Strategy	$T_{60}=0.6$ s	$T_{60}=0.8$ s
Unprocessed	44.16 (14.10)	32.94 (11.96)
LP	73.87 (12.70)	67.65 (10.99)
IRM	79.76 (12.36)	79.30 (10.95)

from the LP criterion at $T_{60}=0.6$ was not significantly different from that achieved with IRM.

Although not evaluated with CI listeners in this study, according to the spectrogram comparisons, the proposed strategy can still be effective only in situations where the SNR level is high (SNR > 20 dB). This is expected as the LP analysis would not be reliable when additive noise is present in addition to reverberation.

IV. CONCLUSIONS

Motivated by intelligibility improvements obtained using the IRM strategy, which can be viewed as an upper bound in performance, a new blind (i.e., no *a priori* knowledge of anechoic signal is required) channel-selection criterion for reverberation suppression in CIs was proposed. In this strategy, the speech-dominant channels are retained according to the RRR criterion obtained solely from the reverberant signal. The proposed method was found to be quite successful in recovering the smeared boundaries of speech yielding substantial speech intelligibility improvements for CI users. This technique can also be implemented at the FFT level of the ACE strategy. Moreover, as a single-microphone method, it is more applicable compared to multi-microphone techniques.

ACKNOWLEDGMENTS

This work was supported by Grant No. R01 DC 010494 (P.C.L.) awarded from the National Institute on 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 Dr. Arlene C. Neuman of the NYU Langone Medical Center for providing the RIRs and thank Dr. Omid Sadjadi and Dr. Nirmal Srinivasan for their help in preparing the manuscript.

¹SNR loss reflects the increase in SNR required attaining 50% correct performance due to reverberation. This is relative to the SNR required by NH adults in anechoic, noise-alone conditions (SRT).

²SPEAR3 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.

³The ACE strategy implemented in the SPEAR3 processor is very similar to that implemented in the 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.

⁴Palomäki *et al.* (2004) introduced a reverberant binary mask which was based on subband Hilbert envelopes of the reverberant speech and was used primarily for automatic speech recognition (ASR). Although resulting in large gains in ASR, it did not improve the intelligibility of reverberant speech for CI users, and caused a decrement in CI users' performance in reverberation. We speculate that this intelligibility degradation was due to aggressiveness of the binary mask.

⁵In this study, we use ARPABET phonetic symbols.

⁶We ran the ACE strategy with and without the LP strategy and compared the required CPU time for both cases. We found that the CPU time for the ACE+LP was 1.5 times of ACE. For example, if the ACE processing requires 8 ms to run, the ACE+LP will run in less than 15 ms which is less

than the delay noticeable by CI users (usually delays greater than 25 ms are noticeable to the CI users).

- 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.
- Brown, G. J., and Pälomäki, K. J. (2006). "Reverberation," in *Computational Auditory Scene Analysis: Principles, Algorithms, and Applications*, edited by D. L. Wang and G. J. Brown (Wiley, New York), pp. 209–250.
- Gaubitch, N. D., Ward, D. B., and Naylor, P. A. (2006). "Statistical analysis of the autoregressive modeling of reverberant speech," *J. Acoust. Soc. Am.* **120**, 4031–4039.
- Huang, Y., Benesty, J., and Chen, J. (2007). "Dereverberation," in *Handbook of Speech Processing*, edited by J. Benesty, M. M. Sondhi, and Y. Huang (Springer, Germany), pp. 929–943.
- IEEE (1969). "IEEE recommended practice for speech quality measurements," *IEEE Trans. Audio Electroacoust.* **AU-17**, 225–246.
- Kiefer, J., Hohl, S., Stürzebecher, E., Pfennigdorff, T., and Gstöettner, W. (2001). "Comparison of speech recognition with different speech coding strategies (SPEAK, CIS, and ACE) and their relationship to telemetry measures of compound action potentials in the Nucleus CI 24M cochlear implant system," *Audiol.* **40**, 32–42.
- Kjellberg, A. (2004). "Effects of reverberation time on the cognitive load in speech communication: Theoretical considerations," *Noise Health.* **7**, 11–21.
- Kokkinakis, K., Hazrati, O., and Loizou, P. C. (2011). "A channel-selection criterion for suppressing reverberation in cochlear implants," *J. Acoust. Soc. Am.* **129**, 3221–3232.
- Krishnamoorthy, P., and Prasanna, S. R. (2009). "Reverberant speech enhancement by temporal and spectral processing," *IEEE Trans. Audio Speech Lang. Process.* **17**, 253–266.
- Loizou, P. C. (2007). *Speech Enhancement: Theory and Practice* (Taylor and Francis, Boca Raton, FL), pp. 1–608.
- MATLAB (2011). version 7.12.0 (The MathWorks Inc., Natick, MA).
- 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.
- 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 Pickett, 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.
- Naylor, P. A., and Gaubitch, N. D. (2010). *Speech Dereverberation* (Springer, United Kingdom), pp. 1–388.
- 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 Parker, 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.
- Sadjadi, O., and Hansen, J. H. (2011). "Hilbert envelope based features for robust speaker identification under reverberant mismatched conditions," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Prague, Czech Republic, pp. 5449–5451.
- Skinner, M. W., 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. J., and Dorman, M. F. (2004). “Performance of subjects fit with the Advanced Bionics CII and Nucleus 3G cochlear implant devices,” *Arch. Otolaryngol. Head Neck Surg.* **130**, 624–628.
- 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., and Lunar, 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.
- Yegnanarayana, B., and Murthy, P. S. (2000). “Enhancement of reverberant speech using LP residual signal,” *IEEE Trans. Speech Audio. Process.* **8**, 267–281.