

Simultaneous suppression of noise and reverberation in cochlear implants using a ratio masking strategy

Oldooz Hazrati,^{a)} Seyed Omid Sadjadi, Philipos C. Loizou, and John H. L. Hansen
Department of Electrical Engineering, The University of Texas at Dallas, Richardson, Texas 75080

(Received 1 May 2013; revised 3 September 2013; accepted 10 September 2013)

Cochlear implant (CI) recipients' ability to identify words is reduced in noisy or reverberant environments. The speech identification task for CI users becomes even more challenging in conditions where both reverberation and noise co-exist as they mask the spectro-temporal cues of speech in a rather complementary fashion. Ideal channel selection (ICS) was found to result in significantly more intelligible speech when applied to the noisy, reverberant, as well as noisy reverberant speech. In this study, a blind single-channel ratio masking strategy is presented to simultaneously suppress the negative effects of reverberation and noise on speech identification performance for CI users. In this strategy, noise power spectrum is estimated from the non-speech segments of the utterance while reverberation spectral variance is computed as a delayed and scaled version of the reverberant speech spectrum. Based on the estimated noise and reverberation power spectra, a weight between 0 and 1 is assigned to each time-frequency unit to form the final mask. Listening experiments conducted with CI users in two reverberant conditions ($T_{60} = 0.6$ and 0.8 s) at a signal-to-noise ratio of 15 dB indicate substantial improvements in speech intelligibility in both reverberant-alone and noisy reverberant conditions considered. © 2013 Acoustical Society of America. [<http://dx.doi.org/10.1121/1.4823839>]

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

Pages: 3759–3765

I. INTRODUCTION

Reverberation, which results from multiple reflections of sounds from objects and surfaces in an acoustic enclosure, causes two types of spectro-temporal smearing effects on speech: (i) self-masking (caused by early reflections) which is the internal smearing of energy within each phoneme and (ii) overlap-masking (caused by late reflections) which is due to the temporal smearing of high energy phonemes that mask their succeeding sounds. Reverberation not only degrades speech intelligibility for cochlear implant (CI) and hearing impaired listeners significantly (Hazrati *et al.*, 2013; Hazrati and Loizou, 2013; Nabelek and Letowski, 1985), but also poses a detrimental effect on the performance of automatic speech recognition (Palomäki *et al.*, 2004), and speaker identification systems (Sadjadi and Hansen, 2012). Unlike reverberation, noise is additive and affects speech in a different and complimentary fashion. Noise masks the weak consonants to a greater degree than the higher intensity vowels, but unlike reverberation this masking does not depend on the energy of the preceding segments (Nabelek *et al.*, 1989).

Hence, the combined effects of reverberation and noise adversely affect speech intelligibility more than either reverberation or noise alone effect (e.g., Nabelek and Mason, 1981). Useful speech-information extraction becomes challenging for CI users due to envelope modulation reductions caused by reverberation. Moreover, CI users rely on a limited number of channels of information which makes speech recognition in noise a formidable task (e.g., Friesen *et al.*, 2001). Degraded temporal envelope information and poor spectral resolution will result in inferior levels of speech

understanding by CI users in environments where noise and reverberation co-exist (Hazrati and Loizou, 2012a).

Binary masks have been widely used for different speech enhancement as well as sound-separation applications resulting in gains in intelligibility and quality of the processed noisy/reverberant speech (Wang, 2005; Kim *et al.*, 2009; Kokkinakis *et al.*, 2011). The ideal reverberant mask (IRM) (Kokkinakis *et al.*, 2011) was proposed as an ideal binary masking strategy based on the signal-to-reverberant ratio (SRR) of the individual frequency channels which suppressed reverberation and improved the intelligibility of the reverberant speech significantly. The IRM strategy retained channels with SRR values larger than a fixed threshold while eliminating other channels.

In noisy environments, the channel selection criterion is based on the signal-to-noise ratio (SNR) of individual channels, and the ideal mask selects target-dominated ($\text{SNR} > 0$ dB) channels while discarding masker-dominated ($\text{SNR} < 0$ dB) channels. Large improvements in intelligibility were reported with the SNR channel-selection criterion in ideal conditions where it was assumed that we had access to the clean and noise signals prior to mixing. Large intelligibility gains were found for both CI listeners (Hu and Loizou, 2008) and normal-hearing (NH) listeners (Brungart *et al.*, 2006).

Use of ideal binary masking, also known as ideal channel selection (ICS), was also found to be successful in removing the negative effects of both noise and reverberation from noisy reverberant speech when SRR criterion was adopted (the mask was applied to noisy reverberant speech as opposed to reverberant speech). ICS resulted in large intelligibility gains for both CI recipients (Hazrati, 2012) as well as NH subjects (Hazrati and Loizou, 2012b) tested under reverberation and additive stationary speech-shaped noise (SSN). The study by Roman and Woodruff (2011) also

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

evaluated the effect of different ideal binary masks on the intelligibility of the noisy reverberant speech and found that a binary mask based on the direct path and early reflections results in large intelligibility gains for NH listeners.

Inspired by the large intelligibility gains achieved using ideal binary masking, or ideal time-frequency (T-F) masking, strategies for tackling the negative effects of reverberation, noise, and their combination, several studies proposed and evaluated T-F mask estimation techniques for noisy or reverberant speech enhancement (Hazrati *et al.*, 2013; Hazrati and Loizou, 2013; Kim *et al.*, 2009). However, no masking strategy has been previously proposed in order to suppress both reverberation and noise simultaneously resulting in intelligibility improvements for CI users.

Considering both reverberation and noise in speech intelligibility enhancement techniques is important, especially for CI users, because most studies only consider the individual effects of noise or reverberation which results in under-estimation of most real-life conditions. A single-channel and non-ideal solution to the problem of noisy reverberant speech enhancement for CI users is proposed in the present study.

Specifically, a “ratio,” also known as soft, T-F masking algorithm is proposed and evaluated in the context of noisy reverberant speech intelligibility improvements. Ratio masking refers to algorithms that decompose the signal into T-F units and weight the units based on a given criterion (e.g., SNR for noise suppression). The weight assigned to each unit can take on a value between 0 (discarding that unit) and 1 (no attenuation). If only 0 and 1 are assigned as weights then the mask is called “binary.” Ratio masks are used here as they result in processed speech with better quality due to less “on” and “off” frequency channels (with less roughness and musical noise) compared to binary-masked speech.

The proposed “blind” ratio masking strategy for simultaneously suppressing noise and reverberation in noisy reverberant speech is evaluated in two different reverberant environments ($T_{60} = 0.6$ and 0.8 s) both in the presence and absence of steady-state noise [reverberant signal to noise ratios¹ (RSNRs) = 15 dB]. The proposed strategy is “blind” meaning that no *a priori* information about clean speech or the room impulse response (RIR) is assumed in advance. Of the two reverberation times² considered, one ($T_{60} = 0.6$ s) is allowable in classrooms in the US according to the ANSI S12.60 (2002) standard while the other ($T_{60} = 0.8$ s) exceeds the ANSI recommended values even for larger classrooms.

II. SIMULTANEOUS SUPPRESSION OF REVERBERATION AND MASKING NOISE

A. Methods

1. Subjects

Seven adult post-lingually deafened CI subjects participated in this study. All subjects were native speakers of American English who received no benefit from hearing aids preoperatively. All subjects were paid for their participation. CI users were fitted with the Nucleus 24 multichannel implant device manufactured by Cochlear Corporation

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	59	M	2/-	21/-	N5	Unknown
S2	53	F	3/2	22/22	N5	Unknown
S3	60	F	2/1	22/20	N5	Hereditary
S4	79	M	7/5	22/22	N5	Hereditary
S5	61	F	2/6	22/21	N5	Unknown
S6	65	M	3/-	21/-	N5	Hereditary
S7	65	F	3/3	22/22	N5	Unknown

which used the advanced combination encoder (ACE) processing strategy. All subjects used their devices routinely and had a minimum of one year experience with their devices. The detailed biographical data of the CI subjects are presented in Table I.

2. Research processor

Subjects were tested using a personal digital assistant (PDA-) based CI research platform (Ali *et al.*, 2013) in a double-walled sound-proof booth (Acoustic Systems, Inc.). The signals were streamed off-line via the PDA platform and sent directly to the subject’s cochlear implant unilaterally. The PDA processor was programmed for individual subjects using their threshold and comfortable loudness levels, and coding strategy parameters. The use of a PDA provides us with the advantage of testing all subjects with the same basic ACE implementation, without concern about special settings (e.g., ADRO, BEAM, etc.) in individual CI devices. The volume of the speech processor was also adjusted to a comfortable loudness prior to initial testing. Institutional review board approval was obtained for the listening tests and participants signed informed consent forms prior to testing. Each CI user initially participated in a short practice session to become familiar with the listening task. During the practice session, the subjects were allowed to adjust the volume to their comfort levels. This volume level was fixed throughout the tests. In order to avoid listener fatigue, participants were given a 15 min break every 60 min during the test session.

3. Stimuli

IEEE sentences (IEEE, 1969) were used as the speech stimuli for testing. There are 72 lists of 10 phonetically balanced sentences in the IEEE database, where each sentence is composed of approximately 7–12 words. The root-mean-square level of all sentences was equalized to the same value.

The reverberant stimuli were generated by convolving the clean signals with measured RIRs recorded in a $10.06 \text{ m} \times 6.65 \text{ m} \times 3.4 \text{ m}$ (length \times width \times height) room (Neuman *et al.*, 2010). The reverberation time of the room was varied from 0.8 to 0.6 s by adding absorptive panels to the walls and floor carpeting. The direct-to-reverberant ratios of the RIRs were -1.8 and -3.0 dB for $T_{60} = 0.6$ and 0.8 s, respectively. The distance between the single-source signal and the microphone was 5.5 m, which was beyond

the critical distance ($\simeq 1$ m) (Naylor *et al.*, 2010). SSN with the same long-term spectrum as the test sentences in the IEEE corpus was used as a continuous (steady-state) masker. The masker was added to the reverberant stimuli at 15 dB RSNR level.

B. Signal processing

In this study, the noise is considered to be additive to the reverberant signal as shown in Fig. 1,

$$x(n) = s(n) * h(n) + y(n), \quad (1)$$

where $x(n)$, $s(n)$, $h(n)$, and $y(n)$ denote corrupted signal (by noise and reverberation), anechoic clean signal, RIR, and additive noise, respectively. The model considered here is valid for environments with a steady-state noise source close to the listener and speech source far from the listener. Therefore, the received signal is affected by reverberation convolutively and noise additively.³

The block diagram of the proposed noisy reverberant speech enhancement algorithm is shown in Fig. 1. The noise power spectral density (PSD), $\hat{Y}(t, f)$, is computed from the first 100 ms of the corrupted signal, as this initial segment of the signal is only corrupted by the additive noise and no convolutive distortions due to reverberation exist. As shown in previous studies (Hazrati *et al.*, 2013), late reverberation tends to fill in the speech gaps and blur vowel/consonant boundaries, thereby degrading speech intelligibility for CI users. Thus, removing the reverberation energy from the gaps can restore onsets/offsets of phonemes and provide speech identification cues to CI users. The PSD of late reflections can be modeled as a delayed and smoothed version of the PSD of reverberant speech as follows (Wu and Wang, 2006):

$$|R_l(t, f)|^2 = \alpha w(t - t') * |X(t, f)|^2, \quad (2)$$

with t and f being the time and frequency indices, respectively. Here, $X(t, f)$ and $R_l(t, f)$ represent the complex fast Fourier transform (FFT) spectra of the noisy reverberant speech and late reflections, respectively, and the asterisk denotes the convolution in time domain. The $w(t)$ is a smoothing function (a Rayleigh function is adopted here), α is a scaling factor set to 0.1, and t' is the time threshold between early and late reflections of the RIR which is usually set to 50 ms for speech and is independent of the reverberation time (T_{60}). The superposition of noise and late reverberation PSDs is considered as the PSD of distortion (caused by both reverberation and noise),

$$|D(t, f)|^2 = |R_l(t, f)|^2 + |\hat{Y}(t, f)|^2, \quad (3)$$

which is used in the ratio mask estimation for each T-F unit. The ratio mask is applied to the T-F representation of the corrupted signal as

$$\hat{S}(t, f) = M(t, f)X(t, f), \quad (4)$$

where $M(t, f)$ is the ratio mask of time frame t and frequency bin f that is computed as

$$M(t, f) = \left(\frac{\lambda(t, f)}{\lambda(t, f) + \delta} \right)^\theta, \quad (5)$$

where δ and θ are constant parameters determined experimentally, and λ is the *a priori* signal-to-distortion ratio. The decision-directed method (Ephraim and Malah, 1984) is used to recursively estimate λ as

$$\lambda(t, f) = \mu \cdot \lambda(t - 1, f) + (1 - \mu) \frac{|X(t, f)|^2}{|D(t, f)|^2}, \quad (6)$$

where $|D(t, f)|^2$ is the PSD of distortion (late reverberation and noise) and μ is a constant (set to 0.98). The parameters δ and θ were set to 2 and 1.5, respectively. The sampling frequency, frame length and frame shift were 22 050 Hz, 8 ms, and 2 ms, respectively to match the parameters of the ACE strategy used with the PDA platform (Ali *et al.*, 2013). The inverse FFT of the $\hat{S}(t, f)$ is computed and the enhanced speech is finally re-synthesized using the overlap-add (OLA) method.

C. Procedure

The subjects participated in a total of thirteen conditions presented in Table II. The noisy reverberant ratio-mask processed signals were tested in three different scenarios where (a) both noise and late reverberation PSDs, (b) only noise PSD, and (c) only late reverberation PSD were estimated and used in the mask estimation. The unprocessed sentences in anechoic ($T_{60} \approx 0.0$ s) quiet conditions were used as a control condition. Twenty IEEE sentences (two lists) were used per condition. None of the lists used was repeated across conditions. The order of the test conditions was randomized across subjects in order to minimize any 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 scored off-line based on the number of words correctly identified. All words were

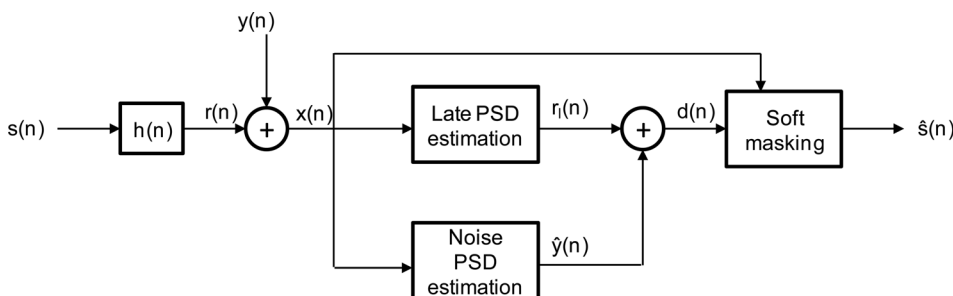


FIG. 1. Block diagram of the proposed noisy reverberant speech enhancement algorithm. Here, $\hat{y}(n)$, $r_l(n)$, and $d(n)$ denote the estimated noise PSD, late reverberation PSD, and the total estimated distortions, respectively.

TABLE II. Listening test conditions. “SM” denotes soft-mask processed stimuli. “SM-noisy reverberant (R=0),” “SM-noisy reverberant (N=0),” and “SM-noisy reverberant” stand for soft-mask processed noisy reverberant speech without late reverberation PSD estimation, soft-mask processed noisy reverberant speech without additive noise PSD estimation, and soft-mask processed noisy reverberant speech, respectively.

Condition	T_{60} (ms)	SNR (dB)
1—Anechoic quiet	0	—
2—Reverberant	600	—
3—Reverberant	800	—
4—Noisy reverberant	600	15
5—Noisy reverberant	800	15
6—SM-reverberant	600	—
7—SM-reverberant	800	—
8—SM-noisy reverberant	600	15
9—SM-noisy reverberant	800	15
10—SM-noisy reverberant (N=0)	600	15
11—SM-noisy reverberant (N=0)	800	15
12—SM-noisy reverberant (R=0)	600	15
13—SM-noisy reverberant (R=0)	800	15

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 sentence lists tested.

III. RESULTS

The individual and mean speech intelligibility scores for both reverberation-alone and noisy reverberant conditions are shown in Fig. 2 (unprocessed as well as ratio-mask processed scores). The speech intelligibility scores improved from an average of 46.42 and 36.59 %, to 60.35 and 55.98 % at $T_{60} = 0.6$ and 0.8 s conditions, respectively. In the presence of noise, speech identification scores improved from an average of 30.89 and 24.96 % to 51.97 and 41.98 %, respectively. The average score in the anechoic quiet condition was 92.46%. An analysis of variance (with repeated measures) confirmed a significant effect ($F[1,6] = 25.59$, $p < 0.005$) of reverberation time, a significant effect of noise ($F[1,6] = 31.19$, $p < 0.005$) and a significant effect of processing ($F[1,6] = 47.03$, $p < 0.005$) on speech intelligibility.

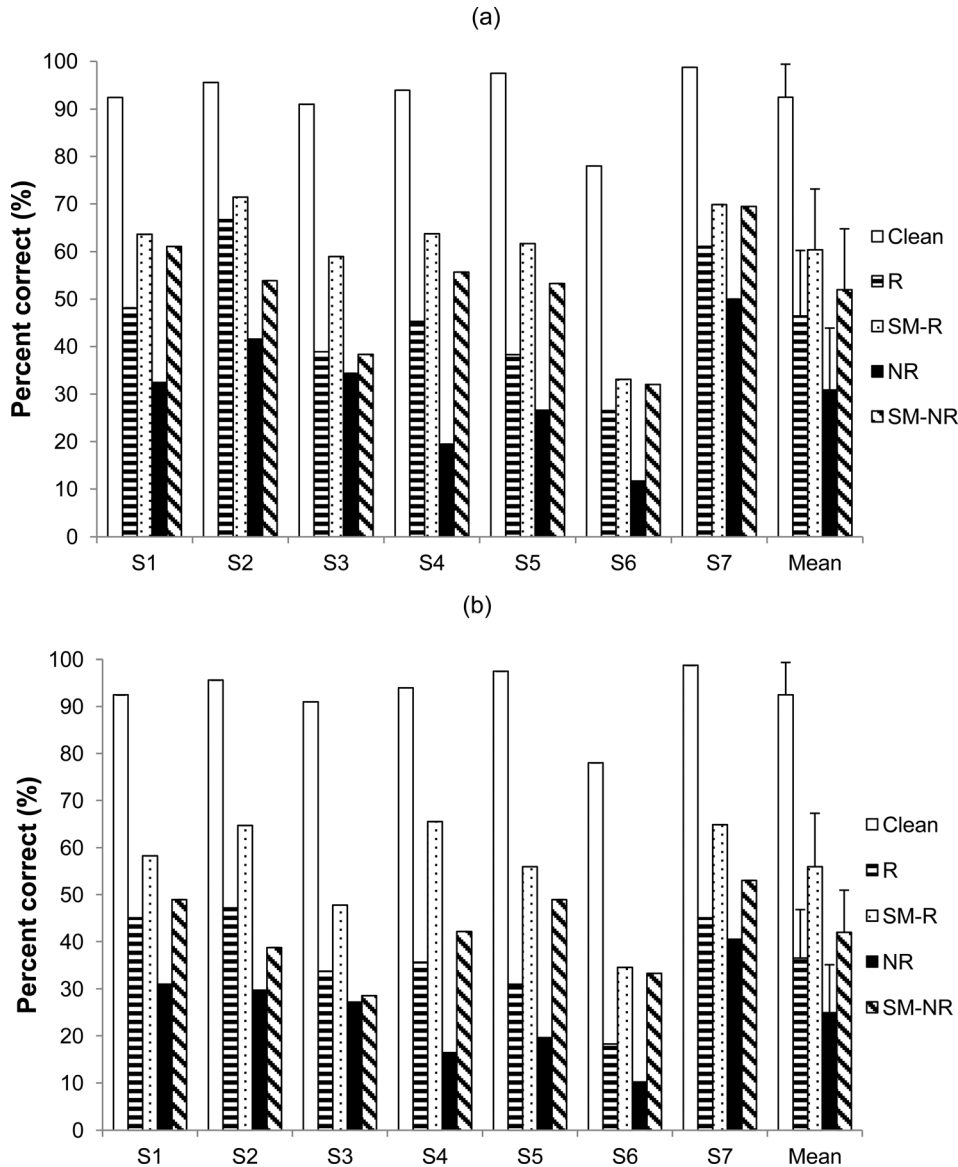


FIG. 2. Individual percent correct scores of seven CI users tested on IEEE sentences using unprocessed and soft-mask processed reverberant and noisy reverberant acoustic inputs (RSNR = 15 dB), (a) $T_{60} = 0.6$ s and (b) $T_{60} = 0.8$ s. “Clean,” “R,” “NR,” and “SM” stand for anechoic clean, reverberant, noisy reverberant, and soft-mask processed conditions, respectively. Error bars indicate standard deviations.

In order to evaluate the impact of individual components (i.e., noise and late reverberation suppression components) of the proposed interference suppression technique, the results obtained through speech intelligibility listening experiments with each stage are presented in Fig. 3. The noisy reverberant speech intelligibility scores improved from an average of 30.89 and 24.96 % to 36.87 and 29.17 % in $T_{60}=0.6$ and 0.8, respectively (RSNR = 15 dB) using only the noise PSD in the ratio mask estimation, and from an average of 30.89 and 24.96 % to 41.67 and 30.62 % in $T_{60}=0.6$ and 0.8, respectively (RSNR = 15 dB), when only reverberation PSD was considered in the mask estimation process.

Post hoc comparisons were performed to assess differences in scores obtained with unprocessed reverberant (noisy reverberant) and ratio-mask processed reverberant (noisy reverberant) signals. The results indicated that in both reverberant-alone and noisy 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 corrected) compared to the unprocessed reverberant and noisy reverberant stimuli.

IV. DISCUSSION

As shown in Fig. 2, the intelligibility scores obtained from the ratio-mask processed corrupted speech surpass those of unprocessed corrupted (by reverberation and noise)

speech stimuli in both reverberant conditions tested. This is because of the assignment of smaller weights to the T-F units in which reverberation and masking noise are more dominant and larger weights to the speech-dominant T-F units. The effect of this ratio-masking on suppressing reverberation and noise from the corrupted speech spectrum is evident in Fig. 4. In this figure, the spectrograms of an IEEE sentence are shown for anechoic quiet, noisy reverberant, and ratio-mask processed noisy reverberant signals.

All gaps and phoneme boundaries are present in the anechoic spectrogram shown in Fig. 4(a). Self- and overlap-masking effects of reverberation as well as the additive noise effects are evident in Fig. 4(b) ($T_{60}=0.6$ s and RSNR = 15 dB). The gaps are filled with noise and reverberation energy and the phoneme onsets/offsets are obscured resulting in masked word recognition cues useful for CI users. The spectrogram of the same noisy reverberant sentence processed with the ratio-masking strategy is shown in panel (c) of Fig. 4. As shown in this figure, the noise and reverberation effects are suppressed to a great extent and the vowel-consonant boundaries are restored. These boundaries will provide CI users with useful acoustic cues for speech identification.

Figure 3 demonstrates the efficacy of the individual components of the proposed algorithm in speech intelligibility gains. As evident from this figure (Fig. 3), although either the noise PSD or reverberation PSD estimation stage has a

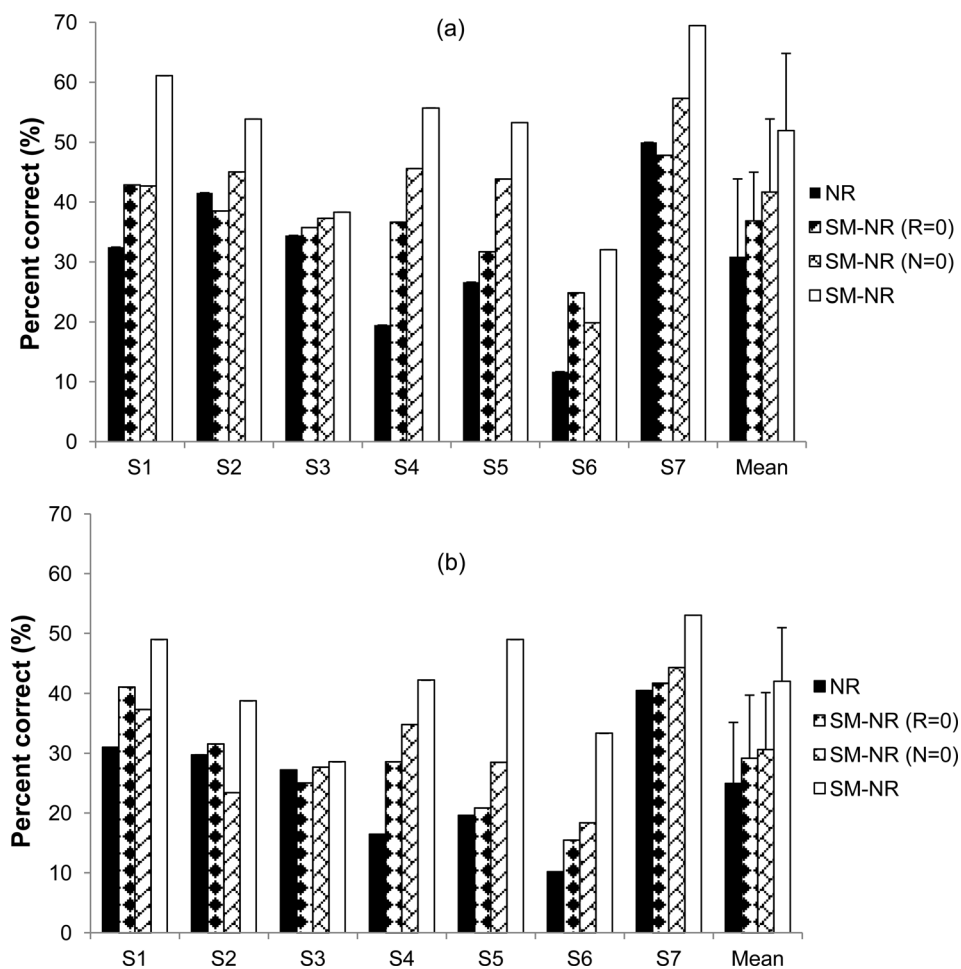


FIG. 3. Individual percent correct scores of seven CI users tested on IEEE sentences using unprocessed and soft-mask processed reverberant and noisy reverberant acoustic inputs (RSNR = 15 dB), (a) $T_{60}=0.6$ s and (b) $T_{60}=0.8$ s. “NR,” “SM-NR (R=0),” “SM-NR (N=0),” and “SM-NR” stand for noisy reverberant, soft-mask processed noisy reverberant speech without late reverberation PSD estimation, soft-mask processed noisy reverberant speech without additive noise PSD estimation, and soft-mask processed noisy reverberant speech, respectively. Error bars indicate standard deviations.

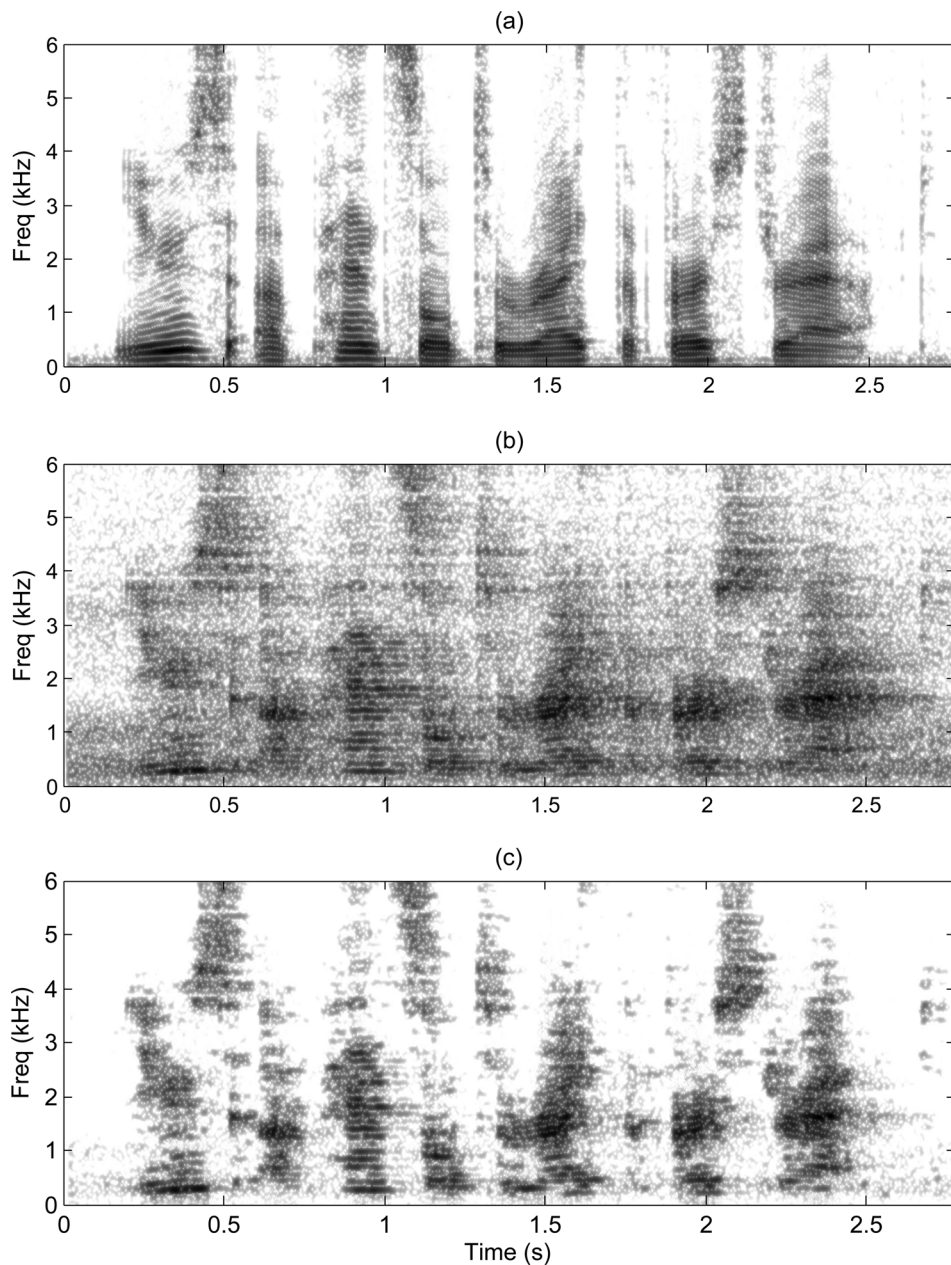


FIG. 4. Spectrograms of the IEEE sentence “use a pencil to write the first draft” uttered by a male speaker. (a) Spectrogram of the anechoic (uncorrupted) sentence, (b) spectrogram of the same sentence corrupted by noise and reverberation (RSNR = 15 dB and $T_{60} = 0.6$ s), and (c) spectrogram of the same noisy reverberant sentence processed by the soft-masking strategy.

beneficial contribution to enhancing the overall intelligibility of noisy reverberant speech for CI users, their combination (where both noise and reverberation PSDs are used in mask estimation) has a significantly greater impact on intelligibility improvements. This is expected as the two components complement each other in a synergistic manner to suppress the combined interference, thereby boosting the speech intelligibility for CI subjects.

It is worth mentioning that the proposed strategy is not specifically developed for CI users and in the context of CI sound processing, thus it may be adopted in other applications such as hearing aids. Moreover, it can suppress not only simultaneous noise and reverberation, but also noise or reverberation alone which makes it more practical in different acoustic environments. Although not evaluated here, non-stationary noise estimation algorithms can also be adopted in situations where the additive noise is non-stationary (Abramson and Cohen, 2007). The noise PSD can be

estimated from the non-speech gaps (e.g., gaps between utterances of speech or speech pauses) with the help of a voice activity detector (Sadjadi and Hansen, 2013).

V. CONCLUSIONS

The combined effects of noise and reverberation are more detrimental to speech intelligibility compared to their individual effects. In this study, a blind single-channel ratio-masking strategy was evaluated for simultaneous suppression of noise and reverberation. This strategy assigns small weights to the masker (noise and reverberation) dominant T-F units while the speech dominant T-F units are given larger weights based on the late reverberation and noise power spectrum estimates. The performance of the proposed method was assessed through listening tests conducted on seven CI listeners. The proposed technique was found to be quite successful in removing noise and reverberation energy

from the gaps of speech yielding significant speech-intelligibility improvements for CI users.

ACKNOWLEDGMENTS

This work was supported by Cochlear Limited and 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 Dr. Nirmal Srinivasan for his help and the CI users for their time.

¹For the noisy reverberant stimuli, the reverberant signal served as the target signal in the SNR computation. Therefore, we refer to these SNRs as RSNRs.

²Reverberation time or T_{60} is the time required for a steady state sound to drop 60 dB below its initial level when the sound source is switched off.

³This is different from the scenario where both speech and noise sources are located far from the receiver (listener) in a reverberant environment (where noise also becomes reverberant). Although both models are valid in real-life situations, the way we have added the noise to the reverberant speech is common practice in engineering literature (Naylor *et al.*, 2010).

Abramson, A., and Cohen, I. (2007). "Simultaneous detection and estimation approach for speech enhancement," *IEEE Trans. Acoust. Speech Signal Process.* **15**, 2348–2359.

Ali, H., Lobo, A. P., and Loizou, P. C. (2013). "Design and evaluation of a personal digital assistant based research platform for cochlear implants," *IEEE Trans. Biomed. Eng.* (in press).

ANSI (2002). S12.60, *Acoustical Performance Criteria, Design Requirements and Guidelines for Schools* (American National Standards Institute, New York).

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.

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

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

Hazrati, O. (2012). "Development of dereverberation algorithms for improved speech intelligibility by cochlear implant users," doctoral dissertation, University of Texas at Dallas.

Hazrati, O., Lee, J., and Loizou, P. C. (2013). "Blind binary masking for reverberation suppression in cochlear implants," *J. Acoust. Soc. Am.* **133**, 1607–1614.

Hazrati, O., and Loizou, P. C. (2012a). "The combined effect of reverberation and noise on speech intelligibility by cochlear implant listeners," *Int. J. Audiol.* **51**, 437–443.

Hazrati, O., and Loizou, P. C. (2012b). "Tackling the combined effects of reverberation and masking noise using ideal channel selection," *J. Speech Lang. Hear. Res.* **55**, 500–510.

Hazrati, O., and Loizou, P. C. (2013). "Reverberation suppression in cochlear implants using a blind channel-selection strategy," *J. Acoust. Soc. Am.* **133**, 4188–4196.

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.

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

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.

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.

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., 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 Mason, D. (1981). "Effect of noise and reverberation on binaural and monaural word identification by subjects with various audiograms," *J. Speech Hear. Res.* **24**, 375–383.

Naylor, P. A., Habets, E. A. P., Wen, J. Y.-C., and Gaubitch, N. D. (2010). "Models, measurement and evaluation," in *Speech Dereverberation*, edited by P. A. Naylor and N. D. Gaubitch (Springer, London), pp. 21–56.

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.

Roman, N., and Woodruff, J. (2011). "Intelligibility of reverberant noisy speech with ideal binary masking," *J. Acoust. Soc. Am.* **130**, 2153–2161.

Sadjadi, S. O., and Hansen, J. H. L. (2012). "Blind reverberation mitigation for robust speaker identification," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Kyoto, Japan, pp. 4225–4228.

Sadjadi, S. O., and Hansen, J. H. L. (2013). "Unsupervised speech activity detection using voicing measures and perceptual spectral flux," *IEEE Signal Process. Lett.* **20**, 197–200.

Wang, D. L. (2005). "On ideal binary mask as the computational goal of auditory scene analysis" in *Speech Separation by Humans and Machines*, edited by P. Divenyi (Kluwer Academic, Norwell, MA), pp. 181–197.

Wu, M., and Wang, D. (2006). "A two-stage algorithm for one microphone reverberant speech enhancement," *IEEE Trans. Audio Speech Lang. Process.* **14**, 774–784.