The impact of reverberant self-masking and overlap-masking effects on speech intelligibility by cochlear implant listeners (L)

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

(Received 27 May 2010; revised 29 June 2011; accepted 29 June 2011)

The purpose of this study is to determine the relative impact of reverberant self-masking and overlap-masking effects on speech intelligibility by cochlear implant listeners. Sentences were presented in two conditions wherein reverberant consonant segments were replaced with clean consonants, and in another condition wherein reverberant vowel segments were replaced with clean vowels. The underlying assumption is that self-masking effects would dominate in the first condition, whereas overlap-masking effects would dominate in the second condition. Results indicated that the degradation of speech intelligibility in reverberant conditions is caused primarily by self-masking effects that give rise to flattened formant transitions. © *2011 Acoustical Society of America*. [DOI: 10.1121/1.3614539]

PACS number(s): 43.71.Ky, 43.66.Ts [RYL] Pages: 1099–1102

I. INTRODUCTION

Acoustic reverberation is known to have a negative impact on speech intelligibility, particularly for hearing-impaired listeners and the elderly (e.g., see Nabelek, 1993). This degradation in speech intelligibility is attributed to two different but interdependent effects: (1) overlap-masking and (2) self-masking effects. Overlap-masking is caused by the overlap of reverberant energy of a preceding phoneme on the following phoneme. This effect is particularly evident for low-energy consonants preceded by high-energy voiced segments (e.g., vowels). The additive reverberant energy fills in the gaps and silent intervals (e.g., stop closures) associated with vocal tract closures. Overlap-masking also occurs in vowels, but it is unlikely that the effect is significant, as the intensity of preceding consonants is much lower than the intensity of the following vowels. Self-masking is caused by the internal smearing of energy within each phoneme. This effect is particularly evident in reverberant sonorant sounds (e.g., vowels), where the formant transitions become flattened. This, in turn, produces confusion between monopthongs and dipthongs (e.g., see Nabelek and Dagenais, 1986; Nabelek *et al.*, 1989). Self-masking is also evident in consonants in the initial position. However, the self-masking effect is substantially smaller when compared to the overlap-masking of consonants in the final position (Nabelek *et al.*, 1989).

The relative impact of self-masking and overlap-masking effects on speech intelligibility by cochlear implant (CI) listeners is unknown. Therefore, it is of interest to investigate whether much of the degradation in speech intelligibility by CI users is caused primarily by self-masking effects, overlap-masking effects, or by both. Isolating the two effects is not straightforward. Vowel and consonant tests cannot provide a satisfactory answer to this question due to the influence of the preceding vowel and consonant context on the intelligibility of reverberant speech. The amount of overlap-masking, for instance, depends largely on the intensity of the preceding phoneme relative to that of the following phoneme, as well as the difference in their spectra. Individual vowel or consonant tests can only examine a limited number of vowel and consonant contexts, and as such, have a limited scope. Sentence intelligibility tests are more appropriate and more reflective of daily reverberant communication settings, wherein the two aforementioned effects co-exist.

To assess the relative impact of self-masking and overlap-masking effects on speech intelligibility by CI users, we take an approach similar to that used by Kewley-Port *et al.* (2007), who assessed the contribution of information carried by vowels versus consonants on speech intelligibility. In their experiments, which were conducted in anechoic settings with sentence materials, they replaced vowel or consonant segments with (normalized-level) noise. In our study, we also use sentence materials, but replace in one condition reverberant consonant segments with clean consonants (C-CLN), and in another condition reverberant vowel segments with clean vowels (V-CLN). The underlying assumption is that self-masking effects will dominate in the first condition, whereas overlap-masking effects will dominate in the second condition. The specific hypothesis investigated in the present study is that the self-masking effects, which are responsible for the flattened formant transitions, will contribute to a larger degree to intelligibility degradation than the overlap-masking effects. The rationale is that CI users can only receive a limited amount of spectral information via their cochlear implant, which in turn renders the perception of formant transitions extremely challenging.

II. METHODS

A. Subjects and material

Six unilateral cochlear implants listeners participated in this study. All participants were American English speaking adults with postlingual deafness who received no benefit from hearing aids preoperatively. Their ages ranged from 47 to 76 yr. They were all paid to participate in this study. All

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

Downloaded 02 Sep 2011 to 129.110.5.92. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

subjects were fitted with the Nucleus 24 multichannel implant device (CI24M), which is manufactured by Cochlear Corporation. All the participants used their devices routinely and had a minimum of five years experience with them. During their visit, the participants were temporarily fitted with the SPEAR3 wearable research processor.

All the parameters in the SPEAR3 processor (e.g., stimulation rate, number of maxima, frequency allocation table, etc.) were matched to each patient's clinical settings. Before participants were enrolled in this study, institutional review board approval was obtained. In addition, informed consent was obtained from all participants before testing commenced. The speech material consisted of sentences taken from the IEEE database (IEEE, 1969). Each sentence was composed of 7–12 words and was produced by a single talker. The root-mean-square amplitude of all sentences was equalized to the same value (65 dB). All the stimuli were recorded at the sampling rate of 25 kHz.

B. Signal processing

The IEEE sentences were manually segmented into two broad phonetic classes: (1) obstruent sounds, which included the stops, fricatives, and affricates, and (2) the sonorant sounds, which included the vowels, semivowels, and nasals. The phonetic segmentation was carried out in a two-step process (see Li and Loizou, 2008). The two-class phonetic segmentation of all the IEEE sentences was saved in transcription files in the same format as TIMIT (.PHN) files and is available in Loizou (2007). Head-related transfer functions (HRTFs) recorded by Van den Bogaert *et al.* (2009) were used to simulate the 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 m × 4.50 m × 3.10 m (length × width × height) and a total volume of 76.80 m^3. The average reverberation time of the experimental room (average in one-third-octave bands with center frequencies between 125 and 4000 Hz) was equal to $RT_{60} = 1.0$ s.

The stimuli were presented in the following conditions. The first processing condition (REV) was designed to simulate the effects of acoustical reverberation encountered in realistic environments. To generate the reverberant (corrupted) stimuli, the premeasured HRTFs were convolved with the speech files from the IEEE test materials using standardized linear convolution algorithms in MATLAB. In the second processing condition, obstruent consonants in the sentences that were corrupted with reverberation were replaced with the corresponding obstruent consonants with no reverberation present. The remaining segments in the sentence were left unmodified, i.e., the sonorant segments were left reverberant. We refer to this condition as the C-CLN (clean consonants) condition. In the third processing condition, reverberant sonorant segments (vowels, semivowels, and nasals) were replaced with the corresponding nonreverberant sonorant segments. The remaining speech segments were left corrupted with reverberation. This condition is referred to as the V-CLN (clean vowels) condition. The replacement of phonetic information in the stimuli was carried out using MATLAB scripts and the existing .PHN files. Unprocessed IEEE sentences in quiet were also tested as the control condition (CLN). Figure 1 illustrates example snapshots of stimulus output patterns (electrodograms) of an IEEE sentence segment. In all panels shown, the vertical axes represent the electrode position corresponding to a specific frequency, whereas the horizontal axes show time progression. As is evident from the electrodograms in Fig. 1, no notable distortions were present at the electrode (envelope) level in either the consonant-to-vowel transitions [(see example in Fig. 1(b)] or vowel-to-consonant transitions [(see example in Fig. 1(c)]. Anecdotal reports by the CI users tested also confirmed that there were no distracting distortions present.

C. Procedure

All stimuli were presented to the CI listeners through the auxiliary input jack of the SPEAR3 processor inside a double-walled sound attenuated booth. The stimuli were presented monaurally to the implanted ear. Prior to testing and following initial instructions, each user participated in a brief practice session to gain familiarity with the listening task and also get acclimatized to the SPEAR3 processor settings. The practice session consisted of two practice runs. Two IEEE lists were used at each practice run. During the practice session, the subjects were allowed to adjust the volume to reach a comfortable level. Subjects participated in a total of four different conditions. Two IEEE lists (20 sentences) were used per testing condition. None of the lists were repeated across conditions. To minimize any order effects the order of the test conditions was randomized across subjects. Subjects were given a 15 min break every 90 min during the testing session. During testing, the participants were instructed to type as many of the words as they could identify via a computer keyboard. No feedback was given. 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. 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.

III. RESULTS AND DISCUSSION

Individual and mean percent correct scores are plotted in Fig. 2. Speech intelligibility is reduced substantially when subjects are tested with sentences corrupted by reverberation. In fact, when compared to the CLN condition (unprocessed speech in quiet), the mean speech intelligibility scores for all CI listeners in the REV condition were approximately 60 percentage points lower (see Fig. 2). An analysis of variance (ANOVA) (with repeated measures) was run to assess the effects of the corrupted phonetic segment (REV, C-CLN, V-CLN) on speech intelligibility. A significant effect ($F_{(2,10)} = 60.7$, $p < 0.005$) was found on intelligibility when reverberation corrupted either the vowel segments alone (C-CLN), consonant segments alone (V-CLN) or both (REV). *Post hoc* comparisons (according to Schèffe) revealed significantly higher ($p < 0.05$) scores in both the

Downloaded 02 Sep 2011 to 129.110.5.92. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

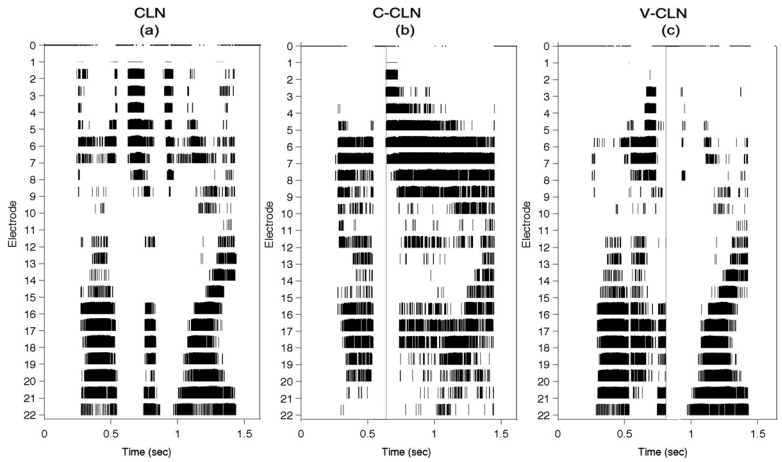

FIG. 1. Electrodograms of the sentence excerpt "Dots of light betrayed." (a) Uncorrupted sentence (CLN), (b) same sentence with clean consonant segments (C-CLN) but reverberant vowel segments (vertical line at 0.65 s depicts a consonant-to-vowel transition), and (c) same sentence with clean vowel segments (V-CLN) but reverberant consonant segments (vertical line at 0.8 s depicts a vowel-to-consonant transition). In each panel, time is shown along the abscissa and the electrode number is shown along the ordinate.

C-CLN and V-CLN conditions when compared against the reverberant condition (REV).

Figure 3 provides useful insights with regards to the dominant self-masking and overlap-masking effects. As illustrated in Fig. 3(b), overlap-masking effects are mostly evident in consonant segments (e.g., see stop consonant /t/ at 0.9–1.2 s), whereas self-masking effects are mostly evident in vowel segments (e.g., see dipthong /aI/ at 0.7–0.9 s). Self-masking effects are dominant in sonorant sounds, such as vowels, semivowels, and nasals, whereas overlap-masking mostly dominates obstruent sounds, such as fricatives, affricates, and stops. As mentioned earlier, self-masking is also present in consonants (in the initial position) and overlap-masking is present in vowels, however their effect is small. Both overlap-masking and self-masking effects pose collectively a huge challenge for CI users who receive their auditory cues mainly from temporal envelope modulations in a limited number of spectral channels (e.g., see Munson and Nelson, 2005).

To assess the relative impact of the overlap-masking and self-masking effects on speech intelligibility, we turn to Fig. 2. First, the mean intelligibility scores obtained with the C-CLN condition were around 12 percentage points higher than the reverberant condition. That is, introducing clean consonants, while preserving reverberation in vowels, provided a small but significant benefit to speech intelligibility. Second, speech intelligibility improved substantially in the V-CLN condition, where reverberant sonorant segments (e.g., vowels) were replaced with the corresponding clean vowel segments. The benefit observed in the V-CLN condition was found to be significant and was equal to around 40 percentage points, when compared to the reverberant condition. In fact, as can be seen in Fig. 2, a 2:1 benefit was observed for sentences that preserved sonorant segments and removed dominant self-masking effects when compared against sentences that preserved only obstruent speech segments. This benefit was similar to that observed by Kewley-Port *et al.* (2007) who assessed the contribution of consonant versus vowel information to sentence intelligibility by normal-hearing and elderly hearing-impaired listeners The larger benefit in speech intelligibility observed in the V-CLN condition (relative to the C-CLN condition) leads us to conclude that reverberant self-masking effects (mostly associated with sonorant sounds) are more detrimental to speech intelligibility than overlap-masking effects. The large benefit observed in the V-CLN condition can be attributed to the fact that the flat $F1$ and $F2$ formant transitions present in the reverberant stimuli were replaced with the natural $F1$ and $F2$ transitions (see Fig. 3). We cannot discard the possibility that better transmission of voicing and duration cues, available in the V-CLN stimuli, also contributed to the improvement in intelligibility. The $F1$ and $F2$ formant transitions in the sonorant speech segments carry not only information about the identity of the vowels and semivowels, but also contain useful co-articulatory information regarding the

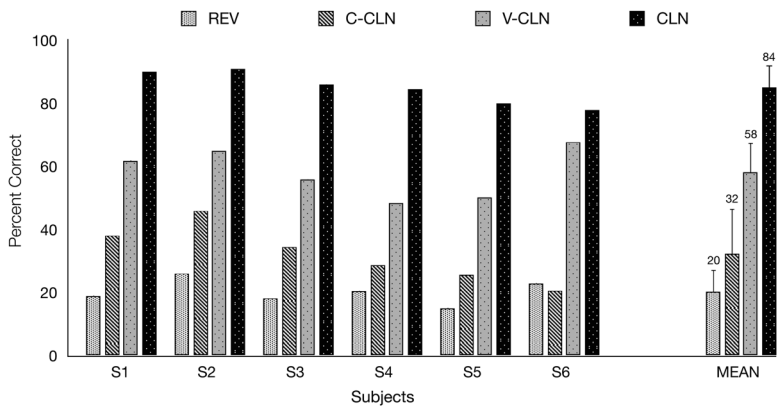

FIG. 2. Percent correct scores of six CI listeners tested with IEEE sentences presented in reverberation. Percent correct scores in the CLN condition (anechoic) are also shown for comparative purposes. Error bars indicate standard deviations.

Downloaded 02 Sep 2011 to 129.110.5.92. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp

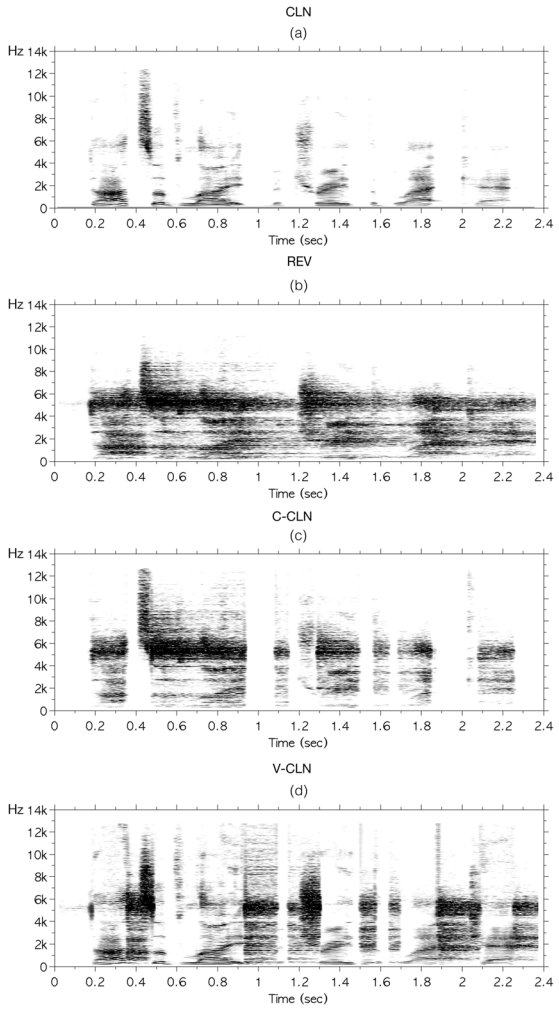

FIG. 3. Spectrograms of the IEEE sentence "Dots of light betrayed the black cat" uttered by a male speaker. (a) Unmodified (uncorrupted) sentence (CLN), (b) sentence corrupted by reverberation (REV), (c) sentence with clean obstruent consonants only (C-CLN), and (d) sentence with clean vowels only (V-CLN).

surrounding consonants (Kewley-Port *et al.*, 2007). The surrounding consonants contain other cues (e.g., voicing, duration) that might be detectable in the V-CLN sentences, however, overlap-masking introduces a large number of manner and place of articulation errors (Nabelek *et al.*, 1989). Despite the fact that the information preserved in the surrounding consonants is not reliable, due to overlap-masking effects, this co-articulatory information seems to be sufficient in as far as allowing the CI users to better understand the intended words uttered. Hence, we can conclude that the degradation observed in speech intelligibility by CI listeners in reverberation is caused primarily by the loss of useful cues (e.g., $F1$ and $F2$ formant movements), which are normally present in the sonorant vowel segments rather than information contained in the obstruent consonant segments.

IV. CONCLUSIONS

The present study assessed the relative impact of self-masking and overlap-masking effects on speech intelligibility by CI listeners in reverberant environments. Based on the intelligibility scores obtained, the following conclusions can be drawn. (1) Reverberation adversely affects sentence recognition by CI users. A reverberation time of 1.0 s resulted in almost a 60 percentage point drop in mean intelligibility scores when compared to speech intelligibility attained in quiet. The perceptual effect of reverberation differed across the vowel and consonant segments of the utterance. The formant transitions were flattened during the vowel segments, whereas reverberant energy leaked from preceding phonemes and filled succeeding gaps present in low-energy obstruent consonants. (2) Significant gains in speech intelligibility in reverberation were observed when introducing clean sonorant segments (e.g., vowels), while still preserving reverberation in obstruent segments (V-CLN condition). In contrast, smaller gains were attained in the C-CLN (clean consonants) condition. This leads us to conclude that the degradation of speech intelligibility by CI users observed in reverberant conditions is caused primarily by self-masking effects, which result in loss of information (e.g., $F1$ and $F2$ formant movements) contained in vowel speech segments.

This study contributes to our understanding of the relative impact of acoustic reverberation on sentence recognition by CI users. Such knowledge is important for the development of future signal processing strategies that aim to enhance the intelligibility of speech in reverberant listening settings. The outcomes of the present study suggest that efforts need to be devoted toward developing speech coding strategies capable of suppressing reverberant energy from the sonorant segments of speech.

ACKNOWLEDGMENTS

This work was supported by Grant Nos. R03 DC 008882 (K.K.) and R01 DC 010494 (P.C.L.) awarded from the National Institute of Deafness and other Communication Disorders (NIDCD) of the National Institutes of Health (NIH).

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

Kewley-Port, D., Burkle, Z., and Lee, J. (**2007**). "Contribution of consonant versus vowel information to sentence intelligibility for young normal-hearing and elderly hearing-impaired listeners," J. Acoust. Soc. Am. **122**, 2365–2375.

Li, N., and Loizou, P. C. (**2008**). "The contribution of obstruent consonants and acoustic landmarks to speech recognition in noise" J. Acoust. Soc. Am. **124**, 3947–3958.

Loizou, P. C. (**2007**). *Speech Enhancement: Theory and Practice* (CRC Press, Boca Raton, FL), Appendix C, pp. 589–599.

Munson, B., and Nelson, P. (**2005**). "Phonetic identification in quiet and in noise by listeners with cochlear implants," J. Acoust. Soc. Am. **118**, 2607–2617.

Nabelek, A. K., and Dagenais, P. A. (**1986**). "Vowel errors in noise and in reverberation by hearing-impaired listeners," J. Acoust. Soc. Am. **80**, 741–748.

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. (**1993**). "Communication in noisy and reverberant environments," in *Acoustical Factors Affecting Hearing Aid Performance*, edited by G. A. Studebaker and I. Hochberg (Allyn & Bacon, Boston), pp. 15–30.

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. **124**, 360–371.

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.

Downloaded 02 Sep 2011 to 129.110.5.92. Redistribution subject to ASA license or copyright; see http://asadl.org/journals/doc/ASALIB-home/info/terms.jsp