

Effects of early and late reflections on intelligibility of reverberated speech by cochlear implant listeners

Yi Hu

*Department of Electrical Engineering and Computer Science,
University of Wisconsin–Milwaukee, Milwaukee, Wisconsin 53211
huy@uwm.edu*

Kostas Kokkinakis^{a)}

*Department of Speech-Language-Hearing: Sciences and Disorders,
University of Kansas, Lawrence, Kansas 66045
kokkinak@ku.edu*

Abstract: The purpose of this study was to determine the overall impact of early and late reflections on the intelligibility of reverberated speech by cochlear implant listeners. Two specific reverberation times were assessed. For each reverberation time, sentences were presented in three different conditions wherein the target signal was filtered through the early, late or entire part of the acoustic impulse response. Results obtained with seven cochlear implant listeners indicated that while early reflections neither enhanced nor reduced overall speech perception performance, late reflections severely reduced speech intelligibility in both reverberant conditions tested.

© 2013 Acoustical Society of America

PACS numbers: 43.71.Gv, 43.71.Ky [AC]

Date Received: September 23, 2013 Date Accepted: November 8, 2013

1. Introduction

Reverberation is defined as the sum total of all sound reflections arriving at a certain point inside an acoustical enclosure after the enclosure has been excited by an impulsive sound signal (Kuttruff, 2000). When listening to speech inside such an enclosure, the sound reaching the listener's ears consists of the direct sound (DS), early reflections (ERs), and late reflections (LRs). These components are shown in Fig. 1. The direct sound travels from the sound source to the listener along a straight line and is not influenced at all by the walls or the ceiling of the room. Early reflections arrive at the listener within approximately the first 50–100 ms after the direct sound and are considered by many, capable of boosting overall speech intelligibility as they can be integrated with the direct sound (Bradley *et al.*, 2003). Late reflections, on the other hand, consist of a dense succession of echoes with diminishing intensity and typically arrive at the listener with a much longer delay after the arrival of the direct component. Late reflections cannot be integrated with the direct sound and are perceived either as separate echoes or as reverberation (Boothroyd, 2004).

It is fairly well understood that acoustic reverberation can 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 (Assmann and Summerfield, 2004). The impact of reverberation on speech perception is largely unnoticed by normal-hearing (NH) listeners, especially in small

^{a)} Author to whom correspondence should be addressed.

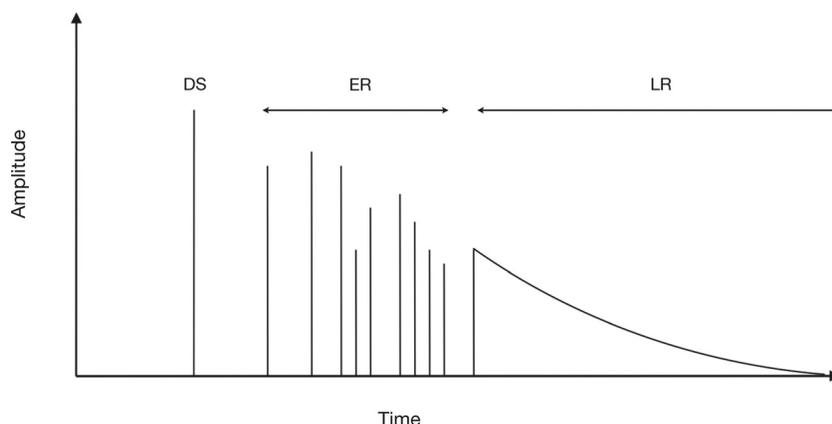


Fig. 1. Schematic representation of an acoustic impulse response showing the DS, the ER part, and the LR components.

and moderately reverberant rooms. However, reverberation negatively affects the perception of consonant and vowel phonemes and has a detrimental overall effect on speech intelligibility by cochlear implant (CI) listeners (Kokkinakis *et al.*, 2011; Hazrati and Loizou, 2012). Early reflections have predominantly been associated with moderate self-masking (coloration) effects due to the internal smearing of energy within each phoneme. Late reflections have been linked to a severe degradation of the low-frequency envelope that is essential to speech intelligibility. LRs produce significant overlap-masking effects due to the overlap of reverberant energy of a preceding phoneme on the following phoneme (Kokkinakis and Loizou, 2011). Such effects are evident in low-energy consonants preceded by high-energy voiced segments (e.g., vowels).

Tackling speech degradation due to acoustic reverberation has recently become an area of intense research activity and this has given rise to several reverberation suppression (dereverberation) strategies which have potential for use in commercially available cochlear implant devices (e.g., see Kokkinakis and Loizou, 2009; Kokkinakis *et al.*, 2011; Hazrati and Loizou, 2013; Roman and Woodruff, 2013). Traditionally, a common strategy to suppress reverberation is to process the reverberant (corrupted) signal with a filter that inverts the reverberation process and recovers the original (anechoic) signal. This process of estimating or inverting the unknown acoustic impulse response between the source and the receiver (ear or microphone) may also be facilitated by resorting to more than one microphones (e.g., see Kokkinakis and Loizou, 2009). In a previous study (Kokkinakis *et al.*, 2011), we used an ideal reverberant (binary) mask (IBM) based on the signal-to-reverberant ratio (SRR) of individual frequency channels. The IBM strategy was shown to successfully suppress additive reverberant energy by retaining only channels with SRR values larger than a fixed threshold while eliminating all other channels. The proposed strategy was found to yield consistent gains in intelligibility (30–55 percentage points) when tested with cochlear implant listeners. A variant of this approach, which operates by assuming no prior knowledge of the anechoic signal, also produced a substantial overall benefit when tested with sentences corrupted with reverberation (Hazrati and Loizou, 2013).

In the Roman and Woodruff (2011) study, the relative contribution of early and late reflections to speech intelligibility was evaluated in twelve young normal-hearing listeners by using three alternative binary masking approaches: the direct sound IBM (IBM-DS), the direct sound and early reflections IBM (IBM-ER), and the reverberant IBM (IBM-R). The direct component was obtained by filtering the target signal through the anechoic impulse response. The direct sound plus early reflections conditions were obtained by filtering the target signal through the first 50 ms of the room impulse response. In each binary masking approach, the selection criterion was

formulated using either the direct, early or reverberant part of the signal (e.g., see [Roman and Woodruff, 2011](#)). Speech reception thresholds were calculated for the aforementioned conditions using HINT sentences corrupted with reverberation equal to 0.4 and 0.8 s. A significant improvement of 8.27 dB was obtained at 0.4 s of reverberation time and a benefit of 11.76 dB was noted at 0.8 s of reverberation when processing the corrupted stimuli with the IBM-ER strategy. No significant improvements were found when resorting to either the IBM-DS or the IBM-R dereverberation strategies. More recently, the same IBM strategy was modified to be a function of a reflection boundary, namely, a pre-determined division point between early and late reflections ([Roman and Woodruff, 2013](#)). Results with normal-hearing listeners in reverberation, suggested that in order to achieve significant intelligibility improvements, either the direct sound or the early reflections should be preserved by the binary mask. These finding suggests that, at least for NH listeners, early reflections are integrated with the direct sound and could be conducive to intelligibility.

The current study builds upon the recent work of [Roman and Woodruff \(2013\)](#) by investigating the impact of early and late reflections on the intelligibility of reverberated speech by cochlear implant listeners. Based on these prior findings, early reflections occurring within the first 50 ms may be beneficial to speech perception by NH listeners. However, the benefit of early reflections in the perception of reverberated speech by CI listeners remains largely unclear. There is also prior evidence to suggest that the overlap-masking artifacts introduced by late reflections (>50 ms) are the main cause for the reduced speech understanding observed by CI listeners in reverberation. Yet, the mechanisms under which the presence of late reflections mask or diminish the benefit due to early reflections are not well understood.

2. Methods

2.1 Participants

Seven post-lingually deafened adult cochlear implant listeners (4 female, 3 male) with a fully inserted (i.e., long electrode array) cochlear implant on one side took part in this study. All seven subjects were native speakers of American English, and had acquired at least 12 months of experience with their device post-implantation prior to testing. More detailed demographic information is provided in [Table 1](#). All subjects were paid an hourly wage. This study was approved by the Human Subjects Committee of the University of Kansas in Lawrence (HSCL). All subjects gave written informed consent prior to the beginning of testing and a case history interview was conducted with each subject to determine eligibility in this study. Only subjects who scored greater than 70% on the consonant-nucleus-consonant test were included.

2.2 Stimuli

Stimuli consisted of sentences from the Institute of Electrical and Electronics Engineers (IEEE) database ([IEEE, 1969](#)). Each sentence is composed of approximately 7–12 words. In total, there are 72 lists of 10 sentences each produced by a single talker. IEEE sentences are specifically designed to have very few contextual cues that could aid in speech understanding. The root-mean-square amplitude of all sentences was equalized to the same value (65 dBA). All stimuli were recorded at the sampling frequency of 16 kHz. MATLAB was used to generate sentence stimuli corrupted with various degrees of reverberation. Acoustic impulse responses (AIRs) recorded by [Rychtarikova et al. \(2009\)](#) were used to simulate two reverberant conditions ($RT60 = 0.3$ and 1.0 s). To obtain measurements of AIRs, the authors used a CORTEX MKII manikin artificial head placed inside a rectangular reverberant room with dimensions $5.50 \text{ m} \times 4.50 \text{ m} \times 3.10 \text{ m}$ (length \times width \times height) and a total volume of 76.80 m^3 . The average reverberation time of the room (averaged in one-third-octave bands with center frequencies between 125 and 4000 Hz) before any modification was equal to $RT60 = 1.0$ s. By increasing the number of acoustic panels and by adding floor carpeting to the room, as well as highly absorbent

Table 1. Participant demographic information.

Subject	Gender	Age at testing	Months of experience with stimulation	Etiology of hearing loss	Cochlear implant type	Consonant-nucleus-consonant word-recognition score in quiet (% correct)
S1	M	54	38	Unknown	Nucleus 5	89%
S2	F	22	60	Noise exposure	Nucleus 5	95%
S3	F	57	27	Unknown	Nucleus 5	86%
S4	F	58	21	Unknown	Nucleus 5	80%
S5	M	56	72	Genetic	AB Harmony	75%
S6	F	35	24	Meniere's	Nucleus Freedom	72%
S7	M	48	20	Noise exposure	Nucleus 5	88%
MEAN		47.1	37.4			84%
SD		13.7	20.7			8%

rectangular acoustic boards, the average reverberation time was reduced to $RT60 = 0.3$ s. This latter value corresponds to a standard well-dampened room and is typical of the reverberation time often encountered in small office spaces.

Prior to generating the stimuli used in this study, the impulse response $h(k)$ for each reverberant condition $RT60 = 0.3$ s and $RT60 = 1.0$ s, was partitioned (divided) into early and late components,

$$h(k) = \begin{cases} h_{\text{early}}(k) & \text{if } 1 \leq k < Tf_s, \\ h_{\text{late}}(k) & \text{if } Tf_s \leq k \leq Lf_s, \end{cases} \quad (1)$$

where k denotes time-index, L represents the total length of the impulse response, f_s is the sampling rate, and T defines the time span after which the late reverberation components begin. This term, dubbed reflection boundary as per the Roman and Woodruff (2013) study, essentially represents a pre-determined division point between early and late reflections. This boundary serves as a reasonable upper bound on the time after which reflections are no longer beneficial for speech perception. The reflection boundary T was set to 50 ms, which corresponds to the first 800 samples for the 16 kHz sampling frequency used. To generate the reverberant stimuli, the AIRs obtained for each reverberation condition were convolved with the speech files from the IEEE test materials using standardized linear convolution algorithms in MATLAB. For each of the two $RT60$ conditions, we generated reverberant sentences, which were corrupted with: (1) only early reflections (ER condition) described in Eq. (2), (2) only late reflections (LR condition) described in Eq. (3), and (3) both the early and late reflection components (REV condition) shown in Eq. (4),

$$x_{\text{early}}(k) = \sum_{k=1}^{Tf_s} h_{\text{early}}(k)s(n-k), \quad (2)$$

$$x_{\text{late}}(k) = \sum_{k=Tf_s}^{Lf_s} h_{\text{late}}(k)s(n-k), \quad (3)$$

$$x_{\text{full}}(k) = x_{\text{early}}(k) + x_{\text{late}}(k) = \sum_{k=1}^{Lf_s} h(k)s(n-k). \quad (4)$$

2.3 Procedure

During testing, the participants were seated inside a double-walled IAC sound-attenuating booth. The sentence stimuli for each condition were delivered to the

implanted ear via direct audio input. The microphones in the speech processor remained deactivated. All stimuli were presented at the subject's most comfortable level. The subjects participated in a total of six listening conditions, each corresponding to a different combination of RT60 and reflection boundary condition. Two IEEE lists (20 sentences) were used per condition. Anechoic (unprocessed) IEEE sentences were also used as the control condition (RT60=0 s). Each participant completed all the conditions in a single session. Participants were given a 15 min break every 60 min during the testing session. Each testing session lasted approximately four hours including breaks.

Following the initial instructions, each listener participated in a brief practice session to gain familiarity with the listening task. 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 participants were encouraged to guess if unsure. 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 were calculated by dividing the number of words correctly identified by the total number of words in the particular sentence list.

3. Results

The mean percent correct scores obtained from all participants for each reflection condition tested, are plotted in Fig. 2. For comparison, the average scores obtained with anechoic (unprocessed) stimuli are also shown. All percent correct scores were rationalized arcsine unit-transformed prior to the statistical analyses. A Kolmogorov-Smirnov test was run to confirm the normality of the transformed percent correct scores. Note that a critical value equal to 0.05 was used as the significance level on the statistical analyses performed.

For RT60=0.3 s, one-way analysis of variance (ANOVA) (with repeated measures) indicated a significant effect of the reflection condition ($F[3, 18]=22.43$, $p=0.03$) on speech intelligibility. *Post hoc* comparisons using Tukey's honest significant difference tests were run to assess significant differences in scores obtained

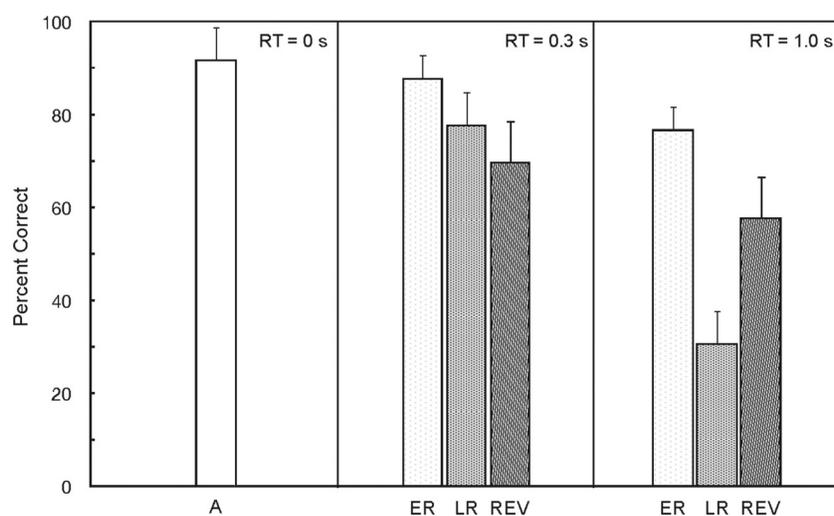


Fig. 2. Mean percent correct scores obtained from seven cochlear implant listeners tested on IEEE sentences corrupted with varying degrees of reflections. For each of the two RT60 conditions, the different reflection conditions are shown along the abscissa. Error bars indicate standard errors of the mean.

between different conditions. Results indicated that performance in the reverberant condition deteriorated significantly ($p=0.04$) relative to the anechoic scores. On average, speech intelligibility of reverberant stimuli was around 20% lower than the performance in the anechoic condition. The scores in the ER and LR conditions were not significantly different from the scores obtained in the anechoic condition.

For $RT60=1.0$ s, the ANOVA indicated a significant effect of the reflection condition ($F[3, 18]=64.55$, $p<0.05$) on speech intelligibility. *Post hoc* comparisons using Tukey's honest significant difference tests indicated significant differences between anechoic scores and the scores in the LR ($p=0.001$) and REV conditions ($p=0.042$). A marginally non-significant difference was noted between the anechoic and ER conditions ($p=0.056$). The scores in the ER condition were significantly higher than those obtained in the LR condition ($p=0.004$) but were not different from the REV condition scores ($p=0.15$). The LR condition scores were significantly lower than the scores obtained in the REV condition ($p=0.02$). Average performance with LR stimuli was 30% lower than the performance noted in the reverberant condition.

4. Summary and discussion

The performance of the CI listeners in reverberation was greatly affected even with a moderate amount of added reverberant energy ($RT60=0.3$ s). The average speech intelligibility scores for all listeners dropped from 90% in the anechoic condition to around 70% for $RT60=0.3$ s in the reverberant condition. In fact, the findings of this study are consistent with the results obtained from an earlier study by Kokkinakis *et al.* (2011), which concluded that there is a very strong, and negative, relationship between speech perception and the amount of additive acoustical reverberation. In line with the Kokkinakis *et al.* (2011) study, speech intelligibility by cochlear implant users in reverberation degraded exponentially as the amount of reverberation increased. In the $RT60=1.0$ s condition, the intelligibility scores were on average between 35 and 60 percentage points lower (REV and LR conditions) when compared to the anechoic listening condition ($RT60=0$ s). In agreement with the overall pattern noted in the $RT60=0.3$ s condition, the scores obtained in the ER condition seem to suggest that our subjects did not benefit from fusion. In other words, the CI listeners could not integrate the early energy with the direct sound in order to further enhance either the perceived loudness or clarity of speech. In some cases there was even a decrement in overall speech perception performance. For $RT60=1.0$ s, the CI listeners scored on average 15% lower when the sentence stimuli were filtered through early reflections only, although the difference between that condition and the anechoic condition was marginally non-significant. The reduced ability observed by CI listeners to perceptually integrate (fuse) the energy of the early reflections in the perceived signal with the direct source can be attributed to (1) the cochlear implant processing strategy (e.g., see Kokkinakis *et al.*, 2011) and (2) the overall reduced short-term temporal integration often associated with subjects suffering from cochlear hearing loss (Moore, 1996).

Intelligibility was reduced considerably in the $RT60=1.0$ s condition when the reverberated stimuli containing only late reflections were presented to the cochlear implant listeners. On average, in the LR condition the subjects scored approximately 60% lower than the anechoic condition, 40% lower than the ER condition and 30% lower than the reverberant condition (see Fig. 2). Note that in the reverberant condition the anechoic sentences were convolved with both the early and late reflection components as described in Sec. 2.2. All the differences in the mean scores across conditions were found to be statistically significant. This would seem to suggest the following: (1) self-masking effects (causing flat formant transitions) that are often associated with early reflections contribute the least to the overall drop in performance observed in reverberation and (2) the main effect of additive reverberant energy on speech perception is largely the result of late reflections present in the reverberant field. These findings agree with prior work, which has suggested that overlap-masking effects due to temporal envelope smearing are believed to be the most damaging to speech perception (e.g., see Bradley *et al.*, 2003; Kokkinakis *et al.*, 2011; Roman and Woodruff, 2013). Although in low reverberation, the early reflections

of reverberated speech have only a neutral effect on speech perception, this is not the case when reverberation increases. In 1.0 s of reverberation, average performance with stimuli corrupted with late reflections was 30% lower than the performance noted in the reverberant condition. This implies that the subjects had some residual ability to fuse the energy of the early reflections with the direct source component. In this highly reverberant condition, the subjects' perception was improved by using early reflections to increase the instantaneous SRR, which 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.

These findings highlight the negative impact of reverberation and shed new light on the limitations that cochlear implant users face in challenging (e.g., reverberant) environments. One of the first steps towards improving speech perception for CI users should be to reduce the level of late reflections to the greatest extent possible. Since late-arriving components of reverberation are essentially (spectrally) equivalent to masking noise, naturally a first step towards restoring speech intelligibility in reverberation would be to apply spectral subtraction strategies typically used for noise reduction.

Acknowledgments

This research was supported in part by a University of Wisconsin–Milwaukee Research Growth Initiative (RGI) grant awarded to Y.H., and by an NIH Clinical and Translational Science Award, Grant No. 5UL1 TR000001, awarded to the University of Kansas Medical Center (K.K.).

References and links

- 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.
- Boothroyd, A. (2004). "Room acoustics and speech perception," *Semin. Hear.* **25**, 155–166.
- Bradley, J. S., Sato, H., and Picard, M. (2003). "On the importance of early reflections for speech in rooms," *J. Acoust. Soc. Am.* **113**, 3233–3244.
- Hazrati, O., and Loizou, P. C. (2012). "The combined effects of reverberation and noise on speech intelligibility by cochlear implant users," *Int. J. Audiol.* **51**, 437–443.
- 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.
- IEEE (1969). "IEEE recommended practice speech quality measurements," *IEEE Trans. Audio Electroacoust.* **AU17**, 225–246.
- 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.
- 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. (2011). "The impact of reverberant self-masking and overlap-masking effects on speech intelligibility by cochlear implant listeners," *J. Acoust. Soc. Am.* **130**, 1099–1102.
- Kuttruff, H. (2000). *Room Acoustics* (Taylor & Francis, New York), pp. 204–250.
- Moore, B. C. (1996). "Perceptual consequences of cochlear hearing loss and their implications for the design of hearing aids," *Ear Hear.* **17**, 133–161.
- Roman, N., and Woodruff, J. (2011). "Intelligibility of reverberant noisy speech with ideal binary masking," *J. Acoust. Soc. Am.* **130**, 2153–2161.
- Roman, N., and Woodruff, J. (2013). "Speech intelligibility in reverberation with ideal binary masking: Effects of early reflections and signal-to-noise ratio threshold," *J. Acoust. Soc. Am.* **133**, 1707–1717.
- Rychtarikova, M., van den Bogaert, T., Vermeir, G., and Wouters, J. (2009). "Binaural sound source localization in real and virtual rooms," *J. Audio Eng. Soc.* **57**, 205–220.