Speech intelligibility enhancement by early reflections

IRIS ARWEILER, JÖRG M. BUCHHOLZ, AND TORSTEN DAU

Centre for Applied Hearing Research, Department of Electrical Engineering, Technical University of Denmark, DK-2800 Lyngby, Denmark

Early reflections in a room can improve speech intelligibility for normal-hearing listeners, because the auditory system integrates them with the direct sound which results in an increased speech level. The present study investigated the underlying mechanisms involved in early reflection processing. Monaural and binaural speech intelligibility tests were performed with 9 normal-hearing listeners in a loudspeaker-based virtual auditory environment where the amplitude of the direct sound and the early reflections could be varied independently. The reflection pattern was taken from a classroom simulated with the room acoustic software Odeon [Odeon (2008)]. The Danish sentence test Dantale II [Wagener, Int. J. Aud. **42**, 10-17 (2003)] was used. The sentences were presented from the loudspeaker at 0° azimuth and speech intelligibility was measured with a diffuse speech shaped noise (SSN). Different signal-to-noise ratios were obtained by changing either the direct sound level or the early reflection level of the speech signal. Increased early reflection levels improved speech intelligibility but the effect was smaller than for increased direct sound levels. No binaural processing of early reflections other than a summation of the signals at the two ears could be observed for the diffuse SSN interferer.

INTRODUCTION

One aspect of spatial hearing is the utilisation of early reflections for improved speech intelligibility in a room. The benefit from early reflections for speech intelligibility has been demonstrated in several studies (e.g. Nabelek and Robinette, 1978; Soulodre *et al*., 1989; Parizet and Pollack, 1992; Bradley *et al*., 2003). However, it is not clear to what extent early reflections can contribute to improved speech intelligibility in realistic acoustic environments. Furthermore, it is not well understood if the early reflection benefit is a monaural or a binaural effect.

It is well known that reflections only increase speech intelligibility if they arrive within a certain time window after the direct sound. Within this time window, which is usually assumed to be about 50 ms for speech sounds (Thiele, 1953; Cremer and Müller, 1982), the reflections are integrated with the direct sound. Lochner and Burger (1964) measured binaural word intelligibility with a loudspeaker setup for several early reflection delays. The direct sound was presented from one loudspeaker and a delayed copy of the direct sound was presented from the other loudspeaker. The level of the single reflection was varied. When the reflection had the same intensity as the direct sound and arrived within 30 ms after the direct sound, the effective level increase of the combined signal was 3 dB due to energy addition of the two single sounds. When the level of the early reflection was 5 dB less than the direct sound, the increase

Proceedings of ISAAR 2009: Binaural Processing and Spatial Hearing. 2nd International Symposium on Auditory and Audiological Research. August 2009, Elsinore, Denmark. Edited by J. M. Buchholz, T. Dau, J. Cristensen-Dalsgaard, and T. Poulsen. ISBN: 87-990013-2-2. The Danavox Jubilee Foundation, 2010.

in effective level of the combined signal was only about 1.2 dB but the integration window was longer (40 ms). However, Lochner and Burger (1964) considered only one reflection and very low sound pressure levels without background noise for their speech intelligibility measurements, which is not a very realistic scenario for typical listening situations.

Soulodre *et al*. (1989) therefore extended Lochner and Burger's (1964) experiment to investigate the combined effect of early reflections and background noise level on speech intelligibility. The tests were conducted at a signal-to-noise ratio (SNR) of 0 dB with the speech and noise levels set to 55 dB(A). They used 13 early reflections arriving within 40 ms after the direct sound. Each of them had an individual sound pressure level of 50 dB(A). According to Lochner and Burger (1958) they should have increased the speech signal by 7 dB but the actual improvement in speech intelligibility only corresponded to a 3 dB increase in direct sound level. Soulodre *et al*. (1989) concluded that the integration of early reflections depends on the SNR. Hence, in the presence of a background noise, speech intelligibility could not be predicted from a simple addition of the early reflection energy to the direct sound energy.

More recently, Bradley *et al*. (2003) presented new results on the importance of early reflections for speech intelligibility. They found that the energy in seven early reflections arriving within 50 ms after the direct sound was as beneficial to speech intelligibility as the energy in the direct sound. They measured word intelligibility in ambient noise that had a constant level of 47.6 dB(A). All signals were presented with eight loudspeakers arranged around the listener in an anechoic room. In one condition, the speech signal consisted of the direct sound only and different SNRs were achieved by varying the level of the direct sound. In the second condition, the speech signal consisted of the direct sound plus early reflections. Different SNRs were achieved by keeping the direct sound level constant and varying the level of the early reflections. With this setup they could directly compare the benefit of increased direct sound energy with the benefit of the same increase in early reflection energy. The benefit from early reflections was found to be similar both in normal-hearing and hearing-impaired listeners.

The present study extended Bradley's *et al*. (2003) investigations. The idea was to use early reflections from a realistic room impulse response, so that there was no need to arbitrarily choose the number and delay times of the early reflections. Furthermore, these reflections were spectrally filtered according to the absorptive properties of the walls in the corresponding room. The reflections and the direct sound were manipulated in the same way as in the two conditions of Bradley's *et al*. (2003) experiment. A third condition was added where the early reflections were not distributed spatially but presented from the same direction as the direct sound. All sounds were presented to the listener in a virtual auditory environment with 29 loudspeakers. Measuring monaural and binaural speech intelligibility for these three conditions should then allow identifying: (i) how useful early reflections are compared to the direct sound, (ii) the impact of the spatial distribution of the reflections, and (iii) whether the benefit from early reflections is a monaural or binaural process.

METHODS

Sound field simulations

The speech intelligibility measurements took place in an acoustically dampened room

(T30 < 100 ms for frequencies f > 200 Hz) with 29 loudspeakers arranged symmetrically around the listener (16 loudspeakers in the horizontal plane at the height of the listener's head, 7 loudspeakers at the ceiling and 6 loudspeakers on the floor). The room impulse response (RIR) used to create a realistic sound field was taken from a classroom modelled with the room acoustic software Odeon (Odeon, 2008) as shown in the left panel of Fig. 1. The classroom had a volume of 170 m3 and the source-receiver distance was 3.5 m. The early reflection pattern (reflectogram) in this classroom is shown in the right panel of Fig. 1. It illustrates the spatial distribution and delay times (relative to the direct sound) of the 20 early reflections used in this experiment. From the RIR, all reflections up to the second order were included, the last reflection arriving about 55 ms after the direct sound. Reflections arriving later than 55 ms after the direct sound were discarded. The direction of each component in the reflectogram was adjusted to match the position of the closest loudspeaker in the 29-loudspeaker array. The RIR was then processed with the LoRA toolbox (Loudspeaker Room Auralisation System, Favrot and Buchholz, 2009), resulting in a 29-channel RIR, which was then convolved with the speech signal and played back via the loudspeaker array. The listener was seated in the middle of the array at a distance of 1.8 m from the loudspeakers in the horizontal plane. Speech intelligibility was measured binaurally, monaurally left and monaurally right for three conditions. In the first condition, only the direct sound of the speech was presented from 0° azimuth (DSonly). In the second condition, the direct sound of the speech was presented from 0° azimuth together with the spatially distributed early reflections from the right panel of Fig. 1 (DSERspatial). Finally, in the third condition, both the direct sound of the speech and the early reflections were presented from 0° azimuth (DSERfrontal). In condition one, the level of the speech signal was varied by changing the level of the direct sound. In conditions two and three, the direct sound level was kept constant and the level of the spatial or the frontal early reflections was varied. In all conditions the level was measured with an omni-directional microphone at the location of the centre of the listener's head with the listener absent.

Speech material

The Danish sentence test Dantale II (Wagener, 2003) was used to measure speech intelligibility. It is based on the Hagerman sentence test (Hagerman, 1982) where each sentence consists of five words with a fixed syntactical structure. The listeners responded via a Matlab user interface (see Fig. 2) by choosing the words they had heard on a hand-held touch screen.

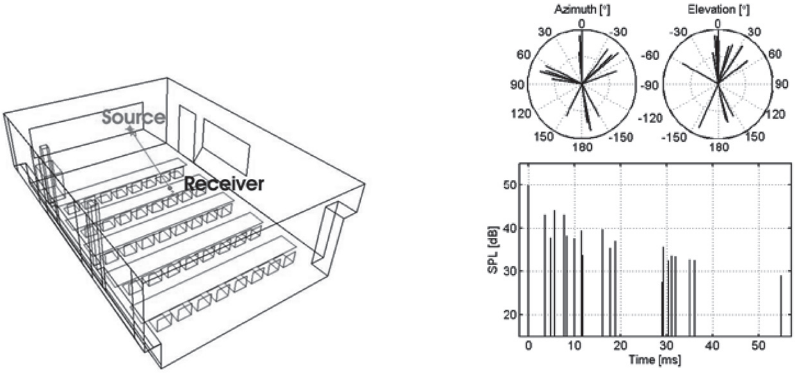
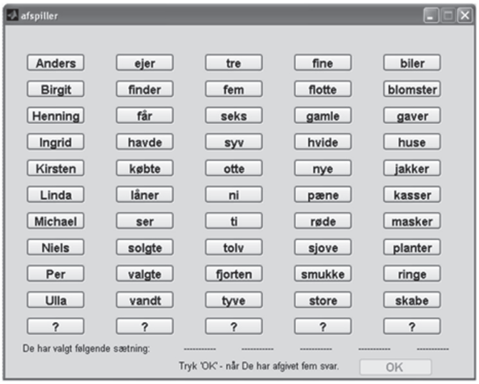


Fig. 1: Left: Classroom with talker (source) and listener (receiver) position as simulated in Odeon. **Right**: Reflectogram for the simulated classroom.

Background noise

A diffuse stationary speech-shaped noise (SSN) created from the sentence material was used as interferer (Wagener, 2003). The SSN was cut in uncorrelated noise signals which were played simultaneously from all loudspeakers. A gated-noise procedure was used, i.e. the noise started 1s before each sentence with a 0.6 s onset ramp and stopped 0.5 s after each sentence with a 0.3 s offset ramp. The interferer was presented at a fixed level of 60 dB SPL, measured with an omni-directional microphone at the location of the centre of the listener's head with the listener absent.



De har valgt følgende sætning: ---------- ---------- ---------- ---------- ----------

Tryk 'OK' - når De har afgivet fem svar.

Fig. 2: Matlab user interface employed in the speech intelligibility test Dantale II.

Listeners

Nine normal-hearing listeners participated in the experiment. Their hearing thresholds were 20 dB HL (re. ISO 389-8) or better for both ears at octave frequencies between 0.25 and 6 kHz. The listeners were between 23 and 46 years old with a median age of 24. The experiments were split in two sessions, each lasting about 1.5 hours per person. Before each session, the listeners performed a training with 30 sentences. All listeners were familiar with the sentence test before. Listeners were paid on an hourly basis for their participation.

Procedure

In the first part of the experiment, the speech reception threshold (SRT) of each listener was determined with 20 sentences using the modified RIR with spatial early reflections. The overall level, i.e. the level of the direct sound plus the early reflections was varied adaptively with a maximum likelihood procedure. Afterwards, in the main experiment, all speech intelligibility scores were measured relative to each individual listener's SRT, i.e., the sensitivity of each listener was normalised. At the SRT, the contribution of spatial early reflections to the overall speech signal was 6 dB. The reference point for all conditions was set to the speech level at the SRT minus 6 dB (i.e. no early reflections). From this reference point, the SNR was increased stepwise by either adding direct sound energy or early reflection energy. For each SNR, the speech intelligibility was measured with 10 sentences per person.

For the monaural speech intelligibility measurements, one ear at a time was closed with an ER2 insert earphone (Etymotic Research) which provided a minimum of 30 dB sound attenuation between 0.125 and 8 kHz. In addition, white noise was presented through the earphone at a level of 75 dB SPL. In this way, the non-test ear was completely masked and no headphones disturbed the sound field at the test ear. Binaural speech intelligibility was measured in a first session followed by a second, monaural session where the listeners either started with the left or the right ear. The insert earphone was not removed until the monaural measurement was finished. The conditions and the order of measurements within each condition were randomized. Before the actual measurement, the listeners were instructed to look at the front loudspeaker and to hold their head upright during sentence presentation.

RESULTS

In Fig. 3 the speech intelligibility scores averaged over 9 listeners, are shown for the different conditions. Error bars indicate ± 1 standard deviation. The left panel shows the binaural speech intelligibility scores, the middle panel the scores measured with the left ear only and the right panel the scores measured with the right ear only. Speech intelligibility scores for the three conditions are indicated by circles (direct sound only), squares (direct sound + spatial early reflections) and diamonds (direct sound + frontal early reflections) respectively. Each level increase corresponds to an increase in SNR. The speech intelligibility for each condition increased with increasing SNR and speech intelligibility scores ranged from about 20% to almost 100% intelligibility. A logistic function p(SNR) was fit to the data given by:

$$p(SNR) = \frac{1-\alpha}{1+\exp(4 * s_{55} * (SRT_{55} - SNR))} + \alpha \qquad \text{(Eq. 1)}$$

where SNR is the signal-to-noise ratio, α is the chance level of 10%, SRT55 is the SNR at 55% correct speech intelligibility and s55 is the slope at SRT55. The parameters s55 and SRT55 were estimated by minimizing the root mean squared error.

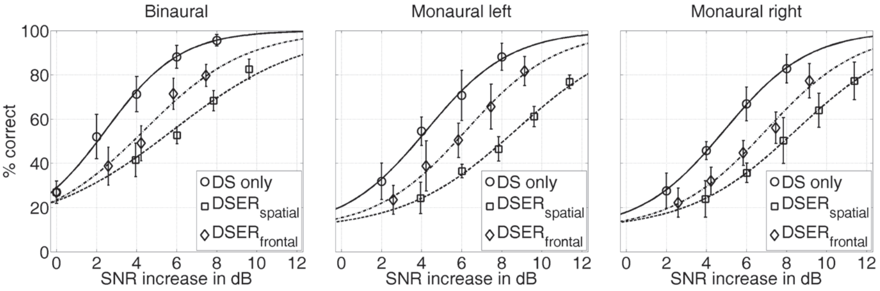


Fig. 3: Mean speech intelligibility scores and fitted psychometric functions for binaural listening (left panel), listening with the left ear only (middle panel) and listening with the right ear only (right panel).

An SNR "increase" of 0 dB corresponds to the reference point (speech level at SRT minus 6 dB). The average SRT determined binaurally with spatial early reflections was -12.9 dB SNR with a standard deviation of 0.76 dB. This corresponds to the binaural intelligibility score at an SNR increase of 6 dB for the DSERspatial condition.

A paired t-test revealed no significant difference between the left and the right ear, except for the DSonly condition at a level increase of 4 dB. Binaural scores were significantly higher than monaural scores, except for the DSERfrontal condition at a level increase of 4.36 dB. The difference between monaural and binaural scores is shown as binaural benefit in Fig. 4. It was calculated by subtracting the monaural SNR at 55% correct speech intelligibility from the binaural SNR at the same intelligibility. The binaural benefit was very similar for all three conditions and was between 1.8 and 2.9 dB.

The speech intelligibility was significantly better when the level of the speech signal was increased by direct sound energy than when it was increased by early reflection energy. Furthermore speech intelligibility increased faster with increasing direct sound energy (steeper slope) than with increasing early reflection energy. Even though speech intelligibility was higher with increased direct sound energy, there was still a benefit from increased early reflection energy, otherwise the intelligibility scores would not have increased with added early reflection energy. In the conditions with early reflections the speech intelligibility depended on the direction of the early reflections. For SNR increases of 6 dB and above, speech intelligibility was significantly higher when the early reflections were presented from the front than when they were spatially distributed.

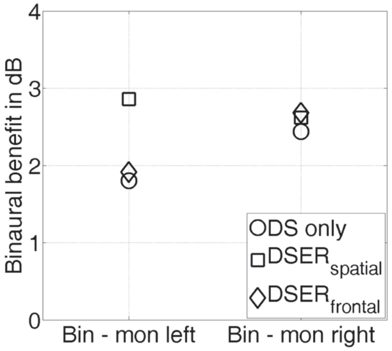


Fig. 4: Binaural benefit expressed as the difference in SNR at 55% intelligibility between monaural (left or right ear) and binaural listening.

DISCUSSION

The results have shown that early reflections improved speech intelligibility, but not to the same extent as the direct sound. Thus, only part of the energy of the early reflections arriving within 55 ms after the direct sound was useful for speech intelligibility. This is in contrast to Bradley at al. (2003), who suggested that all the early reflection energy can be used for speech intelligibility and, in realistic scenarios, can contribute up to 9 dB to the total speech level. However, there were some differences between the two studies. First, a different speech material was used to measure speech intelligibility. The test Bradley *et al*. used (a Fairbank's rhyme test modified by Latham; Latham, 1979) has a very shallow slope between 80% and 100% intelligibility of the psychometric function (the dynamic range in which they measured speech intelligibility). For normal-hearing listeners, an SNR increase of 10 dB is necessary to increase the speech intelligibility from 87% to 99%. The three DSERspatial conditions Bradley *et al*. used, increased the total speech level (and therefore the SNR) by 0, 3 and 6 dB respectively. The first DSERspatial condition (0 dB increase) corresponded to their DSonly condition, therefore speech intelligibility was (nearly) the same in these two conditions. With these two conditions as the starting point on the psychometric function and a very shallow increase of speech intelligibility, the dynamic range used by Bradley *et al*. might not have been sufficient to show the differences between the two conditions (DSERspatial and DSonly).

Second, the sound field simulations were not identical in the two studies. Bradley *et al*. used 7 early reflections and the present study used 20 spectrally filtered early reflections. Figure 5 shows the power spectrum between 100 Hz and 10 kHz for the DSonly (averaged for the left and right ear), the DSERspatial and the DSERfrontal (averaged for the left and right ear) condition recorded with the Head and Torso Simulator (HATS, Brüel & Kjær) and equalized in level. The frequencies above about 1 kHz are attenuated for the conditions with early reflections (dashed and dotted lines), due to the absorption characteristics of the walls in the simulated classroom. According to Pavlovic (1987), for average speech, the 1/3 octave frequency bands between 1 and 4 kHz contribute most to speech intelligibility. Therefore, the reduced high frequency content in the speech signals with early reflections might have led to poorer speech intelligibility than for DSonly. It is not clear if the spectrum of the early reflections that Bradley *et al*. used was changed according to the absorption of the walls or if the reflections were just delayed and attenuated copies of the direct sound.

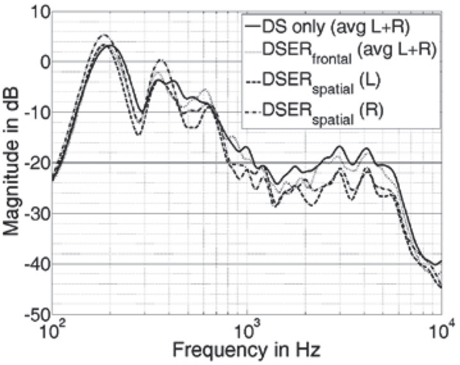


Fig. 5: Power spectrum of the speech signals in the different conditions, recorded with a Head and Torso Simulator (HATS).

The results did not only reveal a difference in speech intelligibility between the DSonly and the DSERspatial/frontal conditions, but also between the DSERspatial and the DSERfrontal conditions, both for binaural and monaural listening. The integration of early reflections with the direct sound might have been facilitated when both came from the same direction, because of the direction-dependent filtering of the outer ear and the head shadow. This is indicated in Fig. 5 where the frequencies above about 2 kHz are more attenuated for the DSERspatial than for the DSERfrontal condition. Furthermore, the characteristics of the playback system might have also played a role here (weak reflections in the acoustically dampened listening room, slight differences between individual loudspeakers), resulting in better speech intelligibility for the DSERfrontal condition. However, the difference between monaural and binaural speech intelligibility were the same for the DSERspatial and the DSERfrontal condition, which indicates that the binaural auditory system could not integrate (spatial) early reflections with the direct sound more efficiently than the monaural system. The binaural benefit of 2-3 dB shown in Fig. 4 can be explained by a summation of the power of the signals at the two ears. A perfect summation would result in a 3 dB benefit, but studies have shown that the benefit is usually a bit less (Pollack, 1948).

CONCLUSIONS

In the present study, it was possible to investigate the benefit from early reflections for speech intelligibility in a realistic listening scenario. Increased early reflection energy improved speech intelligibility, but the improvement was less than for increased direct sound energy. Therefore, treating a room acoustically to enhance early reflections is reasonable, but increasing the level of the talker's voice is even more beneficial. However, for specific listening situations, e.g. when the talker's head is turned away from the listener, early reflections can be relatively more important than the direct sound. No binaural processing of early reflections other than a summation of the signals at the two ears could be found. Thus, it is assumed that the integration of the direct sound with the early reflections takes place at an early stage of the auditory system and that the combined signal is then processed binaurally. The background noise used in this study was a diffuse stationary SSN. Such a noise might hamper binaural processing, because it is uncorrelated at the two ears and therefore difficult (or even impossible) to cancel using binaural cues. Hence, in a follow-up study, directional noise will be used to investigate if binaural listening provides an advantage over monaural listening for spatial early reflections when the noise is more correlated at the two ears. Furthermore it needs to be investigated how hearing-impaired listeners benefit from early reflections. Bradley *et al*. have used impaired listeners with a mild hearing loss at high frequencies (PTA at 3, 4 and 6 kHz: 30.5 dB), who benefitted from early reflections in the same way as normal-hearing listeners. It needs to be tested whether this is also the case for more severe hearing losses.

REFERENCES

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.

Cremer L., and Müller, H. A. (**1982**). *Principles and applications of room acoustics* Vol. 1 Section III. 2.2, (Applied Science Publishers, London).

Favrot, S., and Buchholz, J. B. (**2009**). "Validation of a loudspeaker-based room auralisation system using speech intelligibility measures," 126th convention of the Audio Eng. Soc., Munich.

Hagerman, B. (**1982**). "Sentences for testing speech intelligibility in noise," Scand. Aud. **11**, 79-87.

ISO (**2004**). ISO 389-8. "Acoustics – Reference zero for the calibration of audiometric equipment – Part 8. Reference equivalent threshold sound pressure levels for pure tones and circumaural earphones," International Standardization Organization, Geneva, Switzerland.

Latham, H. G. (**1979**). “Signal-to-noise ratio for speech-intelligibility – Auditorium acoustics design index,” Appl. Acoust. **12**, 253-320.

Lochner, J. P. A., and Burger, J. F. (**1958**). “The subjective masking of short time delayed echoes by their primary sounds and their contribution to the intelligibility of speech,” Acustica **8**, l-10.

Lochner, J. P. A., and Burger, J. F. (**1964**). “Influence of reflections on auditorium acoustics,” J. Sound Vibrat. **1**, 426-454.

Nabelek, A. K., and Robinette, L. (**1978**). “Influence of precedence effect on word identification by normally hearing and hearing-impaired subjects,” J. Acoust. Soc. Am. **63**, 187-194.

Odeon Room Acoustic Software (**2008**). Version 9.1.

Parizet, E., and Polack, J. D. (**1992**). “The influence of an early reflection upon speech-intelligibility in the presence of a background-noise,” Acustica **77**, 21-30.

Pavlovic, C. V. (**1987**). “Derivation of primary parameters and procedures for use in speech-intelligibility predictions,” J. Acoust. Soc. Am. **82**, 413-422.

Pollack, I. (**1948**). “Monaural and binaural threshold sensitivity for tones and for white noise,” J. Acoust. Soc. Am. **20**, 52-57.

Soulodre, G. A., Popplewell, N., and Bradley, J. S. (**1989**). “Combined effects of early reflections and background-noise on speech-intelligibility,” J. Sound Vibrat. **135**, 123-133.

Thiele, R. (**1953**). “Richtungsverteilung und Zeitfolge der Schallrueckwuerfe in Raeumen,” Acustica **3**, 291–302.

Wagener, K., (**2003**). “Design, optimization and evaluation of a Danish sentence test in noise,” Int. J. Aud. **42**, 10-17.