Forum Acusticum 2011 - Aalborg - Denmark

Influence of early reflections on speech intelligibility under different noise conditions

Anna Warzybok
Medizinische Physik, University Oldenburg, Germany

Jan Rennies
Fraunhofer IDMT / Hör-, Sprach- und Audiotechnologie, Oldenburg, Germany

Medizinische Physik, University Oldenburg, Germany

Simon Doclo
Signal Processing, University Oldenburg, Germany

Fraunhofer IDMT / Hör-, Sprach- und Audiotechnologie, Oldenburg, Germany

Birger Kollmeier
Medizinische Physik, University Oldenburg, Germany

Fraunhofer IDMT / Hör-, Sprach- und Audiotechnologie, Oldenburg , Germany

Summary
Early reflections are known to be useful for speech intelligibility due to the fact that they can be partially integrated with the direct speech signal. The present work focuses on the influence of early reflections on speech intelligibility under different noise conditions. More specifically, the influence of a single frontal reflection, arriving within 10 ms to 200 ms after the direct sound, was investigated for diotic noise, diffuse noise and a spatially separated noise source (135°). Binaural signals were generated by convolving speech and noise signals with room impulse responses simulated using the CATT Acoustics software. Speech reception thresholds (SRTs) were measured in eight normal-hearing subjects with an adaptive procedure, using speech material from the Oldenburg Sentence Test and headphone presentation. The reference condition consisted of presenting only the direct sound. Except for the reference condition, the speech level was varied by proportionally changing the energy of the direct sound and the early reflection with respect to the noise energy. The measured SRT data indicate that the early reflection can be fully integrated with the direct sound in a certain time window, resulting in the same SRT as for the reference condition. The length of this integration window is comparable across noise conditions: a single early reflection can be fully integrated within at least 25 ms in diffuse or diotic noise and within at least 50 ms in spatially separated noise. The spatial release from masking was independent from the delay of the first reflection. Hence, the integration process of early reflections in the temporal domain seems to take place independent of the spatial processing.

PACS no. 43.71.Gv, 43.66.Ba

1. Introduction

In a realistic acoustic environment speech intelligibility is mainly determined by background noise, reverberation, and hearing abilities. In literature a number of experimental and theoretical studies have been reported, which investigate the relationship between the acoustical conditions and their influence on speech intelligibility [1, 2, 3].

The influence of background noise can be taken into account by considering the signal-to-noise ratio (SNR), where in this case the signal energy is the speech energy at the receiver position. The total reflected energy reflected (at the room walls or objects) can be divided into two parts, namely the useful part (so-called early reflections) and the detrimental part (so-called late reflections). Early reflections are mostly defined as reflections arriving within the first 50 ms after the direct sound, and are considered to be useful for speech intelligibility as they can be integrated with the

direct sound [3, 4, 5, 6]. Understanding this integration effect in a variety of spatial speech and noise configurations is of great importance for accurate speech intelligibility prediction as well as for signal processing in hearing instruments. The benefit of early reflections for speech intelligibility has been examined in several studies [4, 7, 8]. More recently, Bradley et al. [5] showed that the energy in seven early reflections arriving within the first 50 ms after the direct sound can be as beneficial to speech intelligibility in noise as the energy in the direct sound. Bradley's studies were extended by Arweiler et al. [6], who used early reflections from the first 55 ms of a realistic room impulse response. Contrary to Bradley's investigations, they found that the improvement in speech intelligibility in noise is greater for increased direct sound energy than for increased early reflection energy. In other words, they showed that the energy of the early reflections cannot be fully integrated with the energy of the direct sound.
These contrary findings motivated us to further investigate the influence of early reflections on speech intelligibility in noise. The present study tested how a single frontal reflection can be integrated with the direct sound as a function of delay under different spatial noise conditions.

2. Methods

1.1. Participants

Eight subjects between 22 to 27 years (mean 25.4 years) participated in the speech intelligibility measurements. All listeners had pure-tone thresholds not exceeding 20 dB HL at octave frequencies from 250 to 8000 Hz. The subjects were paid for their participation in the experimental sessions. None of them had an extensive experience with speech intelligibility measurements.

1.2. Procedure: speech material and measurements setup

Speech intelligibility was measured using the Oldenburg Sentence Test [9, 10, 11], which consists of semantically unpredictable sentences of a fixed syntactical structure *'name verb numeral adjective object'*. Each test list contains 20 sentences. The interfering noise was generated by multiple superimpositions of the speech test material. As a consequence, the long-term spectrum of the noise matches the long-term spectrum of the sentences [11]. In the measurements the noise level was kept constant at 65 dB SPL. The speech level was varied adaptively, converging to the threshold of 50% speech intelligibility (so-called Speech Reception Theshold, SRT). The step size of the level change depended on the number of correctly repeated words of the previously presented sentence [12].

The influence of the first reflection on speech intelligibility was investigated for three different noise conditions, namely diotic noise (N_0), diffuse noise (N_d) and a spatially separated noise source located at 135° azimuth (N_{135}). The target speaker (S_0) and the first reflection (R_0) were always located in front of the listener (see Figure 1). Speech intelligibility was measured for six different delays, i.e. 10, 25, 50, 75, 100, or 200 ms, of the first reflection. For each type of noise, the reference condition contained only the direct sound as the speech signal.

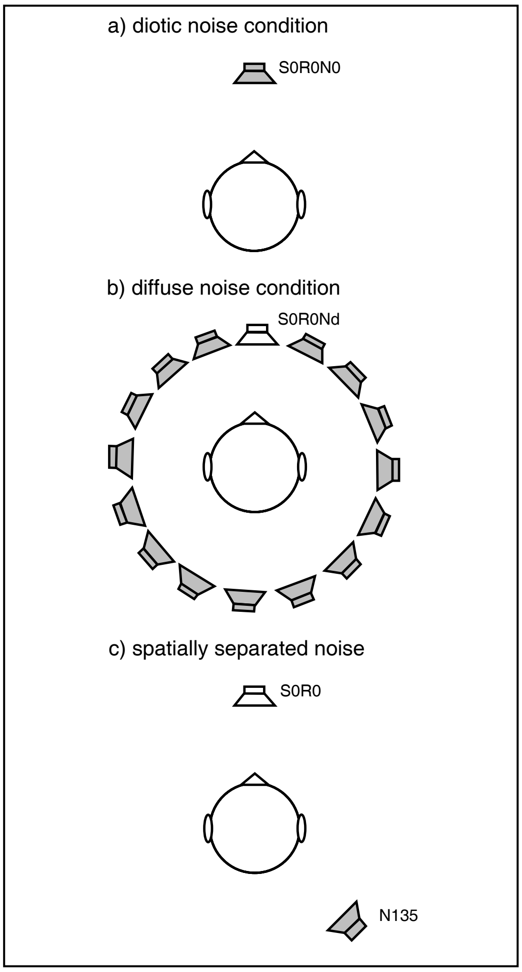

Figure 1. Sketch of the spatial arrangement of target speech, first reflection and noise sources.

To create these spatial configurations, clean speech and noise signals were convolved with binaural room impulse responses (BRIRs) simulated using the CATT Acoustics software. The BRIRs were generated for a room with dimensions of 12 x 24 x 4.4 m³ (width x length x height) and a source-receiver distance of 5 m. The speech and noise sources were modeled as omnidirectional sources and the receiver, located on the central axes of the room, was simulated with a head-and-torso simulator (KEMAR). All sources had an elevation of 1.2 m. In the reference condition all surfaces of the room were non-reflective, such that the resulting BRIR contained only the direct sound without any reflections. To compute the BRIRs for the remaining conditions the absorption coefficient of the ceiling was set at 0.05 for all frequencies. The so simulated BRIRs correspond to the condition where the first reflection arrives at 10 ms and provided the basis for generating BRIRs for other delays, which were generated by shifting the reflection of 10 ms by 662, 1764, 2866, 3969 and 8379 samples. Due to the sampling frequency of 44.1 kHz delay of 25, 50, 75, 100, and 200 ms was introduced.

The simulated signals were presented binaurally over free-field equalized Sennheiser HD200 headphones. The measurement setup was calibrated to dB SPL using a Brüel&Kjær (B&K) 4153 artificial ear, a B&K 4134 1/2 in. microphone, a B&K 2669 preamplifier, and a B&K 2610 measurement amplifier.

A pilot study was conducted to examine the effect of the amplitude of the first reflection on speech intelligibility. Two different settings were used: a) an early reflection with an amplitude originally recorded with KEMAR, and b) an early reflection with an amplitude scaled to the amplitude of the direct sound. The results, presented in Figure 2, showed that there were no statistical differences in SRTs between these two conditions for delays up to 75 ms. For greater delays SRTs for setting ‘b’ were higher than for setting ‘a’. These measurements indicate that the effect of delay of the first reflection with originally recorded amplitude is very small. Only if the first reflection is amplified (i.e., scaled to the same amplitude as the direct sound as in these measurements), a clear (detrimental) effect is observed.

During the measurement, the order of the delays as well as the noise conditions was randomized across subjects. First, test subjects were trained to familiarize them with the stimuli and the task and to account for the training effect [11]. For each listener two practice lists in noise were presented using the original speech material (i.e. without convolution with BRIRs). The first list was presented at a fixed SNR of -2 dB in a closed-set format. The second practice list was used to determine the SRT using an open-set format. In the closed-set format after presentation of a sentence to a subject, a panel containing the 50 words of the base matrix was displayed and the subjects’ task was to indicate the presented words. In the open-set format the subjects’ task was to repeat the words they had understood and the experimenter marked the correctly repeated words on a display.

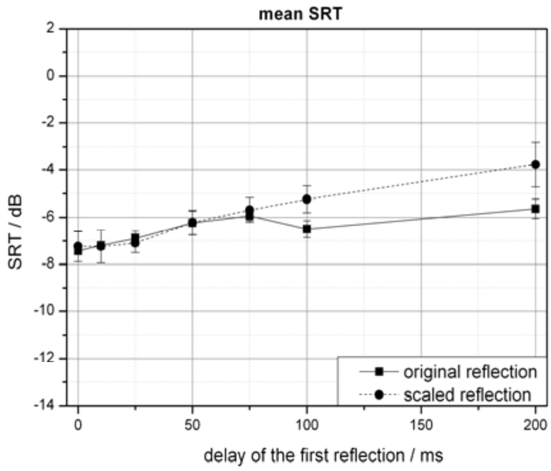

Figure 2. Mean SRTs and corresponding standard deviation for different delays of the first reflection with originally recorded amplitude (squares) and with scaled amplitude (circles) for 8 subjects.

3. Results

The measured SRTs for the different noise conditions are shown as mean SRTs across all eight subjects in Figure 3. Error bars indicate plus and minus one interindividual standard deviation. SRTs obtained for the respective noise condition and delay of the first reflection were analyzed by means of a two-way analysis of variance (ANOVA). Bonferroni tests (level of significance set at $p<0.05$) were used to determine the sources of significant effects indicated by the ANOVA. The ANOVA revealed a significant main effect of noise condition ($F(2,147)= 1852.025$, $p<0.001$) and delay ($F(6,147)= 62.846$, $p<0.001$), but no interaction was found between them ($p=0.07$). Post hoc comparisons showed statistically significant differences between all noise conditions ($p<0.001$). For the delay factor, no significant differences in SRT were found between the reference condition and thresholds for delays of 10 ms and 25 ms. This suggests that a single frontal reflection can be fully integrated with the direct sound up to 25 ms. SRTs as a function of delay varied from -7.2 dB (for delay=10 ms) to -3.8 dB (for delay=200 ms) in $S_0R_0N_0$ condition, from -10.6 dB (for delay=10 ms) to -6.6 dB (for delay=200 ms) in the diffuse noise condition, and from -18.6 dB (delay=10 ms) to -13.6 dB (delay=200 ms) for the spatially separated noise source ($S_0R_0N_{135}$). The mean SRT measured in diotic noise was on average 3.9 dB higher than the mean SRT measured in diffuse noise and on average 11.2 dB higher than in spatially separated noise.

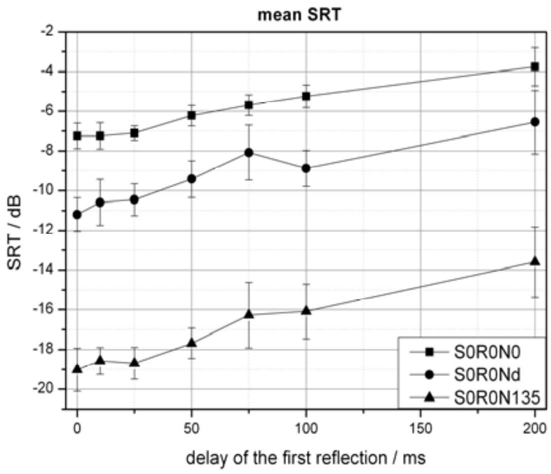

Figure 3. Mean SRTs and corresponding standard deviation for different delays of the first reflection in three noise conditions: diotic noise (squares), diffuse noise (circles) and spatially separated noise at 135° (triangles).

A one-way ANOVA was conducted for each noise condition separately to test if the length of the integration time window was similar for each noise condition. The statistical analyses revealed no significant difference in SRT up to 25 ms for diotic and diffuse noise, but a larger value of up to 50 ms for spatially separated noise.

To examine the effect of delay of the first reflection on the spatial release from masking, the individual difference in SRT between the S0R0N0 and the S0R0N135 condition as well as between the S0R0N0 and the S0R0Nd condition was calculated for each delay. The ANOVA revealed that neither the spatial release from masking caused by separating the direction of the noise source from 0° to 135° nor the diffuseness-related release from masking did not depend on the delay of the first reflection ({$F(6,49)=0.38$, $p=0.99$)} and {$F(6,49)=0.724$, $p=0.63$)}, respectively).

4. Discussion

Integration of the first reflection

The present study indicates that a single reflection can be fully integrated with the direct sound within a certain time window. It was shown that the spatial properties of the noise did not considerably influence the temporal integration of the first reflection. This is in line with Bradley et al. [5], who showed that the increased energy of early reflections can have approximately the same effect on speech intelligibility as an equal increase in the direct sound energy. However, there are some differences between the two studies. First, Bradley used seven spatially separated early reflections arriving within the first 50 ms after the direct sound instead of a single frontal reflection. In our study the single reflection was fully integrated up to 25 ms and not to 50 ms. Bradley et al. measured speech intelligibility at fixed SNRs for which intelligibility scores were ranging from 80% to 94%. In our study speech intelligibility was measured adaptively to obtain 50% intelligibility threshold, because the intelligibility function at this point has the highest steepness, resulting in the highest SRT estimation accuracy. Further experiments using more than one reflection are required to directly compare the outcomes of these studies.

The intelligibility level difference

In literature, the SRT improvement caused by separating the direction of the noise source from the direction of the target sound (the so-called intelligibility level difference) averages about 12 dB in an anechoic condition for noise sources at 105° to 120° azimuth [13, 14, 15]. In our study the intelligibility level difference for a noise source located at 135° was 11.5 dB for the reference condition, i.e. the difference in SRT between S0R0N135 and S0R0N0 condition at delay=0ms. Since both the S0R0N135 and the S0R0N0 condition show a similar dependence on the delay of the first reflection (i.e., upper and lower curve in Fig. 3), the ILD (i.e., the difference between both curves) can be averaged across all delays. This mean ILD (11.1 ± 1.1 dB) is very close to the ILD for the reference condition. This suggests that the temporal integration takes place independent of spatial processing.

The diffuseness of the sound field yielded an SRT improvement of 4.2 dB in the reference condition. A similar improvement was found by vom Hövel [13] who studied the influence of the diffuseness of the sound file on speech intelligibility and showed a benefit of 3 dB for diffuse noise. The mean effect of diffuseness averaged across all delays (difference in mean SRT between diotic and diffuse noise condition) was 3.8 dB ±1.1 dB. This indicates that the SRT improvement related to diffuseness of the noise source is similar for all conditions. Hence, this diffuseness-related special form of spatial processing does also not depend on the delay of the first reflection.

Conclusions

In the present study the influence of a single frontal reflection on speech intelligibility was studied as a function of delay with respect to the direct sound. It was shown than for different spatial noise conditions a frontal reflection arriving within 10 ms to at least 25 ms after the direct sound can be fully integrated with the direct speech signal. Hence the assumption of a time window is supported that integrates early reflections and yields the same SRT as measured only with the direct sound as speech source. The estimated length of the integration window was comparable across noise conditions. Its lower bound was 25 ms in diotic or diffus noise and 50 ms in spatially separated noise. The spatial release from masking was independent from the delay of the first reflection. Hence, it can be assumed that the integration process of early reflections in the temporal domain takes place independent of the spatial processing.

Acknowledgement

This work was supported by the PhD Program “Function and Pathophysiology of the Auditory System” funded by the state of Lower Saxony.

References

[1] A. K. Nabelek, J. M. Pickett: Monaural and binaural speech perception through hearing aids under noise and reverberation with normal and hearing-impaired listeners. J Speech Hear. Res. 7 (1974).

[2] J. S. Bradley: Speech intelligibility studies in classrooms. J. Acoust. Soc. Am. 80 (1986).

[3] J. S. Bradley, R. D Reich, and S. G. Norcross: On the combined effects of signal-to-noise ratio and room acoustics on speech intelligibility. J. Acoust. Soc. Am. 106 (1999).

[4] H. Haas: Über den Einfluss eines Einfachechos auf die Hörsamkeit von Sprache (On the influence of a single echo on speech intelligibility). Acustica 1 (1951) 49-58.

[5] J. S. Bradley, H. Sato, M. Pickard: On the importance of early reflections for speech in rooms. J. Acoust. Soc. Am. 113(6) (2003) 3233-3244.H.

[6] I. Arweiler, J. M. Buchholz, T. Dau: Speech intelligibility enhancement by early reflections. ISAAR (2009), Marienlyst**,** Denmark.J. P. A.

[7] Wallach, E. B. Newman, et al: The precedence effect in sound localization. Am. J. Psychol. 52 (1949) 315-336.

[8] Lochner, J. F. Burger: The influence of reflections on auditorium acoustics. J Sound Vib. 1 (1964).

[9] K. Wagener, T. Brand, B. Kollmeier: Development and evaluation of a German sentence test I: Design of the Oldenburg sentence test. Z. Audiol. 38 (1999a) 4-15.

[10] K. Wagener, T. Brand, B. Kollmeier: Development and evaluation of a German sentence test II: optimisation of the Oldenburg sentence test. Z. Audiol. 38 (1999b) 44-56.

[11] K. Wagener, T. Brand, B. Kollmeier: Development and evaluation of a German sentence test III: evaluation of the Oldenburg sentence test. Z. Audiol. 38 (1999c) 86-95.

[12] T. Brand, B. Kollmeier: Efficient adaptive procedures for threshold and concurent slope estimates for psychophysics and speech intelligibility tests. J. Acoust. Soc. Am. 111 (2002) 2801-2810.

[13] H. vom Hövel: Zur Bedeutung der Übertragungseigenschaften des Außenohres sowie binauralen Hörsystems bei gestörter Sprachübertragung (On the importance of transmission properties of the outer ear and the binaural auditory system for disturbed speech transmission). Ing. Dissertation (1984), RWTH Aachen.

[14] R. Plomp, A. M. Mimpen: Effect of the orientation of the speaker’s head and the azimuth of a noise source on the speech-reception threshold for sentences. Acustica 48 (1981) 325-328.

[15] J. Peissig, B Kollmeier: Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners. J. Acoust. Soc. Am. 101(3) (1997) 1660-1670.