Temporal-spatial processing of a single speech reflection in normal-hearing and hearing-impaired listeners

Jan Rennies¹, Anna Warzybok², Thomas Brand², Birger Kollmeier¹,²

¹ Projektgruppe Hör-, Sprach- und Audiotechnologie, Fraunhofer IDMT, 26129 Oldenburg, E-Mail: {jan.rennies, birger.kollmeier}@idmt.fraunhofer.de
² Carl-von-Ossietzky Universität, 26111 Oldenburg, E-Mail: {a.warzybok, thomas.brand, birger.kollmeier}@uni-oldenburg.de

Introduction

Reflections and reverberation considerably affect speech recognition in rooms. Early reflections of the speech signal have mostly been found to enhance speech intelligibility [1, 2, 3, 4], while late reflections can have detrimental effects [4, 5]. A better understanding of the mechanisms underlying speech perception in noisy and reverberant conditions can be important for various practical investigations, e.g., the improvement of prediction models to be used in room design, de-reverberation and noise suppression algorithms, or the optimization of PA systems and loudspeakers configurations in rooms.

In recent studies [4, 5], the ability of the normal auditory system to integrate a single speech reflection with frontal direct sound has been investigated with respect to reflection delay (0 to 200 ms), reflection azimuth (0° to 315° in steps of 45°), and type of interferer (frontal, diffuse, or lateral). The main findings were that the benefit of early reflections was independent of reflection azimuth, that the binaural gain in the presence of non-frontal interferers was independent of reflection delay for frontal reflections, and that the detrimental effect of a late reflection depended on the reflection azimuth relative to the spatial position of the interferer. The data could be well predicted by a binaural speech intelligibility model [5, 6], when the distinction between early and late reflections was made prior to the model’s binaural processing stage. The present study extended the previous experiments by measurements with hearing-impaired listeners. The data were compared to model predictions to investigate to what extent reductions in speech intelligibility could be explained by the pure tone audiogram, which was the only stage of the model adjusted to individual hearing impairments.

Experiment

The experimental method was the same as described in [4, 5], the only differences being the group of listeners and the fact that only a subset of the conditions was tested. Data of normal-hearing subjects were taken from the previous studies and included in the present study for comparison.

Listeners

Nine hearing-impaired subjects between 40 and 77 years participated in the experiments. All had moderate, sloping sensorineural hearing impairment, which was rather symmetrical between left and right ears. Pure-tone averages (average hearing loss across octave frequencies from 500 to 4000 Hz) ranged from 29 to 58 dB HL and did not differ by more than 7 dB between left and right ear for any subject.

Stimuli and procedure

Speech reception thresholds were measured adaptively using speech material from the Oldenburg sentence test [7]. The signals were presented binaurally over free-field equalized HDA200 headphones. The level of the masking noise was kept constant and the speech level was varied to converge to the signal-to-noise ratio (SNR) corresponding to 50% speech intelligibility. The level of the masking noise was fixed at 65 dB SPL for the normal-hearing listeners [4, 5], but increased for the hearing-impaired subjects by half the hearing loss at 500 Hz followed by an individual adjustment to a comfortable listening level for each listener. This resulted in noise levels between 75 and 80 dB SPL. The long-term spectrum of the noise matched the long-term spectrum of the sentences. To simulate different directions of the reflection and noise the clean speech and noise signals were convolved with binaural head-related impulse responses (HRIRs) simulated using the CATT Acoustic software v8.0a. Signals were always calibrated based on the right-ear signals. In the following, results are always presented as SNRs at the better-ear, i.e., the measured thresholds are shifted for asymmetrical listening conditions to eliminate broadband better-ear effects (see also [4]). The direct sound was always located in front of the listener. The speech reflection of the same amplitude as the direct sound arrived at delays between 10 and 200 ms after the direct sound. The temporal processing of the frontal reflection was investigated for three different noise conditions, namely diotic noise (Exp. I, N₀), diffuse noise (Exp. II, N₀), and lateral noise located at 135° azimuth (Exp. III, N₁₃₅). For each type of noise, the reference condition contained only the direct sound (0 ms delay). The temporal as well as spatial processing of the speech reflection was investigated in diffuse (Exp. IV) and lateral noise (Exp. V). In diffuse noise, the reflection arrived from an azimuth of 135° or 225°. In

Figure 1: Average (thin gray lines) and mean audiograms (thick lines) of the nine-hearing-impaired subjects.
lateral noise, the reflection arrived either from the same hemisphere as the noise source (azimuth of 90° and 135°) or from the opposite hemisphere (azimuth of 225°). The noise source was located at an azimuth of 135°. Table I summarizes the spatial configurations of sources and reflection delays in each experiment. The speech level was always calculated from the root-mean-square value of the entire speech signal comprising direct sound and reflection. This means that when a reflection was added, the level of the direct sound was reduced compared to conditions without a reflection. Listeners were trained prior to the actual measurements to familiarize them with the task and to reduce the effects of training [7].

<table>
<thead>
<tr>
<th>Exp.</th>
<th>$\phi_S / ^\circ$</th>
<th>$\phi_R / ^\circ$</th>
<th>$\phi_N / ^\circ$</th>
<th>$\Delta t / \text{ms}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0, 10, 25, 50, 200</td>
</tr>
<tr>
<td>II</td>
<td>0</td>
<td>0</td>
<td>D</td>
<td>0, 10, 25, 50, 200</td>
</tr>
<tr>
<td>III</td>
<td>0</td>
<td>0</td>
<td>135</td>
<td>0, 10, 25, 50, 200</td>
</tr>
<tr>
<td>IV</td>
<td>0</td>
<td>135, 225</td>
<td>D</td>
<td>10, 50, 200</td>
</tr>
<tr>
<td>V</td>
<td>0</td>
<td>90, 135, 225</td>
<td>135</td>
<td>10, 50, 200</td>
</tr>
</tbody>
</table>

Table 1: Experimental parameters used in the different experiments of this study: speech azimuth ($\phi_S$), reflection azimuth ($\phi_R$), noise azimuth ($\phi_N$), and reflection delay ($\Delta t$). $D$ denotes diffuse characteristics.

**Speech intelligibility model**

Different models were proposed to predict binaural speech intelligibility in noisy rooms. In this study, a model developed by Rennies et al. [5, 6] was used because it had been already proven accurate for predicting SRTs of normal-hearing subjects in the same conditions as considered here [5]. The model is based on the binaural speech intelligibility model (BSIM) presented in [8]. The first stage of the original model consists of bandpass filtering the binaural speech and noise signals using an auditory filterbank. Internal noise is derived from the individual audiograms and added to the external noise to account for hearing impairment (note that this is the only stage of the model adjusted to individual listener properties). In each auditory filter, an equalization-cancelation mechanism [9] is applied to take advantage of spatially separated speech and noise sources. This mechanism aims at optimizing the SNR by amplifying and delaying the left and right ear signals relative to each other followed by subtraction of the two channels from each other. If speech and noise signals differ in their interaural gain and time differences (i.e., if they arrive from different directions), gain and delay parameters can be chosen such that destructive interference is achieved for the noise signals, but less destructive (or even constructive) interference is achieved for the speech signals resulting in an improved SNR. The resulting SNRs are then used as input to the Speech Intelligibility Index (SII) from which SRTs are calculated [6].

The basic model structure of BSIM was extended in [6] to be applicable also speech degraded by reverberation. This was achieved by separating the input speech signals into an early (presumably useful) part and a late (presumably detrimental) part based on the binaural HRIRs. Only the early part was used as speech input, while the late part was added to the external masker. By comparing model predictions to data of normal-hearing listeners, it was demonstrated that it was imperative that the distinction between early and late (i.e., between useful and detrimental) speech components be made prior to the binaural processing stage [5]. In addition, it was shown that the separation of useful and detrimental components based on the classically employed sharp transition at a certain time after the direct sound (e.g., at 50, 80, or 100 ms) was problematic for stimuli consisting only of direct sound and a single reflection. Instead, a smooth transition weighting function of linearly increasing and decreasing ramps was proposed [5]. In the present study, the exact same parameters as used in [5] for normal-hearing listeners were employed (separation time 100 ms, ramp durations 200 ms), i.e., the only difference between the model for normal-hearing subjects of [5] and for the hearing-impaired subjects of this study was the use of individual audiograms. The model is referred to as BSIM-UD100 as in the original studies.

**Results**

**Exps. I-III: Frontal speech reflection in different masking conditions**

SRTs averaged across listeners and the corresponding standard deviations obtained in conditions with frontal direct sound and a frontal reflection are shown in Figure 2 for normal-hearing (left panels) and hearing-impaired subjects (right panels) as a function of reflection delay. Different gray scales represent the different spatial noise conditions.

SRTs of normal-hearing listeners remained constant up to reflection delays of about 25 ms, and then increased with increasing reflection delay. The largest increase compared to the reference condition was observed at the longest delay and amounted to between 4.1 dB ($N_{10}$) and 5.5 dB ($N_{115}$). The increase was very similar for all masker types as indicated by the constant binaural intelligibility differences (BILDs, bottom panel). On average across reflection delays, SRTs
were lower by about 4 dB (N_D) and 8 dB (N_135) compared to the diotic noise condition (N_0), i.e., normal-hearing listeners could make substantial use of the spatial properties of the masker in non-diotic conditions. The patterns of measured SRTs were qualitatively similar in hearing-impaired listeners (right panels). In the S_0N_0 condition with a delay of 0 ms (i.e., the condition without reflection and potential binaural advantage), the mean SRT is 2 dB higher than observed for normal-hearing subjects, reflecting a general increase in SRT. As for normal-hearing listeners, BILDs were independent of reflection delay (bottom right panel). The increase in SRT with increasing reflection delay, however, was considerably larger than in normal-hearing listeners, ranging from 7.7 dB (N_D) to 9.3 dB (N_135). Hearing-impaired subjects had a reduced benefit of only about 2 dB in N_D and 3 dB in N_135 compared to a diotic masker.

**Exp. IV and V: Varying speech reflection azimuth in diffuse and lateral noise**

To investigate not only the temporal but also the spatial processing of a speech reflection, speech intelligibility measurements were conducted for different azimuths of the reflection. The measurements were conducted in the same lateral (N_135) and diffuse noise (N_D) as used in Exp. II and III, respectively. Mean SRTs with corresponding standard deviations measured in hearing-impaired listeners in the presence of a lateral masker (Exp. V) are shown in the left panel of Figure 4. The corresponding predictions of BSIM-UD100 are presented in the right panel. Data and predictions of normal-hearing listeners were shown in [5]. In general, SRTs measured in hearing-impaired listeners showed similar trends as observed in the normal-hearing data: an early lateral reflection (delay 10 or 50 ms) was equally beneficial as a frontal reflection at the same delay, i.e., SRTs did not differ between the different azimuths. The effect of a detrimental reflection (delay 200 ms), however, depended on reflection azimuth relative to the noise azimuth (135°). While a large SRT increase was observed for reflection azimuth from the front (same direction as the direct sound, black line in Figure 4) or from the hemisphere opposite the noise source (light gray line), only a very small SRT increase was found when the late reflection originated from the same hemisphere as the noise source (dark gray lines). The same trends were observed in normal-hearing listeners [5], and were interpreted as an azimuth-dependent suppression of the detrimental effect of a late reflection due to the spatial separation of the reflection from the direct sound. The observed trends were also predicted by BSIM-UD100 (right panel of Figure 4), although the same discrepancies as described above were also revealed in the data of Exp. V: predicted SRTs were generally too low, and the detrimental effect at the largest delay was considerably underestimated.

**Overall model performance**

Figure 5 shows the scatter plot of measured against predicted SRTs. These analyses included all data of Exps. I to V. It can be seen that the prediction accuracy is close-to-optimum for normal-hearing subjects (left) with an R² of 0.95 and a negligible bias. For hearing-impaired subjects (right),

---

**Figure 3:** Same data representation as in Figure 2, including predictions of BSIM-UD100 (dashed lines).

The corresponding predictions of BSIM-UD100 are indicated by dashed lines in Figure 3. As was already shown in [5], predictions agreed very well with measured data for normal-hearing subjects (left panels) as indicated by the coefficient of determination (R²) and the prediction bias. For hearing-impaired subjects, however, some discrepancies were observed (right panels). The model correctly predicted the general SRT increase in the S_0N_0 condition without reflection, and also predicted that the binaural gain was independent of reflection delay. The amount of binaural gain, however, was generally underestimated (by 1 and 2 dB for N_0 and N_135, respectively). The largest discrepancies were observed for the longest delay. The predicted SRT increase compared to the reference condition was the same as for normal-hearing listeners (i.e. between 3 and 4 dB), while the measured increase was more than twice as large (see above). Despite these differences, about 72% of the data’s variance could be explained by the data. The prediction bias was about 2 dB larger than for normal-hearing subjects.
predictions were less accurate ($R^2=0.69$, bias 1.8 dB), but still reasonably good despite some clear outliers.

**Figure 5:** Measured vs. predicted SRTs for normal-hearing (left) and hearing-impaired subjects (right). Different gray scales represent data measured in diotic noise ($N_D$, black), diffuse noise ($N_D$, dark gray), and in lateral noise ($N_{135^\circ}$, light gray).

**Summary and conclusions**

The present study investigated the temporal and spatial processing of a single speech reflection in the presence of different maskers in normal-hearing and hearing-impaired subjects. Normal-hearing listeners could fully integrate a frontal speech reflection with the direct sound up to a delay of about 25 ms, irrespective of the spatial properties of the masker. A similar integration window was observed for hearing-impaired subjects although the SRT increase at a delay of 25 ms was slightly larger than in normal-hearing subjects. The temporal processing of the frontal reflection was independent of the spatial properties of the noise source for both groups of listeners, resulting in a binaural gain independent of reflection delay. Overall, the binaural benefit in maskers spatially separated from direct sound and reflection was reduced in hearing-impaired subjects. A large discrepancy between normal-hearing and hearing-impaired subjects was observed with respect to the detrimental effect of a late reflection, which was about twice as large in hearing-impaired listeners. An extended binaural speech intelligibility model [5, 6] could accurately predict normal-hearing data, but overestimated the remaining binaural gain in hearing-impaired listeners and underestimated the detrimental effect of a late reflection for this group of subjects. While the reduced binaural gain could presumably be predicted by adjusting parameters in the binaural processing stage of the model, the reasons underlying the increased detrimental effects at large reflection delays are still unclear and require further research.

For a reflection spatially separated from the direct sound, an interaction between temporal and spatial processing was observed: For both listener groups, the detrimental effect of a late (detrimental) reflection was reduced when it originated from the same hemisphere as the lateral masker. This effect, which was also predicted by the model, was in line with the interpretation that the binaural auditory system is better at suppressing two noise sources when they are placed close to each other (ideally in the same hemisphere) than when they are spatially separated [10].

Despite the observed discrepancies between data and predictions, the overall prediction performance of the model was still reasonably good also for hearing-impaired listeners. This may indicate that a significant amount of the observed deficits are related to elevated thresholds as indicated by the pure-tone audiograms. Other effects, however, clearly involve other factors of spatial and temporal speech processing and require further research to improve future predictions models for hearing-impaired subjects.

**References**


