Effects of spatial and temporal integration of a single early reflection on speech intelligibility[a)]

Anna Warzybok[b)]
Medical Physics, Institute of Physics, Carl-von-Ossietzky University Oldenburg, D-26111 Oldenburg, Germany

Jan Rennies
Project Group Hearing, Speech and Audio Technology, Fraunhofer Institute for Digital Media Technology IDMT, Marie-Curie-Straße 2, D-26129 Oldenburg, Germany

Thomas Brand
Medical Physics, Institute of Physics, Carl-von-Ossietzky University Oldenburg, D-26111 Oldenburg, Germany

Simon Doclo[c)]
Signal Processing Group, Institute of Physics, Carl-von-Ossietzky University Oldenburg, D-26111 Oldenburg, Germany

Birger Kollmeier[c)]
Medical Physics, Institute of Physics, Carl-von-Ossietzky University Oldenburg, D-26111 Oldenburg, Germany

(Received 30 September 2011; revised 11 September 2012; accepted 5 November 2012)

In order to study the interaction between the intelligibility advantage in rooms due to the presence of early reflections and due to binaural unmasking, a series of speech reception threshold experiments was performed employing a single reflection of the frontal target speech source as a function of its delay ranging from 0 to 200 ms. The direction of the reflection and the spatial characteristic of the interfering noise (diotic, diffuse, or laterally localized) were varied in the experiments. For the frontal reflection, full temporal integration was observed for all three noise types up to a delay of at least 25 ms followed by gradual intelligibility decay at longer delays. At 200 ms delay the reflection introduced additional intelligibility deterioration. For short delays, intelligibility was not reduced when the reflection was spatially separated from the direct sound in the diffuse and lateral noise conditions. A release from the deterioration effect at 200 ms delay was found for all spatially separated reflections. The suppression of a detrimental reflection was symmetrical in diffuse noise, but azimuth-dependent in lateral noise. This indicates an interaction of spatial and temporal processing of speech reflections which challenges existing binaural speech intelligibility models.
 [http://dx.doi.org/10.1121/1.4768880]

PACS number(s): 43.55.Hy, 43.66.Pn, 43.71.Gv [RYL] Pages: 269–282

I. INTRODUCTION

The acoustic design of a listening environment is one of the major factors affecting the perception of speech. In a realistic room speech intelligibility is mainly determined by background noise and reverberation. The total reflected energy (of the walls or objects) can be divided into two parts, namely the early, useful part and the late, detrimental part. While the late reflections create a more or less diffuse sound field which can substantially reduce speech intelligibility (e.g., Steeneken and Houtgast, 1980; George *et al.*, 2010; Rennies *et al.*, 2011), the early reflections can be integrated with the direct sound and thus increase speech intelligibility (e.g., Lochner and Burger, 1964; Bradley *et al.*, 2003; Arweiler and Buchholz, 2011). However, the influence of the early reflections depends on several factors such as their delay and amplitude relative to the direct sound or their direction. Investigating the influence of these factors can help to understand temporal and spatial integration properties of the auditory system. In this study, the interactions of different parameters of a single, strong reflection with respect to binaural speech intelligibility are systematically investigated. The outcomes of the present study may be important for improving existing models of binaural speech intelligibility. Such models are an interesting alternative for fast assessment of speech intelligibility in different acoustic scenarios. Several binaural speech intelligibility models have been developed to predict intelligibility in different rooms and for various noise azimuths (vom Hövel, 1984; Beutelmann and Brand, 2006; Lavandier and Culling, 2010; Beutelmann *et al.*, 2010; Rennies *et al.*, 2011). In general,

[a)]A part of this study was presented at Forum Acusticum 2011, 17 June–1 July, Aalborg, Denmark.
[b)]Author to whom correspondence should be addressed. Electronic mail: anna.warzybok@uni-oldenburg.de
[c)]Also at Project Group Hearing, Speech and Audio Technology, Fraunhofer Institute for Digital Media Technology IDMT, Marie-Curie-Str. 2, D-26129 Oldenburg, Germany.

they can predict intelligibility with high accuracy for various noise azimuths, distances, and rooms by including some of the basic properties of the auditory system in spatial masking and unmasking of spatially arranged target speech and interfering sounds. Moreover, some of them can account for useful and detrimental influence of reflections and reverberation in the speech signal (e.g., vom Hövel, 1984; Lavandier and Culling, 2010; Rennies *et al.*, 2011), e.g., by a temporal integration of the direct sound and the early reflections. However, these models do not explicitly include an interaction between temporal integration and spatial (un)masking that may occur in the auditory system when dealing with early reflections from different incoming directions in the presence of a directional interfering noise. Even though the idealized case of a single strong reflection may be viewed as the prototype condition in which any such interaction between temporal integration and spatial (un)masking might most clearly be revealed, the models listed above have not yet been evaluated in conditions comprising single strong reflections. Therefore, it is currently unknown to what degree the models can predict the interaction of temporal and spatial processing in the binaural auditory system. The present study aims at providing an experimental basis for better understanding the underlying mechanisms. The results can then be used as a benchmark for testing different interaction mechanisms realized in existing binaural speech intelligibility models or their future extensions. The specific implications of the experimental results for existing models are discussed in Sec. IV. The data may also be interesting for other practical applications, for example in the positioning of reflecting surfaces and loudspeakers in a room or in signal-processing strategies such as de-reverberation algorithms. Such algorithms aim at reducing the detrimental effects of late reflections and thereby improving speech intelligibility and sound quality, e.g., in hearing aids or hands-free mobile phones. The present data can be of interest for the improvement of psychoacoustically derived criteria for de-reverberation algorithms (Fielder, 2001).

Several studies have addressed the role of early reflections in speech intelligibility. It is generally agreed that intelligibility is enhanced by the addition of early reflections compared to a situation with only the direct sound when the level of the direct sound is kept constant. This is true as long as the reflections arrive within a certain time window after the direct sound. Lochner and Burger (1964) investigated the integration of a single reflection for speech in quiet. The reflection was a delayed version of the direct sound. They found that up to delays of about 30 ms, the single reflection was fully integrated in terms of energy with the direct sound, resulting in the same intelligibility as measured for a 3-dB increase in the effective speech level. For larger delays, intelligibility was still improved in the presence of the reflection, but the improvement gradually decreased and disappeared at a delay of about 95 ms. When the amplitude of the reflection was reduced by 5 dB, the time window for full integration slightly increased to about 40 ms and the overall benefit was smaller.

The integration effect of a single reflection in the presence of masking noise was investigated by Nábělek and Robinette (1978). Similar to the study of Lochner and Burger (1964), a single reflection arrived at different delays from 0 to 160 ms. For all delays, the speech items were presented at constant speech level in a background of babble noise consisting of eight speakers. They observed constant word intelligibility scores (i.e., full integration of the reflection) up to delays of 20 ms.

In contrast, only partial integration of a single reflection of the same amplitude as the direct sound was found by Parizet and Polack (1992). They showed that for speech presented in pink noise a single reflection arriving 5 or 10 ms after the direct sound could improve speech intelligibility but could not be fully integrated with the direct sound resulting in lower speech intelligibility scores than those obtained when SNR was increased by simply raising energy of the direct sound by 3 dB. Similarly, Soulodre *et al.* (1989) observed a beneficial influence of 13 reflections arriving within 40 ms after the direct sound on speech intelligibility in noise, but no full energetic integration of all reflections with the direct sound.

In contrast, Bradley *et al.* (2003) found full integration of early reflections with the direct sound in the presence of noise. As in the other studies, speech intelligibility was compared between speech consisting of only the direct sound and speech additionally containing early reflections. They used seven reflections all arriving within 50 ms after the direct sound. Each reflection was presented from a different loudspeaker. Although the setup of loudspeakers was symmetrical around the listener, the reflection pattern was not diotic, since each speaker produced a reflection at a different delay and level. The distribution of relative delays and amplitudes of the reflections across speakers was always the same. The signal-to-noise ratio (SNR) was modified by keeping the noise level constant and either changing the level of the direct sound (in the condition without reflections) or the overall level of the early reflections (in the condition with reflections present). Bradley *et al.* (2003) observed the same increase in intelligibility irrespective of whether the additional speech energy was provided as direct sound or as early reflections. In additional experiments, they showed that the observed effects of early reflections could also be found when additional reverberation was added to the stimuli.

In a more recent study, Arweiler and Buchholz (2011) measured speech intelligibility in diffuse noise using a similar paradigm, i.e., they also varied the SNR by either increasing the direct speech energy or the energy of early reflections. They used a realistic impulse response containing 20 reflections within 55 ms after the direct sound. The spectrum of the reflections was different from that of the direct sound since the absorptive characteristics of the room boundaries were preserved. The reflections were presented to the listeners via a 29-loudspeaker setup approximating the original distribution of angles in horizontal direction and elevation. In addition to the realistic sound presentation, Arweiler and Buchholz (2011) also measured intelligibility in a condition in which all reflections were presented from a frontal loudspeaker. Each of these conditions was measured binaurally and monaurally with an insert earphone in either ear, which produced masking noise and thereby prevented binaural speech processing. In contrast to Bradley *et al.* (2003), Arweiler and Buchholz (2011) observed that an increase in direct-speech energy was more beneficial than an equivalent increase in energy of the early reflections. This difference was smaller when all reflections were presented from the frontal speaker, leading Arweiler and Buchholz (2011) to the conclusion that temporal integration of early reflections may have been facilitated when they arrived from the same direction as the direct sound. Intelligibility was always better in binaural than in monaural listening conditions by about 2 and 3 dB. This difference, which could be explained by spatial unmasking in the presence of the diffuse masker, was constant for frontal and spatially distributed early reflections, i.e., the binaural system could not integrate early reflections more efficiently than the monaural system. Arweiler and Buchholz (2011) therefore argued that the integration of early reflections is independent of binaural processing. This hypothesis is tested in the current paper using an even sparser representation of a binaural room impulse response which allows for a better separation between temporal integration of an early reflection and its spatial masking or unmasking produced by different noise maskers.

The above-mentioned studies mostly considered a fixed binaural configuration of direct speech and reflection. Only Arweiler and Buchholz (2011) compared two spatial presentation modes of the reflections (frontal and spatially distributed). Moreover, all studies used a fixed configuration of masking noise, e.g., a fixed dichotic (Nábělek and Robinette, 1978; Parizet and Polack, 1992) or diffuse noise (Soulodre *et al.*, 1989; Bradley *et al.*, 2003; Arweiler and Buchholz, 2011). It is possible (and was also suggested by Arweiler and Buchholz, 2011) that directional noise sources interact in quite a different way with the temporal and spatial integration of early reflections. This motivated us to extend previous findings and to further investigate the influence of an early reflection in different spatial conditions of direct sound, reflection, and masking noise. In particular, the interaction between the delay relative to the direct sound, the azimuth of the reflection, and the spatial noise configuration was addressed. In Expt. I, a diotic listening condition was considered as a reference for the subsequent experiments, examining the integration of a single, frontal reflection with respect to the influence of reflection amplitude and a possible influence of a reflection in the masker. Experiments II and III tested if the integration of a single, frontal reflection depends on spatial noise configurations using a diffuse noise (Expt. II) and a single, lateral noise source (Expt. III). Using the same noise types, Expts. IV–VII investigated a dichotic speech presentation with frontal direct sound and different reflection azimuths.

II. METHODS

A. Subjects

Twelve normal-hearing subjects (four male, eight female) participated in Expts. I to III and ten of them in Expts. IV to VII. The listeners' ages were between 21 and 27 years, with a mean of 24.5 years. All participants had pure tone thresholds not exceeding 20 dB HL (hearing level) at octave frequencies from 125 to 8000 Hz and reported no problems with their listening capabilities. The subjects were paid for their participation. None of them had extensive experience with speech intelligibility measurements.

B. Procedure

Speech intelligibility was measured using the Oldenburg sentence test (Wagener *et al.*, 1999a,b; Wagener *et al.*, 1999c), which consists of sentences of the fixed syntactical structure "*name verb numeral adjective object*." For each word, ten alternatives are available which can be randomly combined resulting in semantically unpredictable sentences. In total the test comprises 45 test lists of 20 sentences, each list containing each of the 50 words exactly twice. The speech items were presented in stationary interfering noise which had been generated by multiple superpositions of the test material so that the long-term spectrum of the noise matched the long-term spectrum of the sentences (Wagener *et al.*, 1999c). The noise was presented at a level of 65 dB SPL (sound pressure level) and started and ended 500 ms before and after the sentence (including 50 ms Hann ramps), respectively. The initial level of the speech signal was also 65 dB SPL, i.e., the initial SNR was 0 dB. The speech level varied during the adaptive measurement procedure to converge to the threshold of 50% speech intelligibility, i.e., to the speech reception threshold (SRT). The task of the subjects was to repeat the words they had understood and an instructor marked the correct responses on a display (not visible to the subjects). The step size of the speech level change depended on the number of correctly repeated words of the previous sentence and a convergence factor (for details, see Brand and Kollmeier, 2002). Each of the tested conditions (see Sec. II C) was measured with one randomly selected test list.

During the measurements, the order of the different experiments was the same for each subject, but within each experiment the different conditions were randomized. The measurements were made in several sessions of 1–2 h each. The overall measurement duration for all experiments was about 7 h per subject. Prior to the first measurement session, two practice lists in noise were presented using the original speech material (i.e., without convolution with impulse responses as described in Sec. II C) to familiarize the subjects with the stimuli and the task and to account for the training effect (Wagener *et al.*, 1999b). The first training list was presented at a fixed SNR of −2 dB, which corresponds to good intelligibility for normal-hearing listeners. The second training list was used to determine the SRT using the adaptive procedure described above. In contrast to the actual measurements, both practice lists were presented in closed-set format, i.e., the subjects' task was to indicate the words they had understood at a panel containing all 50 words of the test. The closed-set format was chosen for the training to ensure that the entire speech material was known to the subjects. In the beginning of each of the remaining measurement sessions, one training list was presented at a fixed SNR of −2 dB in a closed-set format. The data from the training test lists were discarded.

C. Acoustical setup and stimuli

The influence of the single reflection on speech intelligibility was investigated as a function of its delay (Δt) relative to the direct sound in seven different experiments. The experiments varied in reflection azimuth (ϕ_R) and noise type (see Table I for an overview), while the direct sound of the speech was always presented frontally (S_0). Figure 1 illustrates the spatial configurations of the sources used in the experiments.

In Expt. I, the effect of temporal integration of a frontal reflection (R_0) on speech intelligibility was investigated using a frontal noise source (N_0, Expt. I). The reflection had either the same amplitude as the direct sound ($A_R = 1.0$) or an amplitude reduced by a factor of 0.3 ($A_R = 0.3$). The delay was 0, 10, 25, 50, 75, 100, or 200 ms relative to the direct sound. Two different frontal interferers (including or not including a reflection) were tested for short delays of 10, 25, and 50 ms to examine the influence of adding a reflection to the noise. In Expts. II and III, the speech signals (S_0R_0) were presented in different noise interferers, namely in diffuse noise (N_D, Expt. II) and dichotic noise located at an azimuth of 135° (N_{135}, Expt. III). Next, the temporal integration of a spatially separated reflection arriving from 45° was examined in diffuse noise ($S_0R_{45}N_D$, Expt. IV) and in laterally located noise ($S_0R_{45}N_{135}$, Expt. V). Based on the results of Expts. IV and V, a subset of delays (10, 50, and 200 ms) was chosen and speech intelligibility was measured for different reflection azimuths ϕ_R in diffuse noise ($S_0R_{\phi R}N_D$, Expt. VI) and in laterally located noise ($S_0R_{\phi R}N_{135}$, Expt. VII). The reflection was located at an azimuth of 135°, 225°, or 315° in Expt. VI, and at 90°, 135°, 180°, 225°, 270°, or 315° in Expt. VII.

To distinguish between a true binaural mechanism and monaural effects (related to spectral differences), a speech-weighted SNR was calculated for each experimental condition (Greenberg *et al.*, 1993). The speech-weighted SNR is a measure of an effective SNR taking into account the relative contributions of different regions of the frequency spectrum to speech intelligibility. Therefore, it is considered more meaningful for comparisons of different measurement conditions than the broadband SNR, especially with regard to speech intelligibility. To calculate the speech-weighted SNR, speech and noise signals were first divided into 1/3-octave bands. Then, the SNRs in each frequency band were weighted according to their contribution to speech intelligibility. The frequency-weighting function represented the average speech and was taken from Table III of the Speech Intelligibility Index standard (ANSI, 1997).

TABLE I. Experimental parameters used in the different experiments of this study: speech azimuth (ϕ_S), reflection amplitude relative to the direct sound (A_r), reflection azimuth (ϕ_R), noise azimuth (ϕ_N), and delay between direct sound and reflection (Δt). D denotes diffuse characteristics.

Expt.	ϕ_S/°	A_r	ϕ_R/°	ϕ_N/°	Δt/ ms
I	0	1.0, 0.3	0	0	10, 25, 50, 75, 100, 200
II	0	1.0	0	135	10, 25, 50, 75, 100, 200
III	0	1.0	0	D	10, 25, 50, 75, 100, 200
IV	0	1.0	45	135	10, 25, 50, 75, 100, 200
V	0	1.0	45	D	10, 25, 50, 75, 100, 200
VI	0	1.0	90, 135, 180, 225, 270, 315	135	10, 50, 200
VII	0	1.0	135, 225, 315	D	10, 50, 200

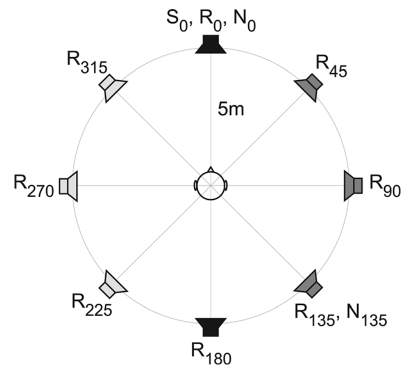

FIG. 1. Spatial configurations of the sources used in the different experiments. The direct sound of the speech material was always presented frontally (S_0); noise was also presented frontally (N_0), laterally (N_{135}), or diffusely (N_D, not shown). Black speakers indicate diotic reflections (R_0 and R_{180}), dark gray speakers indicate a reflection from the same side as the lateral noise source (R_{45}, R_{90}, and R_{135}), and light gray speakers indicate a reflection from the opposite direction (R_{225}, R_{270}, and R_{315}). The azimuth of the single reflection varied in the experiments in steps of 45°.

All measurements were conducted over headphones. To simulate different delays and azimuths of the reflection as well as the different spatial conditions of the interfering noise, the original, clean speech and noise signals were convolved with head-related impulse responses (HRIRs).[1] The HRIRs were simulated using the CATT Acoustic software v8.0a (Gothenburg, Sweden). Speech and noise sources were modeled as omnidirectional sources at the desired positions. The receiver was simulated as a head-and-torso simulator (KEMAR; G.R.A.S., Sound & Vibration, Holte, Denmark). All sources had the same elevation and the source-receiver distance was 5 m. In the reference condition ($\Delta t = 0$ ms), all surfaces of the room were non-reflective such that the resulting HRIRs contained only the direct sound. The HRIR of the reference condition provided a basis for HRIRs of the remaining conditions using a reflection of the same amplitude as the direct sound. To generate different delays, a copy of the direct sound was shifted by 441, 1103, 2205, 3307, 4410, or 8820 samples to introduce the desired delays of 10, 25, 50, 75, 100, or 200 ms (sampling frequency: 44.1 kHz). This ensured that the direct sound and the reflection had the same amplitude and spectral characteristics in experiments using a frontal speech reflection (Expts. I to III). In Expt. I, for measurements investigating the influence of reflection amplitude, the reflection was additionally reduced in amplitude by multiplication with a factor of 0.3. The diffuse noise was generated as the sum of uncorrelated portions of the noise signal convolved with HRIRs of sources spaced at steps of 5° on a circle around the receiver.

The simulated signals were presented binaurally over free-field equalized Sennheiser (Wademark, Germany) HDA200 headphones. The measurement setup was calibrated to dB SPL using a Brüel & Kjær (B&K, Bremen, Germany) 4153 artificial ear, a B&K 4134 $\frac{1}{2}$ in. microphone, a B&K 2669 preamplifier, and a B&K 2610 measurement amplifier. All signals were calibrated such that the signal at the right ear corresponded to the desired level. Except for the reference condition with direct sound only, the energy of the speech signal was always calculated from the root-mean-square value of the sum of direct sound and early reflection. The speech level was varied by identically changing the energy of the direct sound and the early reflection. For example, at the first presentation of the adaptive procedure ($SNR = 0$ dB), the speech level on the right ear was 65 dB SPL for conditions when the reflection was located in the right hemisphere. At the same time, the level at the left ear was slightly lower due to head shadow. For reflections arriving from the left side of the head, the level at the right ear still equaled 65 dB SPL and the level on the left ear was greater than 65 dB SPL. This means that, in conditions with symmetrical noise (frontal or diffuse), the observed influence of reflection azimuth would be asymmetrical due to the calibration with respect to the right ear. To facilitate the interpretation of the results, these effects of calibration were excluded from the analyses. This was achieved by shifting the data collected in conditions with the higher SNR at the left ear ($S_0R_{225}N_D$, $S_0R_{315}N_D$, $S_0R_{\phi R}N_{135}$) by the SNR differences between the left and right ear. In other words, the SRTs reported in the following were calculated as the better-ear SNR at 50% intelligibility.

D. Statistical analysis

The statistical significance of the measured effects was analyzed by means of two-way analyses of variance (ANOVAs) with the factors "reflection delay" and "noise condition" for Expts. I to III, and the factors "reflection delay" and "reflection azimuth" for Expts. IV to VII. Bonferroni *post hoc* tests (level of significance set at 5%) were used to determine the sources of significant effects indicated by the ANOVAs.

III. RESULTS AND DISCUSSIONS

A. Effects of reflection amplitude and masker reflection (Expt. I)

This part of the study focused on two aspects, namely the effect of the amplitude of a single, frontal reflection on speech intelligibility and the influence of an early reflection added to the diotic noise signal (N_0). Lochner and Burger (1964) showed that for measurements in quiet the integration of a single reflection depended on reflection amplitude. To test how this affected the speech intelligibility measurements in noise in the present study two different settings were considered: (a) a single reflection with the same amplitude relative to the direct sound and (b) a single reflection with an amplitude reduced by a factor of 0.3. SRTs were measured at delays of 0, 10, 25, 50, 75, 100, and 200 ms.

Next, the influence of a single reflection on speech intelligibility was considered using two different interferers, namely noise including the same reflection ($A_r = 1.0$) as the frontal speech source ($N_{0,R}$), and noise only consisting of direct sound (N_0). Noise including only the direct sound was used for reflection delays of 10, 25, and 50 ms. On the one hand, including the reflection for the noise source is more realistic since, in real rooms, speech and noise would experience the same reflection pattern when presented from the same source. On the other hand, the superposition of the noise with an early reflection of itself may change its masking properties. This may be related to comb-filter effects which are typically observed with ripple noise (e.g., Yost and Hill, 1978), and which were addressed in several studies (e.g., Blauert, 1971; Hartung and Trahiotis, 2001). When only one, strong reflection follows the direct sound, spectral notches occur at equidistant frequencies, the distance being inversely proportional to the delay of the reflection (Blauert, 1971). Because the focus of the present study was mainly on the integration of an early reflection with the direct speech signal, the influence of a reflection in the masking noise on temporal integration of the early speech reflection was only considered in this experiment, while the masker only consisted of direct sound in the remaining experiments.

1. Results

The mean SRTs with corresponding interindividual standard deviations for setting (a) ($A_R = 1$, black triangles connected with a solid line) and (b) ($A_R = 0.3$, gray squares connected with a dashed line) are shown in the left panel of Fig. 2 as a function of reflection delay.

A one-way ANOVA revealed a significant main effect of reflection amplitude ($p < 0.001$). *Post hoc* tests showed no statistical differences in SRTs between the two conditions for delays up to 75 ms. Remember that the level of the speech signal (calculated from the sum of direct sound and early reflection) was calibrated to the desired level at the right ear (see Sec. II C). Therefore, the overall level of the speech signal was equal in the two conditions. For longer delays, significantly higher SRTs were observed for setting (a) than for setting (b).

The right panel of Fig. 2 shows the mean SRTs and corresponding standard deviations for measurements in noise including a frontal reflection (black triangles) and for a noise source consisting only of the direct sound (gray circles). For each delay between 10 and 50 ms, a separate one-way ANOVA was conducted to test if significant differences occurred at each reflection delay. SRTs measured in noise without a reflection were statistically lower than in noise including a reflection by 1.0 dB ($p = 0.006$) and 0.6 dB ($p = 0.017$) for delays of 10 and 25 ms, respectively, while there was no significant difference between SRTs for a delay of 50 ms ($p = 0.85$). For noise consisting only of the direct sound no differences in SRT were found between the reference condition ($\Delta t = 0$ ms) and conditions with a reflection in the speech signal at delays of 10, 25, and 50 ms.

2. Discussion

a. Effect of reflection amplitude. No differences were observed between SRTs at short delays in the two conditions with different reflection amplitudes. This suggests that the amplitude of a reflection does not affect temporal integration of speech information in noise at short delays. Previous work of Lochner and Burger (1964) showed that the

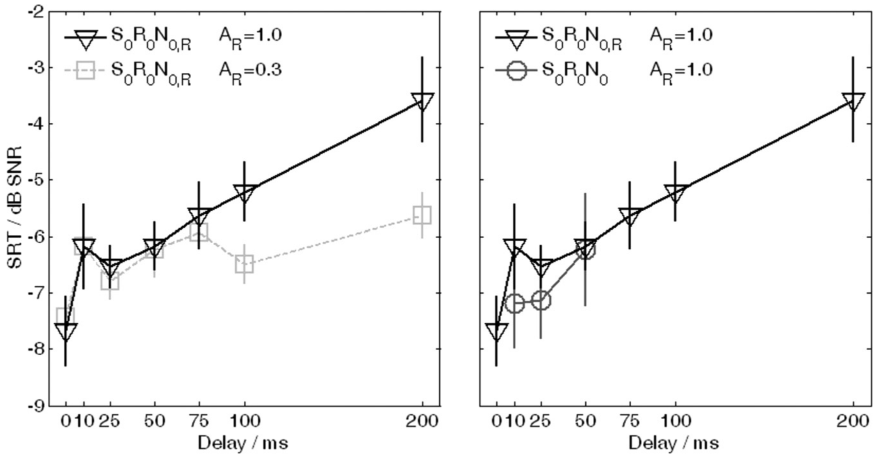

FIG. 2. Mean SRTs and interindividual standard deviations measured with frontally presented direct speech, noise, and reflection of the same amplitude as a direct sound ($A_R = 1$) and of reduced amplitude ($A_R = 0.3$). Right: Mean SRTs and corresponding standard deviation for diotic speech condition (S_0R_0) obtained in noise with ($N_{0,R}$) and without (N_0) speech reflection.

temporal window for full integration of a single reflection differed slightly depending on the amplitude of the reflection. The estimated length of the integration window was 30 ms for a reflection of the same amplitude as the direct sound and 40 ms for a reflection reduced in amplitude by 5 dB, i.e., Lochner and Burger (1964) examined the effect of reflection amplitudes using very similar amplitudes as in this study but speech intelligibility measurements were done in quiet. They found that the energy of both reflections was partially useful for speech intelligibility up to about 95 ms. The data shown in Fig. 2 indicate no such dependence on amplitude for short delays. In addition, a significant difference between SRTs obtained with reflections of different amplitude at a delay of 100 ms was observed, while Lochner and Burger (1964) found no such difference. These differences may be due to the different groups of subjects and experimental details. The measurements of Lochner and Burger (1964) were done without any interfering noise at quite low speech levels. In addition, their speech signals were not diotic since direct sound and reflection were presented from azimuths of +30° and −30°, respectively.

At a large delay of 200 ms (which was not measured by Lochner and Burger, 1964), the data of the present study showed a detrimental effect of the reflection only when the reflection had the same amplitude as the direct sound. This detrimental effect is discussed in detail in Sec. III B. In the present study, this detrimental effect and its dependence on reflection direction, delay, and type of interferer were of particular interest. Therefore, a reflection with larger amplitude ($A_r = 1.0$) was used in all remaining experiments to analyze not only a beneficial effect of a single reflection but also a detrimental effect.

b. Effect of masker reflection. The data of Expt. I showed that the addition of a reflection to the noise masker (same delay and amplitude as the speech reflection) decreased speech intelligibility compared to the condition when the masker consisted of direct sound only. As described in Sec. II C, the overall levels of the noise as well as of the speech signal with and without reflection were equal due to the calibration procedure applied. Therefore, the missing difference in SRTs between the condition containing only the direct sound and conditions in which the signals contained direct sound and a reflection indicates full integration of the early reflection with the direct sound. However, this was only observed for measurements in noise consisting of direct sound only, while intelligibility was deteriorated when adding a single, strong reflection to the noise source. One possible reason for this deterioration could be a change in spectrum of the masker. It is well-known that a strong, early reflection introduces audible changes in noise (Bilsen, 1968; Blauert, 1971; Halmrast, 2001). However, such changes seem to be only relevant for speech for delays shorter than 10 ms resulting in spectral notches of a frequency greater than 100 Hz. Since, in the present study, the shortest reflection delay was 10 ms the role of the spectral changes was probably only small. This is supported by the observation that the speech-weighted SNR did not reveal any differences between the two noise conditions (with and without masker reflection). The exact reason for the observed SRT increase is therefore not clear.

In all remaining experiments (Expts. II–VII), speech and noise were not co-located and no reflection was used for the noise source, i.e., noise only consisted of its direct sound. Therefore, for the comparison of SRTs in diotic (Expt. I) and dichotic noise (Expts. II–VII), the data collected in Expt. I were combined such that, at short delays, the SRTs were taken from the measurement without noise reflection and, at larger delays, SRTs were taken from measurements including a noise reflection. At larger delays it was assumed that the reflection did not affect SRTs, as was already observed for a delay of 50 ms.

B. Influence of a frontal reflection on speech intelligibility under different noise conditions (Expts. I to III)

This part of the study investigated the influence of a single, frontal reflection on speech intelligibility under three different noise conditions, namely diotic noise (N_0, data of Expt. I described in Sec. III B 1), diffuse noise (N_D, Expt. II) and lateral noise located at an azimuth of 135° (N_{135}, Expt. III). The goal was to investigate the integration of the single reflection with the direct sound for different delays and noise conditions. The experimental parameters are summarized in Table I. The underlying reasoning was as follows: If the reflected energy can be fully integrated with the direct sound, no differences in SRTs measured in the reference condition (speech containing only the direct sound) and conditions in which the speech signal is the sum of direct sound and reflection would be expected. If the energy of the reflected sound with the same amplitude as the direct sound was not useful for speech intelligibility at all, but could simply be ignored by the auditory system, a 3-dB reduction in SRT would be expected, since half of the speech energy would then be useless. Differences smaller than 3 dB would suggest that part of the reflected energy could still be useful for speech intelligibility, while differences greater than 3 dB would indicate that the reflection energy could not be used and additionally had a detrimental influence on speech intelligibility (i.e., served as an additional masker).

Furthermore, the hypothesis was tested whether the temporal integration process of an early reflection is independent of spatial processing (referred to as independence hypothesis in the following). This would mean that differences in SRT between the diotic noise condition (Expt. I) and diffuse noise (Expt. II) as well as lateral noise (Expt. III) are independent on reflection delay. In other words, the improvement in SRT caused by separating the direction of the noise source from the target [the so-called binaural intelligibility level difference (BILD)] would be constant as a function of reflection delay. Note that in the data of Expt. I, the effects of a reflection of the masker have been removed as described in Sec. III B 1 to facilitate the comparison to the other noise conditions, in which no noise reflection was present either.

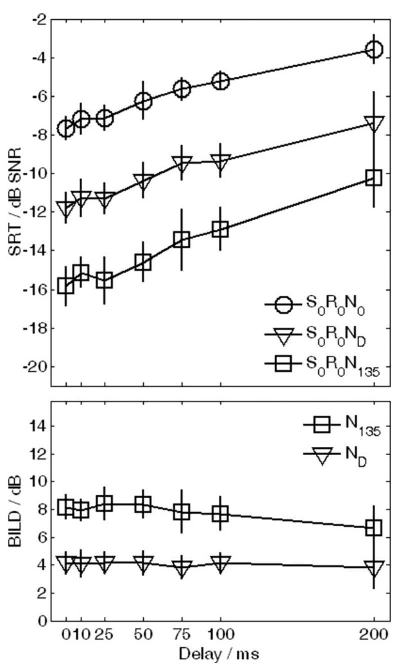

FIG. 3. Mean SRTs and interindividual standard deviations as a function of delay of a frontal reflection of the speech signal in diotic noise (circles), laterally located noise (squares), and diffuse noise (triangles). Direct sound was presented frontally. Bottom: Mean binaural intelligibility level differences for laterally located (squares) and diffuse noise (triangles) as a function of delay of the reflection.

1. Results

The mean data across all subjects of Expts. I to III are shown as different symbols in the top panel of Fig. 3. Error bars indicate plus and minus one interindividual standard deviation. For diotic noise (Expt. I, circles), SRTs generally increased with increasing delay. Mean SRTs varied from -7.7 dB ($\Delta t = 0$ ms) to -3.6 dB ($\Delta t = 200$ ms). For the laterally located noise (Expt. III, squares in top panel of Fig. 3), SRTs were considerably lower than for diotic noise, but the same trend was observed, i.e., SRTs increased from -15.9 dB ($\Delta t = 0$ ms) to -10.3 dB ($\Delta t = 200$ ms). In diffuse noise (Expt. II, triangles), SRTs were between those of Expts. I and II, and again followed the same trend increasing from -11.8 dB ($\Delta t = 0$ ms) to -7.4 dB ($\Delta t = 200$ ms). A two-way ANOVA revealed significant main effects of noise condition $\{F(2,252) = 1733.1,\ p < 0.001\}$ and reflection delay $\{F(6,252) = 53.53,\ p < 0.001\}$, but no interaction was found between them $\{F(12,252) = 1.376,\ p = 0.178\}$ indicating that the binaural advantage (i.e., the difference across the three binaural noise conditions) can be considered as constant across all delays (see below). *Post hoc* comparisons for the factor reflection delay showed no statistically significant differences in SRT between the reference condition ($\Delta t = 0$ ms) and delays of 10 and 25 ms. For each noise condition, a separate one-way ANOVA was conducted to estimate the length of the integration time window and to test differences across noise conditions. The statistical analysis revealed no significant differences in SRT up to 50 ms for diotic and lateral noise, and up to 25 ms for diffuse noise.

The mean BILDs and interindividual standard deviations derived from Expts. I to III are shown in the bottom panel of Fig. 3. Individual BILDs were calculated as the difference between SRTs in the $S_0R_0N_0$ and the $S_0R_0N_D$ condition (BILD_D, squares) as well as between SRTs in the $S_0R_0N_0$ and the $S_0R_0N_{135}$ condition (BILD_{135}, triangles) for each listener and delay. A one-way ANOVA showed that the BILD_D as well as BILD_{135} did not depend on the delay of the single reflection $\{F(6,81) = 0.402,\ p = 0.875$, and $F(6,81) = 1.868,\ p = 0.097$, respectively$\}$. The mean SRT measured in diotic noise was on average 4.1 dB higher than in diffuse noise and 7.9 dB higher than in laterally located noise.

2. Discussion

a. Integration of the frontal reflection. The data of Expts. I to III suggest that a single frontal reflection of the same amplitude as the direct sound can be fully integrated with the direct sound up to about 25 ms in diffuse noise and up to about 50 ms in diotic and laterally located noise. The estimated length of the integration window is similar to that found by Lochner and Burger (1964) in measurements in quiet for a single reflection of the same amplitude as the direct sound. Our findings are also in agreement with data of Nábělek and Robinette (1978) who showed that word identification remained constant over a range of delays from 0 to 20 ms, and with a more recent study of Bradley *et al.* (2003) who observed full integration of seven early reflections in noise within the first 40 ms after the direct sound.

However, Arweiler and Buchholz (2011) found that early reflection energy did not improve speech intelligibility to the same extent as the same amount of direct speech energy. In other words they did not observe a full integration effect of early reflections arriving within 55 ms after the direct sound. The length of the time window of early reflections may be a main reason for this discrepancy. Present data showed a fully integration of an early reflection up to a delay of 25 ms and partial integration at a delay of 50 ms in diffuse noise condition. This indicates that some of the reflections (arriving at 50 ms or later) may not have been fully integrated with the direct sound. Further, in the study of Arweiler and Buchholz (2011), the benefit from early reflections was reduced by absorptions at the walls in the simulated classroom resulting in spectral differences between the direct sound and early reflections. Since in this study the reflection was generated as a copy of the direct sound, the absorptive character of the walls was not included. The third reason might be the low relative level of the direct sound used by Arweiler and Buchholz (2011), which was set at a constant level 6 dB below the individual SRT corresponding to about 20% speech intelligibility. A low level of the direct sound could affect detection of reflections, which could have lead to the smaller benefit observed by Arweiler and Buchholz (2011) compared to the present study. To generalize the present outcomes for real rooms further studies are required including measurements with more than one reflection of the speech as well as of the noise signal.

The increase in SRT calculated as the difference in SRT between the lowest (0 ms) and highest (200 ms) delay was similar in co-located and diffuse noise (4.1 and 4.4 dB, respectively) and slightly higher in lateral noise (5.6 dB). The full integration of reflection energy observed at short delays resulted in no differences in SRTs measured with speech containing only the direct sound and speech being the sum of direct sound and reflection. At larger delays, this was no longer the case. Differences below 3 dB, suggesting partial integration of the reflected energy, were observed for delays up to 100 ms in the diotic and diffuse noise and up to 75 ms in the lateral noise. A detrimental effect of the reflection was found only for a delay of 200 ms in all noise conditions with maximum difference of 5.6 dB in lateral noise (compared to the SRT in a reference condition). This is in line with data of Lochner and Burger (1964) who showed that the single reflection of the same amplitude as the direct sound was partially useful for speech intelligibility in quiet up to delays of 95 ms. This suggests that the temporal integration takes place in a similar way in quiet and in noise.

b. Role of better-ear effects. The diffuse noise condition yielded an average SRT improvement of 4.1 dB relative to the diotic condition ($S_0R_0N_0$). About 2.9 dB of this effect could be described by differences in speech-weighted SNR at the better ear between the two conditions. This better-ear advantage was very similar for all delays ranging from 2.7 dB (75 ms) to 2.9 dB (25 ms), and indicates that most of the benefit in the diffuse noise used in the present study could be explained by better-ear listening. A similar improvement was found by vom Hövel (1984) 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 BILD_{135} across listeners caused by changing the direction of the noise source from 0° to 135° azimuth was 8.2 dB for the reference condition (i.e., the condition containing only the direct sound). This is in good agreement with data from the literature, which showed a decrease in SRT by between 7 and 10 dB for speech presented from the front (0° azimuth) and a noise source between 120° and 150° azimuth in an anechoic condition (Plomp and Mimpen, 1981; vom Hövel (1984); Peissig and Kollmeier, 1997; Beutelmann and Brand, 2006). The decrease in SRT after spatial separation of target and interferer is caused by head shadow and interaural time delays. The unmasking effect can be partially explained by a 4.3-dB higher speech-weighted better-ear SNR in lateral noise (Expt. III) than in the diotic condition (Expt. I). The remaining 3.9 dB of the unmasking effect seem to be related to binaural processing. As in diffuse noise, the speech-weighted SNR advantage was independent of the reflection delay, ranging from 4.2 dB (75 ms) to 4.3 dB (25 ms).

No interaction was found between the factors noise conditions and reflection delays. In consequence, no statistical differences were found between BILDs for the different delays. The mean BILD_D averaged across all delays was 4.1 ± 1.1 dB and was the same as the mean BILD_D for the reference condition (4.1 ± 0.8 dB). The mean BILD_{135} averaged across all delays was 7.9 ± 1.3 dB and again was very close to the mean BILD_{135} obtained for the reference condition (8.2 ± 0.9 dB). This supports the independence hypothesis, i.e., the data of Expts. I to III indicate that the temporal processing of frontal reflections takes place independently of spatial processing and in consequence does not influence binaural release from masking.

C. Spatial separation of direct sound and reflection in lateral and diffuse noise (Expts. IV and V)

This experimental part examined the temporal integration of a spatially separated reflection arriving from 45° in diffuse noise ($S_0R_{45}N_D$, Expt. IV) and in laterally located noise ($S_0R_{45}N_{135}$, Expt. V). The reflection arrived at the same delays as in Expts. I–III (see Table I). According to the independence hypothesis, the temporal integration of an early reflection should not be influenced by its direction. This means that SRTs for short delays should be close to SRTs obtained in Expts. II and III in both noise conditions. However, for a reflection arriving at larger delays a suppression effect is expected. Peissig and Kollmeier (1997) showed that an additional noise source could be suppressed if it arrived from a similar direction as the first interferer. If their observations done in an anechoic environment are valid for the conditions tested here, the detrimental effect of a late reflection (which could be considered as an additional masker) should be reduced. This would result in lower SRTs at long reflection delays than in Expts. II and III for a diffuse and lateral noise masker, respectively.

1. Results

Mean SRTs and the corresponding standard deviations obtained in the measurements with a reflection (45° azimuth) spatially separated from the direct sound (0°) are shown in the upper panel of Fig. 4. Black triangles and squares connected with solid lines represent data for a diffuse (Expt. IV) and a lateral interferer (Expt. V), respectively. For comparison, data of Expts. II and III (frontal reflection) are shown as gray symbols connected with dashed lines. A two-way ANOVA revealed significant differences in SRTs between measurement with a frontal reflection and a spatially separated reflection located at 45° azimuth $\{F(3,280) = 1225.88, p < 0.001\}$. For both interferers, there was little or no increase in SRTs between delays of 10 and 200 ms. In lateral noise, the only significant difference in SRT was found between the reference condition (no delay) and all remaining reflection delays. In diffuse noise, the differences between delays were slightly greater and the following pairs of means were significantly different: 0-ms delay and delays of 10-, 25-, and 50-ms, 10-ms delay and delays from 50 to 200 ms, and 25- and 50-ms delay and a delay of 200 ms. The position of the curves relative to the data for a frontal reflection are similar for the two noise types. The SRTs (except for a delay of 0 ms) at short delays were very close to SRTs measured with a frontal reflection. For delays longer than 50 ms, SRTs were up to 2.5 dB and 5.2 dB lower than for a frontal reflection in diffuse and lateral noise, respectively.

This is also reflected in the bottom panel of Fig. 4 which shows the difference in SRT resulting from moving the reflection from frontal position to an azimuth of 45° in the presence of diffuse (N_D, triangles) and lateral noise (N_{135}, squares). Positive differences indicate that SRTs were lower for R_{45} than for R_0, while negative differences indicate an increase in threshold due to the spatial separation of the direct sound and the reflection. For a delay of 0 ms, the differences were −1.8 and −4.6 dB for diffuse and lateral noise, respectively, while they were 2.6 and 4.1 dB for a delay of 200 ms.

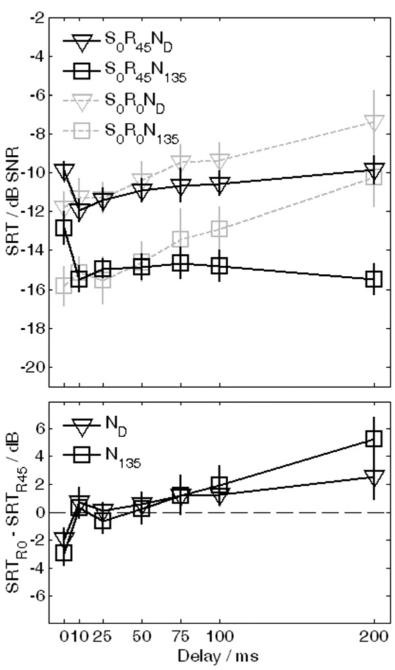

FIG. 4. Mean SRTs and standard deviations as a function of delay of a reflection of the speech signal at an azimuth of 45° in laterally located noise (squares connected with a solid line) and diffuse noise (triangles connected with a solid line). The direct sound was presented frontally. Dashed, gray curves represent data measured with a frontal reflection (Fig. 2). Bottom: Differences in SRT between a frontal and a lateral reflection for lateral (squares) and diffuse noise (triangles) as a function of delay of the reflection.

2. Discussion

The increase in SRT at a delay of 0 ms compared to 10 ms for the spatially separated reflection was most likely due to the fact that, in Expts. IV and V, the HRTF was the sum of the HRTFs for azimuths of 0° and 45°, which led to spectral changes of the speech signal, especially at the right ear. This was confirmed by calculations of the speech-weighted SNR, which was about 3 dB lower for a reflection at 0 ms than 10 ms in diffuse as well as in lateral noise. This approximately corresponds to the observed increase in SRT for a delay of 0 ms.

For short delays, SRTs in both noise conditions were close to SRTs measured with a frontal reflection in Expts. II and III. This is in line with the independence hypothesis assuming that temporal integration of early reflection does not depend on reflection azimuth. Remember that the data are presented as better-ear SNRs, which means differences in SRT due to interaural level differences have been equalized for unsymmetrical conditions (see Sec. II C for details). Significantly lower thresholds in both noise conditions were found in conditions with a spatially separated reflection than with a frontal reflection at delays longer than 75 ms. This shows that the auditory system is less disturbed by a reflection at long delays if its direction is different from the direct sound. This is in line with the hypothesis that the detrimental effect for large delays is reduced if the late reflection is spatially separated from the direct sound. The reduction of the detrimental effect due to spatial separation of the late reflection was more pronounced in laterally located noise (5.2 dB) than in diffuse noise (2.5 dB). The smaller detrimental effect of a late reflection in the lateral noise condition might be related to the fact that the suppressed direction was common for both interferers since the late reflection and the noise source were placed at the same side of the head. Peissig and Kollmeier (1997) showed that SRTs for two interfering noise sources were similar to those for one interfering source as long as both sources were located in the same hemisphere. The larger detrimental effect of a late reflection in diffuse noise might be related to the fact that, in contrast to the lateral noise, the diffuse noise cannot be localized at a certain direction and that, therefore, the simultaneous suppression of the noise and late reflection is less effective. Furthermore, the spatial configuration of a reflection at an azimuth of 45° and a noise source located at 135° might be a special case because both share almost the same interaural time difference (ITD). Therefore, if binaural processing would use the ITD as a main cue, not only the noise source would be suppressed but also the late reflection.

Comparisons of speech-weighted SNRs at a delay of 200 ms between spatially separated and frontal reflection revealed only small differences with a maximum of 0.9 dB in diffuse noise. In lateral noise, speech-weighted SNR was even better for a frontal reflection than for a reflection at an azimuth of 45°. Therefore, the suppression effect seems to be related to binaural processing.

D. Effects of reflection azimuth in lateral and diffuse noise (Expts. VI and VII)

The spatial separation of reflection and direct sound was further investigated in Expts. VI and VII for different reflection azimuths. By analogy to Expts. IV and V, it was expected that the integration process of the early reflection does not depend on reflection azimuth and spatial noise condition. For longer delays a symmetric effect of reflection azimuth was expected for diffuse noise (Expt. VI), with lower SRTs compared to the condition with a frontal reflection. For lateral noise (Expt. VII), the suppression of a detrimental reflection should be greater for azimuths located in the same hemisphere as the noise source than for reflections arriving from the opposite side of the listener's head as the noise source. The following reflection azimuths were used: 0° and 180° (diotic speech), 45° and 90° (azimuths in the same hemisphere as the noise source but not co-located), 135° (reflection co-located with the noise source), and 225°, 270°, and 315° (located in the opposite hemisphere). Measurements were done for a subset of delays (10, 50, 200 ms) based on the results of Expts. IV and V. The experimental conditions are summarized in Table I.

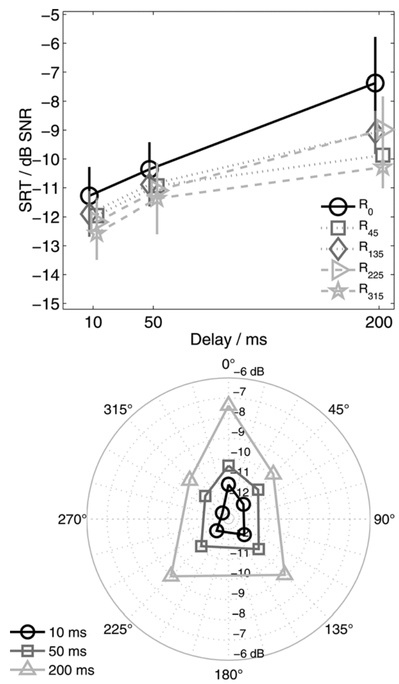

FIG. 5. SRTs measured in Expt. VI (frontal speech, diffuse noise, different reflection azimuths). Top panel: SRTs as a function of reflection delay. Black symbols connected with a solid line indicate a diotic speech signal, dark gray symbols connected with dotted lines indicate a reflection azimuth at the right side of the listeners' head, and light gray symbols connected with dashed lines indicate a reflection from the left side. For clarity, curves are slightly shifted horizontally. Bottom panel: SRTs as a function of azimuthal angle. Circles connected with a solid, black line indicate data for a reflection delay of 10 ms, squares connected with a dark gray line show threshold for a delay of 50 ms, and triangles connected with a light gray solid line indicate data for a delay of 200 ms.

1. Results

The mean SRTs and interindividual standard deviations for a diffuse noise source are shown in Fig. 5. The top panel depicts the thresholds as a function of reflection delay whereas the bottom panel shows the SRTs as a function of reflection azimuth in a polar diagram. Data for R_0 and R_{45} were taken from Expts. II and IV, respectively. In the top panel, the different symbols and gray scales indicate different reflection azimuths (black: frontal, dark gray: right hemisphere, light gray: left hemisphere). In the bottom panel, the different symbols and gray scales represent the different reflection delays (black: 10 ms, dark gray: 50 ms, light gray: 200 ms).

A two-way ANOVA revealed significant effects of reflection azimuth $\{F(4,150) = 15.17,\ p < 0.001\}$ and delay $\{F(2,150) = 125.32,\ p < 0.001\}$ as well as their interaction $\{F(8,150) = 2.88,\ p = 0.005\}$. *Post hoc* comparisons showed statistically significant differences in SRTs between 0° and all other azimuths. For reflection azimuths of 45° and 225°, there were no differences in SRTs compared to the other azimuths (except 0°). However, thresholds for 135° azimuth were significantly different from those for 315° azimuth. Further analyses (separate ANOVAs for each delay) indicated that these differences occurred only for a delay of 200 ms. For all reflection azimuths SRTs increased with increasing delay. Between delays of 10 and 50 ms, the increase was about 1.1 dB for all reflection azimuths. Between 50 and 200 ms, the increase in SRT amounted to about 3 dB (R_0), 1 dB (R_{45} and R_{315}), and 2 dB (R_{135} and R_{225}), i.e., the trends were symmetrical. For short delays (10 and 50 ms), there was no statistically significant difference between thresholds with spatially separated and frontal reflection but, for a long delay of 200 ms, the spatial separation of a reflection lead to about 2.4-dB lower thresholds than in the condition with a frontal reflection.

The corresponding data collected in lateral noise (Expt. VII) are shown in Fig. 6. Data for azimuths of 0° and 45° were taken from Expts. III and V, respectively. The symbols, gray scales, and line styles are the same as in Fig. 5. A two-way ANOVA revealed that both factors reflection azimuth $\{F(7,238) = 13.99,\ p < 0.001\}$ and delay $\{F(2,238) = 198.05,\ p < 0.001\}$ as well as their interaction were significant $\{F(14,238) = 11.77,\ p < 0.001\}$. The mean SRT in diotic speech conditions (R_0, R_{180}) was about 1.6 dB higher than the average of the conditions when the reflection source was opposite the noise source (R_{225}, R_{270}, R_{315}), while the trends always showed an increase in SRTs with increasing delay. This increase was quite similar for azimuths arriving from the left side (R_{225}, R_{270}, R_{315}) and diotic conditions (R_0 and R_{180}) and amounted to on average 0.6 dB between delays of 10 and 50 ms, and 3.9 dB between delays of 50 and 200 ms. When the source of the reflection was at the same side as the noise source (R_{45},R_{90},R_{135}), SRTs were equal to SRTs measured with a frontal reflection (R_0) at delays of 10 and 50 ms, and

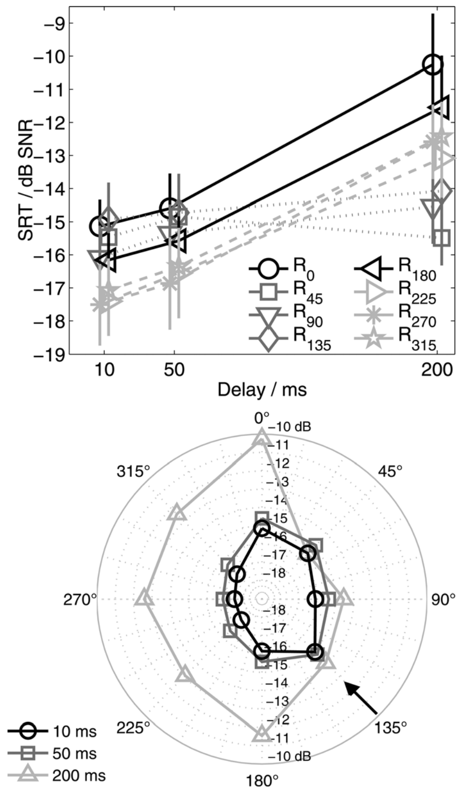

FIG. 6. Same data representation as in Fig. 5, but for lateral noise instead of diffuse noise. The arrow at an azimuth of 135° indicates the direction of the noise source.

did not increase compared to other reflection azimuths when the delay was increased to 200 ms.

A one-way ANOVA was conducted separately for each delay to investigate the influence of the reflection azimuth on speech intelligibility. For a delay of 10 ms, SRTs for an azimuth of 0° were significantly higher than for all azimuths from the left side of the listener (R_{225}, R_{270}, R_{315}). SRTs obtained for azimuths at the right side of the listener (R_{45}, R_{135}) differed significantly from all azimuths on the left side. Exactly the same trends were observed for the 50-ms delay. For a delay of 200 ms, the relations between azimuths were different. SRTs for a frontal reflection were statistically higher than for all other azimuths (except an azimuth of 180°). The location of the reflection in the left hemisphere (R_{225}, R_{270}, R_{315}) had no longer such a beneficial effect on SRTs, which were statistically lower than SRTs measured with a frontal reflection but higher than SRTs for reflections arriving from the right hemisphere. The lowest thresholds at a delay of 200 ms were obtained for reflection azimuths on the right side not co-located with the noise source (R_{45}, R_{90}). The difference in SRT between frontal and spatially separated reflections located at the same side as noise source and a delay of 200 ms was on average 4.4 dB.

2. Discussion

a. Diffuse noise. A spatially separated reflection (without any preference between left and right due to the diffuse noise interferer) resulted in SRTs very close to those obtained with a frontal reflection for short delays of 10 and 50 ms. This indicates that a spatially separated reflection was integrated with the direct sound in the same way as a frontal reflection and that the temporal integration did not depend on the direction of the single reflection. These data measured in diffuse noise at short reflection delays are in line with the independence hypothesis, but are in contrast to data of Arweiler and Buchholz (2011). They measured higher intelligibility scores when 20 early reflections arriving within 55 ms after the direct sound were presented from the front than when they arrived from their original directions distributed around the listener. Only 25% to 34% of the reflection energy was useful for speech intelligibility when reflections were presented spatially distributed and about half of the reflection energy was useful for frontal presentation. Arweiler and Buchholz (2011) showed that most of the differences between spatially separated and co-located reflections could be explained by spectral differences indicated by the speech-weighted SNR. The remaining discrepancy may be related to the fact that the binaural auditory system is unable to integrate more than a single or a low number of spatially distributed reflections. However, further studies are needed to test this hypothesis.

For a long delay of 200 ms and the spatially separated reflections no deterioration was observed, i.e., the increase in SRT did not exceed 3 dB. This could be explained by the binaural system's ability to separate and suppress the late reflection from the target object if this reflection originates from a different direction than the target. The same release from the deterioration effect could be observed as in Expt. IV. The differences in SRTs at long delays between the conditions with spatially separated and frontal reflection could not be explained by analysis of speech-weighted SNR (differences < 0.9 dB). Therefore, the suppression effect observed for late reflections arriving from a different direction as the direct sound seems to be related to binaural processing.

Further investigations should be made with more than one reflection (of the speech signal as well as of the noise signal) and taking into account spectral changes due to the absorption of the walls, ceiling, or floor in order to generalize the present outcomes.

b. Lateral noise. The results of Expt. VII can be interpreted as follows.

For the diotic reflection conditions (i.e., R_0 and R_{180}, solid black lines in the top panel of Fig. 6), the same dependence on the reflection delay was observed as in Expt. III. An advantage of approximately 1 dB was observed for the reflection coming from 180° as compared to 0°. This indicates a small, non-significant front-back advantage, which could not be explained by differences in speech-weighted SNR between the two conditions (advantage only 0.1 dB).

For the cases of contralateral reflections (i.e., R_{225}, R_{270}, R_{315}, dashed light gray lines in the top panel of Fig. 6), a parallel shift to the diotic case was observed, indicating a similar integration of the early single reflection as in the diotic case. The considerable advantage of about 2 dB as compared to the $S_0R_0N_{135}$ condition suggests that speech intelligibility in lateral noise can be improved when a reflection of the speech signal is spatially separated from the direct sound and arrives from a different side of the head as the noise source. Because no differences in speech-weighted SNR were observed between a condition with spatially separated reflections arriving from the left side and a frontal reflection, this effect seems not to be related to any spectral differences. This binaural advantage also holds for delays up to 200 ms where obviously the deterioration effect of a late reflection is overruled in a similar way as already observed in Expt. IV (see Sec. III C 1).

In the conditions with an ipsilateral reflection (i.e., R_{45}, R_{90}, R_{135}, dotted dark gray curves in the top panel of Fig. 6), a comparatively flat function (i.e., no dependence on the delay between early reflection and direct sound) was observed, with SRTs closed to the diotic condition (R_0) at short delays and significantly lower than for a frontal reflection at long delays. This indicates the absence of deterioration at a late reflection delay. At delay of 200 ms, thresholds for ipsilateral reflections were on average 2.0 and 4.4 dB lower than SRTs for contralateral azimuths and a frontal reflection, respectively. This suggests that the spatial separation of direct sound and reflection leads to a suppression of detrimental effects. However, this suppression effect depends on reflection azimuth and is greater for azimuths located on the same side of the listener's head as the noise source. This is in line with observations of Peissig and Kollmeier (1997) made in anechoic conditions. They observed spatial release from masking of 3.5 to 6.2 dB for two interferers located at the same side of the listener and from 2 to 3 dB when the interferers arrived from two different hemispheres. Our analysis revealed no differences or even better speech-weighted SNRs for conditions with a frontal reflection or contralateral reflection compared to the conditions with ipsilateral reflection indicating that the suppression of detrimental effects due to late reflections seems to be solely related to binaural processing. Furthermore, because thresholds for short delays (10 and 50 ms) and azimuths on the side of the noise source remained unchanged compared to the diotic condition, the binaural suppression mechanism seems to be independent of the temporal integration of early reflection.

IV. IMPLICATION FOR MODELS OF SPEECH INTELLIGIBLITY

Speech intelligibility models can be useful for various practical applications such as room acoustical design, audiological diagnostics, or as a tool for the evaluation of signal enhancement algorithms. Their particular advantage is the fast and reproducible estimation of speech intelligibility, in contrast to time-consuming subjective listening tests for which larger numbers of subjects are required to obtain reliable results. However, the applicability of models is generally limited to conditions in which they have been validated. In room acoustics, for example, measures based on useful and detrimental portions of the room impulse response (RIR) are widely applied to characterize the effects of reverberation on sound perception (e.g., Bradley, 1986; ISO, 2009). All of these measures (e.g., clarity, definition, useful-to-detrimental ratio; see Bradley, 1986) are based on the assumption that early reflections enhance speech intelligibility while late reflections cannot be used to recognize speech. The point in time that separates early and late parts varies between studies, and is mostly assumed to be in the range of 50 to 100 ms after the direct sound (Bradley, 1986; ISO, 2009; Rennies *et al.*, 2011). The measures are calculated such that the first part of the RIR is treated as fully useful, and the late part of the RIR is considered fully detrimental. In practice, many RIRs roughly follow a smooth, exponential decay. For such RIRs the exact separation time is not crucial. The RIRs of the present study, however, were considerably different consisting of strong individual components. For such RIRs, a sharp separation between early and late parts may be problematic. Depending on the exact limit for the useful part of the RIR, the measures would consider the entire (for reflection delays below the limit) or half of the speech signal (for larger delays) as useful. The predicted effects on speech intelligibility would therefore be rather "binary." This is at odds with the results of the present study, which showed that there is no single fixed point in time which can separate signal energy into either fully useful or fully detrimental, when a single, strong reflection is used. Instead, the data suggest that up to at least 25 ms, the reflection energy arriving from the same direction as the direct sound could be fully integrated with the direct speech signal. For intermediate delays between about 25 and 100 ms, there was a time window in which the reflection energy was still partially useful for speech intelligibility. For reflections arriving later than 100 ms after the direct sound, a detrimental effect of the reflection co-located with the direct sound was observed. This suggests that a smooth transition between early and late part of the RIR rather than a fixed limit should be employed to account for the present data.

In addition, the above-mentioned room acoustical measures are essentially monaural or single-channel measures. The present data show that such measures cannot explain all observed effects, not even when accounting for interaural differences in the spectra as indicated by the speech-weighted SNR. Instead, models accounting for binaural processing are required. Several such models were described in the literature and could successfully predict speech intelligibility in various listening scenarios (e.g., vom Hövel, 1984; Beutelmann and Brand, 2006; Lavandier and Culling, 2010; Beutelmann *et al.*, 2010; Lavandier *et al.*, 2010; Rennies *et al.*, 2011). Some of these models account for useful and detrimental parts of the (binaural) RIRs (e.g., vom Hövel, 1984; Lavandier *et al.*, 2010; Rennies *et al.*, 2011) and are therefore generally applicable for reverberant speech. However, the separation into useful and detrimental parts was always inspired by room acoustical measures and, consequently, was based on a sharp separation limit. Further investigations are required to test if the suggested models can account for the data of the present study, e.g., when a smooth transition is introduced.

A second property of the mentioned binaural models worth investigating is the interaction of the binaural processing stage and the stage mimicking the effects of reverberation. In some of the models (e.g. vom Hövel, 1984; Rennies *et al*., 2011) the binaural front end is comparatively independent of the speech intelligibility prediction stage at the backend [which describes the sensitivity to the (monaural) reverberation process in a more or less appropriate way]. However, such an approximate independence between binaural processing and reverberation processing seems to be valid only for the case of the single reflection from the front as discussed in Expts. I to III. Therefore, the suppression effect observed in our study for spatially separated reflections arriving 200 ms after the direct sound would not be predicted by these models and hence provides a challenge for them. The data of the present study may serve as a basis to model the interaction between temporal processing and spatial (un)masking within future versions of current models of binaural speech intelligibility.

V. CONCLUSIONS

(i) For a single, frontal reflection of the speech signal, a full temporal integration is observed yielding a constant intelligibility up to a delay of at least 25 ms. The integration gradually decays at longer delays. A deterioration effect is observed at a long delay of 200 ms. This pattern is similar for a frontal noise source, a lateral noise source, and diffuse noise, although a considerable spatial unmasking is found for the non-frontal interferers. Spatial unmasking does not depend on the delay of the frontal speech reflection.

(ii) Using a lateral, localized noise source, intelligibility depends on the direction of the reflection of the speech relative to the noise: Reflection azimuths different from the azimuth of the noise source and the direct speech result in SRTs close to SRTs with a frontal reflection for short delays, but a suppression of the detrimental reflection occurs for long delays. This suppression effect depends on reflection azimuth, and leads to best speech intelligibility for late reflections arriving from a similar direction as the noise source.

(iii) For a single reflection of the speech not co-located with the direct speech source and a diffuse noise masker, temporal integration of the reflection is similar to that observed for a frontal reflection at short delays. A release from the deterioration effect is observed at a delay of 200 ms, possibly because the auditory system is capable of blocking out the direction of the detrimental reflection.

(iv) Even though the condition with one reflection employed here has only limited relevance for real acoustic scenarios, the gradual decrease of temporal integration of the reflection, the deterioration effect and its release for late reflections provide a challenge for current models of speech intelligibility in real environments.

ACKNOWLEDGMENTS

This study was supported by the Ministry of Science and Culture of Lower Saxony, Germany (PhD program "Hearing," Fraunhofer Project Group Hearing, Speech and Audio Technology), and the Deutsche Forschungsgemeinschaft (SFB/TRR 31). Three anonymous reviewers are kindly acknowledged for all their constructive remarks on a previous version of this manuscript, which have improved its quality considerably.

[1]Several studies (e.g., vom Hövel, 1984; Kondo, 2012) found that speech intelligibility in binaural listening conditions did not differ between measurements using individual and non-individual HRIRs. Thus, we assume that results presented here are valid.

ANSI (**1997**). S3.5-1997, *Methods for the Calculation of the Speech Intelligibility Index* (Acoustical Society of America, New York).

Arweiler, I., and Buchholz, J. M. (**2011**). "The influence of spectral characteristics of early reflections on speech intelligibility," J. Acoust. Soc. Am. **130**, 996–1005.

Beutelmann, R., and Brand, T. (**2006**). "Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners," J. Acoust. Soc. Am. **120**, 331–342.

Beutelmann, R., Brand, T., and Kollmeier, B. (**2010**). "Revision, extension, and evaluation of a binaural speech intelligibility model," J. Acoust. Soc. Am. **127**, 2479–2497.

Bilsen, F. A. (**1968**). "On the interaction of a sound with its repetitions," Ph.D. thesis, Delft University of Technology, The Netherlands, pp. 1–93.

Blauert, J. (**1971**). "Localization and the law of the first wavefront in the median plane," J. Acoust. Soc. Am. **50**, 466–470.

Bradley, J. S. (**1986**). "Predictors of speech intelligibility in rooms," J. Acoust. Soc. Am. **80**, 837–845.

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.

Brand, T., and Kollmeier, B. (**2002**). "Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests," J. Acoust. Soc. Am. **111**, 2801–2810.

Fielder, L. D. (**2001**). "Practical limits for room equalization," *111th AES Convention*, New York, preprint 5481.

George, E. L. J., Goverts, S. T., Festen, J. M., and Houtgast, T. (**2010**). "Measuring the effects of reverberation and noise on sentence intelligibility for hearing-impaired listeners," J. Speech Lang. Hear. Res **53**, 1429–1439.

Greenberg, J. E., Peterson, P. M., and Zurek, P. M. (**1993**). "Intelligibility-weighted measures of speech-to-interference ratio and speech system performance," J. Acoust. Soc. Am. **94**, 3009–3010.

Halmrast, T. (**2001**). "Sound coloration from (very) early reflections," *Meeting of the ASA*, Acoustical Society of America, June 4, Chicago, IL.

Hartung, K., and Trahiotis, C. (**2001**). "Peripheral auditory processing and investigations of the 'precedence effect,'" J. Acoust. Soc. Am. **110**, 1505–1513.

ISO (**2009**). ISO 3382-1:2009, Acoustics—Measurement of room acoustic parameters–Part 1: Performance spaces (International Organization for Standardization, Geneva, Switzerland).

Kondo, K. (**2012**). *Subjective Quality Measurement of Speech: Its Evaluation, Estimation, and Applications* (Springer, Berlin), Chap. 6, pp. 47–73.

Lavandier, M., and Culling, J. F. (**2010**). "Prediction of binaural speech intelligibility against noise in rooms," J. Acoust. Soc. Am. **127**, 387–399.

Lavandier, M., Culling, J. F., and Jelfs, S. (**2010**). "Prediction of reverberant speech intelligibility against multiple noise interferers in rooms: Binaural useful-to-detrimental ratios (A)," J. Acoust. Soc. Am. **128**, 2361.

Lochner, J., and Burger, J. (**1964**). "The influence of reflections on auditorium acoustics," J. Sound Vib. **1**, 426–454.

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

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

Peissig, J., and Kollmeier, B. (**1997**). "Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners," J. Acoust. Soc. Am. **101**, 1660–1670.

Plomp, R., and Mimpen, A. M. (**1981**). “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**, 325–328.

Rennies, J., Brand, T., and Kollmeier, B. (**2011**). “Prediction of the influence of reverberation on binaural speech intelligibility in noise and in quiet,” J. Acoust. Soc. Am. **130**, 2999–3012.

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

Steeneken, H. J. M., and Houtgast, T. (**1980**). “A physical method for measuring speech-transmission quality,” J. Acoust. Soc. Am. **67**, 318–326.

vom Hövel, H. (**1984**). “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),” Doctoral dissertation, RWTH Aachen, Germany.

Wagener, K., Brand, T., and Kollmeier, B. (**1999a**). “Entwicklung und Evaluation eines Satztests für die deutsche Sprache II: Optimierung des Oldenburger Satztests (Development and evaluation of a German sentence test II: Optimization of the Oldenburg sentence test),” Z. Audiol. **38**, 44–56.

Wagener, K., Brand, T., and Kollmeier, B. (**1999b**). “Entwicklung und Evaluation eines Satztests für die deutsche Sprache III: Evaluation des Oldenburger Satztests (Development and evaluation of a German sentence test III: Evaluation of the Oldenburg sentence test),” Z. Audiol. **38**, 86–95.

Wagener, K., Kühnel, V., and Kollmeier, B. (**1999c**), “Entwicklung und Evaluation eines Satztests für die deutsche Sprache I: Design des Oldenburger Satztests (Development and evaluation of a German sentence test I: Design of the Oldenburg sentence test),” Z. Audiol. **38**, 4–15.

Yost, W. A., and Hill, R. (**1978**). “Strength of the pitches associated with ripple noise,” J. Acoust. Soc. Am. **64**, 485–492.