Acta Acustica 2023, **7**, 11

https://doi.org/10.1051/aacus/2023005

Available online at:
https://acta-acustica.edpsciences.org

Topical Issue - Auditory models: from binaural processing to multimodal cognition

SCIENTIFIC ARTICLE OPEN ACCESS

Fast processing models effects of reflections on binaural unmasking

Norbert F. Bischof*, Pierre G. Aublin, and Bernhard U. Seeber

Audio Information Processing, School of Computation, Information and Technology, Technical University of Munich, Arcisstrasse 21, 80333 Munich, Germany

Received 2 June 2021, Accepted 28 February 2023

Abstract – Sound reflections and late reverberation alter energetic and binaural cues of a target source, thereby affecting its detection in noise. Two experiments investigated detection of harmonic complex tones, centered around 500 Hz, in noise, in a virtual room with different modifications of simulated room impulse responses (RIRs). Stimuli were auralized using the Simulated Open Field Environment's (SOFE's) loudspeakers in anechoic space. The target was presented from the front (0°) or 60° azimuth, while an anechoic noise masker was simultaneously presented at 0°. In the first experiment, early reflections were progressively added to the RIR and detection thresholds of the reverberant target were measured. For a frontal sound source, detection thresholds decreased while adding early reflections within the first 45 ms, whereas for a lateral sound source, thresholds remained constant. In the second experiment, early reflections were removed while late reflections were kept along with the direct sound. Results for a target at 0° show that even reflections as late as 150 ms reduce detection thresholds compared to only the direct sound. A binaural model with a sluggishness component following the computation of binaural unmasking in short windows predicts measured and literature results better than when large windows are used.

Keywords: Binaural unmasking, Binaural hearing, Dynamic scenes, Reverberation

1 Introduction

In most real-life listening situations, we are not only receiving the direct sound, but also reflections of the sound sources as multiple delayed and modified versions – in rooms and also on the street, where sound is reflected off buildings, automobiles and trees [1]. These reflections alter the interaural phase (IPD) and level differences (ILD) of the direct sound as a function of time. Such changes in the interaural cues can be helpful for detecting a target signal, which is based on mainly two components: the better-ear signal-to-noise ratio (SNR) and the binaural masking level difference (BMLD) [2]. Correlation changes have long been known to improve detection of a target sound in noise [3–5]. Usually, the BMLD is calculated for a situation with a diotic noise masker and a dichotically out-of-phase target relative to a reference situation, where noise and target are presented diotically, as first described by Hirsh [6]. On the other hand, also a better monaural SNR at one of the ears caused by the directional dependence of the ear signals can improve detection thresholds in noise [2, 7, 8]. Both mechanisms are frequency dependent. BMLDs, like the sensitivity to interaural phase changes, are more effective at frequencies below 1.5 kHz, whereas better-ear SNR benefits are more pronounced at higher frequencies.

Several studies investigated detection thresholds and BMLDs for different sound sources in the presence of noise and predicted the binaural benefit. Robinson and Jeffress [9] measured BMLDs as a function of interaural correlation of the masker for a 500 Hz tone which was either binaurally in phase (S_0) or antiphasic (S_π). With increasing interaural correlation of the noise masker, BMLDs increased for a phase-shifted signal and decreased for an in-phase signal. van der Heijden and Trahiotis [10] as well as Bernstein and Trahiotis [11] confirmed these findings and showed that the BMLD is independent on center frequency of the signal and on whether binaural information is presented in the temporal fine structure or in the envelope. In these studies, though, interaural parameters were not changed over time.

Bernstein and Trahiotis [12] proposed a cross-correlation model following Colburn [13] using the mean and variance of the interaural correlation to predict the measured detection thresholds. Another model approach to predict the BMLD contribution is the equalization and cancellation (EC) theory [14]. Both ear signals are temporally aligned and scaled so that the interferer can be optimally cancelled. By subtracting both ear signals, the remaining energy describes the binaural benefit of the listener.

A changing interaural correlation over time, e.g. by incoming reflections, also affects the detection of a target signal in noise. Previous studies showed that for time

*Corresponding author: ga69pov@mytum.de

varying interaural cues, the binaural benefit is reduced in the presence of noise, suggesting a sluggish integration process [15–17]. Grantham and Wightman [15] showed that for a sine tone in the presence of a broadband noise masker with modulated IPD, the BMLD decreased with increasing modulation frequency and becomes absent for modulation frequencies above 2 Hz. The noise masker was modulated between binaurally in-phase to binaural antiphasic. A significant reduction in unmasking was already observed for a modulation frequency of 0.5 Hz. Breebaart et al. [18] also investigated the contribution of time varying interaural cues on binaural detection and proposed a model similar to the EC theory to predict measured thresholds. Their model uses the difference in intensity of the left and right ear signals after peripheral processing as a detection variable. It predicts most of their data better than a cross-correlation model.

All previously mentioned models are based on the whole signal and do not evaluate changes over time in a dynamic manner. Only a few models have been proposed for signal detection in temporally changing binaural and reverberant conditions. Breebaart et al. [19] proposed a binaural processing model to predict detection thresholds for time varying interaural conditions using the same temporal resolution to extract interaural intensity and time differences. Their model does not explicitly account for binaural sluggishness which is expected to influence detection thresholds of temporally changing stimuli. Braasch [20] proposed a binaural model for detection in reverberation. It uses a 50 ms Hanning window to extract monaural and binaural contributions before both are added together, but there is no additional component taking sluggishness into account.

Binaural unmasking is seen as one contributor to binaural speech intelligibility in noise and reverberation. Beutelmann et al. [21] used in their speech model the EC block proposed by Durlach [14] along with the monaural SNR to derive the maximally possible unmasking for the given interaural difference. Lavandier and Culling [22] decomposed the binaural advantage into two separate blocks, the better-ear SNR and a BMLD estimation adopted from Zurek et al. [2]. These models, however, estimate the binaural benefit by averaging across the whole signal and by using the full room impulse response (RIR) (e.g. Rennies et al. [23]) and do not specifically take into account temporal information from the incoming reflections. To consider such temporal changes, Vicente and Lavandier [24] recently proposed a speech intelligibility model which estimates the monaural SNR benefit in short time blocks of 24 ms whereas BMLDs are derived in a much longer 300 ms time window to explicitly consider a sluggish behavior of the binaural auditory system.

A coarse temporal consideration of reflections is often done for speech intelligibility, where early and late reflections are considered separately. Early reflections are commonly described to be useful whereas late reflections affect intelligibility detrimentally [25, 26]. Bradley [27] found 80 ms to be the time when reflections turned from useful into detrimental for predicting the loss of speech intelligibility due to reverberation. Warzybok et al. [28] measured speech reception thresholds in the presence of a single reflection of the same level as the direct sound for varying time delays of the reflection. They neither observed a significant difference in speech intelligibility with a single frontal reflection nor with a single lateral reflection for short time delays, concluding that temporal integration of speech information in early reflections with the direct sound is independent of reflection azimuth. For larger delays they observed a moderate decrease in speech intelligibility, suggesting a partial integration of the reflection with the direct sound. For a delay of 200 ms, the detriment in speech intelligibility compared to only direct sound exceeded 3 dB, indicating a deteriorating effect of late frontal reflections on speech intelligibility. Nevertheless, increasing reverberation [29] or modulated noise maskers [23] remain problematic for speech intelligibility models. Although the useful-to-detrimental approach is established in speech intelligibility models, a fixed temporal boundary generic to different room acoustic conditions has been hard to find [30]. Interestingly, little research has addressed the underlying question, what makes reflections useful or detrimental in a reverberant listening situation for binaural detection, a prerequisite for understanding speech in such situations.

In order to bridge the gap between concepts of established detection models and known speech intelligibility models, the current study deliberately goes one step back to investigate more in detail the effects of early and late reflections as well as the sluggishness integration of time varying cues in a pure detection experiment of a reverberant target signal in noise. In contrast to speech intelligibility in complex listening situations, there is no across-frequency integration in a tone-in-noise detection experiment and cognitive effects are minimized. Self-masking of speech due to the temporal smearing of phoneme information by reverberation is not relevant. With this approach, the current study focusses on the fundamental binaural concepts to better understand the perception of sound sources in reverberant situations.

To investigate the contribution of early and late reflections in a classical detection paradigm, two experiments are conducted with a 50 Hz harmonic complex tone centered around 500 Hz and accompanied by simulated reflections of a room as a target signal, and an anechoic noise masker played from a single loudspeaker in the front. Experiments are conducted in an anechoic chamber and stimuli were spatially auralized via the 36 horizontal loudspeakers of the Simulated Open Field Environment (SOFE, v4) [31]. To solely focus on the effect of reflections on target detection, the noise masker was anechoic. The first experiment investigated the contribution of early reflections. Detection thresholds were measured by successively adding more reflections to the direct sound. The second experiment addressed the contribution of late reflections in the same listening environment. Early reflections were successively removed from the room impulse response of the target sound. A modeling approach was investigated which evaluates the BMLD in a short analysis window with sluggishness considered later, conceptually when objects are formed. This is conceptually different to speech intelligibility

models, notably Vicente and Lavandier [24], which explicitly considered sluggishness through a slow evaluation of IPDs by using a large time frame for BMLD estimation. The proposed alternative approach with a sluggish integration only after fast extraction of the BMLDs is able to better predict the measured detection thresholds of the reverberant harmonic complex tone in noise.

2 Experiment 1: Contribution of early reflections to binaural unmasking

2.1 Experimental setup

Experiments were conducted in the SOFE [31] in the anechoic chamber at the Technical University of Munich. The stimuli were presented via the SOFE's 36 horizontally arranged loudspeakers (Dynaudio BM6A mkII, Dynaudio, Skanderborg, Denmark) placed in 10°-spacing. The loudspeakers were mounted on a custom 4.8 m × 4.8 m squared holding frame in a height of 1.4 m. The loudspeaker at 0°, in front of the listener who was centered in the array, had a distance of 2.4 m to the listener's position. Loudspeaker-individual finite-impulse response equalization filters of length 512 taps (at f_s = 44100 Hz, time-shifted in a 1024 taps filter) were used during playback to compensate for the loudspeakers' frequency and phase response and the difference in time-of-arrival across loudspeakers.

2.2 Simulated room configuration

A non-rectangular virtual room was simulated with two different absorption coefficients $\alpha_1 = 0.1$ and $\alpha_2 = 0.5$. Figure 1 illustrates the virtual room including the simulated listener position and the two simulated source positions at 0° and 60° at a distance of 5 m from the virtual listener position facing in the direction of the abscissa. Direct-to-reverberant ratios (DRRs) were derived for the 0° and 60° source position to −11.8 dB and −12.3 dB, respectively, for $\alpha_1 = 0.1$, and to −4.2 dB and −4.9 dB for $\alpha_2 = 0.5$. The reverberation time RT_{60}, was 736 ms and 302 ms for α_1 and α_2, respectively. In the room simulation, only specular reflections were simulated. To avoid standing waves and strictly repetitive reflection times, the room corners were shifted by up to 50 cm from a rectangular configuration, avoiding parallel walls, which results in a more natural temporal and spatial jittering of the room reflections. This approach makes stimuli reproducible and the contribution of specific reflections interpretable since the impulse response remains deterministic. The exact corner coordinates are listed in Table A1 of the appendix.

All surfaces of the room were covered with the same theoretical material, having either an absorption coefficient of 0.1 or 0.5 for each octave frequency band from 125 Hz to 4 kHz. Room impulse responses (RIRs) were generated using the SOFE [31, 32], which is based on the image source method [33]. Specular reflections were simulated up to the 100th order while all image sources with more than seven invisible parents in a row or a level 80 dB below the direct sound were ignored. For the first experiment, RIRs for two

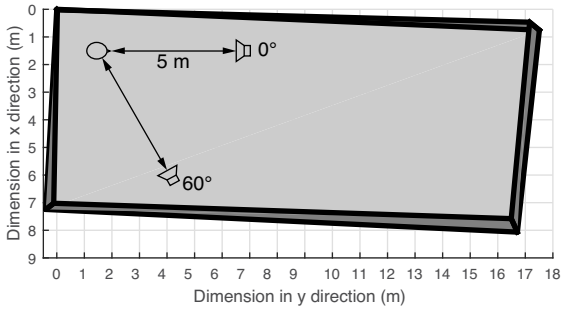

Figure 1. Sketch of the simulated room in topview with source and receiver positions. The room is 3 meters high. All room corners were shifted by up to 50 cm to prevent strictly parallel walls in the room. This avoids standing waves and strictly repetitive reflection times, and thus introduces a more natural temporal jittering of the room reflections. The receiver was located near the corner in the figure's top left with a distance of 1.5 m from the walls and at a height of 1.4 m, corresponding to approximately the seating height of a person. Both simulated sound sources had a fixed distance of 5 m to the listener's position. The condition with the frontal target will be denoted as S_0, the one at 60° as S_{60}.

absorption coefficients ($\alpha_1 = 0.1$ and $\alpha_2 = 0.5$) and two source positions (0° and 60°) were generated. To test the effect of early reflections, the RIRs were truncated after 15 ms (only direct sound), 20 ms, 45 ms, 75 ms, 150 ms, 250 ms, and 500 ms. The direct sound started at approximately 14.5 ms since the sound propagation time of the 5 m distance from the source to the receiver is taken into account. To keep the overall stimulus level constant across conditions, the whole RIRs were scaled for each truncation condition. This results in a decrease of the direct sound when adding reflections, but the ratio between the direct sound and individual reflections is kept the same. One could also interpret this scaling approach as considering the energy of direct sound with reflections as useful, and since it is kept constant, a threshold change indicates a beneficial or detrimental effect of the added reflections irrespective of total target energy.

To illustrate the different modifications of the RIR, Figure 2 shows schematically the truncated RIRs used in the first experiment and the cut RIRs of the second experiment described later.

To reproduce the simulated room over the SOFE's 36 horizontally arranged loudspeakers, the direct sound and all reflections were encoded with 2D 17th-order Ambisonics ([34], p. 61, Eq. (4.19)) and decoded with *max* r_E weighting [35] to maximize the energy vector $\vec{r_E}$ of the sound field. This results in 36 room impulse responses, one for each loudspeaker per tested condition.

2.3 Stimuli

Since BMLDs are known to be more salient at frequencies below 1.5 kHz, a harmonic complex tone consisting of the 7th to 13th harmonic to 50 Hz fundamental frequency (350 Hz to 650 Hz) was used to generate the target stimulus

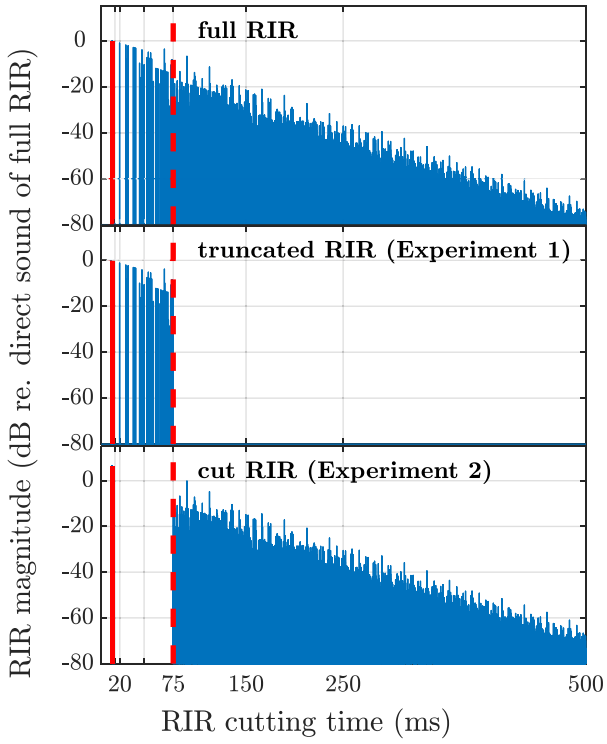

Figure 2. Schematic representation of the complete RIR (top row), the modified RIR of the first experiment (2nd row), truncated after 75 ms, and the modified RIR of the second experiment (bottom row), where reflections were zero'ed after the direct sound and up to 75 ms. In all experimental conditions the direct sound is always present in the RIR and not cut out. RIRs were scaled to keep the overall stimulus level constant across conditions, preserving the ratio between the direct sound and individual reflections.

centered around 500 Hz. This harmonic complex tone was used to excite almost three successive auditory filters according to the Bark scale. Since for truly resolved harmonics, reflections will only affect each harmonic's energy and phase, this harmonic complex tone with unresolved harmonics was chosen to provide envelope fluctuations in each auditory filter. The level of each harmonic was set such that each auditory filter, with a width defined on the Bark scale, received identical energy. The target stimulus was convolved with the truncated room impulse responses for each of the 36 loudspeakers, resulting in 36 loudspeaker signals. As stated above, the level at the listener's position (sum across all loudspeaker channels) of the reverberant signals was then normalized across different truncation conditions by scaling the truncated RIR and keeping the ratio between direct sound and reflections constant. The reverberant harmonic complex tone had an effective duration of 500 ms, defined as the envelope exceeding 90% of its maximum [36], with 10 ms Gaussian rise and fall times.

Uniform exciting noise, which is designed to have the same energy in each auditory filter, was used as masker [37]. The noise was band-limited from 250 Hz to 750 Hz, to ensure masking all components of the harmonic complex tone target stimulus without becoming too loud. It had an overall duration of 900 ms with 30 ms Gaussian rise and fall times. The target stimulus was placed time-centered within the noise masker. The noise source had a sound pressure level of 60 dB at the listener's position. The noise was chosen to be anechoic and not filtered with the room impulse response to avoid interaction by reflections of the noise masker. It was played from a single loudspeaker at 0°, in front of the listener, leading to binaurally highly correlated noise with an interaural correlation coefficient of 0.99. The correlation coefficient was determined from binaural recordings of the noise stimulus at the listener's position with the HMS II.3 artificial head with an anatomically formed pinna (Type 3.3) according to ITU-T P.57 (HMS II, Head acoustics GmbH, Herzogenrath, Germany).

2.4 Participants

Eight participants (3 female) volunteered for the experiment. Participant's age ranged from 21 to 29 years (mean: 25 yr.; SD: 2.3). All participants had normal hearing thresholds with a hearing loss less than 15 dB up to 8 kHz as assessed with a clinical audiometer (Madsen Astera2, GN Otometrics A/S, Taastrup, Denmark). All participants gave written consent and were not paid for participating in the experiment. The study was approved by the ethics committee of the TUM, 65/18S.

2.5 Procedure

The participants sat in the completely darkened anechoic chamber in the center of the loudspeaker array. The detection threshold of the harmonic complex tone in noise was determined with a three-interval three-alternative-forced-choice method (3I-3AFC) using a two-down/one-up adaptive staircase procedure [38] tracking the 71% point of the psychometric function, similar to Kolotzek and Seeber [36]. Participants listened to three intervals of the anechoic uniform exciting bandpass noise, separated by an inter-stimulus-interval of 500 ms. To one of these intervals the reverberant target harmonic complex tone was added. After the stimulus presentation (3.7 s duration), the listeners' task was to indicate which interval differed from the others by pressing the corresponding number on a keyboard. Depending on their response, the overall level of the harmonic complex tone was adjusted. The initial level was set to 65 dB SPL at the listeners' position with an initial step size of 5 dB. After the first reversal, the step size was decreased to 2 dB. From the fourth reversal onwards, it was further decreased to the final step size of 1 dB. Twelve reversals were measured at the final step size and the mean of the last ten reversals was used to calculate the detection threshold of the harmonic complex tone in noise.

The experiment was blocked by the absorption coefficient α. The order of the blocks was randomized between subjects. The combination of used RIR truncation time and target location was randomized within each block. Before a new random test condition started, the previous one had to be finished (blocked by track), i.e. tracks were not interleaved to avoid potential issues with spatial

attention due to the changing target location. Each subject finished one track for each condition, completing the 28 tracks on average in 2 hours.

2.6. Results

Medians and quartiles of the measured thresholds for both absorption coefficients ($\alpha_1 = 0.1$ and $\alpha_2 = 0.5$) and both source positions (0° and 60°) are shown in Figure 3. For a sound source positioned at 0° in front of the listener, thresholds decrease with an increasing number of reflections, which suggests that adding early reflections helps to detect the harmonic complex tone from the front in noise. A similar behavior can be seen for both absorption coefficients. Even when adding only a few early reflections (e.g. truncation after 20 ms), thresholds decrease by more than 5 dB compared to only the direct sound (truncation after 15 ms). Interestingly, such an improvement with increasing number of reflections cannot be observed for a target sound source at 60°. Here, thresholds are 15 dB lower for only the direct sound compared to a target positioned at 0°, because of spatial masking release. When adding early reflections, there is no additional benefit. A slight negative effect can be observed when adding reflections later than 150 ms. Here, thresholds for both absorption conditions increase by 1 to 2 dB and a similar behavior for both absorption coefficients can be observed.

Repeated measures analysis of variance (rmANOVA) with target position, absorption coefficient and truncation as within-subjects variables was performed on the measured data. In the following, p-values together with partial eta-squared (η_p^2) values as an effect size measure are given for all significant effects. The main effects of target position [$F(1, 7) = 1027$, $p < 0.001$; $\eta_p^2 = 0.99$] and truncation [$F(6, 42) = 38$, $p < 0.001$; $\eta_p^2 = 0.85$], and the two-way interactions of position and truncation [$F(6, 42) = 113$, $p < 0.001$; $\eta_p^2 = 0.94$] and of position and absorption [$F(1, 7) = 13$, $p < 0.01$; $\eta_p^2 = 0.65$] and the three-way interaction [$F(6, 42) = 2.7$, $p < 0.05$; $\eta_p^2 = 0.28$] are significant. Since there is no significant main effect of the absorption coefficient and only the two-way interaction of position and absorption is significant, but not the interaction of absorption and truncation, and since the effect size measure of the three-way interaction is small with $\eta_p^2 = 0.28$ compared to the other effects, the different absorption coefficients seem not to affect the binaural benefit. The significant interaction of absorption and position can be explained by the difference in thresholds for short truncations between the two different target positions (see Fig. 3 solid versus dashed lines for truncation 15 ms to 45 ms). To further analyse the interactions, a two-tailed t-test post-hoc analysis with Tukey-Kramer correction was performed. For a sound source at 60°, no pairwise comparison of the different truncation times reaches significance for both absorption coefficients, which suggests that there is no further unmasking benefit from the reflections for a lateral target position.

For the target position at 0°, Tukey-Kramer corrected two-tailed t-test pairwise comparisons show a significant difference between 15 ms and all other truncation times ($p < 0.001$), between 20 ms and all other truncation times ($p < 0.05$) and between 45 ms and 150 ms ($p < 0.05$). No other combination reaches significance. This indicates that the binaural benefit from adding early reflections for a target position at 0° increases up to a truncation time of about 45 ms. Adding later reflections after 150 ms will not further improve the detection of the target harmonic complex tone in noise from the front.

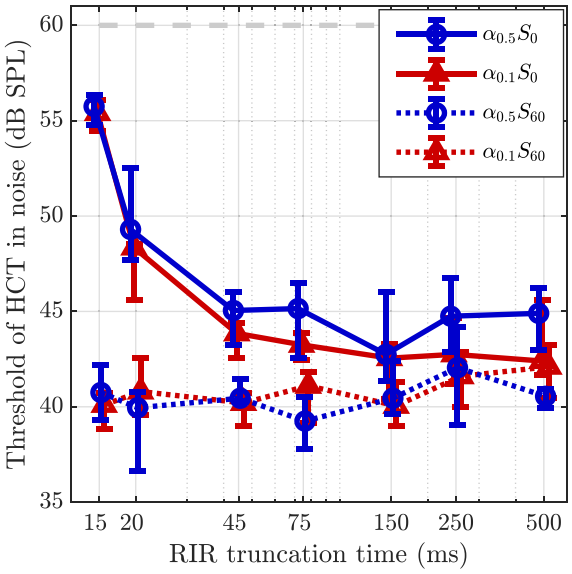

Figure 3. Measured binaural detection thresholds of a reverberant harmonic complex tone for different truncations of the RIR in the presence of an anechoic bandpass noise with 60 dB SPL from the front are shown (experiment 1). With increasing RIR truncation time the amount of late reflections increases, while a 15 ms window corresponds to only the direct sound without any reflections. Solid lines indicate thresholds for a source at 0°, dashed lines for a source at 60°. Blue circles show the median thresholds of the tested participants for an absorption coefficient of 0.5 and red triangles for an absorption coefficient of 0.1. Errors are given as upper and lower quartiles.

3 Experiment 2: Unmasking in the absence of early reflections

The aim of the second experiment is to focus on the effect of late reflections on binaural unmasking of a target sound source in noise. It was shown for speech intelligibility that late reflections can harm the intelligibility [22, 26, 28]. These studies found that reflections arriving within the first 80 to 100 ms after the direct sound can be integrated with the direct sound, whereas later reflections will rather harm intelligibility and can be therefore interpreted as being energetically added to the masking background noise. The main question in this experiment is whether late reflections will also hinder the simple detection of a reverberant target tone in the presence of noise and if also here late reflections will add additional energy to the masking signal. The experiment was similar to the first one, but early reflections were

increasingly removed from the RIR while late reflections were kept along with the direct sound.

3.1 Stimuli

The second experiment used the same room and absorption coefficients, but only the target source position at 0° in front of the listener since there was no change in threshold for a source positioned at 60°. In contrast to the first experiment, early reflections were removed from the RIR so that, besides the direct sound, only reflections after a certain time were kept. These times correspond to the same truncation times as in experiment 1 (15 ms, 20 ms, 45 ms, 75 ms, 150 ms, 250 ms, and 500 ms) with all reflections between the direct sound and the truncation time being removed from the RIR. Therefore, 500 ms corresponds to only the direct sound and 15 ms corresponds to the complete impulse response. The longer the cutting time condition, the larger the gap between direct sound and incoming reflections (see Fig. 2 earlier).

The same harmonic complex target stimulus as in experiment 1 was convolved with the cut impulse responses and the level was normalized across different cutting conditions. The noise masker had the same frequency range and duration as in experiment 1.

3.2. Procedure

The same eight volunteers finished the second experiment in about 1 hour. The experimental procedure followed that of experiment 1. Trials were blocked by the absorption coefficient and randomized between subjects. Within each block, RIR truncations were randomized, but tracks were not interleaved. Each subject finished one track for each condition, resulting in 14 tracks for each subject.

3.3. Results

Thresholds obtained from the second experiment are summarized in Figure 4. Removing the first early reflections does not seem to have an impact on the thresholds, as they remain fairly constant between 15 ms and 20 ms truncation time for both α. However, as more and more early reflections are removed, thresholds start to increase from 45 ms to 150 ms for $\alpha = 0.5$, and stay constant thereafter on the same level reached by only the direct sound (500 ms). For $\alpha = 0.1$, a different behavior can be observed. Thresholds for 45 ms truncation time decrease first and start to increase for truncation times larger than 150 ms. Unlike in the first experiment, absorption influences measured thresholds, since the truncation time from which thresholds start to increase, is different for both absorptions.

An rmANOVA with absorption coefficient and truncation time as within-subject variables was performed on the measured thresholds. Besides a significant main effect of truncation time [$F(6, 42) = 185$, $p < 0.001$; $\eta_p^2 = 0.96$] also the main effect of absorption coefficient [$F(1, 7) = 334$, $p < 0{,}001$; $\eta_p^2 = 0.98$] and the interaction between truncation time and absorption coefficient [$F(6,42) = 56$, $p < 0.001$; $\eta_p^2 = 0.89$] become significant. In a post hoc analysis with Tukey-Kramer correction, pairwise comparison between $\alpha = 0.1$ and $\alpha = 0.5$ shows no significant difference for only direct sound (500 ms) and for 20 ms. All other truncation times are significantly different between absorption coefficients ($p < 0.05$ for 15 ms, else $p < 0.001$) which suggests that late reflections in more reverberant situations ($\alpha = 0.1$) strongly affect detection thresholds. A pairwise comparison of the measured thresholds between different truncation times shows no significant difference when removing the very first reflections for both absorption coefficients (15 ms vs. 20 ms for $\alpha = 0.1$ and 15 ms vs. 20 ms and 45 ms for $\alpha = 0.5$). The decrease in detection thresholds observed for $\alpha = 0.1$ between 20 ms and 45 ms is significant ($p < 0.001$). For an absorption coefficient of 0.5, very late reflections (truncation times larger than 150 ms) do not change detection thresholds compared to only the direct sound, since thresholds are not significantly different from each other.

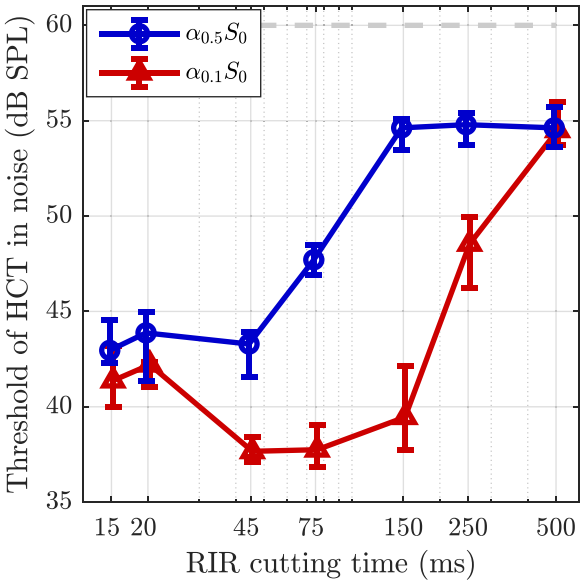

Figure 4. Measured binaural thresholds of a reverberant harmonic complex tone for different time conditions of cut early reflections from the RIR in the presence of an anechoic noise with 60 dB SPL (experiment 2). Both sound sources were colocated at 0°. The blue circles show median thresholds of the tested participants for an absorption coefficient of 0.5, the red triangles for an absorption coefficient of 0.1. Errors are given as upper and lower quartiles.

4 Short vs long window binaural processing for detection of reverberant signals

4.1 *Short window, fast binaural processing model (DynBU$_{\text{fast}}$)*

The aim of the current modeling approach is to predict the overall benefit (binaural unmasking) when detecting a signal with dynamically changing binaural cues over time in noise using a fast BMLD formation (DynBU$_{\text{fast}}$). Starting point of the current approach was the model proposed by

Lavandier and Culling [22], published in the Auditory Modelling Toolbox (AMT) [39]. In the model, the overall binaural benefit is divided into two parts: the better-ear SNR and the BMLD. Both parts are extracted for each critical band separately. The EC formula used in the current model approach was adopted from Lavandier and Culling [22] and is shown in equation (1), where f_i denotes the center frequency of a particular auditory filter, ϕ_T is the interaural phase difference of the target, ϕ_M is the interaural phase difference of the noise masker and ρ_M denotes the interaural coherence of the noise masker:

$$\mathrm{BMLD}(f_\mathrm{i}) = \max\left(k - \frac{\cos\left(\phi_\mathrm{T}(f_i) - \phi_M(f_i)\right)}{k(f_i) - \rho_M(f_i)}\right) \quad (1)$$

$k(f_i)$ can be derived by $k(f_i) = \left(1+\sigma_\epsilon^2\right)\exp\left((2\pi f_i)^2\sigma_\delta^2\right)$ according to the formula given in Lavandier and Culling [22], with $\sigma_\epsilon = 0.25$ and $\sigma_\delta = 0.105 \times 10^{-3}$ ms [39].

The overall structure of the $\mathrm{DynBU_{fast}}$ model approach is shown in Figure 5a. Both, noise and target signal, are filtered with a Gammatone filter bank, according to the Bark scale, separately for the left and the right ear. The output of the Gammatone filter bank is then split into short 12 ms time frames using a Hanning window ("analysis window") with 50% overlap of successive time frames. The effective window length of the Hanning window is therefore 6 ms measured by exceeding 6 dB of its maximum. The time constants were optimized as described in the next section. For each frequency band and time window, the interaural phase difference of the target and the masker noise as well as the interaural coherence of the noise masker are derived using the interaural cross correlation. The extracted interaural cues are used to compute the BMLD according to equation (1), for each auditory filter and time analysis window. Time frames in which the level of the target signal is below hearing threshold are ignored in order to avoid calculation artefacts during fade in and fade out of the signal. The main difference to former models is that the BMLD contribution is derived in short 12 ms time windows before taking sluggishness into account. In the $\mathrm{DynBU_{fast}}$ approach, an IIR exponential decay filter with a time constant of 225 ms simulating the sluggishness of the auditory system is applied only after formation of the BMLD contribution, i.e. on the short time BMLD values. The time constant of the filter corresponds to the time it takes to drop from 1 to $1/e$ in the impulse response and is derived in the next section. An exponential decay was used to weight recent incoming cues more strongly. Thereafter, the BMLD contribution is transformed to decibels [22].

In addition to the BMLD, the better-ear SNR is derived from the binaural ear signals. Similar to the processing of the BMLDs, the signal-to-noise ratios for both ear signals are computed separately in each short time analysis window and for each auditory filter. The better SNR across both ear signals is chosen. To account for temporal integration, the intensity SNR is also filtered with an exponential integration filter [37] with a time constant of 90 ms (see next section) and transformed to decibels. Both BMLD and better-ear SNR are summed for each time frame and for each critical band, resulting in an overall SNR benefit. To model a simple detection process, the frequency band with the highest overall binaural benefit in each time frame is selected followed by selecting the maximum of the time series. Although the target stimulus was centred around 500 Hz, due to the stimulus covering almost three critical bands, detection could have occurred in any of the three auditory filters.

$\mathrm{DynBU_{fast}}$ as well as $\mathrm{DynBU_{slow}}$ (see Sect. 4.3) are implemented in MATLAB (Mathworks, Natick, MA) and are available together with the data and code to generate all figures of this manuscript at DOI: 10.5281/zenodo.7643249 [40]. $\mathrm{DynBU_{fast}}$ is also available as *bischof2023* and data as *exp_bischof2023* in the AMT [39].

4.2 Estimation of optimal time constants

The optimal combination of the three used time constants, for the short time analysis window, the sluggish integration of BMLDs and the intensity integration, was found by minimizing the root-mean-squared error (RMSE) to the experimental data presented before. The RMSE to the experimental data from Sections 2 and 3 was computed for every combination of the three time constants in the model: Four short time analysis windows (from 6 to 48 ms), 63 sluggishness time constants (from 10 ms to 350 ms) and 43 intensity integration time constants (from 10 ms to 250 ms). Figure 6 shows the RMSE for all tested combinations of sluggishness and intensity integration separately for each analysis window size. The optimal combination of the three parameters was chosen by finding the local minimum of the RMSE across all parameters (crosses in Fig. 6). The lowest RMSE was found for an analysis window of 12 ms in combination with a sluggishness time constant of 225 ms and an intensity integration of 90 ms, resulting in an RMSE of 1.33 dB. With longer analysis windows, the RMSE increases to 1.88 or 2.43 dB for 24 and 48 ms, respectively. With an analysis window of 6 ms in combination with a sluggishness time constant of 295 ms and an intensity integration of 220 ms the RMSE slightly increased to 1.39 dB. The optimized time constants are already included in Figure 5.

4.3 Long window, slow binaural processing approach ($\mathrm{DynBU_{slow}}$)

The $\mathrm{DynBU_{fast}}$ approach is compared to a slow binaural processing model ($\mathrm{DynBU_{slow}}$), which differs only in a few details. $\mathrm{DynBU_{slow}}$, shown in Figure 5b, uses two different time frames, a fast 12 ms frame to extract the better-ear SNR identical to the $\mathrm{DynBU_{fast}}$ approach, and a 225 ms frame to compute the BMLD contribution directly after the Gammatone filter bank. Since the temporal integration is already done by computing the signal level in the longer BMLD frame, no additional sluggishness filter is applied after extracting the BMLD. The final threshold prediction is identical to the $\mathrm{DynBU_{fast}}$ approach.

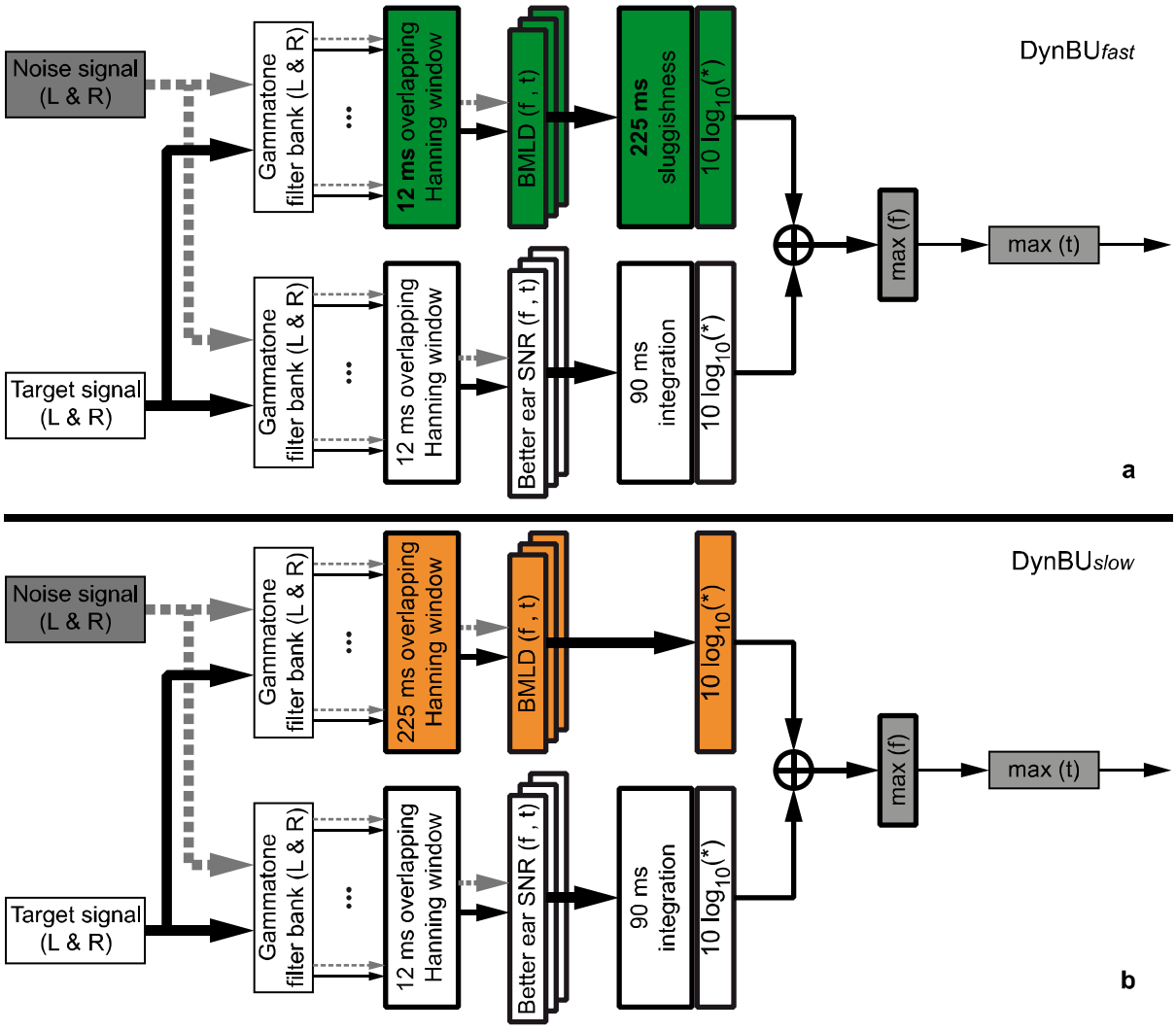

Figure 5. Block-diagram of the short-window, fast processing approach ($\text{DynBU}_{\text{fast}}$) is shown in the top panel a. The left and right ear signals are first bandpass filtered using a Gammatone filter bank parametrized along the Bark scale. The time signal of each filter is windowed with 12 ms overlapping Hanning windows, resulting in an effective window length of 6 ms. The interaural cross-correlation of the interferer (ρ_i) as well as the interaural phase difference of target and interferer (ϕ_t and ϕ_i) are extracted for each filter and each time window to calculate binaural unmasking according to formula (1). A 225 ms exponential decay filter is subsequently used to account for sluggishness of binaural processing. Binaural unmasking and the better-ear SNR are added for each frequency band and time frame, followed by selecting the maximum of the binaural benefit across frequency bands per time frame. The binaural benefit for signal detection is estimated by selecting the maximum binaural benefit over time. The $\text{DynBU}_{\text{slow}}$ approach (see Sect. 4.3) differs from $\text{DynBU}_{\text{fast}}$ only by using a 225 ms window directly after the Gammatone filterbank to derive the BMLDs without integration afterwards. The $\text{DynBU}_{\text{slow}}$ approach is shown in lower panel b of the Figure.

4.4 Evaluation

4.4.1 Current experimental conditions

Both model approaches were first evaluated against the above gathered experimental results. In-situ binaural recordings were used as input signals for the model and were normalized to an initial SNR of 0 dB. The signals of the anechoic noise masker and of the reverberant target were recorded with an artificial head at the listener's position in the SOFE (see Methods). The model predictions for all tested RIR conditions and source positions are shown together with the experimental results in Figure 7.

Predictions with the $\text{DynBU}_{\text{fast}}$ approach follow the measured data well across almost all conditions. The data from the first experiment with collocated target and masker at 0° (panel a and b) can be predicted well with the fast BMLD extraction. The root mean square error (RMSE) of the predictions to the experimental data is 1.14 dB and 1.45 dB for an absorption coefficient of 0.1 and 0.5, respectively. The Pearson's correlation coefficient expresses a high correlation ($\rho = 0.99$; $\rho = 0.96$) for both absorption coefficients. With the $\text{DynBU}_{\text{slow}}$ approach, the overall binaural benefit is underestimated for an absorption coefficient of 0.5. This can also be seen in the high RMSE of 3.84 dB, which is more than twice the RMSE of the

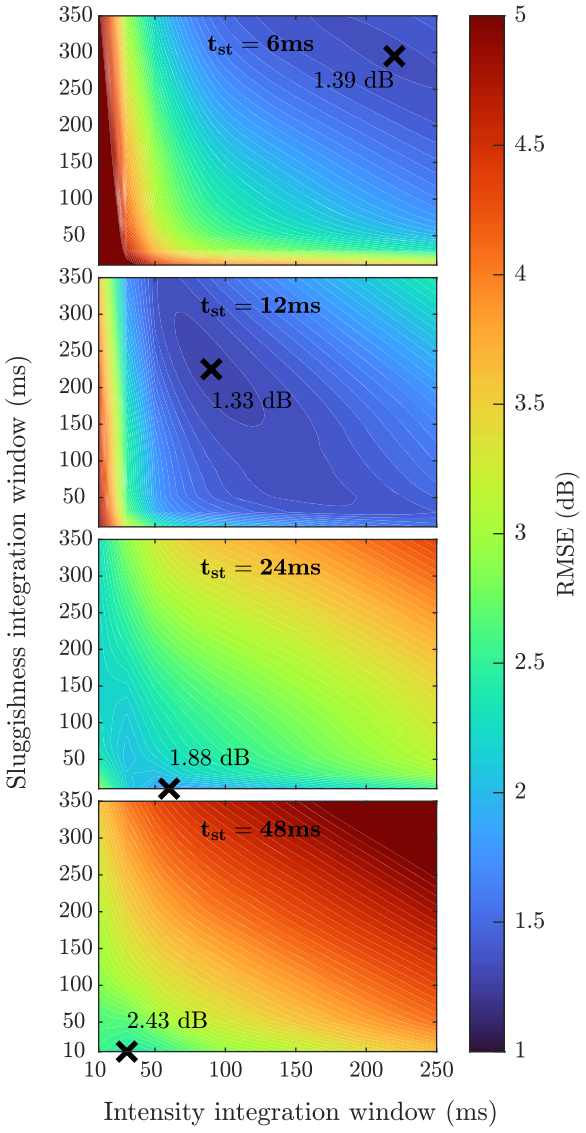

Figure 6. Root-mean-square errors in dB are shown color-coded for all combinations of sluggishness and intensity integration time constants for each tested analysis window duration (see top of each panel). The cross indicates the local minimum along with the corresponding RMSE value.

DynBU$_{fast}$ approach in this condition. For an absorption coefficient of 0.1, the RMSE is 1.80 dB for the DynBU$_{slow}$, slightly higher than for DynBU$_{fast}$. The correlation for DynBU$_{slow}$ vs the data is nevertheless high for both absorption coefficients ($\rho = 0.98$ for $\alpha = 0.1$; $\rho = 0.93$ for $\alpha = 0.5$).

Predictions for the N_0S_0 condition also differ between fast and slow BMLD formation for the second experiment (panel e and f). When early reflections are successively cut out, the difference between a sluggish integration before or after the formation of the BMLD contribution is clearly visible for truncation times larger than 75 ms, especially for higher reverberation ($\alpha = 0.1$). Here, DynBU$_{slow}$ leads to an underestimation of the measured thresholds whereas DynBU$_{fast}$ matches the measured thresholds well. This can also be observed in the RMSE and the correlation of the predictions to the measured data. While the RMSE is 0.90 dB and 1.18 dB for the DynBU$_{fast}$ approach, errors increase for the DynBU$_{slow}$ approach to 2.92 dB and 2.46 dB for $\alpha = 0.1$ and $\alpha = 0.5$, respectively. This is mainly due to the huge underestimation of unmasking for late incoming reflections in the DynBU$_{slow}$ approach. DynBU$_{fast}$ predictions are highly correlated with the measured threshold data ($\rho = 0.99$) for both absorption conditions, whereas the DynBU$_{slow}$ approach shows lower correlation of 0.96 and 0.87 respectively. One reason for the better performance with the DynBU$_{fast}$ approach is that faster interaural correlation changes, caused by late incoming reflections, are established in short time frames and are only averaged afterwards. Fluctuations in the interaural correlation are smeared over time when using a longer time window for BMLD estimation. For a target sound source located at 60° for a frontal noise masker (panels c and d), the overall performance of both model approaches does not differ much. The RMSE is 1.74 dB and 1.45 dB for DynBU$_{fast}$ and 2.05 dB and 1.46 dB for DynBU$_{slow}$ for $\alpha = 0.1$ and $\alpha = 0.5$ respectively. Pearson's correlation coefficients are low for an azimuth of 60° and stay in the range of 0.11 to 0.24 for both model approaches. The low ρ values here can be explained by considering that across truncation time there is no change that can be predicted. The RMSEs and Pearson's correlation coefficients are summarized in Table 1 for both experiments and modelling approaches.

The overall trend and most of the tested conditions can be predicted quite well. The overall average error of the model predictions to the measured data is 1.3 dB for the DynBU$_{fast}$ approach and 2.5 dB for the DynBU$_{slow}$ approach. Some conditions, though, cause difficulties for both approaches: adding only very early reflections to a lateral sound source (panels c and d at 20 ms cutting time) or cutting out early reflections from a frontal sound source (panel f at 20 and 45 ms cutting time) results in an overestimation of the overall binaural benefit with both model approaches. This is likely caused by a better-ear SNR contribution which is discussed next.

4.4.2 Better-ear and BMLD contribution in the DynBU$_{fast}$ approach

To better understand the contributions of the better-ear SNR and the BMLD components, Figure 8 shows them individually for the experimental conditions shown in Figure 7. Values are presented as SNR to show the contribution independent of masker level and to facilitate comparison with the BMLD literature.

For experiment 1 with collocated target and masker at 0° (panels a and b), the BMLD contribution increases from 0 dB for only the direct sound (15 ms) to 12 dB (for $\alpha = 0.1$) when adding early reflections up to 75 ms, or 11 dB for $\alpha = 0.5$. The short-time better-ear SNR dominantly contributes to the detection threshold for only direct sound and very early reflections. The better-ear contribution of 4.1 dB for an approximately N_0S_0 condition indicates a benefit from short-time listening into the gaps.

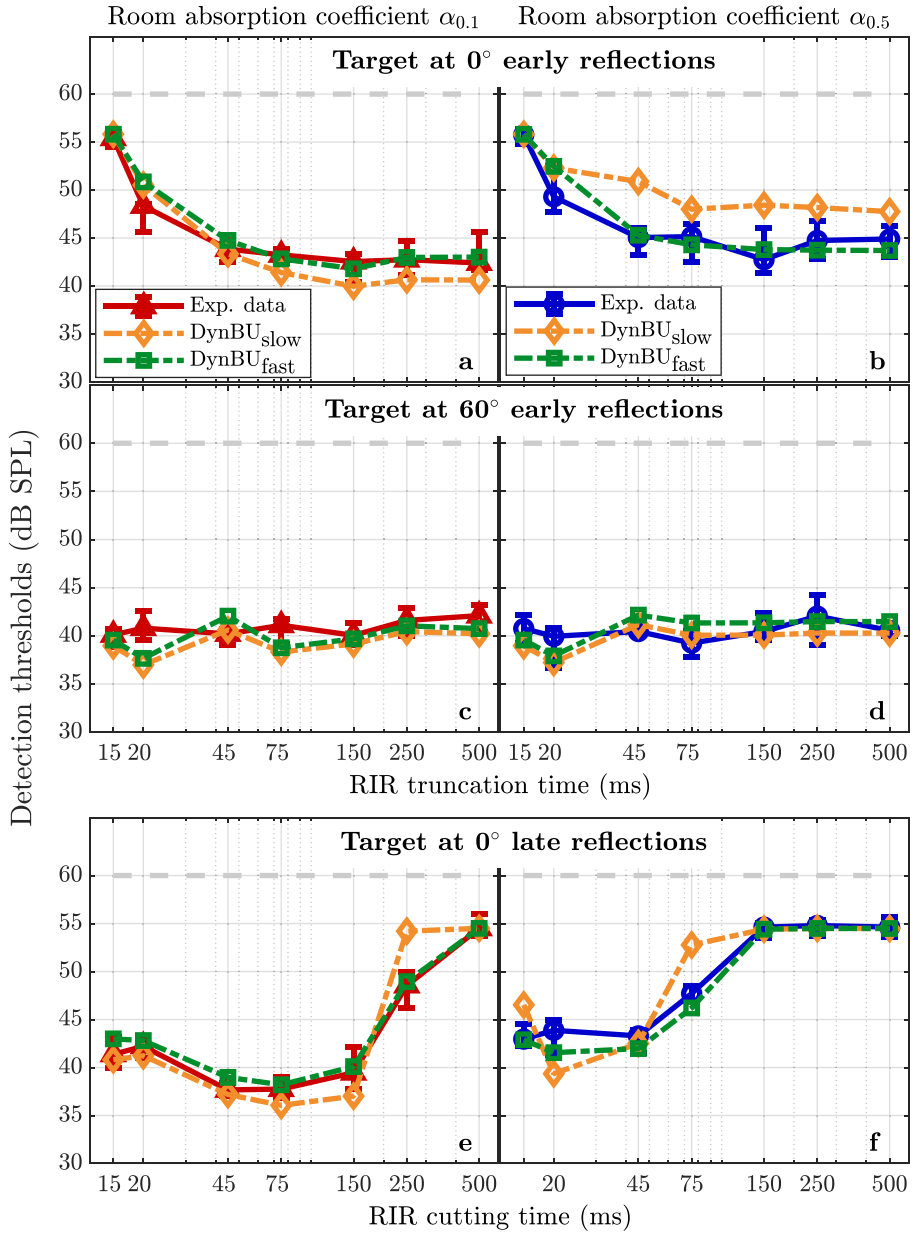

Figure 7. Predictions of the short-window model approach (DynBU$_{fast}$; green dashed lines with squares) with a sluggishness integration of short-time BMLDs are shown along with predictions using a 225 ms window for BMLD extraction (DynBU$_{slow}$; orange dashed lines with diamonds) and measured results. Data are given as a function of cutting time of the room impulse response. The left column (panels a, c and e) shows predictions for an absorption coefficient of $\alpha = 0.1$ and the right column (panels b, d and f) for $\alpha = 0.5$. The first row (panels a and b) shows the prediction for sound source and noise being co-located in the front of the listener (N_0S_0), the second row (panels c and d) for a sound source at 60° (N_0S_{60}). The third row (panels e and f) shows predictions for the second experiment (N_0S_0 with only late reflections). The experimentally measured binaural unmasking (solid lines) are replotted from Figure 2, panels a–d, and Figure 4, panels e and f, for comparison.

For a target sound source at 60° with a frontal noise masker (panels c and d), the overall benefit is dominated by the BMLD contribution of about 14 dB whereas better-ear SNR is on average 6 dB, for all tested conditions. The slight overestimation of the overall detection benefit with very early reflections is caused by the better-ear contribution, while the BMLD contribution stays constant across truncation times.

When early reflections are successively cut out from the full RIR (panels e and f), the overall detection benefit is dominated by the BMLD contribution at least up to a cutting time of 75 ms. For $\alpha = 0.1$ (panel e), the BMLD contribution caused by late reflections, arriving 150 ms after the direct sound, is still larger than the better-ear contribution, but the ratio declines for later arriving reflections. For only the direct sound (cutting time 500 ms) and with late

Table 1. Root-mean-squared errors (RMSE) and correlation coefficients (ρ) of the $\text{DynBU}_{\text{fast}}$ and $\text{DynBU}_{\text{slow}}$ predictions to the experimental data for all tested conditions.

	$\text{RMSE}_{\text{fast}}$	ρ_{fast}	$\text{RMSE}_{\text{slow}}$	ρ_{slow}
Exp. 1, S_0 $\alpha = 0.1$ early reflections	1.14 dB	0.99	1.80 dB	0.98
Exp. 1, S_0 $\alpha = 0.5$ early reflections	1.45 dB	0.96	3.84 dB	0.93
Exp. 1, S_{60} $\alpha = 0.1$ early reflections	1.74 dB	0.11	2.05 dB	0.22
Exp. 1, S_{60} $\alpha = 0.5$ early reflections	1.45 dB	0.21	1.46 dB	0.24
Exp. 2: S_0 $\alpha = 0.1$ late reflections	0.90 dB	0.99	2.46 dB	0.96
Exp. 2: S_0 $\alpha = 0.5$ late reflections	1.18 dB	0.99	2.92 dB	0.87

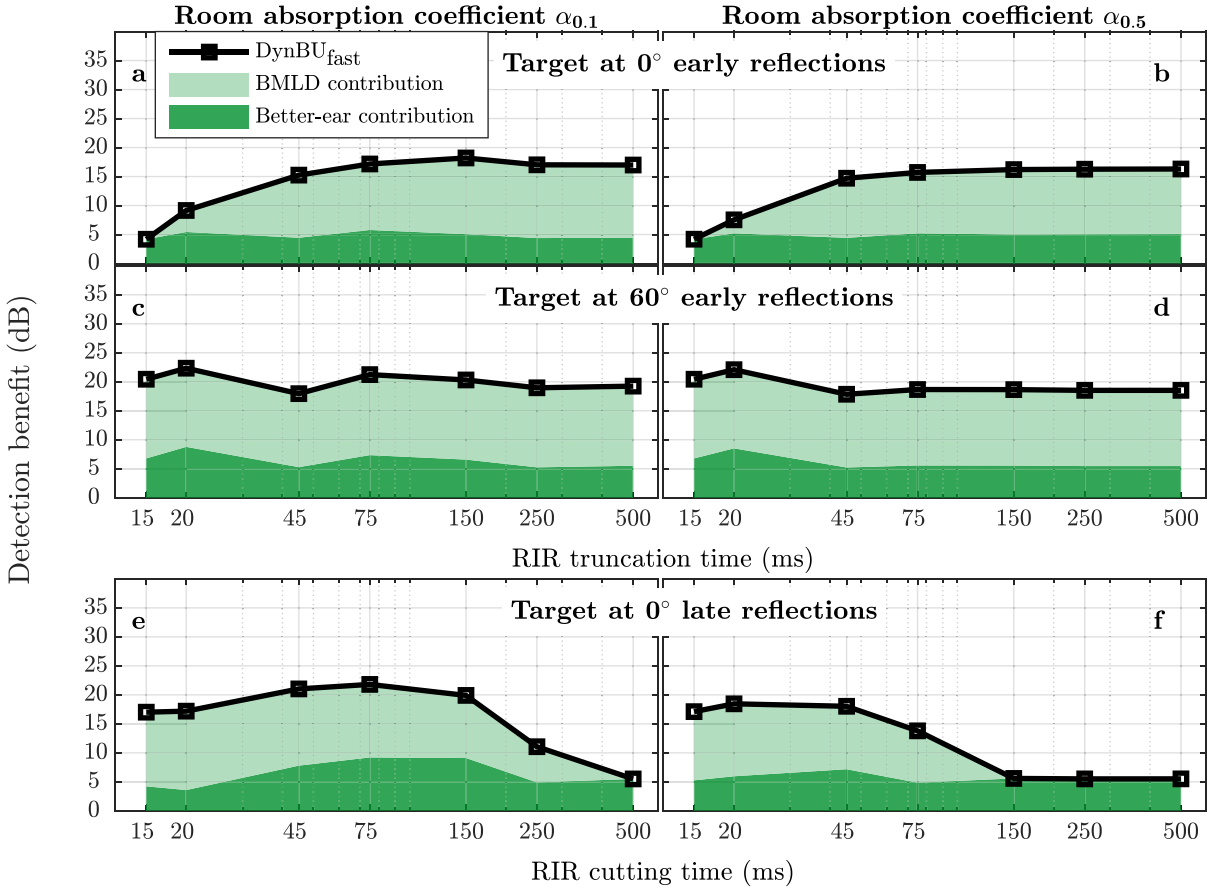

Figure 8. Contributions of better-ear SNR (dark green shaded area) and BMLD (light green shaded area) to the overall predicted binaural benefit of the $\text{DynBU}_{\text{fast}}$ approach (the sum of both contributions, indicated with black squares). The overall prediction is plotted as detection benefit for the different experimental conditions of Figure 7 derived as the difference between masker level (60 dB SPL) and predicted detection thresholds in dB SPL of the target harmonic complex tone. The left column (panels a, c and e) shows data for $\alpha = 0.1$ and the right column (panels b, d and f) for $\alpha = 0.5$.

reflections in the less reverberant situation, the overall benefit is exclusively driven by the better-ear SNR contribution.

The significant decrease in detection threshold by removing early reflections between 20 ms and 45 ms for $\alpha = 0.1$ (panel e) can be traced back to the better-ear SNR contribution, since it increases while the binaural contribution stays fairly constant. The late reflections arriving before 250 ms might carry enough energy to increase the short-time better-ear SNR compared to the full impulse response. Also here, the relative BMLD contribution is higher than the better-ear contribution as long as reflections are present carrying enough energy to decorrelate the target signal.

4.4.3 Evaluation on binaural detection experiments in the literature

To further evaluate the differences between the $\text{DynBU}_{\text{fast}}$ and the $\text{DynBU}_{\text{slow}}$ approach, two additional data sets from the literature were used. Braasch [20] measured detection thresholds of a reverberant broadband

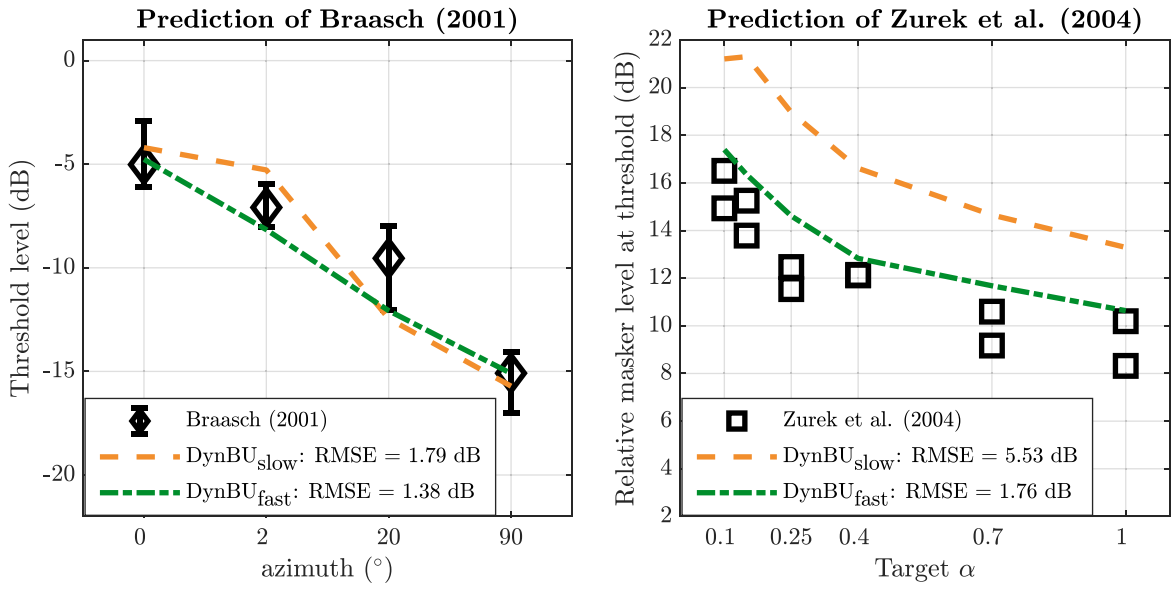

Figure 9. Predictions of the DynBU$_{fast}$ (green dashed-dotted line) and DynBU$_{slow}$ (orange dashed line) modelling approaches for two former detection experiments in reverberant environments. The left panel shows predictions for detection data by Braasch [20] for a reverberant broadband noise target at different azimuth positions in the presence of another broadband noise masker at 0°. The right panel presents predictions of the data of two subjects collected by Zurek et al. [2] for a 3rd-octave bandpass noise target at 0° in reverberant space with different absorption coefficients when an anechoic broadband noise masker was presented from +60°.

noise signal at different azimuth angles, 0°, 2°, 20° and 90°. Another broadband noise was used as masker located at 0°. Both noises, target and masker, had a frequency range of 200 Hz to 14 kHz and were presented from a distance of 2 m to the virtual listener position. A rectangular room (5 m × 6 m × 3 m) was simulated using the mirror image technique [38], but reflections formed temporally repetitive patterns. Measured thresholds of stimuli with all binaural cues available are replotted from Figure 5.9 in Braasch [20] and are shown with the predictions in Figure 9 in the left graph. The thresholds predicted with the DynBU$_{fast}$ approach match the measured data of Braasch [20]. Only for a target source located at 2° or 20°, thresholds are slightly overestimated, but still inside the across-subject variance. The RMSE of the predicted benefit against the provided measured data is 1.38 dB. Figure 9 also shows predictions of the DynBU$_{slow}$ approach. The overall decrease of the binaural benefit with increasing azimuth angle can also be predicted, but overall binaural unmasking is less consistent, resulting in an RMSE of 1.79 dB.

The second data set for comparing both model approaches is taken from a study by Zurek et al. [2]. The room simulated in this study was also rectangular (4.8 m × 6.6 m × 2.6 m), with the virtual listener placed near the middle, 2.8 m from the right wall and 2.5 m from the rear wall. The listener was turned by 20° to the left. They used a 3rd-octave bandpass noise with a center frequency at 500 Hz as target stimulus and a continuous broadband noise as masker. Detection thresholds of the reverberant target at 0° in 1 m distance to the listener were measured in an anechoic noise masker at 60° azimuth and 1 m distance for different absorption coefficients. Binaural room impulse responses were derived with a spherical head model with 8.75 cm head radius. Their threshold data, relative to averaged thresholds measured only presenting to the left or right ear, are replotted from Figure 7e in Zurek et al. [2] and are shown with the model predictions in the right panel of Figure 9. The DynBU$_{fast}$ approach predicts their results across all tested absorption coefficients well with a slight underestimation of the binaural benefit resulting in an overall RMSE of 1.78 dB. The DynBU$_{slow}$ approach captures the trend of a decreasing binaural benefit with increasing absorption coefficient, but errors increase with more reverberation (RMSE = 5.53 dB). Results indicate that using a fast BMLD extraction followed by sluggish integration is beneficial for the prediction in highly reverberant conditions. This is also in line with results from the current study, showing that the DynBU$_{fast}$ approach predicts the benefit caused by late reflections in highly reverberant situations better.

5 General discussion

This study investigated how early and late room reflections affect the detection of a harmonic complex tone in the presence of a noise masker in free-field listening conditions. Almost all former studies conducted their tone-in-noise detection experiments with headphones using HRTFs. In the current study, two experiments were conducted in a simulated room with two different absorption coefficients auralized via multiple loudspeakers in free field. Listeners detected a reverberant harmonic complex tone, centered around 500 Hz and located at 0° or 60°, in an anechoic uniform exciting noise masker presented from the frontal

loudspeaker in the anechoic setup. Experiment 1 focused on the effect of early reflections on detection by subsequently adding reflections to the direct sound of the target, whereas experiment 2 investigated the influence of late reflections by subsequently cutting out early reflections from the full room impulse response. Two modelling approaches were compared, one approach where interaural cues for BMLD computation were extracted on a larger time frame (225 ms; $\mathrm{DynBU_{slow}}$), and a suggestion for a dynamic approach operating on short time frames for BMLD computation with binaural sluggishness taken into account only afterwards ($\mathrm{DynBU_{fast}}$). The $\mathrm{DynBU_{fast}}$ approach excels over the $\mathrm{DynBU_{slow}}$ approach when predicting detection thresholds of a reverberant harmonic complex tone in noise presented from the front collected in this study and for predicting various literature data. The results suggest that a fast extraction of the binaural benefit with sluggishness applied only afterwards matches detection thresholds more precisely than a slow extraction of BMLDs, especially in higher reverberation and non-standard situations with only late reflections.

5.1 Effects of early reflections on signal detection in noise

Results of experiment 1 show that early reflections improve detection thresholds of a low frequency harmonic complex tone in static noise if the target sound source is collocated with the masker at 0°. In this condition, the direct sound does not provide advantageous binaural information to unmask the target signal (comparable with an $\mathrm{N_0S_0}$ condition in a classical BMLD experiment). Adding early reflections up to 75 ms decreases the interaural correlation of the target which results in an increased binaural benefit in noise from the front. Noteworthy here is that the ratio between direct sound and individual reflections is kept the same since the whole RIRs were scaled to ensure an overall constant sound pressure level across conditions. This suggest that early reflections can be seen as useful and contribute to the binaural decorrelation which improves detectability. Adding later reflections does not further decrease the interaural correlation, which might explain the constant thresholds obtained when adding additional reflections after 75 ms. To illustrate these observations, Figure 10 shows the time course of the interaural correlation (IC) of the reverberant target signal located at 0° for an absorption coefficient of 0.1 and different RIR truncations. Panel a) shows the IC over time when early reflections are successively added (experiment 1). The IC decreases for truncation times up to 75 ms, and remains constant when later reflections are added. Figure 10, panel b, shows the IC over time of all cutting time conditions in experiment 2. Here, late reflections decorrelate the frontal target signal, reaching the maximum decorrelation if reflections arrive within 150 ms.

Zurek et al. [2] measured detection thresholds of a 1/3 octave narrowband noise with a broadband noise masker in simulated reverberation. Monaural thresholds in the anechoic condition were compared to binaural thresholds in reverberation. Their results for collocated target and masker at 0° suggest that reverberation does not have a significant impact on detection thresholds. This is in contrast to the results of the current study, which clearly shows that adding early reflections to a frontal target with a collocated anechoic masker leads to a significant decrease in detection thresholds. Late reflections do not contribute further to unmasking because the IC does not decrease further (see Fig. 10, panel a). One reason for this different outcome might be that Zurek et al. [2] used reverberant target and masker stimuli in a steady state condition without a build-up of incoming reflections, resulting in a decorrelation of both the noise and masker signals. In the current study, only the target sound was reverberant, potentially emphasizing the unmasking effects of reflections. The current study likely shows the maximum benefit of early reflections under binaurally optimal circumstances.

Zurek et al. [2] also tested different absorption coefficients. For a frontal target sound source with a collocated masker, binaural detection thresholds did not differ for absorption coefficients in the range of 0.1 to 1. This result is in accordance with our findings. In the first experiment of the current study no significant difference can be found across different absorption conditions.

Braasch [20] measured detection thresholds of a broadband noise target at different azimuth angles, simulated with head-related transfer functions and played via headphones, in the presence of a broadband noise masker in the front in a simulated reverberant room as well as in anechoic space. Detection thresholds decreased with increasing azimuth of the target sound source, in accordance with the current findings. However, thresholds differed for an anechoic versus a reverberant lateral target with a frontal noise masker, which we did not observe. Here, thresholds were not significantly different for a lateral target position when comparing the direct sound (anechoic) to the full RIR condition. The differences might stem from an additional detrimental effect of reflections from the reverberant noise masker used in his study [20].

5.2 Effects of late reflections on signal detection

The results of experiment 2 demonstrate that isolated late reflections can also improve detection thresholds: reflections arriving 60 ms after the direct sound lowered detection thresholds significantly compared to the direct sound only condition. These isolated late reflections decorrelate the target signal and therefore increase binaural unmasking as analyzed in Figure 9 (panel b). As expected, the later the reflections arrive, the later the decorrelation of both ear signals starts. However, reflections arriving 235 ms after the direct sound also decrease the IC for the last 200 ms of the stimulus. The unmasking process for detecting a longer harmonic sound can thus benefit from the decorrelation by late reflections. For speech, such a benefit would presumably be available if phonemes are voiced on the same fundamental frequency for long enough that the late reflections can still contribute energy to the harmonics. This might be the case when singing, and also for musical instrument

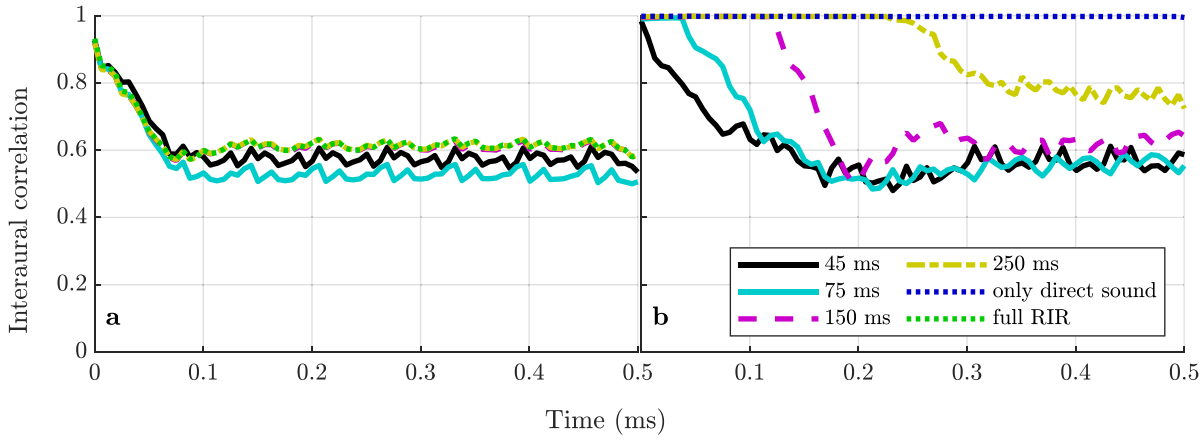

Figure 10. Short-term interaural correlation (IC) of the reverberant target signal located at 0° depending on the time point of the signal. The parameter varied between curves is the cutting time condition of the RIR, given as truncation time in panel a (experiment 1), and as the cutting time up to which the RIRs were zeroed in panel b (experiment 2). When adding early reflections up to a truncation time of 75 ms (panel a), the IC decreases, and it remains at a constant, low value when later reflections are added. Late reflections decorrelate the fronal target signal, but the maximum decorrelation requires reflections to arrive within 150 ms (panel b).

sounds. For regular speech, the spectral speech content changes at the syllable rate of 3–4 Hz, thus preventing the add-on of similar harmonic energy from late reflections. For larger frequency changes this might limit the unmasking benefit and the reflections will interfere with the newly incoming speech sounds also in terms of the information they carry, leading to the "detrimental window" concept for late reflections which function like interfering noise. Such a segmentation in useful and detrimental energy was proposed by Bradley [27] who showed that reflections arriving after 80 ms do not contribute to speech intelligibility in rooms. Srinivasan et al. [41] measured, like most studies, speech reception thresholds and compared a full room impulse response with two truncated versions, one including only early reflections within 50 ms and one with only late reflections arriving after 50 ms. They observed lower thresholds for the condition with only early reflections compared to that with only late reflections, especially when target and noise masker were collocated in the front. Comparable results were found by Lochner and Burger [26] and Leclère et al. [30], all agreeing on a useful window size in the range of 50 to 80 ms. Late reflections can also contribute to speech intelligibility. Rennies et al. [42] used a single late reflection 200 ms after the direct sound with the same amplitude as the direct sound but with an IPD of 180°. Listeners' speech reception thresholds decreased compared to only the direct sound if the single reflection contained binaurally favorable information (e.g. IPD of 180°). However, a single late reflection of equal amplitude to the direct sound is likely perceived as a separate sound event.

5.3 Contribution of monaural cues, better-ear SNR and BMLD

Listening into the gaps of a slowly fluctuating noise masker might play an important role especially in a nearly monaural listening situation with an anechoic target collocated with the masker at 0° [43]. Schubotz et al. [43] measured monaural speech detection in maskers with varying spectro-temporal features and mentioned that overall masking can be mainly explained by short-term energetic masking. Braasch [20], for example, used separate monaural and binaural detection stages for the detection algorithm. Breebaart et al. [19] also used monaural and binaural channels which are processed by a central processor afterwards. In the current DynBU$_{\text{fast}}$ model approach there is no separate monaural processing stage. Interestingly also in a nearly monaural listening situation, the current model approach provides accurate predictions although there is no separate monaural path to derive the absolute SNR. It seems that the short-time better-ear SNR, which is derived across both ears and therefore binaural, is sufficient to also consider monaural benefits because it also takes into account hearing into gaps. A better-ear SNR derived over 200 ms would lead to less unmasking since it introduces more temporal smearing, which would, however, underestimate the measured threshold in an N_0S_0 condition. The importance of short-time better-ear SNR can be seen for very early reflections (up to 20 ms cutting time). Here, the better-ear contribution dominates the overall detection threshold. For larger cutting times the BMLD contribution increases further while the better-ear contribution stays fairly constant. This might be because the early reflections from the floor and the ceiling of the room carry similar binaural information as the direct sound and therefore influence the better-ear advantage more strongly whereas later reflections provide more differing binaural information.

For a lateral target, the BMLD contributes dominantly to the overall benefit across all truncation conditions, because early reflections from floor and ceiling reinforce binaural unmasking, unlike in the N_0S_0 condition. Early reflections also increase the better-ear SNR, which results in a slight overestimation of the measured threshold at 20 ms truncation time.

BMLDs also contribute dominantly to the detection benefit for late incoming reflections especially for lower absorption coefficients, suggesting that late reflections coming from different directions in strongly reverberant situations decorrelate the signal sufficiently. With less reverberation, however, late reflections do not carry enough energy to decorrelate the signal to a sufficient extent, and the almost constant better-ear contribution dominates.

5.4 *Optimal time constants for the DynBU$_{\text{fast}}$ approach*

The time constants in the DynBU$_{\text{fast}}$ approach were found in a least-squares optimization. These optimal time constants are in accordance with time constants proposed in the literature. The optimal short-time analysis window was found to be 12 ms (effective length of 6 ms). Bernstein et al. [44] found that interaural changes in time and intensity can be processed on a short timescale of about 10 ms which is in agreement with the estimated 12 ms time frames in the current paper. The estimated integration time for sluggishness of 225 ms is well in agreement with previous research [15–17, 30, 44]. Intensity integration is often assumed to take around 200 ms [37] which is longer that the 90 ms estimated here. Viemeister and Wakefield [45] assumed that a long-term integration does not necessarily occur in the auditory process. They suggested also shorter time windows in their multiple-look model, which would support the assumption of 90 ms intensity integration.

5.5 Fast versus slow BMLD extraction for a binaural detection model

Incoming reflections will cause ongoing changes of the binaural cues, affecting the unmasking of a sound source in noise as a function of time. The present article questions if such changes need to be taken into account with a dynamic model. Former detection models [15–17] have processed a long integration window to account for sluggishness. Those detection models considering temporally changing signals [19, 20] do not explicitly consider binaural sluggishness which is expected to influence detection thresholds. The proposed model approach in the current study tries to include and discuss the sluggish integration for detecting a reverberant signal in noise.

Recent models focus especially on speech intelligibility in reverberant listening situations [21, 24, 46]. These models use two different time constants. Binaural unmasking is usually derived from a larger time frame (200 to 300 ms) whereas the monaural contribution is derived on much shorter time frames. Hauth and Brand [46] recently extended the model from Beutelmann et al. [21] by introducing a binaural temporal window of 200 ms. They extract the EC parameters within 23 ms short time block but average these parameters across 200 ms by taking the median. The averaged parameters are then used in the EC-process to derive the binaural benefit effectively on 200 ms time frames, i.e. the binaural contribution is computed from already integrated parameters. Vicente and Lavandier [24] recently followed a related approach. They divide the input signal into 300 ms time frames to derive the binaural benefit and take sluggishness into account in one step. The better-ear contribution is instead computed in "fast" 24 ms time frames. Both models introduce a sluggish component through the integration of binaural cues in a long-time window before computing the binaural benefit, assuming that the auditory system is not able to process fast changes of these cues. This differs from the approach suggested in the present paper which computes BMLDs in short analysis windows and averages afterwards. Because the BMLD computation is a non-linear operation, changing the order yields different, and, as shown here, better results.

The current approaches and some former speech intelligibility models compute the benefit from separate presentations of masker and target signal and are thus not functional models as a classical EC model approach. Since the DynBU$_{\text{fast}}$ approach incorporates an EC-based computation, which leads to similar results as the full EC implementation from Durlach [14], the extension to a functional model using the mixed ear signals should be a formal step. Wan et al. [47] proposed a short-time version of the EC model to predict speech intelligibility with speech maskers using the EC process with a sliding window of 20 ms length, which is in agreement with the current data. However, Wan et al. [47] only used low-reverberant signals whereas the present paper also describes the positive effects of early and late reflections especially in highly reverberant environments. These effects can be accurately predicted with the DynBU$_{\text{fast}}$ model approach.

Using short evaluation time frames for BMLD contribution is also motivated in the literature which shows that the auditory system can process interaural changes in time and intensity on a short timescale [44] of about 10 ms in certain situations. Siveke et al. [48] used a noise stimulus with modulated binaural coherence and ITDs at the same time (*Phasewarp stimulus*) and contrasted it with modulation detection in monaural noise. With increasing modulation frequency, the sensitivity to detect a modulation decreases for both, the Phasewarp stimulus and monaural modulation in the same manner. They concluded that there is no indication for additional binaural sluggishness. However, the results might be affected by across-frequency processing. While interaural cues can be extracted on a short time basis, the localization of a tone needs an auditory object to be formed and followed, which might explain the sluggish behavior observed in some studies. Building up an auditory object takes time [49–51] and attaching a location to it might happen at a low rate. The conceptual advantage of a fast extraction is that fine temporal information is binaurally compared only within a short analysis window, reducing any requirement for a "storage".

6 Conclusion

The current study investigated the effect of room reflections on binaural unmasking of a low frequency harmonic complex tone in anechoic noise. The following main findings can be drawn from the current study:

- Early reflections up to 45 ms can improve binaural detection thresholds for a target in the front in the presence of a collocated, anechoic noise masker, consistent with a decorrelation imposed on the target.
- For a lateral sound source position at 60° and a masker from the front, neither early nor late reflections contribute to further increase binaural unmasking.
- In the N_0S_0 condition, in the absence of early reflections and reverberation in the masker, listeners are still able to benefit from isolated late reflections up to 250 ms RIR cutting time, leading to significantly decreased detection thresholds. This is consistent with a sufficient decorrelation evoked by late reflections for a frontal target in almost diotic noise.
- Detection studies on tone-in-noise in free field can only be found sparsely in the literature. The current study can therefore also be seen as a step from basic headphone experiments into the direction of hearing research in real world scenarios. The current free field results are in agreement with results from former studies conducted via headphones.
- A model approach computing the BMLD and better-ear detection cues in short time analysis windows (12 ms) followed by an integration to account for sluggishness and intensity integration, respectively, can predict the measured detection thresholds especially in high reverberation and with isolated late reflections more accurately than when BMLDs are derived from a large time window, which tends to underestimate thresholds.
- Even for almost monaural listening situations with an anechoic target and masker collocated at 0°, the current model approach provides accurate predictions without a separate monaural path, as used in other detection models.

Conflict of interest

The authors declared no conflict of interests.

Acknowledgments

This study was funded by TUM and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Projektnummer 352015383 – SFB 1330 C5. The rtSOFE system was funded by BMBF 01 GQ 1004B.

The authors thank two anonymous reviewers and the editor for their helpful comments on an earlier version of the manuscript.

Data availability statement

The $\text{DynBU}_{\text{fast}}$ and the $\text{DynBU}_{\text{slow}}$ model as well as the experimental data and code to generate all figures of this manuscript are available at https://doi.org/10.5281/zenodo.7643249 [40]. The DynBUfast model and experimental data will be available as *bischof2023* and *exp_ bischof2023*, respectively, in the auditory modeling toolbox (AMT) version 1.3 [39]. The real-time Simulated Open Field Environment (rtSOFE) software package (1.1) is available at https://doi.org/10.5281/zenodo.5648305 [32].

References

1. H. Kuttruff: Room acoustics. CRC Press, Boca Raton, Florida, 2017.
2. P.M. Zurek, R.L. Freyman, U. Blalkrishnan: Auditory target detection in reverberation. Journal of the Acoustical Society of America 115 (2004) 1609–1620.
3. L.A. Jeffress, H.C. Blodgett, B.H. Deatherage: The masking of tone by white noise as a function of interaural time or phase displacement. Journal of the Acoustical Society of America 25 (1953) 190.
4. S. van de Par, A. Kohlrausch: A new approach to comparing binaural masking level differences at low and high frequencies. Journal of the Acoustical Society of America 101 (1997) 1671–1680.
5. S. van de Par, A. Kohlrausch: Dependence of binaural masking level differences on center frequency, masker bandwidth, and interaural parameters. Journal of the Acoustical Society of America 106 (1999) 1940–1947.
6. I.J. Hirsh: The influence of interaural phase on interaural summation and inhibition. Journal of the Acoustical Society of America 20 (1948) 536–544.
7. T. Biberger, S.D. Ewert: The effect of room acoustical parameters on speech reception thresholds and spatial release from masking. Journal of the Acoustical Society of America 146 (2019) 2188–2200.
8. B.A. Edmonds, J.F. Culling: The spatial unmasking of speech: evidence for better-ear listening. Journal of the Acoustical Society of America 120 (2006) 1539–1545.
9. D.E. Robinson, L.A. Jeffress: Effect of varying the interaural noise correlation on the detectability of tonal signals. Journal of the Acoustical Society of America 65 (1963) 1947–1952.
10. M. van der Heijden, C. Trahiotis: Binaural detection as a function of interaural correlation and bandwidth of masking noise: Implications for estimates of spectral resolution. Journal of the Acoustical Society of America 103 (1998) 1609–1614.
11. L.R. Bernstein, C. Trahiotis, Accounting for binaural detection as a function of masker interaural correlation: effects of center frequency and bandwidth, Journal of the Acoustical Society of America 136 (2014) 3211–3220.
12. L.R. Bernstein, C. Trahiotis: An interaural-correlation-based approach that accounts for a wide variety of binaural detection data. Journal of the Acoustical Society of America 141 (2017) 1150–1160.
13. H.S. Colburn: Theory of binaural interaction based on auditory-nerve data. II. Detection of tones in noise. Journal of the Acoustical Society of America 61 (1977) 525–533.
14. N.I. Durlach: Equalization and cancellation theory of binaural masking-level differences. Journal of the Acoustical Society of America 35 (1963) 1206–1218.
15. D.W. Grantham, F.L. Wightman: Detectability of a pulsed tone in the presence of a masker with time-varying interaural correlation. Journal of the Acoustical Society of America 65 (1979) 1509–1517.
16. I. Holube, M. Kinkel, B. Kollmeier: Binaural and monaural auditory filter bandwidths and time constants in probe tone detection experiments. Journal of the Acoustical Society of America 104 (1998) 2412–2425.
17. B. Kollmeier, R.H. Gilkey: Binaural forward and backward masking: evidence for sluggishness in binaural detection. Journal of the Acoustical Society of America 87 (1990) 1709–1719.

18. J. Breebaart, S. van de Par, A. Kohlrausch: The contribution of static and dynamic varying ITDs and IIDs to binaural detection. Journal of the Acoustical Society of America 106 (1999) 979–992.
19. J. Breebaart, S. van de Par, A. Kohlrausch: Binaural processing model based on contralateral inhibition. I. Model structure. Journal of the Acoustical Society of America 110 (2001) 1074–1088.
20. J. Braasch: Auditory localization and detection in multiple-sound-source scenarios. PhD thesis. Ruhr-Universität Bochum, Bochum, Germany, 2001.
21. R. Beutelmann, T. Brand, B. Kollmeier: Revision, extension, and evaluation of a binaural speech intelligibility model. Journal of the Acoustical Society of America 127 (2010) 2479–2497.
22. M. Lavandier, J.F. Culling: Prediction of binaural speech intelligibility against noise in rooms. Journal of the Acoustical Society of America 127 (2010) 387–399.
23. J. Rennies, A. Warzybok, T. Brand, B. Kollmeier: Modelling the effects of a single reflection on binaural speech intelligibility. Journal of the Acoustical Society of America 135 (2014) 1556–1567.
24. T. Vicente, M. Lavandier: Further validation of a binaural model predicting speech intelligibility against envelope-modulated noises. Hearing Research 390 (2020) 107937.
25. R.Y. Litovsky, H.S. Colburn, W.A. Yost, S.J. Guzman: The precedence effect. Journal of the Acoustical Society of America 106 (1999) 1633–1654.
26. J.P.A. Lochner, J.F. Burger: The influence of reflections on auditorium acoustics. Journal of Sound and Vibration 1 (1964) 426–454.
27. J.S. Bradley: Predictors of speech intelligibility in rooms. Journal of the Acoustical Society of America 80 (1986) 837–845.
28. A. Warzybok, J. Rennies, T. Brand, S. Doclo, B. Kollmeier: Effects of spatial and temporal integration of a single early reflection on speech intelligibility. Journal of the Acoustical Society of America 133 (2013) 269–282.
29. J. Rennies, T. Brand, B. Kollmeier: Prediction of the influence of reverberation on binaural speech inteligibility in noise and in quiet. Journal of the Acoustical Society of America 130 (2011) 2999–3012.
30. T. Leclère, M. Lavandier, J.F. Culling: Speech intelligibility prediction in reverberation: Towards an integrated model of speech transmission, spatial unmasking, and binaural de-reverberation. Journal of the Acoustical Society of America 137 (2015) 3335–3345.
31. B.U. Seeber, S. Kerber, E.R. Hafter: A system to simulate and reproduce audio-visual environments for spatial hearing research. Hearing Research 260 (2010) 1–10.
32. B.U. Seeber, T. Wang: real-time Simulated Open Field Environment (rtSOFE) software package (1.1). Zenodo (2021). https://doi.org/10.5281/zenodo.5648305.
33. J.B. Allen, D.A. Berkley: Image method for efficiently simulating small-room acoustics. Journal of the Acoustical Society of America 65 (1979) 943–950.
34. F. Zotter, M. Frank: Ambisonics – a practical 3D audio theory for recording, studio production, sound reinforcement, and virtual reality. Germany, Springer, Heidelberg, 2019.
35. J. Daniel: Représentation de champs acoustiques, application à la transmission et à le reproduction de scènes sonores complexes dans un context multimèdia. Université de Paris, Paris, France, 2001.
36. N. Kolotzek, B.U. Seeber: Spatial unmasking of circular moving sound sources in the free field, in: M. Ochmann, M. Vorländer, J. Fels (Eds.), Proc. 23rd International Congress on Acoustics integrating 4th EAA Euroregio, Deutsche Ges. für Akustik e.V. (DEGA), Germany, 2019:7640–7645.
37. H. Fastl, E. Zwicker: Psychoacoustics: Facts and models. Springer, Heidelberg, Germany (2007).
38. H. Levitt: Transformed up-down methods in psychoacoustics. Journal of the Acoustical Society of America 49 (1971) 467–477.
39. P. Majdak, C. Hollomey, R. Baumgartner: AMT 1.3: A toolbox for reproducible research in auditory modeling. Acta. Acustica 6 (2022) 19.
40. N.F. Bischof, B.U. Seeber: Dynamic Binaural Unmasking model with fast cue extraction (DynBU_fast) to predict the better-ear and binaural benefit for detecting a dynamic sound source in noise (1.0). Zenodo (2023). https://doi.org/10.5281/zenodo.7643249.
41. N.K. Srinivasan, M. Stansell, F.J. Gallun: The role of early and late reflections on spatial release from masking: Effects of age and hearing loss. Journal of the Acoustical Society of America 141 (2017) EL185–EL191.
42. J. Rennies, A. Warzybok, T. Brand, B. Kollmeier: Spatial-temporal integration of speech reflections, in: S. Spors, F.-H. Wurm (Eds.), Fortschritte der Akustik - DAGA '19.Rostock, Germany: Deutsche Ges. für Akustik e.V. (DEGA), 2019: 840–843.
43. W. Schubotz, T. Brand, B. Kollmeier, S.D. Ewert: Monaural speech intelligibility and detection in maskers with varying amount of spectro-temporal speech features. Journal of the Acoustical Society of America 140 (2016) 524–540.
44. L.R. Bernstein, C. Trahiotis, M.A. Akeroyd, K. Hartung: Sensitivity to brief changes of interaural time and interaural intensity. Journal of the Acoustical Society of America 109 (2001) 1604–1615.
45. N.F. Viemeister, G.H. Wakefield: Temporal integration and multiple looks. Journal of the Acoustical Society of America 90 (1991) 858–865.
46. C.F. Hauth, T. Brand: Modeling sluggishness in binaural unmasking of speech for maskers with time-varying interaural phase differences. Trends in Hearing 22 (2018) 1–10.
47. R. Wan, N.I. Durlach, H.S. Colburn: Application of a short-time version of the Equalization-Cancellation model to speech intelligibility experiments with speech masker. Journal of the Acoustical Society of America 136 (2014) 768–776.
48. I. Siveke, S.D. Ewert, B. Grothe, L. Wiegrebe: Psychophysical and physiological evidence for fast binaural processing. Journal of Neuroscience 28 (2008) 2043–2052.
49. S. Anstis, S. Saida: Adaptation to auditory streaming of frequency-modulated tones. Journal of Experimental Psychology: Human Perception and Performance 11 (1985) 257–271.
50. A.S. Bregman: Auditory streaming is cumulative. Journal of Experimental Psychology: Human Perception and Performance 4 (1978) 380–387.
51. S. Deike, P. Heil, M. Böckmann-Barthel, A. Brechmann: The build-up of auditory stream segregation: a different perspective. Frontiers in Psychology 3 (2012) 1–7.

Appendix A

Table A1. x, y and z coordinates of the room corners of the simulated room shown in Figure 1. The corner indexes are given clockwise starting in the corner near the subject on the floor (1–4) followed by the corners of the ceiling (5–8). The coordinates are given in meters.

	x	y	z
S_1	0	0	0
S_2	0.77	17.49	0.15
S_3	8.06	16.71	0.19
S_4	7.24	−0.39	0.31
S_5	0.01	0.02	3.12
S_6	0.46	17.14	2.84
S_7	7.58	16.49	3.04
S_8	7.01	−0.12	3.35

Cite this article as: Bischof N.F. Aublin P.G. & Seeber B.U. 2023. Fast processing models effects of reflections on binaural unmasking. Acta Acustica, **7**, 11.