The effect of early and late reflections on binaural unmasking

Norbert Kolotzek, Pierre G. Aublin*, Bernhard U. Seeber
Audio Information Processing, Technical University of Munich, 80333 Munch
{norbert.kolotzek, seeber}@tum.de
**pierre.aublin@inserm.fr*

Introduction

In many listening situations we are surrounded not only by the direct sound of a source but also by sound reflections form walls of a room, buildings, automobiles and trees [1]. These sound reflections will arrive at the ears at different times so that interaural level and phase differences will likely change over time. Such changes in the interaural cues result also in time-variant changes in the interaural correlation, which is known to improve detection of a target sound source in noise from the front (N0). The benefit in detection stems partly from a better-ear monaural signal-to-noise ratio but also from the Binaural Masking Level Differences (BMLD) [2]. This can also be seen for circular moving signals in the free field for quite slow angular velocities as shown by Kolotzek and Seeber [3]. Several former studies focused on detection experiments of a variety of different sound sources in noise and aimed to predict the binaural benefit. van der Heijden and Trahiotis [4] as well as Bernstein and Trahiotis [5] measured detection thresholds of a 500 Hz sine tone, interaurally phase shifted by 180°, as a function of the interaural correlation of a broadband noise masker. Both studies showed that binaural unmasking decreases with increasing interaural correlation of the masker noise.

Binaural unmasking is also important to predict e.g. binaural speech intelligibility in noise and reverberation. Some of the existing models [6, 7] use BMLD and better-ear SNR estimates to predict the overall speech intelligibility, but they do not specifically take into account temporal information from the incoming reflections. For speech intelligibility, a common approach to account for effects of early and late reverberation is to define early reflections as useful within the first 50 to 80 ms, whereas late reflections affect intelligibility detrimentally and are added to the energy of the masking sound [8, 9]. To inform models of speech intelligibility by results from pure detection tasks, the current study investigates the effects of early and late reflections when dealing with time varying cues in a detection task.

Methods

Simulated room

A nearly rectangular virtual room with randomly shifted corners was simulated, which reduces room modes from strictly repetitive reflection times. Two different absorption coefficients, $\alpha_1 = 0.1$ and $\alpha_2 = 0.5$, and two target positions at 0° and at 60° to the left at 5 m distance to the listener were simulated. The virtual listener was placed 1.5 m from the left and the rear wall, the masker position was always at 0° in front of the virtual listener in a simulated distance of 5 m. Room impulse responses (RIRs) were generated using the Simulated Open Field Environment (SOFE) [10], and reflections were simulated up to 100th order. Direct sound and reflections were rendered using 17th-order Ambisonics [11] with *max* r_E decoding [12] and played over the SOFE's 36 horizontally arranged loudspeakers.
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 (full RIR).

Stimuli

As a target stimulus, a harmonic complex tone (HCT) consisting of the 7th to 13th harmonic to 50 Hz fundamental frequency (350 Hz to 650 Hz) was used. The level of each harmonic was set such that each auditory filter received the same energy (equally exciting). The target stimulus was convolved with the truncated room impulse responses for each of the 36 loudspeakers. The level at the listener's position of the reverberant signals was normalized across different truncation conditions.

Uniform exciting noise was used as masker [13]. The noise was band-limited from 250 Hz to 750 Hz, to ensure masking all components of the HCT target stimulus without becoming too loud. It had an overall duration of 900 ms with 30 ms Gaussian rise and fall times. The noise source had a sound pressure level of 60 dB at the listener's position. The noise was kept anechoic and not filtered with the room impulse response to avoid additional influence of the reflections of the noise masker. Therefore, it was just played from a single loudspeaker at 0° leading to highly binaurally correlated noise (almost N0).

Procedure

The participants sat in the completely darkened anechoic chamber in the center of the loudspeaker array. The detection threshold of the HCT in noise was determined with a three-interval three-alternative-forced-choice method (3I-3AFC) using a two-down/one-up adaptive staircase procedure [14] tracking the 71% point of the psychometric function. 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 HCT was added. The listeners' task was to indicate which interval differed from the others by pressing the corresponding number on a keyboard. The overall level of the HCT was then adjusted depending on a right or wrong answer. The mean of the last ten reversals at the final step size of 1 dB was used to calculate the detection threshold of the HCT in noise. The experiment was blocked by the absorption coefficient α. Each subject finished one track for each condition combination of target position, truncation time and absorption coefficient, resulting in 28 tracks for each subject. Subjects finished the experiment on average in 2 hours.

Participants

For this preliminary experiment, three male participants volunteered. Their age ranges from 23 to 29 years (mean: 27 yr, std: 3,5). All participants had normal hearing thresholds with a hearing loss less than 15 dB up to 8 kHz as assessed with a clinical audiometer. They gave written consent and were not payed for participating in the experiment. The study was approved by the ethics committee of the TUM, 65/18S.

Experimental results

Target at 0° collocated with the noise masker(N0S0)

Figure 1 shows the benefit in detection thresholds from adding more early reflections relative to the measured threshold with a target without any reflections. Results are for the target at 0° and are presented as medians with quartiles. This threshold difference relative to an anechopic target is shown for both tested absorption coefficients. With an increasing amount of reflections, this difference in threshold also increases. This observation suggests that adding early reflections might be helpful to detect the target sound source in noise. The benefit from early reflections does not differ between absorption coefficients. Even when adding only very early reflections (e.g. truncation after 20 ms), the benefit compared to an anechoic target (only direct sound, truncation after 15 ms) is already at about 5 dB. It seems that already early reflection decorrelate the target sound in a sufficient manner. Interestingly, adding late reflections arriving after 150 ms do not negatively affect this benefit caused by early reflections. The difference in threshold compared to a target without reflections, is quite constant and does not decrease.

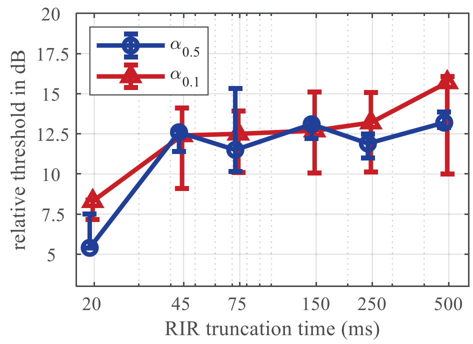

Figure 1:. Benefit in detection threshold of a reverberant harmonic complex tone (HCT) from 0° in the front for different truncations of the RIR in the presence of a collocated anechoic bandpass noise<f with 60 dB SPL from the front are shown compared to the detection threshold measured with only the direct sound (15 ms: anechoic target). With increasing RIR truncation time, the amount of late reflections increases. Blue circles show the median threshold benefit across 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.

Target at 60° with the noise masker at 0°(N0S60)

The difference in threshold relative to a target sound source without reflections for a lateral sound source located at 60° is shown in Figure 2. From these preliminary data it seems that adding early reflections does not change the detection threshold relative to that for only the direct sound, since the this difference in threshold is around 0 dB for truncation times up to 150 ms. Also for a lateral sound source, the difference in threshold relative to an anechoic target seems to be independent from the amount of absorption at least for adding reflections up to 150 ms. For reflections arriving later than 150 ms, there is a slight detriment in detection threshold visible, since the change in threshold relative to an anechoic target becomes slightly negative (-2.5 dB) at least for higher reverberation ($\alpha = 0.1$).

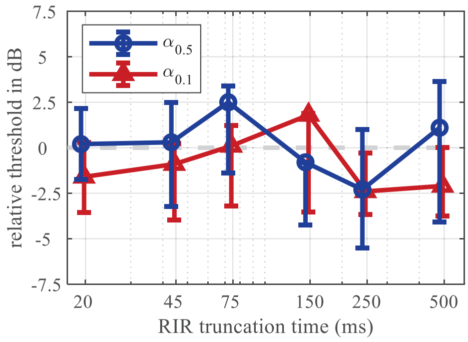

Figure 2:. Change in detection threshold of a reverberant harmonic complex tone (HCT) from 60° relative to the detection threshold for only the direct sound (15 ms: anechoic target) is shown for different truncations of the RIR. Detection was measured in the presence of an anechoic bandpass noise with 60 dB SPL from the front. Blue circles show the median threshold change 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.

Conclusions

The current study investigated the effect of room reflections on binaural unmasking of a low frequency harmonic complex tone in anechoic noise. The data from the experiment suggest that early reflections up to 45 ms can improve binaural detection thresholds for an N0S0 condition due to the decorrelation of the target signal caused by incoming early reflections. Furthermore, late and more diffuse reflections might not destroy the unmasking effect of the early reflections.

For a lateral sound source at 60° with an anechoic noise masker from the front, early reflections seem to not have an effect on the detection threshold. For a lateral sound source a slight detrimental effect can be observed for reflections arriving after 150 ms for higher reverberant environments.

Acknowledgements

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.

Literature

[1] Kuttruff, H. Room acoustics, CRC Press, Boca Raton, Florida, 2017.

[2] Zurek, P.M., Freyman, R.L. and Blalkrishnan, U.: Auditory target detection in reverberation. J. Acoust. Soc. Am. 115 (2004), 1609-1620.

[3] Kolotzek, N and Seeber, B.U.: Spatial unmasking of circular moving sound sources in the free field. Proc. 23rd International Congress on Acoustics, integrating 4th EAA Euroregio 2019, Aachen, Germany, German Acoustical Society e.V. (DEGA) (2019) 7640-7645.

[4] van der Heijden, M. and Trahiotis, C.: Binaural detection as a function of interaural correlation and bandwidth of masking noise: Implications for estimates of spectral resolution. J. Acoust. Soc. Am. 103 (1998) 1609-1614.

[5] Bernstein, L.R. and Trahiotis, C.: Accounting for binaural detection as a function of masker interaural correlation: Effects of center frequency and bandwidth. J. Acoust. Soc. Am. 136 (2014) 3211-3220.

[6] Beutelmann, R., Brand, T. and Kollmeier, B.: Revision, extension, and evaluation of a binaural speech intelligibility model. J. Acoust. Soc. Am. 127 (2010) 2479-2497.

[7] Lavandier, M. and Culling, J.F.: Prediction of binaural speech intelligibility against noise in rooms. J. Acoust. Soc. Am. 127 (2010) 387-399.

[8] Lochner, J.P.A. and Burger, J.F.: The influence of reflections on auditorium acoustics. Journal of Sound and Vibration 1 (1964) 426-454.

[9] Litovsky, R.Y., Colburn, H.S. , Yost, W.A. and Guzman, S.J.: The precedence effect. J. Acoust. Soc. Am.106 (1999) 1633-1654.

[10] Seeber, B.U., Kerber, S. and Hafter, E.R.: A system to simulate and reproduce audio-visual environments for spatial hearing research. Hear. Res. 260 (2010) 1-10.

[11] F. Zotter F. and Frank, M.: Ambisonics - A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality, Heidelberg, Germany, Springer, 2019.

[12] Daniel, J.: 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

[13] Fastl, H. and Zwicker, E.: Psychoacoustics Facts and Models, Berlin Heidelberg, Springer, 2007.

[14] Levitt, H.: Transformed Up-Down Methods in Psychoacoustics. J. Acoust. Soc. Am. 49 (1971) 467-477.