Audio Engineering Society

Conference Paper

Presented at the Conference on
Spatial Reproduction
2018 August 6 – 9, Tokyo, Japan

Influence of the frequency dependent directivity of a sound projector on the localization of projected sound

Sebastian Merchel, Tom Wühle, Felix Reichmann, and Ercan Altinsoy

Dresden University of Technology, Chair of Acoustic and Haptic Engineering

Correspondence should be addressed to Sebastian Merchel (`sebastian.merchel@tu-dresden.de`)

ABSTRACT

This study investigates the localization in a sound projection scenario using a realistic surround audio scene (street with traffic noise). Real sound projectors, e.g., phased arrays, are characterized by a physically limited focusing ability. This leads to multiple sound propagation paths to the listener: direct sound via the shortest path and projected sound via reflections. The spectral characteristic of the direct sound is varied in a listening test to simulate realistic directivity limitations at low and high frequencies. The attenuation of the direct sound at mid-range frequencies, which is just necessary so that the position of the auditory event is localized in the direction of the projection, is determined.

1 Introduction

Multichannel spatial sound reproduction is a challenging task, partly because of the need for many distributed loudspeakers. In many scenarios, the placement of speakers at arbitrary positions is not possible because of functional or aesthetical restrictions, the latter also known as 'wife acceptance factor'. A solution is the invisible projection of sound on walls or other reflective surfaces. The aim is to create auditory events from various directions using this reflected sound. To this end, so called sound projectors are applied (e.g., [1, 2]). Figure 1 shows a typical sound projection scenario. A sound beam is created by a phased speaker array. A schematic directivity pattern of such an array is shown in the background for mid frequencies (gray). The main lobe is steered towards an acoustically hard room boundary, where it is reflected before it reaches the listener. This projected sound (green) is used to create an auditory event in the direction of the last reflection behind the listener. However, a speaker array radiates additional (usually unwanted) side lobes, inter alia, in the direction of the listener. This direct sound (red) arrives earlier at the ears than the projected sound because of the shorter propagation path. In real listening rooms the delay is typically below the echo threshold. If the direct sound would be undamped, it would dominate the localization because of the law of the first wave front, also called precedence effect [3]. For very short delays, summing localization can occur [4]. However, the directivity of the speaker array reduces the amplitude of the direct sound compared to the projected

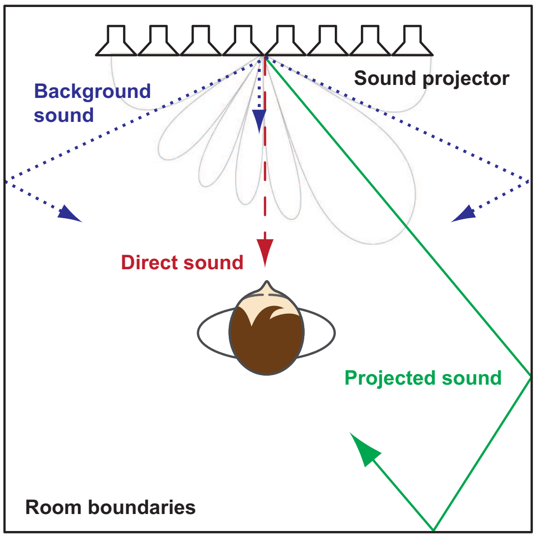


Fig. 1: Typical sound projection scenario using wall reflections: a speaker array is applied to create an auditory event behind a listener using a projected sound beam (green). Direct sound (red) arrives somewhat earlier at the listing position because of the limited focusing ability of the array. A schematic directivity pattern is shown for mid frequencies (gray). In a real surround scenario simultaneous reproduction from multiple directions overlays which is illustrated as background sound (blue).

sound. If this damping (between main lobe and side lobes) is strong enough, the projected sound can inhibit the preceding directed sound [5]. The spatial location of the auditory event is then dominated by the projected sound [6]. Unfortunately, the focusing ability of real sound projectors is physically limited and frequency dependent [7]. A schematic damping characteristics in the direction of the direct sound is shown in gray in the background of Figure 2. Significant damping can be achieved at mid-range frequencies (MF). Damping at low frequencies (LF) is limited because of limited array lengths. Damping at high frequencies (HF) is restricted by strong grating lobes. These HF lobes can be shifted to some extend in frequency by changing the speaker spacing [7]. There are further methods to increase damping and influence the directivity, e.g., by using logarithmic speaker spacing [8], delaying and shading of driving signals [9] or numerical beamforming techniques [10]. The question raises: what is the optimal directivity of a sound projector? Which directivity should be used as a target, when designing such systems? Performance parameters, like beamwidth uniformity or smoothness and flatness of off-axis frequency response, which are used to rank classical line arrays for public address systems [11], might not be suitable [12]. This pilot study aims at determining critical parameters in the frequency response of the direct sound of a speaker array, which are relevant for the localization in a projection scenario.

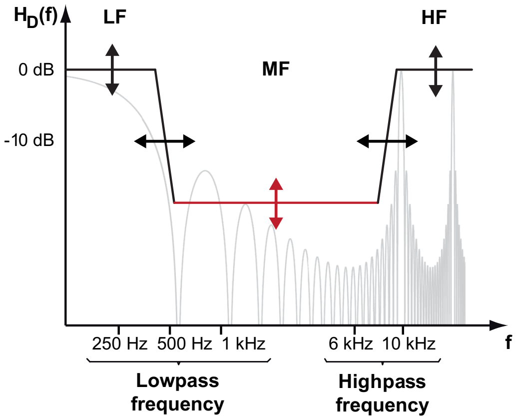


Fig. 2: Schematic filter characteristic to simulate the typical damping of the direct sound in a real projection scenario (gray). A simplified model with three frequency bands is used. A damping of 0 dB corresponds to equal amplitude of direct and projected sound at the listener position.

2 Setup

To discuss the effect of the frequency dependent directivity of a sound projector, a perceptual experiment was conducted. Therefore, the typical damping characteristic of the direct sound shown in Figure 2 (gray) was abstracted as a filter characteristic $H_\mathrm{D}(f)$ using three frequency bands (LF, MF and HF). A damping of 0 dB corresponded to equal amplitude of direct and projected sound at the listener position. The two corner frequencies between the three bands were varied independently to represent different array lengths and speaker spacings. Three lowpass frequencies at 250 Hz, 500 Hz and 1 kHz were selected to represent different array sizes of approximately 2 m, 1 m and 50 cm at an exemplary main lobe deflection of 30°. Two highpass frequencies at 6 kHz and 10 kHz were chosen. For each combination of lowpass and highpass frequencies a bandpass resulted for MF with varying band width. Linkwitz-Riley filters of 8th order were applied by cascading two 4th-order Butterworth filters to implement steep, 48 dB/octave (160 dB/decade) slopes [13]. An advantage of this filtering approach was the resulting flat amplitude response if all bands had the same level. The levels of the LF and HF bands were controlled independently. A damping of 0 dB and 10 dB was selected for both. The level of the MF band (highlighted in red in Figure 2) was varied by the test participant. The filtering of the direct sound was implemented in PureData. It resulted in three general filter characteristics with varying crossover frequencies:

- '0 dB LF+HF' represents only damping at MF by the test participant,
- '-10 dB LF' represents a fixed damping of 10 dB at LF and adjustable damping at MF and
- '-10 dB HF' represents a fixed damping of 10 dB at HF and adjustable damping at MF.

For the listening test, the individual portions of the sound field needed to be controlled separately. Therefore, a multichannel loudspeaker setup in the anechoic chamber of the Chair of Acoustic and Haptic Engineering was used. The different sound paths were replaced with individual loudspeakers (EVENT 20/20 bas) placed on a circle with a radius of 3 m as shown in Figure 3. The direct sound (red) was radiated in front of the listener. It was a modified version of the projected sound filtered with $H_D(f)$ as discussed above. The projected sound (green) was reproduced at an horizontal angle of 135° (referred to the line of sight) in the back of the listener. It was unfiltered, which corresponds to an equalization of the projecting main lobe to a flat amplitude response $H_P(f)$. To represent a typical scenario, the projected sound was delayed by a fixed lag of 10 ms, which corresponds to a path difference of 3.4 m. Additionally, three frontal loudspeakers (at -45°, 0° and 45°) were used to reproduce a realistic background sound (blue) comparable to an audio scene in a surround mix. Signal processing and data acquisition was realized on a PC using PureData. An RME Fireface UCX soundcard was applied as audio interface.

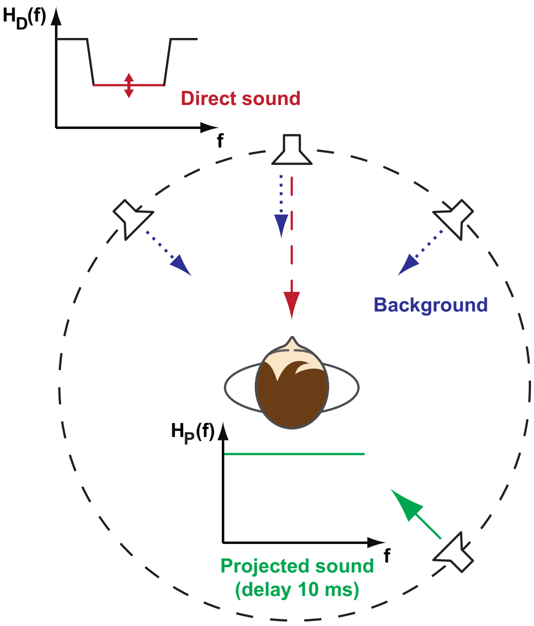


Fig. 3: Experimental setup with four loudspeakers (radius = 3 m) in an anechoic environment to simulate different sound paths in a projection scenario. The direct sound (red) was filtered with $H_D(f)$. The MF band of the direct sound was adjusted by the participant during each experimental run.

3 Stimuli and Participants

A steady real life stimulus was selected for projection in this pilot study: a mono recording of a diesel car in idle mode [14]. Figure 4 shows the A-weighted spectrum of the projected sound at the listener position. It had a broad frequency spectrum so that all filtering variants of the direct sound described above resulted in a change of the signal. The strongest part of the stimulus was a modulated knocking between 1 kHz and 2 kHz, which was always in the MF band. The stimulus length was cut to 30 s. To enhance the realism of the scenario, a steady background sound was selected: a multichannel recording of a street with homogeneous traffic noise and constant voice babbling [15]. The resulting spectrum of the overall background sound at the listener position is added in Figure 4. The relative level was selected to be lower than the projected and unfiltered direct sound in the narrowest mid-range frequency band (1 kHz to 6 kHz). The aim was to avoid strong

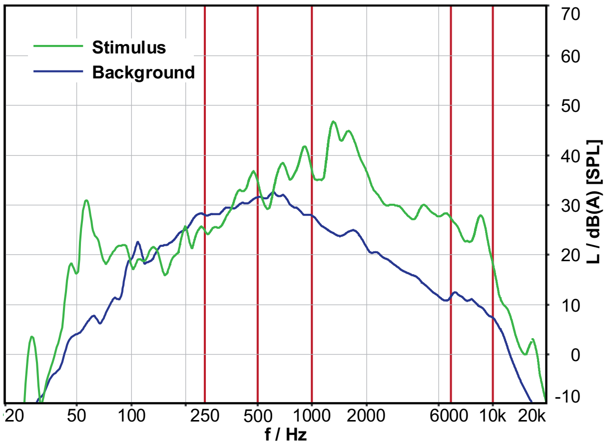


Fig. 4: A-weighted spectrum of the stimulus (projected sound) and sum of the background sounds at the listener position in the absence of the head. Marked in red are the filtering frequencies, used to generate the direct sound from the stimulus. The FFTs were calculated with 65536 samples using 50% overlapping Hann windows and plotted with 1/24th octave intensity averaging.

simultaneous masking of mid-range frequencies in the direct sound by the background sound. The sum of background, projected and direct sounds was adjusted to a sound pressure level of 72 dB(A) at the listener position (without attenuation of the direct sound) to simulate a medium listening level.

Twenty normal hearing participants voluntarily participated in this experiment (19 male and 1 female). Most of them were students between 22 and 33 years old (mean 25 years).

4 Experimental Design

In the current experiment, multiple perceptual attributes may vary simultaneously while changing the direct sound (e.g., timbre and localization) [16, 6]. With such concurrently changing perceptual dimensions, it is difficult to utilize criterion free methods (e.g., alternative forced choice procedures) [17]. However, method of adjustment procedures were found to be suitable, if the test participant is able to focus on one specific perceptual attribute [18]. For this reason, method of adjustment procedures have been utilized frequently in spatial audio experiments [19, 20, 17, 5, 21] and was chosen to measure the projection effect threshold in the present study.

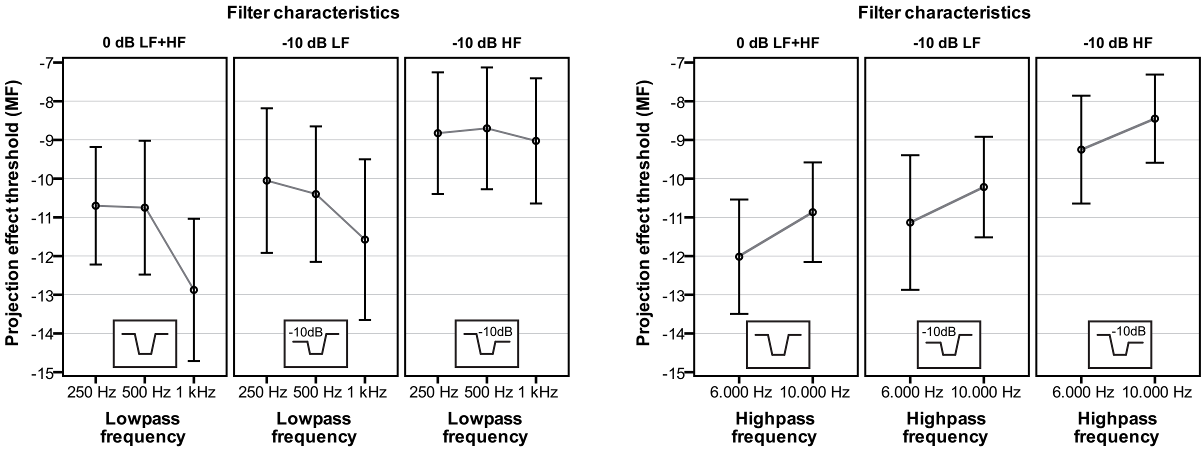
The task of the participant was to “adjust the minimal necessary level of the frontal [direct] sound, so that the perceived location of the car [projected] sound stayed in the back right direction”. For the adjustment a rotary knob was used that was infinitely adjustable and that possessed neither visual indicators, such as tick marks, nor mechanical grating (Griffin Technology, PowerMate). No information about the projection context of the car sound was given. It was also not communicated that the level adjustment influenced the mid-range frequency band of the direct sound only. The level adjustment was allowed between -30 dB and 0 dB with a minimum step size of 1 dB. The initial MF damping for each run was varied randomly within the allowed dynamic range.

The combination of filtering the direct sound with three lowpass frequencies, two highpass frequencies and three filter characteristics resulted in 18 different conditions that needed to be assessed individually. The test scenarios were presented to each participant in random order. The stimuli were looped and repeated as often as necessary, until the participant was satisfied with the adjustment for a scenario. The participant was free to take as much time as necessary for the decision. Before the test began, a training with four conditions was conducted to familiarize the test subject with the stimuli and test procedure. The total duration of the experiment was approximately 20 minutes.

5 Results and Discussion

The adjusted minimal MF damping that is required to keep the localization in the direction of the projetion, is labeled ’projection effect threshold (MF)’. This means that for less attenuation of the direct sound, the main auditory event is no longer localized in the direction of the reflection. It should be noted that this MF threshold is only valid for a specific filter characteristics, e.g., no damping at LF and HF (’0 dB LF+HF’).

For the statistical analysis SPSS software was used. The data were checked to be valid for conducting a three-factor repeated-measures analysis of variance (ANOVA) [22, 23]. Significant effects were found for all three main factors: lowpass frequency ($p = 0.003$), highpass frequency ($p = 0.026$) and filter characteristics ($p < 0.001$). Additionally, the interaction between all factors became significant ($p = 0.005$).



(a)

(b)

Fig. 5: Projection effect threshold (direct sound level MF attenuation which was just necessary so that the position of the auditory event stayed in the direction of the projected source) for different filter characteristics of the direct sound. Mean values and 95% confidence intervals are plotted for varying the (a) lowpass frequency and (b) highpass frequency of the filter.

Figure 5 shows the mean values of the corresponding projection effect thresholds (MF) with 95% confidence intervals. In Figure 5(a) the lowpass frequency is varied. Sub-figures are plotted for the three filter characteristics. For no damping at LF and HF ('0 dB LF+HF', left sub-figure) the strongest overall MF attenuation is necessary to keep the main auditory event in the back right of the listener and hence the lowest projection effect threshold (MF) can be seen. The strongest attenuation of -13 dB, is needed at the highest lowpass frequency (1 kHz). Post-hoc pairwise comparisons (Bonferroni corrected) confirmed a significant difference between the 1 kHz condition and the two lower lowpass frequencies. This seems reasonable, because an increase of the lowpass frequency reduces the bandwidth of the mid-range frequency band. Therefore, from a spectral perspective, less localization cues are influenced when reducing the MF level. This follows the model that the localization of an auditory event is determined by the (weighted) sum of localization cues in different frequency bands [24]. In the current scenario, LF and HF cues pull the auditory event toward the front because of the first arrival of sound from that direction. At MF the attenuation of the direct sound reduces or overcomes this effect. Below the projection effect threshold, MF cues 'win' the localization competition and shift the auditory event towards the projection in the back right.

If the LF band in the direct sound is additionally attenuated, it would be expected that less attenuation is needed in the MF band. A tendency in this direction can be seen in the middle of Figure 5(a) ('-10 dB LF'). However, the effect is not significant. One could argue that the background sound in the current scenario has much energy below 500 Hz (compare Figure 4) and is, therefore, already partially masking much of the direct sound even if it is not attenuated. Additionally, very low frequencies do not contribute strong localization cues at all [25]. Either way, attenuating LF is difficult using real sound projecting speakers because of physical limitations.

If the HF band is attenuated ('-10 dB HF') a comparatively stronger effect can be seen in the right sub-figure of Figure 5(a). The MF attenuation reduces to approximately -9 dB. Interestingly, the separation frequency between the MF band and the not attenuated LF band does not influence the projection effect threshold in this case. HF cues seem to affect the localization stronger compared to LF cues using the applied filtering approach. However, the abstraction of narrow grating lobes with a general highpass filter is rather rough (compare Figure 2). Therefore, the described simplified filtering approach might overestimate the influence of high frequencies. However, the selected filter structure can be regarded as a worst case scenario. The resulting data is than representing a limiting case. Attenuating HF in the direct sound with real sound projectors is possible, e.g., using irregular speaker arrays or shadowing.

An other practical option is to shift the frequency of the first grating lobe towards higher frequencies by reducing the speaker distance. The influence of the resulting shift of the highpass frequency on the projection effect threshold (MF) is shown in Figure 5(b). For a higher highpass frequency, slightly less MF attenuation is necessary to maintain the projection effect. Shifting the highpass frequency from 6 kHz to 10 kHz increases the MF bandwidth. Therefore, more localization cues are effected by adjusting the MF level and less attenuation is necessary. This holds true independent of the general filter characteristics. Sub-figures are plotted for additional attenuation of LF and HF. With varying filter characteristics the same overall tendency can be observed in (b) compared to (a): Reducing LF in the direct sound by 10 dB (middle sub-figure) results in a slight (but not significant) increase of the projection effect threshold. Again, reducing HF by 10 dB (right sub-figure) leads to a more distinct effect. Approximately 2.5 dB less MF attenuation is necessary compared to the filter condition without LF and HF damping (left sub-figure).

The projection effect thresholds determined in this study are in the range between -13 dB and -8.5 dB, depending on the spectral characteristic of the direct sound. However, these values are only valid for the exemplary broadband stimulus applied in this pilot study. In an earlier study a comparable setup and delay was applied [16]. The direct sound filtering was even more simplified with a single lowpass or highpass at 1.6 kHz. The participants were asked to adjusted the overall level of the direct sound. Projection effect thresholds in the range of -8 dB to -12 dB have been determined for speech and music stimuli. A noise stimuli was tested in addition and seemed to be less critical, resulting in thresholds of -5 dB to -7 dB. Another study measured projection effect thresholds for different stimuli including speech. The same simplified filtering approach was used, but the corner frequencies were varied [19]. The results showed that with increasing bandwidth of the direct sound, the projection effect threshold decreased and more attenuation was necessary. A condition with almost unfiltered direct sound and a delay of 10 ms was included. For this case, a threshold of approximately -15 dB was reported for a speech stimulus. This suggest that speech, or other stimuli, might be more critical compared to the stimulus used in this study.

6 Summary and Outlook

In this study, the localization in a sound projection scenario using a realistic surround audio scene (street with traffic noise) was investigated. The projection of a car sound in idle mode behind a listener was simulated. Real sound projectors, e.g., phased arrays, are characterized by a physically limit focusing ability. This leads to multiple sound propagation paths to the listener: direct sound via the shortest path and delayed projected sound via reflections. The spectral characteristic of the disturbing direct sound was varied in a listening test to simulate realistic directivity limitations at low and high frequencies. The attenuation of the direct sound at mid-range frequencies, which was just necessary so that the position of the auditory event was localized in the direction of the projection in the back right, was determined. This attenuation is called projection effect threshold. It was found that widening the attenuated mid-range frequency band increased the projection effect threshold. This means that less attenuation of the direct sound is necessary if a wider frequency range can be focused with the sound projector. Additional damping of the direct sound at high frequencies (e.g., above 6 kHz) further reduced the necessary attenuation at mid-range frequencies. Additional damping at low frequencies (e.g., below 500 Hz) did not significantly influence the localization in the projection scenario.

Further studies are necessary to systematically investigate the localization with other stimuli using the filtering approach described here. Another filtering option would be to apply the actual characteristics of realistic sound projectors using simulations or measurements. Further, scattering and/or absorption at the reflections and distance dependent attenuation should be taken into account for all sound paths. In this study, additional background sounds were reproduced. This might have resulted in partial masking of the direct sound and could explain the small influence of damping low frequencies or changing the lowpass frequency from 250 Hz to 500 Hz. A repetition of this experiment with controlled level of the background sound or with different realistic background scenarios would give more insight. Additionally, this pilot study focused on spectral characteristics only. Further studies focusing on temporal aspects of the interacting sounds in a projection scenario are conducted at the moment.

References

[1] Katayama, M., Takumai, S., and Imaoka, K., "Speaker Device and Audio Signal Processing Method," US Patent 9,674,609 B2, 2017.

[2] *IKO - by IEM and Sonible*, http://iko.sonible.com, accessed: Nov. 24th, 2017.

[3] Litovsky, R. Y., Colburn, H. S., Yost, W. A., and Guzman, S. J., "The precedence effect," *The Journal of the Acoustical Society of America*, 106(4), pp. 1633–1654, 1999.

[4] Zagala, F., Linke, J., Zotter, F., and Frank, M., "Amplitude Panning between Beamforming-Controlled Direct and Reflected Sound," in *Audio Engineering Society 142nd Convention*, 2017.

[5] Meyer, E. and Schodder, G. R., *Über den Einfluss von Schallrückwürfen auf Richtungslokalisation und Lautstärke bei Sprache*, Vandenhoeck & Ruprecht, Göttingen, 1952.

[6] Wühle, T. and Altinsoy, M., "Investigation of auditory events with projected sound sources," in *173rd Meeting of Acoustical Society of America and 8th Forum Acusticum*, 2017.

[7] Olson, H., *Acoustical Engineering*, D. Van Nostrand Company INC., Princeton, New Jersey, Toronto, New York, London, 1957.

[8] van der Val, M., Start, E., and de Vries, D., "Design of Logarithmically Spaced Constant Directivity Transducer Arrays," *Journal of the Audio Engineering Society*, 44(6), pp. 497–507, 1996.

[9] Keele jr., D., "Implementation of Staright-Line and Flat-Panel Constant Beamwidth Transducer (CBT) Loudspeaker Arrays using Signal Delays," in *Audio Engineering Society Convention 113*, 2002.

[10] Mabande, E. and Kellermann, W., "Towards Superdirective Beamforming with Loudspeaker Arrays," in *Conf. Rec. International Congress on Acoustics*, 2007.

[11] Keele jr., D. B., "A Performance Ranking of Seven Different Types of Loudspeaker Line Arrays." in *Audio Engineering Society Convention 129*, 2010.

[12] Schmeder, A., "An exploration of design parameters for human-interactive systems with compact spherical loudspeaker arrays," in *Ambisonics Symposium, Graz*, 2009.

[13] Linkwitz, S. H., "Active Crossover Networks for Noncoincident Drivers," *Journal of the Audio Engineering Society*, 24(1), pp. 2–8, 1976.

[14] *Engine Sound*, http://archive.hoerspielbox.de/2-4-10038.mp3, accessed: Jan. 22nd, 2018.

[15] *Street Recording Warsaw*, http://16wersow.pl, accessed: Jan. 22nd, 2018.

[16] Wühle, T., Merchel, S., and Altinsoy, M., "Evaluation of auditory events with projected sound sources using perceptual attributes," in *Audio Engineering Society 142nd Convention*, 2017.

[17] Bech, S., "Spatial aspects of reproduced sound in small rooms," *The Journal of the Acoustical Society of America*, 103(1), pp. 434–445, 1998.

[18] Cardozo, B. L., "Adjusting the Method of Adjustment: SD vs DL," *The Journal of the Acoustical Society of America*, 37(5), pp. 786–792, 1965.

[19] Wühle, T., Merchel, S., and Altinsoy, M., "Untersuchungen zur Wahrnehmung projizierter Schallquellen," in *DAGA 2016 - 42nd German Annual Conference on Acoustics*, 2016.

[20] Wühle, T., Merchel, S., and Altinsoy, E., "Evaluation of the frequency dependent influence of direct sound on the effect of sound projection," in *AES Conference on Spatial Audio Reproduction*, Tokyo, Japan, 2018.

[21] Wallis, R. and Lee, H., "Vertical Sterephonic Localization in the Precsence of Interchannel Crosstalk: The Analysis of Frequency-Dependent Localization Thresholds," *Jounal of the Audio Engineering Society*, 64(10), pp. 762 – 770, 2016.

[22] Field, A., *Discovering statistics using SPSS*, volume 3, Sage, 2009.

[23] Bortz, J., *Statistik für Human- und Sozialwissenschaftler*, volume 6, Springer, 2005.

[24] Blauert, J., editor, *The Technology of Binaural Listening*, Springer, 2013.

[25] Blauert, J., *Spatial Hearing: The Psychophysics of Human Sound Localization*, MIT Press, 1997.