PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Perceptual relevance of speaker directivity modelling in virtual rooms

Henning STEFFENS[1]; Steven VAN DE PAR[2]; Stephan D. EWERT[3]

[1] CvO University, Acoustics group, Oldenburg, Germany

[2] CvO University, Acoustics group, Oldenburg, Germany

[3] CvO University, Medical Physics, Oldenburg, Germany

ABSTRACT

Sound sources like a human speaker, musical instruments and loudspeakers typically radiate sound with a frequency-dependent directivity pattern. If a human speaker faces the receiver, more high-frequency components reach the receiver as when the speaker turns in the opposite direction. This is perceptually easily detectable, particularly in anechoic conditions. In rooms, early reflections and late diffuse reverberation, which both can be differently affected by the directivity pattern of the source, occur in addition to the direct sound. While source directivity appears relevant for distinct early reflections, it is unclear to which extent the perceptual effects average out for later reflections and diffuse reverberation.

Here we assess to what extent directivity of a human speaker should be modelled for early reflections and diffuse reverberation. For this we used a hybrid approach to simulate and auralize binaural room impulse responses [Wendt *et al.*, JAES, 62, 11 (2014)], combining an image source model (ISM) for early reflections with a feedback delay network for diffuse reverberation. The approach was extended by a directivity filter based on a spherical head model. Listeners distinguished between source orientations solely based on ISM-directivity and detected average source directivity filtering even in reverberant conditions.

Keywords: Source directivity, room acoustics simulation, psychoacoustics

1. INTRODUCTION

Directivity of a human speaker in a room can be easily demonstrated when the speaker turns in the opposite direction of a listener. The directivity pattern of the head and mouth radiate high frequencies more directed to the front while low frequencies are radiated omni-directionally. Depending on the acoustic properties of the room, head rotation of the source causes changes in the spectral pattern of the early reflections in addition to changes in the direct sound, most notably in the high frequencies. For later reflections and diffuse reverberation, the effect of source directivity can be assumed to be more spatially averaged, due to the high number of spatially more even distribution of reflections in the acoustic sound field of the room.

While interaural time differences (ITD), interaural level differences (ILD) and the direct-to-reverberant ratio (DRR) influence the perception of a source's location and distance, interaction effects between spectral and spatial cues for source directivity perception are still open questions.

Perception of sound source orientations was investigated by Neuhoff and colleagues. In (1) listeners had to indicate the orientation of a loudspeaker in front of them in a semi-anechoic room with broadband noise as test signal. Results showed that test subjects were sensitive to the facing angle of a directional sound source with better performances at closer distance and the source directly facing the listeners.

Nakano *et al*. (2) performed experiments with a human speaker in real and anechoic environments. With an average horizontal orientation error of 24.5° in anechoic and 33.8° in realistic (unfurnished) conditions, they concluded that it is more difficult to perceive orientation of a speaker in a real

[1] henning.steffens@uni-oldenburg.de
[2] steven.van.de.par@uni-oldenburg.de
[3] stephan.ewert@uni-oldenburg.de

environment than in an anechoic room. Their results match with Kato *et al.* (3) and they found no significant training effects for central frontal positions.

Edlund *et al.* (4) found that judging different facing angles of a human speaker is also possible with the presence of background noise in furnished, asymmetric rooms and tends to be easier under certain geometrically conditions (i.e. reflections of a nearby window) than in anechoic rooms.

In the above studies, important perceptual cues for the orientation of the sound source are changes in the coloration of the direct sound or ILDs introduced by the effect of directivity on sound reflections.

Modelling the acoustic properties of source directivity within computer-generated auralizations had been identified as an important factor for a perceptually plausible experience. Otondo (5) showed that the directivity of a musical instrument has a direct influence on the distribution of acoustical parameters in a room, which is perceptually perceivable. This is supported by studies from Vigeant and colleagues (6). A recent study by Budnik *et al.* (7) showed significant improvements when comparing artificial-head-based binaural recordings of human speech in a classroom scenario with room simulations including realistic speaker directivity with an omni-directional source.

While source directivity appears relevant for distinct early reflections, it is unclear to which extent the perceptual effects average out for later reflections and diffuse reverberation and could therefore be simplified for source directivity in room acoustics simulations. The current study assessed this question by conducting two listening experiments with a human speaker as sound source in two simulated rooms. Source directivity was simulated with a simple directional filter approach.

2. METHODS

A hybrid approach proposed by Wendt *et al.* (8; room acoustics simulator, RAZR) to simulate and auralize binaural room impulse responses (BRIRs) was used in the current study. RAZR combines an image source model (ISM) by Allen and Berkley (9) for calculation of early reflections with a feedback delay network (FDN) by Jot and Chaigne (10) for diffuse reverberation. This method achieves perceptually plausible simulations with relatively fast computation time. The auralization uses headphones and considers the rotation of the listener's head.

The framework was extended with the option to include rotation of a virtual source as well as a source directivity filter to account for the spectral directivity pattern of the source.

2.1 ISM Source Directivity

In a closed-form ISM for a shoebox-shaped room, the exit angles of each ISM source can be gained by calculating the vector between source and receiver from their relative positions and orientation. As all walls are orthogonal and can be chosen to be axis aligned, either the x, y or z dimension has to be considered for each reflection. This is done by mirroring the initial vector depending on the ISM order. From this the source exit angles can be easily obtained - for further information see (11).

The directivity of a human speaker was modelled with a simplified rotationally-symmetric source filter based on the spherical head model (SHM) by Duda *et al.* (12). Originally intended for the receiver to approximate head related transfer functions, here it is used to approximate the directivity of the mouth of the speaker as a sound source located on a sphere representing the head.

For the technical implementation a head-shadowing filter found in (13) was used. Figure 1 shows the frequency responses for different input angles. Note that the filter response is normalized to be flat at zero degree (frontal direction of the source). For the back side of the source an attenuation of about 20 dB is observed 10 kHz.

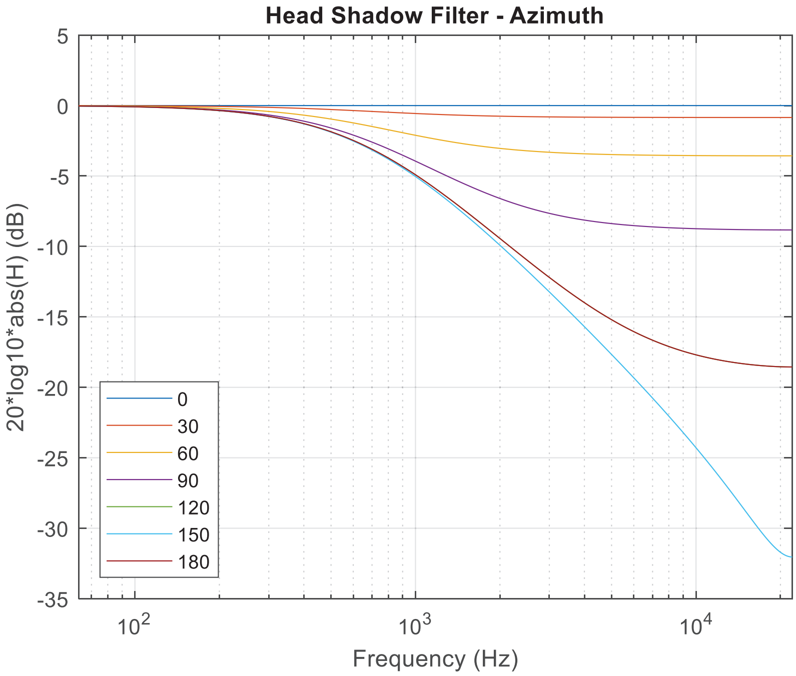


Figure 1 – Frequency responses for different azimuth angles of the normalized SHM-based head-shadow filter.

2.2 Average Source Directivity Filter

Given an increasing number of reflections from many directions overlay for later reflections and late reverberation, it is interesting to consider the radiated power of the source averaged across all directions, expressed as an average source directivity filter. For this, the power spectral densities (PSD) of the source directivity filter were calculated for azimuth angels between zero and 180° in 2.5° steps over frequency.

$$S_{xx}(f)_{\text{azimuth}} = |H(f)_{\text{SHM}}|^2{}_{\text{azimuth}} \tag{1}$$

These PSDs were weighted over a sphere that was subdivided into ring surfaces with a width of 2.5°.

$$A_{\text{Sphere}} = \int dA_{\text{i}\,2.5} = \int_0^{\pi}\int_0^{2\pi} \sin(\theta)d\theta d\varphi \tag{2}$$

$$dA_{\text{i}\,2.5} = r^2 \sin(\theta)d\varphi \tag{3}$$

$$\hat{S}_{xx}(f)_{\text{weighted}} = S_{xx}(f)_{\text{azimuth}} * dA_{\text{i}\,2.5} \tag{4}$$

The weighted PSDs per ring surface were summed and divided by the whole sphere surface for calculation of an average transfer function:

$$\hat{H}(f)_{\text{Average}} = \sqrt{\left(\sum \hat{S}_{xx}(f)_{\text{weighted}}\right)/A_{\text{Sphere}}} \tag{5}$$

Finally, the inverse Fourier transformation of the average transfer function was calculated. The resulting frequency response of average source directivity filter is shown in Figure 2.

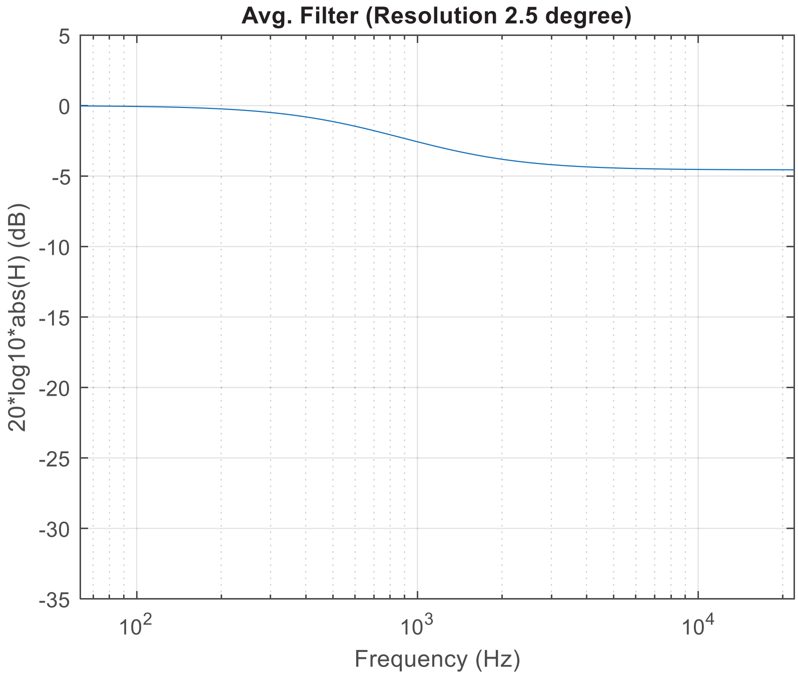


Figure 2 – Frequency response of the average source directivity filter based on the SHM.

2.3 Virtual Rooms

Two virtual shoebox-shaped rooms were used for the experiments: A small room ("Room L") with a shorter reverberation time and a large room ("Room A") with properties summarized in Table 1. The setup of the virtual rooms with source and receiver locations is shown in Figure 3.The distance of source and receiver equals the respective critical distances (DRR of 0 dB for an omni-directional source) in the rooms.

Table 1 – Virtual Rooms

Name	Size [x, y, z], m	Volume, m^3	T60, s	Crit. distance, m
Room L: "Lab."	4.97, 4.12, 3	61.4	~ 0.3	~ 1
Room A: "Aula"	12, 30, 10	3600	~ 4.7	~ 2.5

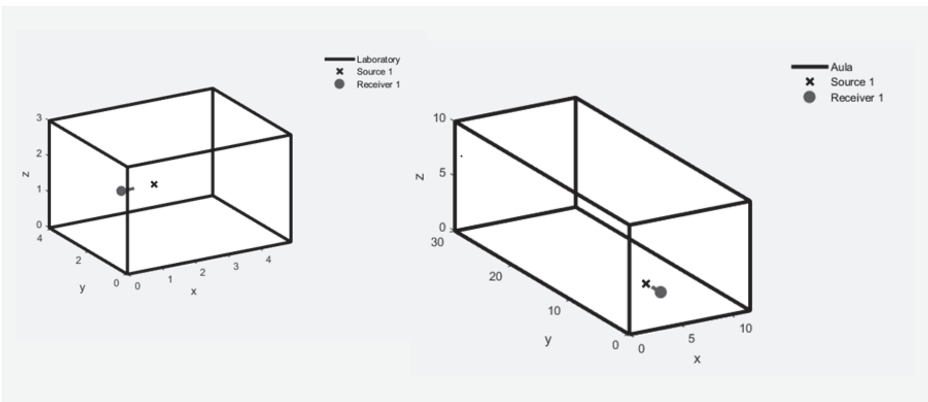


Figure 3– Setup of virtual rooms, left: Room L, right: Room A.

3. LISTENING TEST I – IDENTIFICATION

A listening test with simulated ISM source directivity was conducted to investigate if listeners were able to distinguish between different virtual source orientations. The tasked was to identify the correct orientation of the source among four possible orientations (“Front”, “Left”, “Back”, “Right”). If the speaker was facing the listener, “Front” would have been the correct answer.

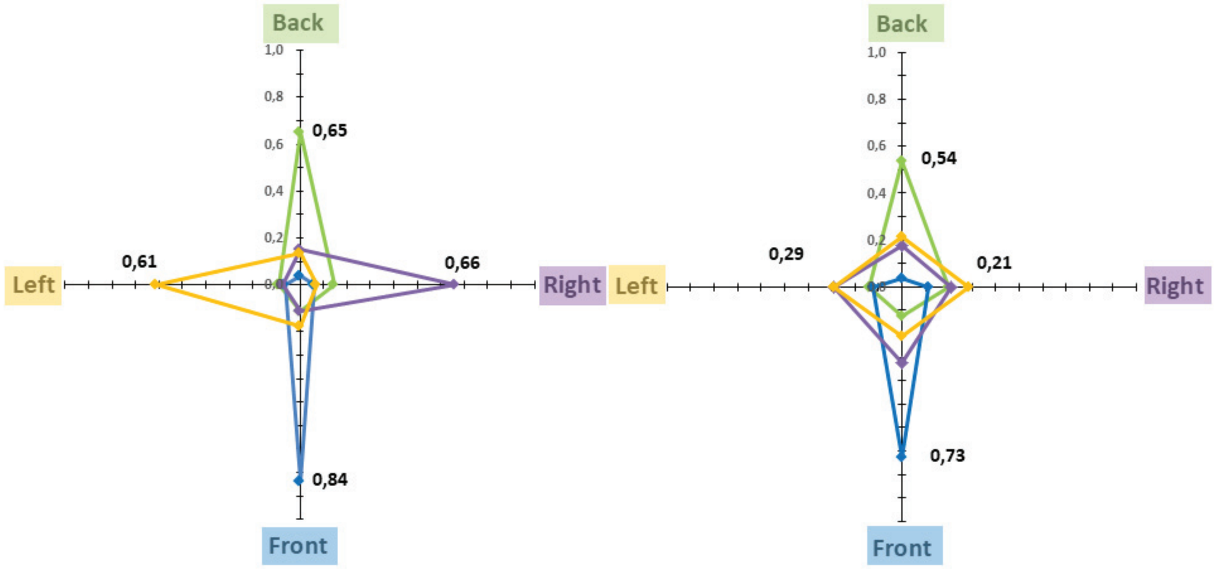
Stimuli were played successively over headphones and the listeners had to provide a response via four GUI buttons to continue. Eight trials per randomized orientation and virtual room were presented and no feedback was provided. Five training trials per orientation and room (different sentence, with feedback) were given before the test phase. Ten normal hearing listeners participated in the test (5 female, 5 male, age 24-38, mean age 28.5).

3.1 Stimuli

An example sentence from the German Oldenburg Sentence Test (OLSA) (14) -“Britta bekommt elf nasse Steine”-, spoken by a male speaker was auralized in central position via a third-order ISM in the above mentioned approach in two virtual rooms (Table 1) for each orientation. For the training phase a different sentence spoken by the same speaker was used.

3.2 Results

The graphs in Figure 4 show the distributions of all answers derived from the confusion matrices below, with a value of 1 representing 100 percent correct answers for one condition.



Room L	Front	Left	Back	Right
"Front"	**0,84**	0,18	0,11	0,11
"Left"	0,06	**0,61**	0,09	0,08
"Back"	0,04	0,14	**0,65**	0,15
"Right"	0,06	0,08	0,15	**0,66**

Room L

Room A	Front	Left	Back	Right
"Front"	**0,73**	0,21	0,13	0,33
"Left"	0,13	**0,29**	0,14	0,29
"Back"	0,04	0,21	**0,54**	0,18
"Right"	0,11	0,29	0,20	**0,21**

Room A

Figure 4 – Identification listening test results, left: Room L, right: Room A.

There were more correct identifications for each orientation in Room L than in Room A. In both virtual rooms the ‘Front’ condition was identified best whereas the identification rates of the ‘Left’/’Right’ conditions differ highly between rooms. Incorrect answers are evenly distributed.

3.3 Discussion

The results indicate that listeners are able to distinguish between different source orientations in the two simulated rooms with the proposed source directivity filter approach. Identification of ‘Left/Right’ orientations benefits from the small Room L with strong and shortly arriving (about 13 ms) early reflections, whereas in the large and more reverberant Room A, reflections are stronger attenuated (and need longer to reach the listener) because of the larger traveling distances.

For further statistical analysis, a resampling-bootstrap method was applied (15, 16), where it is assumed that all samples stem from the same underlying distribution and the difference in means reflects the sampling variation for that distribution.

The frequency of observed differences that were as large as, or larger than the actual difference in means derived from randomly sampling the combined distribution of the Room L and Room A conditions was calculated (1000 iterations). Here identification rates for 'Left' & 'Right' showed to be significantly different ($p < 0.05$) between rooms.

4. LISTENING TEST II – DETECTION

In a second listening test, the question was approached if the simulation of source directivity can be simplified by using the average source directivity filter at higher ISM orders, where a number of spatially distributed reflections overlap. The hypothesis was that the effect of individually simulated directivity of each image source is not perceivable for higher ISM orders compared to applying the averaged source directivity.

Listeners judged in an ABX experiment whether the third stimulus out of three (ABA, ABB) was the same as the first or the second one (X = A or X =B). Listeners had to provide their response via two GUI buttons to continue. Sixteen trials per virtual room over two runs were presented with feedback and no training phase. Ten normal hearing listeners participated in the test (2 female, 8 male, age 25-38, mean age 29.8).

4.1 Stimuli

The "Right"-stimulus from the identification test was chosen for this experiment as reference. The general auralization parameters were the same as in the identification experiment. In the test stimulus, instead of correctly filtering all ISM stages with the source directivity filter (2.1), the average source directivity filter (2.2) was utilized, starting from the third ISM order. This effectively applied the average directivity to the later reflections and diffuse reverberation while maintaining the correct directivity in the direct sound and the early reflections.

4.2 Results

Figure 5 shows the percentages of the correct answers (median and interquartile ranges) for detecting the reference and test stimulus both virtual rooms. A percentage of 50% indicates chance level. The median results are slightly above 60% and below the binomial significance level indicated by the horizontal dashed line.

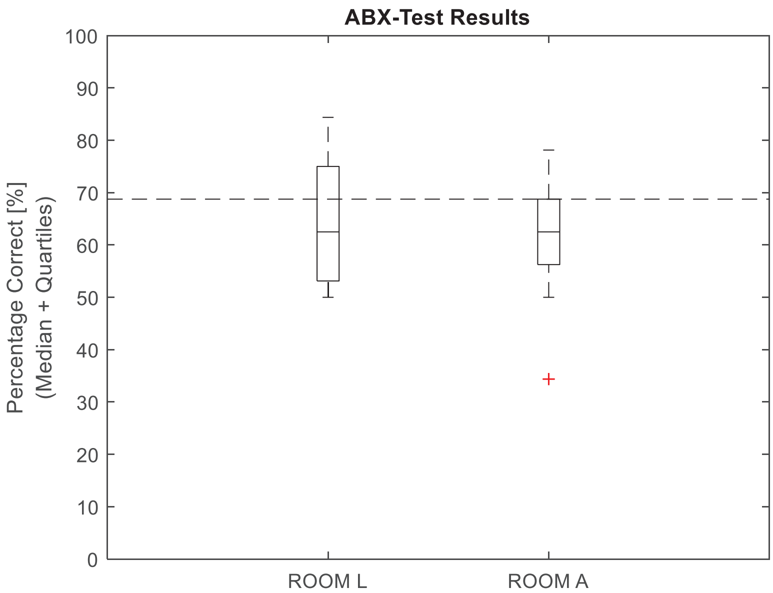


Figure 5 – Detection listening test results. The dashed horizontal line indicates the binomial significance level.

To further analyze the results, a method proposed by Boley and Lester (17, 18, 19) was used to determine the sensitivity index d’ as well as the 95% confidence interval (CI) per listener. The number of listeners that detected a difference in the ABX experiment was determined by d’ ± 95% CI > 0. Two listeners (4 & 7) were able to detect a difference in both conditions. One listener (9) did not detect any difference in both conditions. Results of the remaining listeners show no clear tendency towards one particular condition.

Table 3 – Calculated d’ ± 95% confidence per listener

Listener No.	Room L	Room A
1	0.880±1.614	1.270±1.053
2	2.360±1.547	1.780±6.252
3	0.000±8.840	2.230±0.381
4	1.190±0.775	1.490±0.758
5	2.880±0.324	1.490±1.577
6	0.600±1.715	0.920±0.808
7	2.720±0.258	1.830±0.561
8	1.460±0.878	1.490±1.577
9	0.520±6.439	0.000±1.053
10	1.510±6.254	1.270±1.053
Σ(d’±95% > 0)	5	4

4.3 Discussion

Overall the results do not allow to state that the effect of individual directivity against averaged source directivity for the third-order image sources and late reverberation is not perceivable. Five listeners (Room L) and four listeners (Room A) out of ten distinguished both conditions. However, a slight tendency that detection is more difficult in the larger room A was observed.

It should be noted that the here employed virtual setup and stimuli just reflect one specific condition out of a multitude of potential naturally occurring room conditions and source and receiver locations. For the current two conditions, about half of the listeners could not detect any difference between the more exact and approximated average simulation of the effect of source directivity in the third-order reflections and late reverberation.

5. SUMMARY

A hybrid approach to simulate and auralize BRIRs was extended with a simplified source directivity filter. Two experiments were performed to investigate perceptual effects of source directivity modelling.

Despite the very straightforward approach, experimental results show that a perception of source directivity can be achieved by a simple spectral filter within the ISM. The best performance was observed in the smaller room with strong early reflections from nearby walls.

A spatially averaged source directivity filter applied to later reflections and diffuse reverberation was still detected by half of the listeners. While perceptually detectable in an ABX comparison test, future research has to show which cues were used for detection and how relevant directivity modelling for higher-order reflections is regarding perception of source orientation and evaluation of, e.g., speech intelligibility in simulated room conditions.

REFERENCES

1. Neuhoff J, Rodstrom M, Vaidya T. The audible facing angle. Acoustics Research Letters Online. 2001;2(4):109-114.

2. Nakano A, Nakagawa S, Yamamoto K. Auditory perception versus automatic estimation of location and orientation of an acoustic source in a real environment. Acoustical Science and Technology. 2010;31(5):309-319.
3. Kato H, Takemoto H, Nishimura, Mokhtari, P. Spatial acoustic cues for the auditory perception of speaker's facing direction. Proc ICA 2010; 23-27 August 2010; Sidney, Australia 2010. p. 3279-86.
4. Edlund J, Heldner M, Gustafson J. On the effect of the acoustic environment on the accuracy of perception of speaker orientation from auditory cues alone. In: INTERSPEECH 2012 : vol2 [Internet]. Portland, USA: Curran Associates, Inc.; 2012. p. 1482–5.
5. Otondo F, Rindel JH. The influence of the directivity of musical instruments in a room. Acta Acustica united with Acustica. 2004 May 1;90(6):1178-1184.
6. Vigeant M, Wang L. Subjective evaluation of auralizations created from multi-channel anechoic recordings of a talker in motion. The Journal of the Acoustical Society of America. 2005;117(4):2465-2465.
7. Budnik A, Steffens H, Ewert SD, van de Par S, Blau M. Dynamische binaurale Auralisierung von Klassenraumakustik - Einfluss der Quellenrichtcharakteristik. Fortschritte der Akustik - DAGA 2019; Rostock, Germany 2019. p. 1450-53.
8. Wendt T, van de Par S, Ewert SD. A Computationally-Efficient and Perceptually-Plausible Algorithm for Binaural Room Impulse Response Simulation. Journal of the Audio Engineering Society. 2014;62(11):748-766.
9. Allen J, Berkley D. Image method for efficiently simulating small-room acoustics. The Journal of the Acoustical Society of America. 1979;65(4):943-950.
10. Jot J-M, Chaigne A. Digital delay networks for designing artificial reverberators. AES 90th Convention; Paris, France 1991.
11. Brinkmann F, Erbes V, Weinzierl S. Extending the closed form image source model for source directivity. Fortschritte der Akustik - DAGA 2018; Munich, Germany 2018. p.1298–1301.
12. Brown C, Duda R. A structural model for binaural sound synthesis. IEEE Transactions on Speech and Audio Processing. 1998;6(5):476-488.
13. Zölzer, U. DAFX. Chichester: J. Wiley and Sons; 2011
14. Wagener, K., Kühnel, V., Kollmeier, B.; 1999 Entwicklung und Evaluation eines Satztests für die deutsche Sprache I: Design des Oldenburger Satztests. Zeitschrift für Audiologie/Audiological Acoustics, 38(1):4-15.
15. Simon JL, Bruce P. The new biostatistics of resampling. M.D. computing : computers in medical practice. 1995;12(2):115-21, 141
16. Chernick MR. Bootstrap Methods: A Guide for Practitioners and Researchers, Second Edition. 10.1002/9780470192573; 2008.
17. Boley J, Lester M. Statistical analysis of ABX results using signal detection theory. AES 127th Convention; New York, NY, USA 2009.
18. Macmillan N, Creelman C. Detection theory. Mahwah, N.J.: Lawrence Erlbaum Associates; 2005.
19. Kaplan H, Macmillan N, Creelman C. Tables of d′for variable-standard discrimination paradigms. Behavior Research Methods & Instrumentation. 1978;10(6):796-813.