Proceedings of ICSA 2019

5th International Conference on Spatial Audio
September 26th to 28th, 2019, Ilmenau, Germany

DOI: 10.22032/dbt.39966

DOI (proceedings): 10.22032/dbt.39936

Full Reviewed Paper at ICSA 2019

Presented * by VDT.

Perceptual Evaluation of Spatial Resolution in Directivity Patterns 2: coincident source/listener positions

Matthias Frank[1], Manuel Brandner[1]
[1] *Institute of Electronic Music and Acoustics, University of Music and Performing Arts Graz, Austria*
Email: frank@iem.at, brandner@iem.at

Abstract

The incorporation of source directivity is important for a plausible and authentic auralization. While high-resolution measurement setups and data exist, it is yet not clear how detailed the directivity information has to be measured and reproduced with regard to perception. In particular, when source and listener are at the same location, resulting in a high direct-to-reverberant energy ratio, the precise shape of the directivity pattern might not yield perceptual differences. The paper approaches this question by a listening experiment in a virtual environment with generic directivity patterns and coincident position of listener and source. The experiment compares different spatial resolutions (spherical harmonic orders) of the directivity patterns for multiple virtual listener/source positions/orientations and levels of direct sound for speech and noise. The virtual environment employs a higher-order image-source model and binaural, dynamic Ambisonic playback. The results show that the exact shape of the directivity pattern is often perceptually irrelevant, while the preservation of the direct-to-reverberant energy ratio is more important.

1. Introduction

Plausible and authentic auralization of sound sources in rooms benefits from the incorporation of source directivity and variable source orientation [1]. This is mainly due to the natural perception of distance that is controlled by the direct-to-reverberant energy ratio (DRR) [2, 3]. High-resolution measurement of source directivity is typically done with surrounding microphone arrays of up to 64 microphones at the same time [4] and directivity patterns are often represented in spherical harmonics to facilitate simple rotation. A high resolution is sometimes necessary to compensate for imprecise centering [5, 6], even for sources with low spatial resolution in their directivity patterns. Our previous study [7] revealed that perception of spatial resolution in directivity patterns is limited to spherical harmonic orders around 4 for large distances between source and receiver in a stimulated concert hall. In such cases, the DRR is typically negative.

However, for the auralization of one's own voice or when playing an instrument oneself [8–10], direct sound dominates. This paper investigates how precise directivity patterns are perceived in such cases, i.e. to which order a higher-order directivity pattern can be reduced to still be perceptually indistinguishable from a reference. The reference directivity pattern is highly directive as it appears for large brass instruments at high frequencies. The investigation employs a high level of direct sound as it appears in human speaking/singing and further, reduced levels to represent instruments with less direct sound at the player's ears. The virtual room in which the directional source is playing is simulated by an image-source model without late diffuse reverberation. These settings are chosen to simulate the most sensitive case, whereas a practical application might be less critical.

The paper first introduces setup and conditions of the listening experiment. The following section presents the experimental results. The results are subsequently compared to technical measures that are related to room acoustics and properties of the directivity patterns. Finally, the investigation is summarized and compared to our previous results in [7] for non-coincident listener and source.

© Verband Deutscher Tonmeister e.V., 2019

2. Setup and Conditions

The parameters of the room simulation were identical to those used in [7]: The room had a size of 30 m × 20 m × 10 m and a reverberation time of 1.9 s between 200 Hz and 2 kHz, and was doubled/halved for frequencies below 100 Hz and above 4 kHz, respectively. The simulation employed a 7th-order image-source model (236 reflections) implemented in the IEM `RoomEncoder` VST plug-in[1]. The headphone playback employed 7th-order head-tracked [11] binaural Ambisonics [12] using the IEM `BinauralDecoder`. Note that the rotation of the source was linked to the head rotation.

The direct sound was not generated by the `RoomEncoder` plug-in, as this would result in an infinitely high level for coincident source and receiver position. Therefore, it was realized as omnidirectional sound inside the listener's head with a specific level that should correspond to direct sound level at a speaker's own ears. The level is based on a measurement of a B&K HATS 4128 using its mouth simulator, its ears, and two omnidirectional microphones at 1 m and 25 mm distance (mouth reference point, MRP) from the mouth in an anechoic chamber, respectively. Fig. 2 shows that the level at the ears is roughly 20 dB less than at the MRP. These results are similar to findings in [8] and the deviations can be explained by different distances of the MRP. The level in 1 m distance is again about 10 dB less than at the ears. A broad-band level difference of 10 dB was used to calibrate the direct sound and the image-source model for the experiment and is denoted as 0 dB direct sound level in the remainder of this paper. In order to represent instruments with less direct sound at the player's ears, reduced levels {-10, -20} dB were also evaluated.

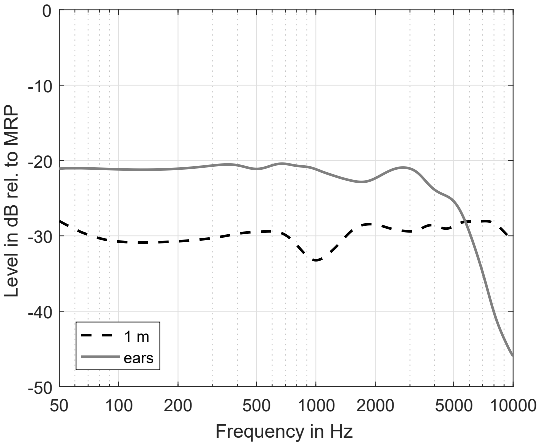


Fig. 2: Sound pressure levels in 1 m in front of the HATS and at its ears relative to the mouth reference point (MRP).

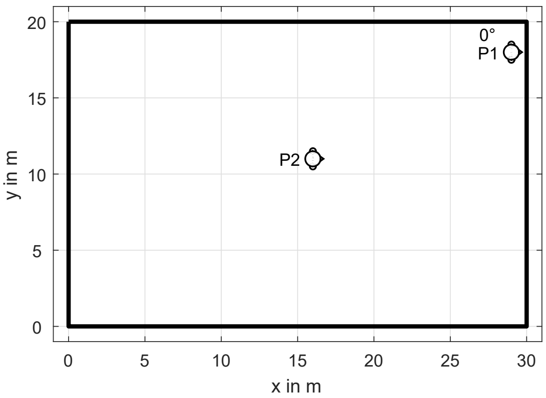
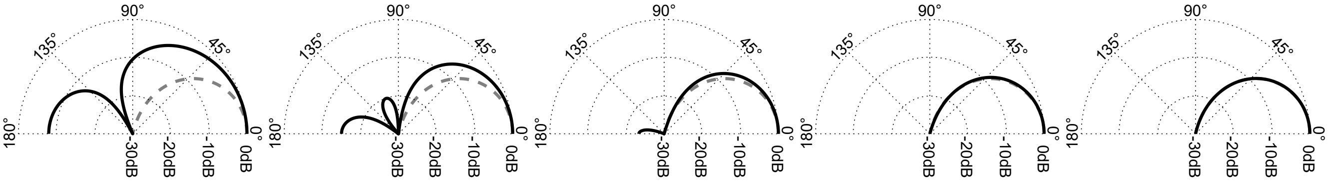


Fig. 3: Listener/source position in the horizontal cross section of the simulated room. Indicated listener/source orientation is defined as 0°.

The source and the listener were positioned coincidentally with a height of 4 m above the floor at positions P1 and P2. P1 was close to a wall to provoke a strong first reflection that could interfere with the direct sound, when the source/listener was facing the wall (0° orientation), cf. Fig. 3. In contrast, the second orientation (180°) at P1 yielded weaker reflections. For P2, which was close to the center of the room, the reflection pattern was less orientation-dependent. Thus, there was only one orientation evaluated at P2.

The reference directivity was a 7th-order inphase [13] design, resulting in no side lobes and a relatively narrow main lobe, cf. Fig. 1. This directivity is similar to that of larger brass instruments, e.g. trombones or tubas, at high frequencies [14]. Typical directivity patterns of other instruments can be assumed to be less directive. In the experiment, the reference directivity pattern was reduced to orders 0 to 5 by simple truncation, as our previous study [7] revealed truncation to be perceptually better than preservation of nulls. Orders higher than 5 were excluded, as they were perceived as identical to the reference in preliminary tests. All resulting directivity patterns were diffuse-field equalized. The experiment employed two different sounds: (a) continuous pink noise for maximum sensitivity to coloration and (b) male English speech [15] that facilitates better spatial perception and familiarity.

Overall, there were 18 = 2 (sounds) × 3 ({0, -10, -20} dB direct sound level) × 3 (2 orientations at P1 + 1 orientation at P2) trials with multi-stimulus comparisons. The listeners task was to compare the similarity of the 6 (0th to 5th order truncation) stimuli to the corresponding 7th-order reference on a continuous scale from *very different* to *identical*. Note that the playback level in each trial was adjusted reversely to the level of the direct sound in order to achieve similar loudness between the trials.



(a) truncated to order 1 (b) truncated to order 2 (c) tunrcated to order 3 (d) truncated to order 4 (e) truncated to order 5

Fig. 1: On-axis equalized directivity patterns of beams in the experiment; gray dashed line indicates 7th-order inphase beam as reference.

---

[1] freely available at plugins.iem.at

3. Results

There were 10 experienced listeners (average age 31 years) who spent about 21 min each on the entire experiment. Based on the 10 values for each condition, the results of the experiment are presented as median values and corresponding confidence intervals in Figs. 4 and 5 for noise and speech, respectively. The gray level of the markers and lines in the figures indicates the level of the direct sound. Obviously, the similarity to the reference increases with the truncation order and also with the level of the direct sound for both sounds and all positions/orientations.

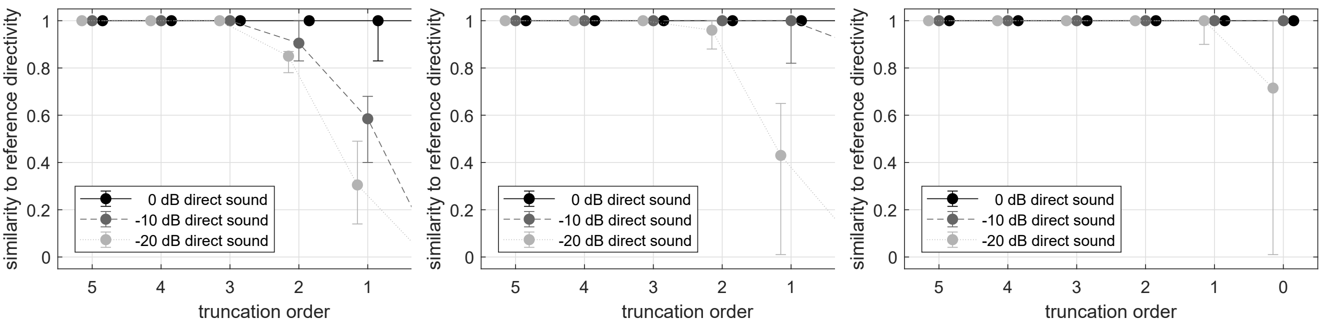
As we were interested in the spatial resolution required for perceptually indistinguishable auralization in comparison to the reference, Tab. 1 provides a suitable and easy-to-read representation of the results: For each condition, the table shows the minimum required order to yield indistinguishable results in terms of a Wilcoxon signed rank test with Bonferroni-Holm correction.

The influence of the direct sound can be seen clearly: While at the lowest level (-20 dB) orders around 2 are required, results are perceptually indistinguishable from the reference already for an order of 0 at the highest level (0 dB) for all conditions except speech at P1 and 0° orientation. This indicates that for dominant direct sound, the exact control of the reflections by the directivity pattern is not important as long as the direct-to-reverberant energy ratio is preserved. This seems to be already assured by the diffuse-field equalization of the truncated directivity patterns.

The sensitivity of the noise conditions increases with the proximity and orientation towards the walls: The central position P2 is most distant to all walls and it requires only an order of 1 or 0 for -20 dB or -10 dB direct sound, respectively. When facing the close wall at P1 and 0° orientation, orders of 3 and 2 are required for the same level of direct sound. In contrast, there is no dependency on the listener/source position and the orientation for speech, except for the increased sensitivity at P1 with 0° orientation. The increased sensitivity of noise in comparison to speech at P1 for 0° orientation and -20 dB direct sound is due to a strong comb filter. As listeners reported after the experiment, the truncation led to different strength of comb filters for noise, while it led to different level and density of reverberation for speech.

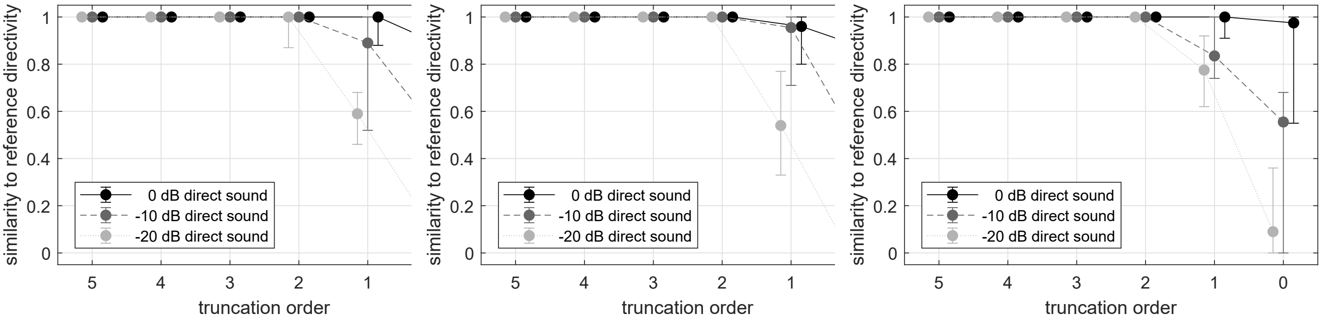
Tab. 1: Minimum required order to be indistinguishable from reference at 5% level with Bonferroni-Holm correction.

	0 dB direct sound			-10 dB direct sound			-20 dB direct sound		
sound	P1, 0°	P1, 180°	P2, 0°	P1, 0°	P1, 180°	P2, 0°	P1, 0°	P1, 180°	P2, 0°
noise	0	0	0	2	1	0	3	2	1
speech	1	0	0	1	1	1	2	2	2



(a) position P1, orientation 0° (towards the wall) (b) position P1, orientation 180° (into the room) (c) position P2, orientation 0° (into the room)

Fig. 4: Medians and 95% confidence intervals of perceived similarity to auralization using $7^{\text{th}}-$order reference directivity for noise at different listening/source positions/ and orientations for different levels of direct sound.



(a) position P1, orientation 0° (towards the wall) (b) position P1, orientation 180° (into the room) (c) position P2, orientation 0° (into the room)

Fig. 5: Medians and 95% confidence intervals of perceived similarity to auralization using $7^{\text{th}}-$order reference directivity for speech at different listening/source positions and orientations for different levels of direct sound.

© Verband Deutscher Tonmeister e.V., 2019

4. Technical Measures

This section calculates some technical measures in order to generalize the experimental results for application on different room settings and directivity patterns.

The first kind of technical measure is the direct-to-reverberant energy ratio (DRR) and it depends on the combination of the directivity pattern, its orientation, the listener/source position, the direct sound level and the room. Note that in our calculation of DRR, the first reflections also contributed to the reverberant energy. Tab. 2 shows the resulting values in dependence of the direct sound level. Naturally, the DRR increases with the level of direct sound. The 0° orientation at P1 results in values about 16 dB lower than for the 180° orientation and P2 because it yields a strong first reflection from the nearby wall. In this case, the DRR increases for truncated orders due to a reduction of the reflection from the wall, i.e. the diffuse-field equalized directivity patterns radiate more energy into all other directions away from the wall. A similar, however weaker behavior can be seen at P2. In contrast, order truncation of the directivity pattern reduces the DRR for the 180° orientation at P1. Here, the lower-order patterns lead to an increase of the energy from the nearby wall that in turn reduces the DRR values.

Tab. 2 relates to the experimental results by printing values in bold that resulted in indistinguishable results for speech. For reference DRR values around 40 dB, deviations of around 4 dB were not perceivable. For values around 30 dB, deviations must not exceed 2 dB to remain perceptually irrelevant. Similar sensitivity can be found for DRR values around 0 dB. The tendency that sensitivity decreases towards higher DRR agrees with literature [16]. However, there are exceptions, where the threshold is smaller (P2, 0° with -20 dB direct sound: below 1 dB). This might be due to the different strategies for creating the stimuli: In [16], the direct sound was attenuated/boosted and the rest of the impulse response was kept identical. In our experiment, the modification of the directivity patterns modified the impulse response but the direct sound remained the same. In this way, we did not directly modify the level ratio between direct sound and reverberation, but the level of each reflection in the impulse response.

Tab. 3: Side lobes, beam width, and front-to-back energy ratio of the tested directivity patterns.

directivity order	side lobe in dB	width in °	F/B-R_{25} in dB
7 (ref)	$-\infty$	71	19.8
5	-49.1	72	19.8
4	-34.2	74	19.9
3	-23.4	81	20.1
2	-15.1	99	15.9
1	-8.0	147	9.5
0	0	360	0

The second kind of technical measures is independent of the room and the listener/source position because it solely depends on the directivity pattern itself. The measures are (a) side lobe: level of the strongest side lobe in dB, (b) width: aperture angle of the cap exceeding -6 dB relative to the maximum at the 0° direction in °, and (c) F/B-R_{25}: front-to-back ratio in dB, with lower dynamic limitation at -25 dB relative to the maximum [7].

Tab. 3 shows the above-mentioned measures for the reference directivity and the directivities truncated at different orders. For -20 dB direct sound, the minimum required order for speech was 2. In this case, a side lobe attenuation of around 15 dB was not distinguished from the reference, a widening of the beam of 28° or 39%, and a F/B-R_{25} difference of 3.9 dB. For noise under the most sensitive conditions, the required 3^{rd} order resulted in a side lobe attenuation of around 23 dB, a widening of the beam of 10° or 14%, and a F/B-R_{25} difference of 0.3 dB. Speech at -10 dB and all position/orientations, as well as at 0 dB at P1 with 0° orientation, required an order of 1, resulting in a side lobe attenuation of 8 dB, a widening of the beam of 76° or 107%, and a F/B-R_{25} difference of around 10 dB. All other conditions with 0 dB direct sound did not require any modeling of the reference directivity except for diffuse-field equalization.

Tab. 2: Direct-to-reverberant energy ratio of the tested directivity patterns at the listener's ears in dB for all listen/source positions and orientation. Values that resulted in indistinguishable results for speech are printed bold.

directivity	0 dB direct sound			-10 dB direct sound			-20 dB direct sound		
order	P1, 0°	P1, 180°	P2, 0°	P1, 0°	P1, 180°	P2, 0°	P1, 0°	P1, 180°	P2, 0°
7 (ref)	22.4	38.7	38.9	12.4	28.7	28.9	2.4	18.7	18.9
5	**22.5**	**38.7**	**38.9**	**12.5**	**28.7**	**28.9**	**2.5**	**18.7**	**18.9**
4	**22.7**	**38.7**	**38.9**	**12.7**	**28.7**	**28.9**	**2.7**	**18.7**	**18.9**
3	**22.9**	**37.9**	**38.7**	**12.9**	**27.9**	**28.7**	**2.9**	**17.9**	**18.7**
2	**24.5**	**36.1**	**39.0**	**14.5**	**26.1**	**29.0**	**4.5**	**16.1**	**19.0**
1	**27.8**	**33.9**	**40.8**	**17.8**	**23.9**	**30.8**	7.8	13.9	20.8
0	34.1	**34.1**	**42.4**	24.1	24.1	32.4	14.1	14.1	22.4

© Verband Deutscher Tonmeister e.V., 2019

5. Conclusion

This paper evaluated the perceptual effect of reducing the spatial resolution (maximum spherical harmonics order) in directivity patterns for coincident source and listener position in a virtual room. For maximum sensitivity, the room simulation employed a higher-order image-source model without late diffuse reverberation and used dynamic binaural playback including head tracking that also controlled the orientation of the source. For the same reason, the reference directivity pattern was highly directive and the level of the direct sound was high, such as in human speech. The direct sound was played back omnidirectonal, i.e. inside the listener's head and the evaluation also included conditions with reduced direct sound to simulate other instruments.

In comparison to our previous experiment [7] with non-coincident listener/source positions, the perceptual influence of the reduction in spatial resolution was less pronounced, i.e. lower spherical harmonic orders were required to produce perceptually indistinguishable results from the reference. This could be attributed to the dominance of the direct sound in the new experiment. Thereby, reducing the direct sound increased the minimum required orders from 0 to 2, on average. This result agrees with the literature [16], where the sensitivity of the direct-to-reverberant energy ratio (DRR) is highest for values around 0 dB and decreases towards large absolute values of the DRR. Although the reduction of the spatial resolution yields an increase in beam width and reduction of side-lobe attenuation, the DRR is often well preserved, especially at the central listener/source position and direct sound levels as in human speech. In such cases, the diffuse-field equalization of the reduced-order directivity patterns might already be good enough. However, when facing a nearby wall and with less direct sound, the preservation of the directivity pattern is more important. The perceptual effect of the order reduction seems to be signal-dependent: coloration for noise, level and density of reverberation for speech.

Acknowledgments

This work is supported by the project Augmented Practice-Room (1023), which is funded by the local government of Styria via Zukunftsfonds Steiermark (future fond of Styria). The authors thank all listeners for their participation in the experiments and the reviewers for their helpful comments.

ZUKUNFTSFONDS STEIERMARK

Das Land
Steiermark

References

[1] B. N. J. Postma, H. Demontis, and B. F. G. Katz, "Subjective Evaluation of Dynamic Voice Directivity for Auralizations," *Acta Acustica united with Acustica*, vol. 103, no. 2, pp. 181–184, Mar. 2017.

[2] A. Kolarik, S. Cirstea, and S. Pardhan, "Discrimination of virtual auditory distance using level and direct-to-reverberant ratio cues," *The Journal of the Acoustical Society of America*, vol. 134, no. 5, pp. 3395–3398, 2013.

[3] F. Wendt, F. Zotter, M. Frank, and R. Höldrich, "Auditory Distance Control Using a Variable-Directivity Loudspeaker," *MDPI Applied Science*, vol. 7, no. 7, 2017.

[4] F. Hohl, "Kugelmikrofonarray zur Abstrahlungsvermessung von Musikinstrumenten," Master's thesis, TU Graz, 2009.

[5] D. Deboy, "Tangential Intensity Algorithm for Acoustic Centering," in *Fortschritte der Akustik, DAGA*, Düsseldorf, 2011.

[6] I. B. Hagai, M. Pollow, M. Vorländer, and B. Rafaely, "Acoustic centering of sources measured by surrounding spherical microphone arrays," *Journal of the Acoustical Society of America (JASA)*, vol. 130, no. 4, 2011.

[7] M. Frank and M. Brandner, "Perceptual Evaluation of Spatial Resolution in Directivity Patterns," in *Fortschritte der Akustik, DAGA*, Rostock, Mar. 2019.

[8] C. Pörschmann, "One's Own Voice in Auditory Virtual Environments," *Acta Acustica united with Acustica*, vol. 87, no. 3, pp. 378–388, 2001.

[9] J. S. Brereton, D. T. Murphy, and D. M. Howard, "The Virtual Singing Studio: A loudspeaker-based room acoustics simulation for real-time musical performance," in *Proceedings of the Baltic Nordic Acoustics Meeting (BNAM2012)*, 2012, pp. 18–20.

[10] J. M. Arend, T. Lübeck, and C. Pörschmann, "A Reactive Virtual Acoustic Environment for Interactive Immersive Audio," in *Audio Engineering Society Conference: 2019 AES International Conference on Immersive and Interactive Audio*, Mar. 2019.

[11] M. Romanov, P. Berghold, M. Frank, D. Rudrich, M. Zaunschirm, and F. Zotter, "Implementation and Evaluation of a Low-Cost Headtracker for Binaural Synthesis," in *Audio Engineering Society Convention 142*, May 2017.

[12] C. Schörkhuber, M. Zaunschirm, and R. Höldrich, "Binaural Rendering of Ambisonic Signals via Magnitude Least Squares," in *Fortschritte der Akustik - DAGA*, Munich, March 2018.

[13] J. Daniel, "Représentation des champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia," Ph.D. dissertation, Université Paris 6, 2001.

[14] J. Pätynen and T. Lokki, "Directivities of Symphony Orchestra Instruments," *Acta Acustica united with Acustica*, vol. 96, no. 1, pp. 138–167, 2010.

[15] EBU, "EBU SQAM CD: Sound Quality Assessment Material recordings for subjective tests," 2008. [Online]. Available: https://tech.ebu.ch/publications/sqamcd

[16] E. Larsen, N. Iyer, C. R. Lansing, and A. S. Feng, "On the minimum audible difference in direct-to-reverberant energy ratio," *The Journal of the Acoustical Society of America*, vol. 124, no. 1, pp. 450–461, 2008.

© Verband Deutscher Tonmeister e.V., 2019