Götz, Georg; Pulkki, Ville

Simplified source directivity rendering in acoustic virtual reality using the directivity sample combination

Published in:
Proceedings of 147th AES Convention

Published: 01/01/2019

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Götz, G., & Pulkki, V. (2019). Simplified source directivity rendering in acoustic virtual reality using the directivity sample combination. In *Proceedings of 147th AES Convention* Article 10286 Audio Engineering Society. http://www.aes.org/e-lib/browse.cfm?elib=20659

Audio Engineering Society

Convention Paper

Presented at the 147th Convention
2019 October 16 – 19, New York

Simplified Source Directivity Rendering in Acoustic Virtual Reality using the Directivity Sample Combination

Georg Götz[1] and Ville Pulkki[1]

[1]*Aalto Acoustics Lab, Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland*

Correspondence should be addressed to Georg Götz (`georg.gotz@aalto.fi`)

ABSTRACT

This contribution proposes a simplified rendering of source directivity patterns for the simulation and auralization of auditory scenes consisting of multiple listeners or sources. It is based on applying directivity filters of arbitrary directivity patterns at multiple, supposedly important directions, and approximating the filter outputs of intermediate directions by interpolation. This reduces the amount of required filtering operations considerably and thus increases the computational efficiency of the auralization. As a proof of concept, the simplification is evaluated from a technical as well as from a perceptual point of view for one specific use case. The promising results suggest further studies of the proposed simplification in the future to assess its applicability to more complex scenarios.

1 Introduction

From our everyday experiences we know that acoustic sources sound different depending on their orientation with respect to ourselves. For example, speech sounds different when people are directly facing us compared to the case when they are talking towards the opposite direction. Since the plausibility of an auditory illusion, for example in virtual reality systems, is strongly interconnected with the listener's expectations towards impressions from the real world [1], the aforementioned effects should be modelled carefully. In this paper, we want to introduce a simplified rendering of source directivity for acoustic virtual reality scenes that benefits the real-time capability of such systems.

1.1 Sound Source Directivity Patterns

The previously described phenomenon, that is inherent to all real world sound sources, can be described by means of sound source directivity patterns. Various measures are used to quantify the underlying physical relationships [2, 3]. For instance, the directivity factor relates the sound pressure at arbitrary positions on a sphere around a sound source to the respective sound pressure at a freely chosen reference direction [2].

In the past, directivity patterns of various sound sources have been investigated. For instance, multiple authors addressed directivity patterns of voice [4, 5, 6, 7, 8] or loudspeakers [9, 10, 11]. A rather omnidirectional pattern is observable for voice at low frequencies, which gets increasingly directional towards the front of a talker for higher frequencies [4, 5, 6, 7, 8]. Similarly, loudspeakers exhibit an increasing attenuation of sound towards the sides [9], with additional deviations from the on-axis response caused by properties of the diaphragm [10, 11] or diffraction effects due to the enclosure geometry [12].
Moreover, a thorough overview of musical instrument

directivity patterns can be found in previous work of Meyer [13] or Pätynen and Lokki [14]. In comparison to speech and loudspeakers, directivity patterns of musical instruments can be more irregular since they are substantially influenced by the instrument structure and the underlying mechanism of sound production [14]. For example, string instruments have a complex mechanism of sound production involving the bow movement and the instrument's body, whereas for brass instruments a more straightforward concentration of sound towards the bell can be observed [14].

1.2 Source Directivity in Room Simulation and Auralization

It is evident from the previous examples, that the generally frequency-dependent and possibly irregularly-shaped source directivity pattern is an integral part of modelling sound sources. Its physical importance for room simulation and auralization is also emphasized by a previous study of Otondo and Rindel [15], where an influence of the source directivity on room acoustic parameters like the sound pressure level, clarity factor (C80), lateral energy fraction (LF) and early decay time (EDT) was observed.
Additionally, a proper incorporation of source directivity patterns was also shown to be important from a perceptual point of view, since listeners were able to detect strong variations of directivity patterns in auralizations of reverberant rooms [16]. Furthermore, Postma et al. [17] studied dynamic scenarios and observed that employing dynamic source directivity patterns can improve the plausibility and listener envelopment of an auralization compared to the use of omnidirectional or static directivity patterns.

Various attempts have been made to include source directivity patterns in the simulation and auralization of auditory scenes in virtual acoustics [18, 19, 20, 21]. Reproducing the source directivity by means of multichannel recordings to drive several virtual sources as proposed by Otondo and Rindel [21] was demonstrated to enable promising results, because also variations of the directivity pattern with time can be captured. However, conducting a multichannel recording for every scene to be auralized causes an uneconomic measurement effort for virtual reality scenarios. This approach is therefore not feasible for the auralization of scenes with multiple sound sources.
Instead, it is beneficial to apply directional filters [22] to render arbitrary source directivity patterns in an auralization with exchangeable source signals. For example, Savioja et al. [18] present the integration of directional filtering into a hybrid room simulation and auralization system. This approach provides the maximum amount of freedom, since directivity filters can either be optimized towards efficiency or towards a close match with measured directivity patterns. However, the modelling of directivity patterns with detailed frequency responses requires higher-order filters and thus considerably increases the computational load of applying them during the auralization.

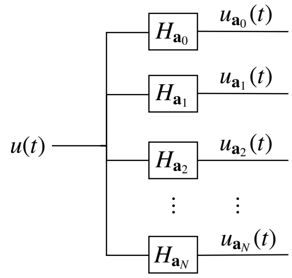

Fig. 1: Using N distinct directivity filters $H_{\mathbf{a}_i}$ to simultaneously auralize the source directivity pattern for N different directions $\mathbf{a}_i$. This modelling approach is just how Savioja et al. [18] incorporated the source directivity into their system.

1.3 Simultaneous Rendering with a Multitude of Sources and Listeners

Applying a detailed directivity filter in real-time for a single source-to-listener combination might nowadays not be too computationally demanding any more. However, extending a system to enable the simultaneous auralization for a multitude of sources and listeners may easily exceed the available computational resources when an accurate directivity rendering is desirable. In such scenarios, multiple, simultaneously working directivity filters are required, as depicted in Figure 1.
For example, consider a scenario with an arbitrary sound source and multiple listeners at several positions scattered around it. For an accurate simulation, this would require a distinct directional filter with a possibly rather detailed frequency response for every listener. An even more drastic increase of required directional filters can be observed for an auralization based on the image-source method. Even for a single listener, the

rendering of every image-source would require a distinct and potentially costly directional filter of the same directivity pattern, because various directions would need to be simulated due to different reflection paths and hence source-exit angles [19].

Summarizing the aforementioned scenarios, it becomes obvious that the computational load of directivity filters becomes increasingly critical, if scenarios with a large number of sources or listeners shall be simulated or auralized. This is especially true in the context of real-time applications.
Therefore, this contribution presents a simplified source directivity rendering for the simulation and auralization of auditory scenes consisting of a multitude of listeners or sources with similar source directivity patterns. By applying the proposed simplification, the amount of required filtering operations to model the sound source directivity can be reduced considerably, enabling an increased computational efficiency of the auralization system.

2 Simplified Rendering: Directivity Sample Combination (DISCO)

Directivity patterns can be fairly irregular and therefore current virtual reality systems (e.g. [18]) model the source directivity with distinct directivity filters for every direction that is considered in the auralization. As elaborated before, this is by no means computationally efficient for the simultaneous auralization of scenes with multiple sources and listeners, especially regarding the real-time capability of a system. However, most variations of the directivity pattern are caused by distinct properties of the source, for example the direction of an instrument's bell, the loudspeaker membrane axis or the direction into which a person is talking.
We therefore propose the Directivity Sample Combination (DISCO), where a limited set of time-invariant directivity filters is determined before the auralization to capture the most characteristic acoustic properties of the directivity pattern. By doing so, the amount of directivity filters is reduced considerably, because only directivity filters for those sampled characteristic directions are required and intermediate directions are approximated by interpolation between the available filter outputs.

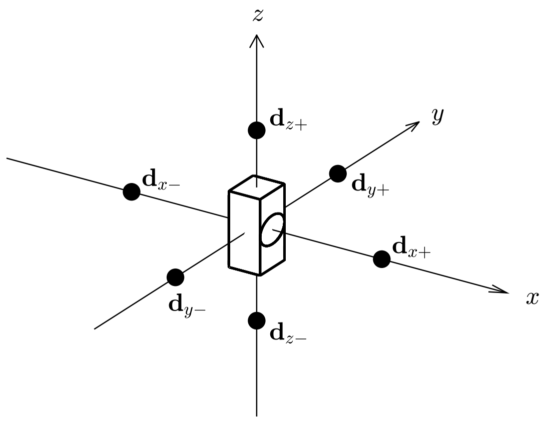

Fig. 2: Exemplary sampling of a loudspeaker directivity pattern at six directivity samples. The directions of the directivity samples are denoted by the coordinate unit vectors $\mathbf{d}_i$.

2.1 Fundamental Idea

In principle, the DISCO simplification allows to sample the directivity pattern at arbitrary, supposedly characteristic directions of the pattern. However, it is still an open research question, which directions of a pattern are actually the perceptually most characteristic ones to enable a plausible auralization. For now, we propose to employ a sampling of the directivity pattern at uniformly distributed directions, such as the surface midpoints of platonic solids (e.g. tetrahedron, cube, octahedron, ...) centred at the source position.

In this paper, we are choosing a cube for the sake of simplicity. This is equivalent to sampling the directivity pattern at six directions, as it is exemplarily illustrated for a loudspeaker in Figure 2. It is assumed that those directivity samples provide enough information to model some general features of rather simple directivity patterns, for example the increasing shadowing towards the sides of a loudspeaker.
Directivity filters are then derived for those six directions. By applying these filters to a dry source signal, the directivity characteristics of the sound source are reproduced exactly for the sampled directions.
To calculate the signals that would be obtained from directivity filters of intermediate directions, interpolation based on the basic idea of vector base amplitude panning (VBAP) [23] is applied. VBAP was originally used to pan a signal to an arbitrary direction

in between loudspeaker triplets by applying gains derived from a vector base formulation of the loudspeaker positions [23]. In the DISCO simplification, directivity samples are used as the vector base to approximate the directivity pattern at intermediate directions as a linear combination of them.

2.2 Mathematical Formulation

In the following, directions of directivity samples are denoted by unit vectors $\mathbf{d}_i = \begin{bmatrix} d_{i,1} & d_{i,2} & d_{i,3} \end{bmatrix}^\mathsf{T}$. Directions to be auralized are denoted by unit vectors $\mathbf{a}_i = \begin{bmatrix} a_{i,1} & a_{i,2} & a_{i,3} \end{bmatrix}^\mathsf{T}$. In both cases, i serves as a descriptive indexing variable.

Since the vector base formulation uses only three of the six directivity samples at a time, it is first necessary to determine into which octant of the Cartesian coordinate system $\mathbf{a}_i$ is pointing. This can be done by applying the dot product with the axis directions and checking for the signs of the results. As soon as the octant is found, the three respective directivity samples that delimit this octant are used as the vector base. For example, to approximate an intermediate direction $\mathbf{a}_i$ in the first octant, the directivity samples of the positive axis directions would be used, i.e. $\mathbf{d}_{x+}$, $\mathbf{d}_{y+}$ and $\mathbf{d}_{z+}$.

The signals produced by the directivity filters of the three determined directivity samples are denoted by $u_1(t)$, $u_2(t)$ and $u_3(t)$. To approximate the output $u_{\mathbf{a}_i}(t)$ of a hypothetical directivity filter at the intermediate direction $\mathbf{a}_i$, they are weighted with the respective factors $\beta_{1,i}$, $\beta_{2,i}$ and $\beta_{3,i}$ (with the index i dropped from now on for readability reasons):

$$\widetilde{u}_{\mathbf{a}}(t) = \beta_1 u_1(t) + \beta_2 u_2(t) + \beta_3 u_3(t)\,. \tag{1}$$

Preliminary values for the weights are determined just as in VBAP [23]

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix} = \begin{bmatrix} d_{1,1} & d_{2,1} & d_{3,1} \\ d_{1,2} & d_{2,2} & d_{3,2} \\ d_{1,3} & d_{2,3} & d_{3,3} \end{bmatrix}^{-1} \mathbf{a}\,, \tag{2}$$

with the three directivity samples $\mathbf{d}_1$, $\mathbf{d}_2$ and $\mathbf{d}_3$ as columns of the matrix. The preliminary values $\hat{\beta}$ are normalized, such that their sum equals 1:

$$\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} = \frac{1}{\hat{\beta}_1 + \hat{\beta}_2 + \hat{\beta}_3} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix}\,. \tag{3}$$

This prevents the weighting result $\widetilde{u}_{\mathbf{a}}(t)$ from exceeding the signals produced by the individual directivity filters.

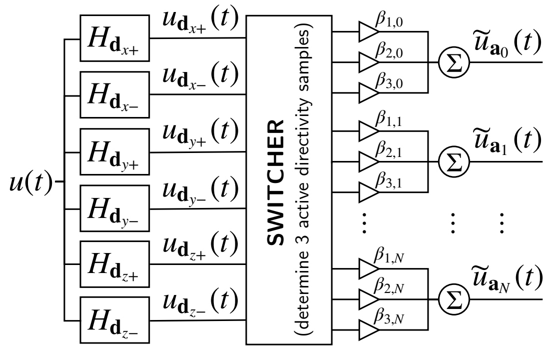

Fig. 3: Applying the DISCO simplification with six directivity samples: only six directivity filters $H_{\mathbf{d}_i}$ are required to approximate the source directivity pattern during a simultaneous auralization of N different directions $\mathbf{a}_i$.

2.3 Properties of the DISCO simplification

Figure 3 shows the application of the DISCO simplification to the introductory scenarios, which were previously described in Section 1.3. Again, a sound source directivity pattern is simultaneously auralized at N different directions $\mathbf{a}_i$. However, compared to N filtering operations that were required in Figure 1, the DISCO simplification reduces the amount of directional filters to six per directivity pattern to be auralized. This means that only six directional filters would be required for the scenario with multiple listeners. The same applies for the scenario of multiple sources with identical source directivity patterns as it is the case during a room simulation based on the image-source model. Obviously, this simplification does not provide any advantage in scenarios where multiple sources with different directivity patterns need to be auralized.

Note, that the weighting has the same desirable property as VBAP: in case the direction to approximate equals one of the directions of the directivity samples, only one of the three weights is non-zero and has a value of 1. In other words, the directivity filter outputs of the directivity samples are always exactly reproduced by this weighting.

2.4 Possible Extension to More Directivity Samples

In principle, the basic idea of the DISCO simplification allows for an extension to more than six directivity

samples or a sampling of the directivity pattern at more irregular directions compared to the sampling at the surface midpoints of platonic solids as it is proposed here. This could be beneficial in cases where the directivity pattern to auralize is considerably irregular or has more than six characteristic directions that shall be sampled. Of course, this would require a more sophisticated determination of the three active directivity samples, as a simple octant check will then not be sufficient any more. Referring back to the original VBAP paper [23], we propose to use similar approaches.
However, for this first proof of concept of the DISCO simplification, only the proposed formulation with six directivity samples at the surface midpoints of a cube was evaluated. This approach already yielded sufficient results for the tested scenarios, which will be discussed in the following Sections 3 and 4.

3 Technical Evaluation: Resulting Directivity Patterns

A measured directivity pattern of a Genelec 8020 loudspeaker is compared to the resulting directivity pattern that is obtained when applying the DISCO simplification to just that pattern. The loudspeaker directivity pattern was measured with a 10° resolution and the DISCO simplification is applied as described in Section 2. This means that for the DISCO directivity pattern, all directions between the six directivity samples are obtained by applying the weighting. FIR filters with the measured responses as tap weights are used as directivity filters. Since the measured directivity pattern only had a resolution of 10°, this resolution is also approximated with the DISCO simplification.
Figure 4 shows the measured pattern ("Reference") and the resulting DISCO pattern on the horizontal plane with an elevation angle of $\vartheta = 90°$ (using the elevation convention $0° \leq \vartheta \leq 180°$). It visualizes the directivity index [2] as a contour plot for the full 360° turn, i.e. $-180° \leq \varphi \leq 180°$, where $\varphi = 0°$ refers to the front of the sound source. The reference direction for the directivity index calculation is set to $(\vartheta_{\text{ref}} = 90°, \varphi_{\text{ref}} = 0°)$, which coincides with the main axis in front of the speaker. All magnitude responses used for the plots are smoothed in 1/3 octave bands.

3.1 Comparison of the Directivity Patterns

As one would expect from a professional studio loudspeaker, the frequency response does not change con-

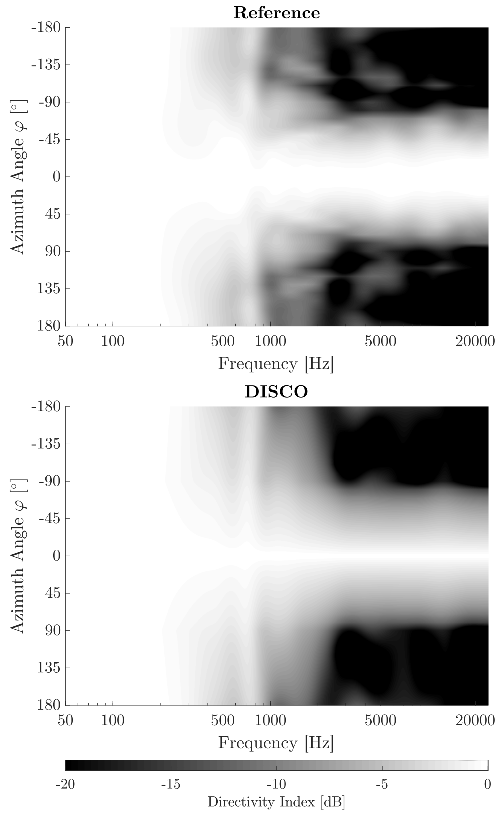

Fig. 4: Comparison of a measured Genelec 8020 loudspeaker directivity pattern with its simplified counterpart employing the DISCO simplification. The plots visualize the directivity index on the horizontal plane ($\vartheta = 90°$) with the reference direction for the directivity index calculation set to $(\vartheta_{\text{ref}} = 90°, \varphi_{\text{ref}} = 0°)$.

siderably for a deviation of almost 30° in both directions from the main axis. In the range between 30° and 90° off the main axis, the absolute difference compared to the main axis increases due to effects of the enclosure geometry or the loudspeaker's diaphragm. The deviations are however rather irregular and occur only at distinct frequencies and directions. Towards the sides of the loudspeaker (±90°), the attenuation compared to the main axis increases and eventually exceeds 15 dB for frequencies above approximately 3000 Hz. From there on, the pattern transitions into the shadow region of the loudspeaker, which can be observed between 90° and 180° off the main axis. In the shadow region, some peaks are visible again, which are probably caused by interactions of the radiated sound with the loudspeaker enclosure. At low frequencies, the pattern is almost omnidirectional.

At a first glance, the DISCO pattern does not differ considerably from the reference pattern. The plot also exhibits a smooth frontal region with an increasing attenuation towards the sides of the loudspeaker and a pronounced shadow region. However, whereas the reference pattern displays almost no attenuation compared to the main-axis in quite a big range from almost $\varphi = -30°$ to $\varphi = 30°$, the pattern of the DISCO simplification immediately shows an attenuation off the main-axis.
Besides, it can be seen that most of the details or irregularities of the reference directivity pattern are not observable in the simplified pattern any more. This can be observed for the frontal directions as well as for the shadow region. Of course, this is not unexpected, since the DISCO simplification samples the reference pattern only at the azimuth angles $\varphi = 0°$, $\pm 90°$ and $\pm 180°$ for the cross section considered here. As a consequence, all detailed structures with respect to the direction-dependence that are observable between the directivity samples in the measured pattern are smoothed out in the simplified pattern.

3.2 Error between the Measured and Simplified Pattern

The absolute error between both patterns is depicted in Figure 5. From this plot, it can also be seen that detailed structures in the directivity pattern are smoothed out, because they show up as absolute differences between the measured and the simplified pattern. Moreover, some additional error peaks in the shadow region become visible here, which were hidden due to the limited colour bar of the previous plots. Further analysis of single directions showed that these peaks are also caused by irregularities of the reference pattern that are smoothed out by the DISCO simplification.
As already mentioned in Section 2, the directivity samples are perfectly reproduced by the DISCO simplification. This can also be observed from the figure, since no errors are depicted along the five horizontal lines at $\varphi = 0°$, $\pm 90°$ and $\pm 180°$.
It should furthermore be noted, that for the full rotation on the horizontal plane with a constant elevation of $\vartheta = 90°$, no contributions of the z-axis directivity samples are required. Therefore, the simplification only

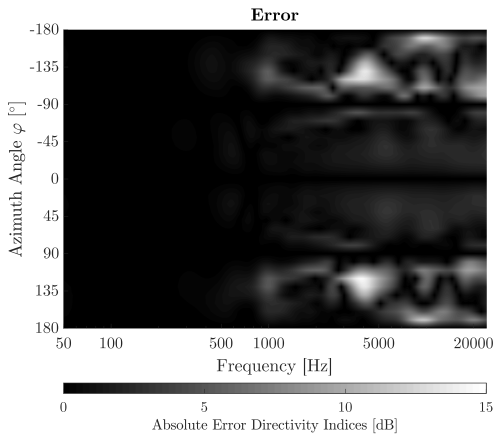

Fig. 5: Absolute error between the directivity indices of the measured and the DISCO-simplified loudspeaker directivity pattern (evaluation at $\vartheta = 90°$; $\vartheta_{\mathrm{ref}} = 90°, \varphi_{\mathrm{ref}} = 0°$).

uses two directivity samples at a time to approximate a certain direction, which means that in total only four of the six directivity samples are used in this case.

A similar smoothing of detailed structures can be observed when investigating other cross sections of the directivity pattern. However, as the plots do not provide any additional insights, they are left out here for the sake of brevity.

4 Perceptual Evaluation: Proof of Concept

So far it remained unclear whether the observed errors introduced by the DISCO simplification have an effect on the perception of the resulting auralization. Therefore, an elementary listening test was conducted to substantiate the basic concept of the DISCO simplification.

4.1 Test Concept and Setup

The idea of this listening test was to evaluate the simplified directivity pattern isolated from the rest of a potential room simulation and auralization system. Therefore, an entirely anechoic scene was auralized with the previously described measured loudspeaker directivity pattern and the respective pattern obtained by applying the DISCO simplification.

The listening test took place in a designated listening booth, where the participants were seated during the whole test. Binaural stimuli were played back over headphones to the participants while a matching animation of either a rotating loudspeaker or a Lego minifigure was shown. The animation was displayed on a computer screen as a part of the graphical user interface. Only source rotations along the horizontal plane with an elevation angle of $\vartheta = 90°$ were investigated and the static listener was directly facing the frontal radiation direction $\varphi = 0$ of the sound source.
Binaural signals for the whole 360° turn were obtained with an angular resolution of 1° and a smooth rotation was enabled by crossfading in between. Since the resolution of the measured pattern was only 10°, impulse responses for intermediate directions were calculated by bilinear interpolation between the available measurements for the auralization of the measured directivity pattern. This did however not affect the auralization of the DISCO pattern, because all directivity samples are already available in the measured pattern. An arbitrarily fine resolution can then be approximated with the weighting procedure described in Section 2.2.

The listening test followed the basic multi stimulus test with hidden reference and anchor (MUSHRA) test paradigm as specified by recommendation ITU-R BS.1534-3 [24]. During the test, the participants were asked to grade the plausibility of the audio playback in connection with the scenario illustrated by the matching animation. The test subjects were told, that the reference stimulus is supposed to be rated with the highest possible rating of 100 and that the plausibility of the other systems shall be graded in relation to this. Table 1 summarizes the four different systems under test. Prior to the test it was hypothesized, that the previously elaborated loss of details in the directivity pattern caused by the DISCO simplification degrades the listening experience and hence reduces the plausibility of the playback. The "Cardioid" system was included in the test as a comprehensible but rather rudimentary modelling approach to provide an additional indication on how the DISCO simplification performs perceptually and to bring the listening test results into a more tangible order.
The playback could at any time be continuously switched to another system by pressing the respective button in the GUI without stopping the animation or the audio playback. If the participant did not stop the playback with the stop button, the rotation would continue in an endless loop. If the playback was stopped, a subsequent continuation would reset both the animation and the audio playback to its initial position (source orientation $\varphi = 0$).

Four different source signals were used in the test, namely a speech, a saxophone, a pink noise and an FX signal. The order of the 12 test items (4 source signals $\times$ 3 repetitions) as well as the arrangement of the different systems in the GUI was randomized. A short training of three test items, with different source signals than in the main test, was included prior to the main test to familiarize the subjects with the test design. The entire test session took less than 30 minutes for all participants. In total, 14 subjects participated in the test. All of them are researchers in the field of acoustics and therefore considered as experienced listeners.

4.2 Test Results

A post-screening was conducted. Following recommendation ITU-R BS.1534-3, the ratings of two subjects were excluded from the analysis, because they rated the hidden reference with a score of less than 90 for two or more ($\geq 15\,\%$) of the test items. Furthermore, ratings of the three repetitions were averaged. The entire following statistical analysis (normal distribution tests, homogeneity of variances tests, significance tests and calculation of confidence intervals) used the remaining 12 averaged plausibility ratings per source signal and system combination.
Shapiro-Wilk tests were performed on the residuals[1] and the null-hypothesis of normal distributions of the residuals was rejected with $p < 0.05$. This was additionally confirmed with normal quantile-quantile plots. Similarly, the homogeneity of variances assumption was violated, which was tested with the Brown-Forsythe test ($p < 0.01$). Therefore, the assumptions of a repeated-measures analysis of variance (rmANOVA) are not met and the statistical analysis of the listening test instead employs non-parametric tests as recommended by ITU-R BS.1534-3. Since the Friedman test does also require the here violated homogeneity of variances, the Wilcoxon signed-rank test was used instead for the significance testing of the results. Pairwise comparisons between the median ratings of the reference,

[1] In conformity with recommendation ITU-R BS.1534-3, tests for normality are not performed on the actual ratings, but rather on the residuals or errors. This way, only one Shapiro-Wilk test per signal is performed. Further details on this can be found in [25].

System	Description
Ref *(Reference)*	Identical with the measured loudspeaker directivity pattern of the technical evaluation in Section 3.
DISCO	Identical with the DISCO pattern of the technical evaluation in Section 3 (applying the DISCO simplification to the reference, using the basic configuration as described in Section 2).
Cardioid	Artificially generated and frequency-independent directivity pattern with a cardioid shape. This pattern can be used to model the increasing shadowing towards the rear of a sound source in a rudimentary way.
Only Frontal *(Anchor)*	Artificially generated and frequency-independent directivity pattern with a dichotomous shape. For listening positions in front of the source ($0^\circ \leq |\varphi| \leq 90^\circ$), no attenuation is applied at all. For positions in the back ($90^\circ \leq |\varphi| \leq 180^\circ$), the sound is fully attenuated.

Table 1: Systems under test during the perceptual evaluation of the DISCO simplification.

DISCO and cardioid system were conducted. Figure 6 visualizes the median values of the obtained ratings along with their bootstrapped 95 % confidence intervals ($N_{\text{boot}} = 10000$) and the results of the significance tests. For all tested signals the median value of the DISCO system is lower than the one of the reference system. This difference is significant for all signals with $p < 0.01$. Moreover, all medians of the cardioid system ratings are lower than the respective median values of the DISCO system. This difference is significant for all signals with $p < 0.001$.
Even though significant differences between the medians of the reference and the DISCO system were observed, it is obvious from the plots, that the medians do not differ considerably. The lowest ratings of the DISCO system are observable for the noise signal. This is an expected result, since the whole audible frequency range is evaluated with this signal and hence more losses of details may be exposed. However, all medians of the DISCO system are still within the upper fifth of the scale, i.e. in the “Excellent” rating range of the MUSHRA recommendation.

5 Discussion

Even though a significant difference between the plausibility ratings of the measured directivity pattern and its DISCO simplification was observed in the listening test, the difference was fairly small and might be observable only because of the rather conservative test design. A direct comparison by continuous switching between systems as it was possible in the present listening test facilitates the detection of nuanced differences. This was also informally reported by the subjects after the listening tests, who said, that they indeed perceived a difference between the two best-rated systems but it was not degrading their listening experience substantially. Thus, the decrease in plausibility of the rendered auralization is assumed to be marginal. Compared to the frequency-independent cardioid pattern, the DISCO-simplified pattern performs considerably better.
The observations suggest, that the DISCO simplification may be an adequate method of decreasing the computational complexity of the auralization while still enabling a plausible auditory illusion, that is comparable to the computationally more expensive full directivity pattern and significantly better than a rudimentary frequency-independent approximation.

However, at this point it remains unclear, how the DISCO simplification performs when simulating reverberant scenes. As demonstrated by Otondo and Rindel [15], the directivity pattern has a considerable influence on room acoustic parameters. Therefore, a future listening test should also evaluate the simplification for simulations of reverberant scenes.
Furthermore, with the loudspeaker directivity pattern used in this proof of concept, only a single directivity pattern was evaluated so far. For more irregular directivity patterns, it is however unlikely that the DISCO simplification performs equally well with just six directivity samples as it was proposed and evaluated in this paper. In these cases it would be necessary to extend the DISCO simplification to more than six directivity samples. With this in mind, it would be desirable to find out, which directions of a source directivity pattern are perceptually characteristic and how they can be determined, ideally with an automatic approach. A

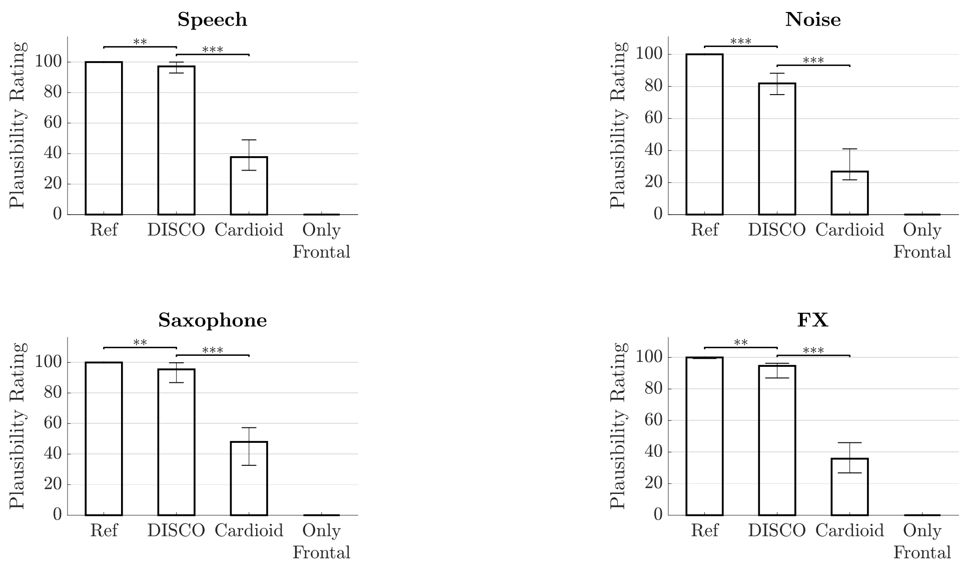

Fig. 6: Listening test results: comparing the plausibility of a measured loudspeaker directivity pattern, its DISCO simplification and two frequency-independent patterns. Bars indicate medians of the plausibility ratings among the participants including bootstrapped 95 % confidence intervals ($N_{\text{boot}} = 10000$). Asterisks denote significant differences determined with the Wilcoxon signed-rank test ($**$: $p < 0.01$, $***$: $p < 0.001$). No significance tests were performed between the "Cardioid" and the "Only Frontal" conditions.

future study should then also find a relation between the complexity of patterns and the amount of required directivity samples to render plausible auralizations.

6 Summary

This paper presents a simplified rendering of source directivity patterns during the auralization of virtual acoustic scenes consisting of multiple sources and listeners. By applying the proposed directivity sample combination (DISCO), the amount of filtering operations to auralize a source with a full-sphere frequency-dependent directivity pattern simultaneously for multiple listeners can be reduced considerably. Furthermore, the savings can be beneficial for auralization systems based on the image-source model, where a multitude of image-sources with the same directivity pattern have to be auralized for one or more listeners.
During the technical evaluation of the DISCO simplification, a loss of directional details in the resulting directivity pattern was observed. This is due to the reduction of a full-sphere directivity pattern to a limited amount of only six directivity samples and the necessary interpolation for intermediate directions that comes along with it. A perceptual proof of concept revealed significant but small differences in the plausibility ratings between a measured loudspeaker directivity pattern and its DISCO simplification.
Future studies should address the determination of perceptually important directions of source directivity patterns and the applicability of the DISCO simplification to more irregular directivity patterns or reverberant scenes.

7 Acknowledgements

The project has received funding from the Academy of Finland, project no. 317341, and from Nordic Sound and Music Computing Network (NordicSMC), project no. 86892.

References

[1] Lindau, A. and Weinzierl, S., "Assessing the Plausibility of Virtual Acoustic Environments," *Acta Acustica united with Acustica*, 98(5), pp. 804–810, 2012.

[2] Mechel, F. P., *Formulas of Acoustics*, Springer, Berlin; New York, 2nd edition, 2008.

[3] Pollow, M., *Directivity Patterns for Room Acoustical Measurements and Simulations*, volume 22 of *Aachener Beiträge zur Technischen Akustik*, Logos Verlag, Berlin, 1st edition, 2015.

[4] Flanagan, J. L., "Analog Measurements of Sound Radiation from the Mouth," *Journal of the Acoustical Society of America*, 32(12), pp. 1613–1620, 1960.

[5] Monson, B. B., Hunter, E. J., and Story, B. H., "Horizontal directivity of low- and high-frequency energy in speech and singing," *Journal of the Acoustical Society of America*, 132(1), pp. 433–441, 2012.

[6] Chu, W. T. and Warnock, A. C. C., "Detailed Directivity of Sound Fields Around Human Talkers," Technical report, National Research Council Canada, 2002.

[7] Kocon, P. and Monson, B. B., "Horizontal directivity patterns differ between vowels extracted from running speech," *Journal of the Acoustical Society of America*, 144(1), pp. EL7–EL12, 2018.

[8] Dunn, H. K. and Farnsworth, D. W., "Exploration of Pressure Field Around the Human Head During Speech," *Journal of the Acoustical Society of America*, 10(3), pp. 184–199, 1939.

[9] Toole, F. E., "Loudspeaker Measurements and Their Relationship to Listener Preferences: Part 2," *Journal of the Audio Engineering Society*, 34(5), pp. 323–348, 1986.

[10] Frankort, F. J. M., "Vibration Patterns and Radiation Behavior of Loudspeaker Cones," *Journal of the Audio Engineering Society*, 26(9), pp. 609–622, 1978.

[11] Wolff, I. and Malter, L., "Directional Radiation of Sound," *Journal of the Acoustical Society of America*, 2(2), pp. 201–241, 1930.

[12] Vanderkooy, J., "A Simple Theory of Cabinet Edge Diffraction," *Journal of the Audio Engineering Society*, 39(12), pp. 923–933, 1991.

[13] Meyer, J., *Acoustics and the Performance of Music: Manual for Acousticians, Audio Engineers, Musicians, Architects and Musical Instruments Makers*, Modern Acoustics and Signal Processing, Springer, New York, 5th edition, 2009.

[14] Pätynen, J. and Lokki, T., "Directivities of Symphony Orchestra Instruments," *Acta Acustica united with Acustica*, 96(1), pp. 138–167, 2010.

[15] Otondo, F. and Rindel, J. H., "The Influence of the Directivity of Musical Instruments in a Room," *Acta Acustica united with Acustica*, 90(6), pp. 1178–1184, 2004.

[16] Wang, L. M. and Vigeant, M. C., "Evaluations of output from room acoustic computer modeling and auralization due to different sound source directionalities," *Applied Acoustics*, 69(12), pp. 1281–1293, 2008.

[17] Postma, B. N. J., Demontis, H., and Katz, B. F. G., "Subjective Evaluation of Dynamic Voice Directivity for Auralizations," *Acta Acustica united with Acustica*, 103(2), pp. 181–184, 2017.

[18] Savioja, L., Huopaniemi, J., Lokki, T., and Väänänen, R., "Creating Interactive Virtual Acoustic Environments," *Journal of the Audio Engineering Society*, 47(9), pp. 675–705, 1999.

[19] Brinkmann, F., Erbes, V., and Weinzierl, S., "Extending the closed form image source model for source directivity," in *Fortschritte der Akustik - DAGA 2018*, pp. 1298–1301, Munich, Germany, 2018.

[20] Wabnitz, A., Epain, N., Jin, C., and Van Schaik, A., "Room acoustics simulation for multichannel microphone arrays," in *International Symposium on Room Acoustics, ISRA 2010*, pp. 1–6, Melbourne, Australia, 2010.

[21] Otondo, F. and Rindel, J. H., "A New Method for the Radiation Representation of Musical Instruments in Auralizations," *Acta Acustica united with Acustica*, 91(5), pp. 902–906, 2005.

[22] Karjalainen, M., Huopaniemi, J., and Välimäki, V., "Direction-Dependent Physical Modeling of Musical Instruments," in *15th International Congress on Acoustics (ICA 95)*, pp. 451–454, Trondheim, Norway, 1995.

[23] Pulkki, V., "Virtual Sound Source Positioning Using Vector Base Amplitude Panning," *Journal of the Audio Engineering Society*, 45(6), pp. 456–466, 1997.

[24] Rec. ITU-R BS.1534-3, "Method for the subjective assessment of intermediate quality level of audio systems," Standard, International Telecommunication Union - Radiocommunication Sector, Geneva, Switzerland, 2015.

[25] Kozak, M. and Piepho, H.-P., "What's normal anyway? Residual plots are more telling than significance tests when checking ANOVA assumptions," *Journal of Agronomy and Crop Science*, 204(1), pp. 86–98, 2018.