Author(s): Mikko-Ville Laitinen, Archontis Politis, Ilkka Huhtakallio, Ville Pulkki

Title: Controlling the Perceived Distance of an Auditory Object by Manipulation of Loudspeaker Directivity

Year: 2015

Version: Final published version

Please cite the original version:

Mikko-Ville Laitinen, Archontis Politis, Ilkka Huhtakallio, Ville Pulkki. Controlling the Perceived Distance of an Auditory Object by Manipulation of Loudspeaker Directivity. JASA Express Letters, 137, 6, EL462-EL468, June 2015. DOI: 10.1121/1.4921678

This publication is included in the electronic version of the article dissertation:
Politis, Archontis. Microphone array processing for parametric spatial audio techniques.
Aalto University publication series DOCTORAL DISSERTATIONS, 195/2016.

Controlling the perceived distance of an auditory object by manipulation of loudspeaker directivity

Mikko-Ville Laitinen, Archontis Politis, Ilkka Huhtakallio, and Ville Pulkki
Department of Signal Processing and Acoustics, School of Electrical Engineering, Aalto University, P.O. Box 13000, FI-00076 Aalto, Finland
mikko-ville.laitinen@aalto.fi, archontis.politis@aalto.fi, ilkka.huhtakallio@aalto.fi, ville.pulkki@aalto.fi

Abstract: This work presents a method to control the perceived distance of an auditory object by changing the directivity pattern of a loudspeaker and consequently the direct-to-reverberant ratio at the listening spot. Control of the directivity pattern is achieved by beamforming using a compact multi-driver loudspeaker unit. A small-sized cubic array consisting of six drivers is assembled, and per driver beamforming filters are derived from directional measurements of the array. The proposed method is evaluated using formal listening tests. The results show that the perceived distance can be controlled effectively by directivity pattern modification.

[QJF]
Date Received: February 5, 2015 **Date Accepted:** May 12, 2015

1. Introduction

Perceiving the location of a sound source is a fundamental property of the human hearing system. In addition to the perceived source direction, human hearing is sensitive to the source distance. Hence spatial sound rendering systems should be able to reproduce the perception of distance convincingly. Although auditory distance perception has been studied to some degree (see Zahorik *et al.,* 2005 for a thorough review), there are not many studies on how to actually render auditory distance with loudspeakers in normal rooms. The aim of this study is to propose such a method.

A number of acoustic cues are known to be important for the perception of source distance (Zahorik *et al.,* 2005): sound power variation following the inverse squared-distance law, the direct-to-reverberant energy ratio (D/R-ratio), spectral changes, and binaural cues. Spectral cues result from high-frequency sound attenuation further away from the listener, and they are perceived at distances greater than about 15 m. Binaural cues result from near-field excess inter-aural differences at distances smaller than about 1 m. As these cues are significant only at very small and large distances, they are not deemed crucial for typical spatial-sound rendering systems.

Controlling the sound power in a distance-dependent manner is trivial; a simple gain modification following the inverse-distance law between a point source and the listening spot can be employed. On the contrary, controlling the D/R-ratio in normal rooms is not straightforward because the room reverberation itself dominates the resulting D/R-ratio. In theory, dereverberation techniques could be applied (Mourjopoulos, 1994), but typically they work only at a single point, and they are sensitive to changes in the room acoustics. Moreover, it is not fully understood how hearing actually detects D/R-ratio changes (Zahorik *et al.,* 2005). Both monaural (Lounsbury and Butler, 1979) and binaural (Bronkhorst, 2002) features have been proposed to be dominant in distance perception. Monaural processing could utilize phase coherence of partials over frequency (Griesinger, 2010; Laitinen *et al.,* 2013), whereas binaural processing could utilize, e.g., fluctuations of the inter-aural level differences over time (Catic *et al.,* 2013). Furthermore, it has been proposed that directions and delays of early reflections provide meaningful cues (Pellegrini, 2002).

This paper proposes a method for distance rendering that is based on controlling the sound power and D/R-ratio cues by controlling the gain and the directional pattern of a loudspeaker unit. It is assumed that other features, such as exact delays and directions of reflections or near-field effects, are far less prominent in a domestic environment. The proposed method can be seen as a loudspeaker version of the binaural D/R-ratio manipulation presented in Zahorik (2002). A related distance control system proposed recently in Jeon *et al.* (2014) is based on controlling the inter-aural cross correlation of the listener, related to the D/R-ratio, through a stereophonic speaker pair. That work does not consider the room reverberation, and hence it differs from the proposed approach. A more closely related system has been presented in Härmä *et al.* (2008), where a highly directional loudspeaker is used to isolate the direct sound from the room effect, and a 5.1 loudspeaker layout to control reverberation. Contrary to these approaches, we propose to manipulate the received D/R-ratio by controlling the directionality of a single loudspeaker unit and thus affecting the amount of received reverberation.

2. Controlling the perceived distance by changing the directivity pattern

The proposed method is based on a simplified model of the reverberant field as a mixture of direct sound and isotropic diffuse reverberation. In this model, the power of the reverberant sound P_{diff} and the power of the direct sound P_{dir} result in a specific D/R-ratio $\Gamma = P_{\mathrm{dir}}/P_{\mathrm{diff}}$. The diffuse power can be assumed to be independent from the source distance, while the direct sound power can be determined based on the inverse-distance law. More specifically, if a source moves from some reference distance r_0 to a new one r, the amplitude of the direct sound is modified by a gain $g = r_0/r$ and hence its power by $P'_{\mathrm{dir}} = (r_0/r)^2 P_{\mathrm{dir}}$. That modification results in a new D/R-ratio $\Gamma' = (r_0/r)^2\Gamma$. By expressing the change in the D/R-ratio in dB as $\Delta\Gamma = 10\log_{10}(\Gamma'/\Gamma)$ and the new distance in units of the reference distance $r = kr_0$, we get

$$\Delta\Gamma = -20\log_{10}k, \tag{1}$$

which shows, for example, that the D/R-ratio is increased or decreased by 6 dB for halving ($k = 1/2$) or doubling ($k = 2$) the distance, respectively.

Altering the gain of a signal reproduced by a single loudspeaker unit does not affect the D/R-ratio. However, effective D/R-ratio modification in a real room can be realized by altering the directivity pattern of the loudspeaker. We assume axisymmetric patterns $d(\theta)$, where θ is the angle between the look direction of the loudspeaker and the listener. Furthermore, it is assumed that the reference D/R-ratio corresponds to omnidirectional behavior of the loudspeaker. Then its change due to applying a more directional pattern can be parameterized by (a) the directional gain $d(\theta)$ and (b) the directivity factor Q of the loudspeaker, expressing the reduction of the radiated power of the directive loudspeaker compared to an omnidirectional one. Using this parameterization, the direct sound power becomes $P'^2_{\mathrm{dir}} = d^2(\theta)P_{\mathrm{dir}}$ and the diffuse power $P'_{\mathrm{diff}} = P_{\mathrm{diff}}/Q$. The overall modification then is

$$\Delta\Gamma = 10\log_{10}(Qd^2(\theta)). \tag{2}$$

It is noted that the directivity factor is $Q \geq 1$, meaning that if the loudspeaker is oriented toward the listening spot and only the directivity factor is modified, only positive increases in the D/R-ratio can be achieved (e.g., by switching from an omni to a cardioid pattern). Reduction of the D/R-ratio, however, can be achieved by steering the pattern to some other angle so that direct sound power is reduced as well. Finally, combining Eqs. (1) and (2), we assume that the perceived distance can be controlled by the directivity pattern according to

$$Qd^2(\theta) = 1/k^2. \tag{3}$$

According to these assumptions, typical use cases may require higher directionality than first-order patterns (such as the cardioid) to allow control over a range of halving to doubling the perceived distance. In this study, second-order cardioid patterns with $Q = 5$ are used in the listening tests. The realization of the patterns is described in Sec. 3.

3. Array assembly and directivity control

Higher-order directivity patterns can be generated by a compact array of drivers and appropriate filtering of the signal distributed to them. The filter design can be based on a model of the array or on anechoic measurements of the array response at multiple directions. The design and assembly of the array used in this work, along with the associated filter design, is presented in the following.

3.1 Loudspeaker design

A cubic loudspeaker with one loudspeaker driver on each face was built (see Fig. 1). The design criteria were to minimize the box size and the driver separation while maintaining a simple construction and low-frequency extension down to 200 Hz. Due to the requirement of independent operation of each source, the back radiation from the drivers needed to be isolated from the rest.

The inner volume of the cube was split into six equally sized and shaped parts. The box was designed for a Peerless P830970 2 in. full range driver. The dimensions of the box were determined so that the drivers could be mounted on the faces of the cube and the corners of the magnet assemblies would be separated by 4 mm from each other leaving space for the walls. Each back enclosure was pyramid-shaped with the volume of approximately 150 cm^3 and the apex of each enclosure met at the center of the cube. The inner structure was made by first folding the pyramid shapes out of rigid cardboard and then reinforcing and sealing the structure by laminating a layer of glass-fiber cloth with polyester resin mixed with sandblasting sand and glass micro balloons. The procedure produced thin enough walls and rigid construction without leakage between the chambers. The cube faces, with a mounting hole for the drivers, were made out of 6 mm plywood and glued to each part separately. The stand for the cube was made out of 3 mm threaded rod for open, minimalistic, and adjustable mount that could be attached to a basic microphone stand. A Crown CT8150 amplifier was used to drive the cube loudspeaker. The measured on-axis frequency response of a single driver is presented in Fig. 1 with its respective directivity pattern in Fig. 2(a).

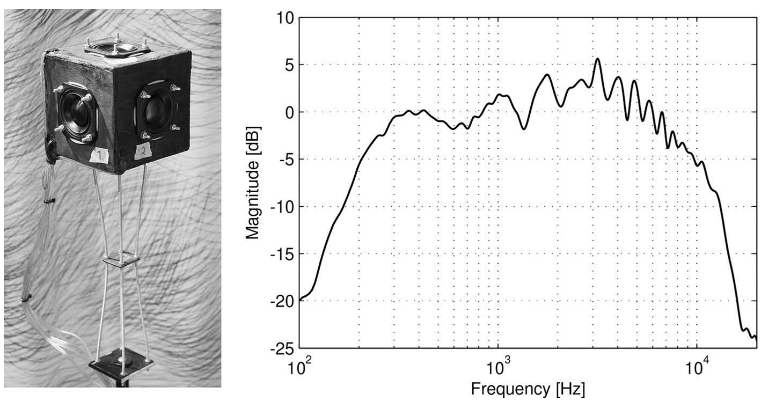

Fig. 1. (Color online) Left: The custom loudspeaker mounted on a microphone stand. Right: The on-axis magnitude response of a single driver.

3.2 Beamforming pattern design

A second-order cardioid pattern was used as the beamforming target, described by the relation $d(\theta) = (1/4)(1 + \cos\theta)^2$. It consists of a single-null on the back and a directivity factor of $Q = 5$, theoretically able to achieve a D/R-ratio increase of 7 dB compared to an omnidirectional speaker. Realization of the patterns was achieved by applying one filter per driver element. The filters were optimized to achieve the second-order cardioid directivity on the far-field, and they were based on free-field measurements of the array response in an anechoic chamber. More specifically, 252 impulse responses were measured from each driver in elevation angles from $-25°$ to $+60°$ with seven equally spaced microphones, for a full rotation of 360° in 10° intervals. After the transfer-function matrix of responses from the drivers to microphones was obtained, a regularized least-squares inversion of the response matrix in the frequency domain was applied to achieve the frequency-independent target pattern at the measurement directions, similar to, e.g., Farina *et al.* (2010) for the equivalent task of microphone array beamforming. Three sets of filters were generated, one for each distance control target of near, mid, and far. The near condition was realized with the second-order cardioid oriented at the listener. The far condition was realized with the same pattern but steered at the opposite direction of the listener to minimize the D/R-ratio. The mid condition was realized using an omnidirectional pattern, which was produced using the same filtering procedure. The achievable polar patterns after filtering are presented in Fig. 2, at octave-band intervals. It is obvious that the target patterns are approximated faithfully up to around 2 kHz for all three cases, above which sidelobes start to appear due to spatial aliasing. However, even in the aliased region, the patterns retain significant directionality to achieve D/R-ratio control. A D/R-ratio measurement of the room impulse response (RIR), using the final directivity patterns and an omnidirectional microphone at the listening spot, showed D/R-ratio differences of -6.6 dB between mid and far, and 4.1 dB between mid and near.

4. Subjective evaluation

Formal listening tests were carried out to assess the performance of the distance-rendering method proposed in this paper. In addition, the proposed method is compared to a basic distance-rendering method that applies only the gain change.

4.1 Listening test procedure

Thirteen listeners participated in the test, excluding the authors of this paper. The listening test was performed in a fairly dry listening room in accordance with the ITU-R BS 1116-1 standard, having the dimensions of $6.3 \times 5.5 \times 2.7$ m. The listening position was near the back wall, slightly offset from the center line of the room. The distance to the front wall was 5 m, while the cube loudspeaker was at 2.3 m in front of the listener. There was an acoustically transparent, visually opaque curtain between the loudspeaker and the listener, at the distance of 0.8 m.

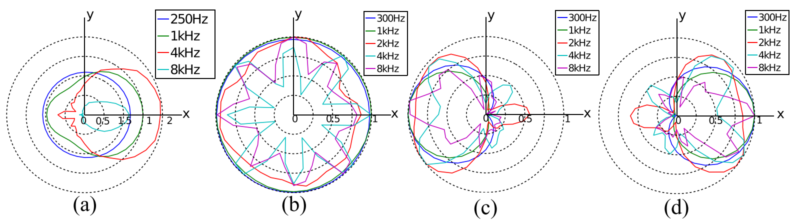

Fig. 2. (Color online) Magnitude of directivity patterns of (a) front driver, (b) omnidirectional, (c) back-facing cardioid, (d) front-facing cardioid, at various frequency bands.

The task of the listener was to indicate the perceived distance, using a graphical user interface. Possible distances were from 0 to 6 m, at steps of 0.1 m. The maximum distance was set slightly higher than the largest possible real distance in the room. It was assumed that listeners would not perceive sources much further away than this distance as they were all familiar with the room.

The listening test was carried out as a multi-stimulus test where listeners had to rate three samples simultaneously. The listeners could switch between the different samples freely. These samples corresponded to the three distance-rendering targets: near, mid, and far. There were two reproduction methods to be evaluated: the proposed method gain + directivity (see Sec. 2) and the reference method gain. The method gain denoted controlling only the gain with an omnidirectional pattern. The gains were +6, 0, and −6 dB for near, mid, and far, respectively. The method gain + directivity controlled both the gain and the directivity, using the patterns depicted in Figs. 2(d), 2(b), and 2(c), for near, mid, and far, respectively. The gains were set in a way that the A-weighted loudness of each case matched those of the gain method. Each method was evaluated with program material of three different recordings with varying captured reverberation characteristics: anechoic speech, low-reverberation acoustic guitar, and a semi-reverberant drums recording. Each material was repeated three times during the test. The samples were band-passed between 200 Hz and 3 kHz, where the directivity was reliably controlled, truncated to about 30 s and looped. These test sets were randomly interleaved producing altogether 18 test sets to be evaluated by each listener. The listening test was conducted in two parts; a training session followed by the actual listening test. The duration of the training sessions was 5 min, and the duration of the actual listening test was 15 min on average.

4.2 Results

Repeated-measures analysis of variance (RM-ANOVA) was used to obtain a statistical analysis of the results. The univariate approach was taken, and the correction for the violation of sphericity was applied when required. The correction method applied was Greenhouse–Geisser when $\epsilon < 0.75$ and Huynh–Feldt when $\epsilon > 0.75$. The repetitions of a same combination were pre-averaged before the analysis. Two-way RM-ANOVA was applied separately to the results of the different methods. The within-subjects factors were *program material* and *target*. RM-ANOVA revealed that target was the only significant factor with both methods, gain + directivity: $F(1.369, 16.429) = 214.814$, $p < 0.05$, and gain: $F(1.143, 13.72) = 23.357$, $p < 0.05$. Figure 3 shows mean plots with 95% confidence intervals for the factor target.

The results indicate that both methods provide a change in the perceived distance. However, these changes were small when using only the gain (0.5 m to each direction on average), whereas there was significant difference when using both the gain and the directivity (1.7 m to each direction on average). Two inter-related conclusions can be deduced from the results. First, the listeners perceived the sound clearly closer and significantly further than the actual distance of the loudspeaker, in the near and far cases. Second, taking the perceived distance of the mid case as the reference, effective rendering at closer than half the reference distance was achieved with the closer condition, in accordance with the theory of Sec. 2 and the realized directivity patterns.

5. Discussion

Correspondence between the proposed D/R-ratio control of Sec. 2 and the true effect on D/R-ratio depends on how well the acoustics of the room follow the basic direct/diffuse model. In the tested room, the measured D/R-ratio differences were lower on both directions than the theoretical ones; however, the listening tests showed a significant perceptual effect on both cases. This indicates that even though the acoustics of the room will presumably affect the performance of the system, the method seems robust to discrepancies from the model. D/R-ratio measurements from binaural RIRs may reveal a closer link among the model, the acoustics, and the perceived distance

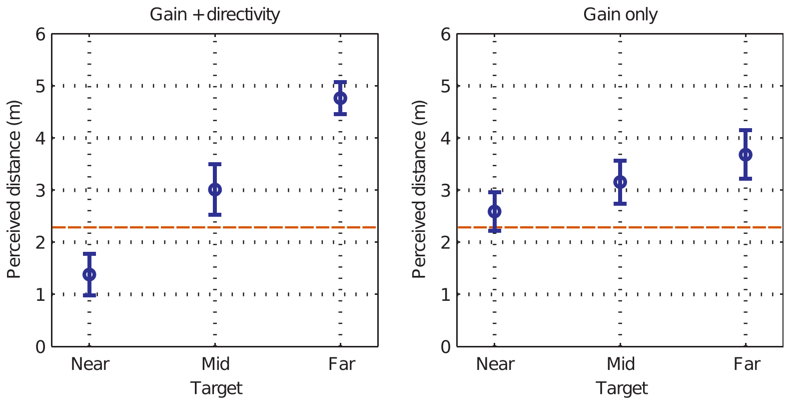

Fig. 3. (Color online) Perceived distance as a function of target distance for the different rendering methods: gain + directivity (left) and gain (right). Mean values and the 95% confidence intervals are shown. The orange dashed line indicates the actual distance of the loudspeaker at 2.3 m.

and are expected to differ from the present measurements due to the head and pinna effects. In addition, it should be noted that since monaural characteristics of reverberation also contribute to the perceived distance (see Sec. 1), the reverberation contained in the input signal might affect the result. However, no such significant differences were found in these listening tests.

In a practical implementation of the proposed method, higher-quality loudspeaker enclosures following a similar cabinet segmentation for compactness should be easy to assemble with CNC-routers or three-dimensional printers. Furthermore the filter design can be further optimized to avoid aliasing by simply using the driver's inherent directionality at high frequencies. The filter design process itself can be simplified by avoiding anechoic measurements of the array and using a theoretical model of its response. However, beamforming performance in this case may be reduced, depending on how close the real array conforms to the model.

6. Conclusion

A method is presented to control the perceived distance of a sound source with a single loudspeaker unit. We assume that the D/R-ratio is the dominant distance cue with typical sound reproduction setups in domestic environments. Based on listening tests, it is shown that manipulation of the directivity pattern along with a single distance-dependent gain proves to be effective and able to render sources closer than the distance of the loudspeaker, while manipulation of the gain alone is ineffective due to no D/R-ratio change. A practical implementation is demonstrated with the construction of an inexpensive small loudspeaker box able to generate second-order directivity patterns based on filters derived from anechoic measurements of each driver.

Acknowledgments

This work was funded by the Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany. The authors thank Tapani Pihlajamäki for helpful discussions.

References and links

Bronkhorst, A. W. (**2002**). "Modeling auditory distance perception in rooms," in *Forum Acusticum* (Sevilla, Spain).

Catic, J., Santurette, S., Buchholz, J. M., Gran, F., and Dau, T. (**2013**). "The effect of interaural-level-difference fluctuations on the externalization of sound," J. Acoust. Soc. Am. **134**, 1232–1241.

Farina, A., Capra, A., Chiesi, L., and Scopece, L. (**2010**). "A spherical microphone array for synthesizing virtual directive microphones in live broadcasting and in post production," in *40th International Conference of AES*, Tokyo, Japan.

Griesinger, D. (**2010**). "Phase coherence as a measure of acoustic quality. I: The neural mechanism," in *20th International Congress on Acoustics*, Sydney, Australia.

Härmä, A., van de Par, S., and de Bruijn, W. (**2008**). "On the use of directional loudspeakers to create a sound source close to the listener," in *124th International Convention of AES*, Amsterdam, The Netherlands.

Jeon, S.-W., Park, Y.-C., and Youn, D. H. (**2014**). "Auditory distance rendering based on ICPD control for stereophonic 3D audio system," IEEE Signal Process. Lett. **22**, 529–533.

Laitinen, M.-V., Disch, S., and Pulkki, V. (**2013**). "Sensitivity of human hearing to changes in phase spectrum," J. Audio Eng. Soc. **61**, 860–877.

Lounsbury, B. F., and Butler, R. A. (**1979**). "Estimation of distances of recorded sounds presented through headphones," Scand. Audiol. **8**, 145–149.

Mourjopoulos, J. (**1994**). "Digital equalization of room acoustics," J. Audio Eng. Soc. **42**, 884–900.

Pellegrini, R. S. (**2002**). "Perception-based design of virtual rooms for sound reproduction," in *22nd International Conference of AES*, Espoo, Finland.

Zahorik, P. (**2002**). "Direct-to-reverberant energy ratio sensitivity," J. Acoust. Soc. Am. **112**, 2110–2117.

Zahorik, P., Brungart, D. S., and Bronkhorst, A. W. (**2005**). "Auditory distance perception in humans: A summary of past and present research," Acta Acust. Acust. **91**, 409–420.