Binaural auralization based on spherical-harmonics beamforming

W. Song[a], W. Ellermeier[b] and J. Hald[a]

[a]Brüel & Kjær Sound & Vibration Measurement A/S, Skodsborgvej 307, DK-2850 Nærum, Denmark

[b]Institut für Psychologie, Technische Universität Darmstadt, Alexanderstraße 10, D-64283 Darmstadt, Germany

wksong@bksv.com

The binaural auralization of a 3D sound field using spherical-harmonics beamforming (SHB) techniques was investigated and compared with the traditional method using a dummy head. Psychoacoustic attributes of multi-channel reproduced sounds were measured in a listening experiment to validate the method subjectively. The results show that subjective ratings of the width, spaciousness and preference of different audio reproduction modes auralized based on SHB were not significantly different from those obtained for dummy head measurements. Thus binaural synthesis using SHB may be a useful tool to reproduce a 3D sound field binaurally while saving considerably on measurement time because head rotation can be simulated based on a single recording.

1 Introduction

Multi-channel audio has been increasingly used in automotive audio, home entertainment, and mobile phone applications, and there is a growing need for evaluating the subjective effects of such setups in listening experiments or for predicting them using objective measures. Rumsey [1] provided a framework for conceptualizing spatial attributes, which separates descriptions of sources, groups of sources, environments, and global scene parameters. Recent empirical studies [2, 3, 4] investigated the identification and quantification of auditory attributes of reproduced sounds in multi-channel setups, and described the relationship between specific auditory attributes and overall preference.

It has been shown that head rotation improves sound source localization, especially for sources located in the median plane [5, 6, 7]. Since localization may influence the judgment of other spatial auditory attributes, it appears reasonable to allow subjects to turn their head during listening tests, which involve assessing spatial sound attributes. This requires measuring binaural room impulse responses (BRIRs) at different head rotation angles, and therefore is a very time-consuming process. By contrast, beamforming [8] measures a sound field with an array of microphones in a "single shot", and can by means of computation steer its beam toward a particular direction. Furthermore, beamforming typically results in the sound pressure contribution toward the focused direction at the center of the array in the absence of the array, and this can be easily transformed to a pair of binaural signals [9] by incorporating binaural technology [10]. Due to these features, beamforming may be utilized to greatly improve the efficiency of BRIR measurements when compared to traditional dummy head measurements.

Therefore, the current study reports on an experiment to investigate the validity of using spherical-harmonics beamforming (SHB) [11, 12, 13, 14, 15] when auralizing a 3D sound field. The goals of this study are twofold:

1. To develop a binaural auralization method of a 3D sound field dependent on the listener's head rotation using SHB. To that effect, a procedure for estimating the BRIRs of individual loudspeakers in a room will have to be suggested using SHB [16].

2. To validate the proposed auralization method by obtaining subjective estimates of auditory attributes, such as width, spaciousness, and preference, in a listening experiment. Syntheses based on dummy head measurements and on SHB will be contrasted with respect to their subjective effects. Furthermore, the subject's head movement shall be controlled in such a way that they either rotate (with a head tracking system) or fix their head during listening tests.

2 Method

2.1 Subjects

Sixteen normal-hearing listeners between the age of 27 and 55 (15 male, 1 female) participated in the experiment. The subjects' hearing thresholds were checked using standard pure-tone audiometry in the frequency range between 0.25 and 6 kHz.

2.2 Apparatus and stimuli

2.2.1 Experimental setup

The experiment was carried out in a small listening room with sound-isolating walls and ceiling. Subjects were instructed to look straight ahead, and were not allowed to move their head in the fixed-head condition. They were instructed to rotate their head continuously within $\pm 30^\circ$ while listening to stimuli in the rotating-head condition. Their head movement was monitored through a window placed between the control room and the listening room.

Subject's head rotation was measured by a head tracker (Polhemus Fastrak) connected to a computer using an RS-232 connection. The receiver was attached to the headphones, and the transmitter was positioned on the table in front of the listeners. The update rate of the head tracker was 120 Hz. A real-time convolution software (customized for this kind of experiment by AM3D A/S) was employed to convolve the program materials with the selected BRIRs according to the subject's head rotation and to switch between different BRIR databases corresponding to different reproduction modes (see 2.2.3). The processed BRIRs had a length of 500 ms, and contained impulse responses from -30° to $+30^\circ$ of head rotation with an angular step size of 2°. In total there were 6 reproduction modes and 2 processing modes, which led to 12 BRIR databases, and they were loaded to the real-time convolution software before the listening experiment started. Two types of databases corresponding to the two different head motility conditions were generated, and the type of database was selected by the listening test program. The maximum response time of the real-time convolution software to movements of the listener's head was 15 ms at a 44.1 kHz sampling rate, which is sufficient for the current investigation.

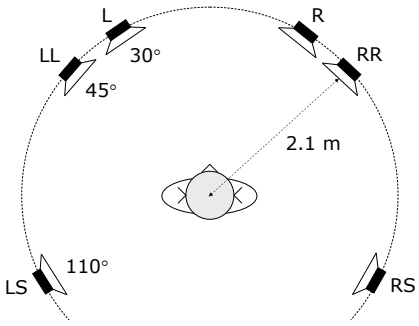

Figure 1: The loudspeaker configuration in the multichannel setup: left (L), right (R), left-of-left (LL), right-of-right (RR), left surround (LS), and right surround (RS).

Name	w	s	Speakers
phantom mono (PM)	0	0	L,R
weak stereo (s)	0.5	0	L,R
stereo (S)	1	0	L,R
wide stereo (WS)	1	0	LL,RR
weak surround (snd)	1	0.5	L,R,LS,RS
surround (SND)	1	1	LL,RR,LS,RS

Table 1: List of reproduction modes

2.2.2 Program materials

Two musical program materials, i.e. one pop and one classical, were selected from commercially available CDs. The classical music has a duration of 5:46 and the pop song of 4:41 min. The musical excerpts were repeated until the subjects completed their judgment of all reproduction modes presented on a given trial. The two program materials were selected to investigate whether their different musical content, spatial information, and recording techniques influenced the perception of spatial attributes as well as of overall quality as a function of the various reproduction modes.

2.2.3 Reproduction modes

The following equations were used to calculate the input of the four loudspeakers from the stereo program materials:

$$Y_L = X_L + (1-w)X_R \quad (1)$$

$$Y_R = X_R + (1-w)X_L \quad (2)$$

$$Y_{LS} = (X_L - X_R)s \quad (3)$$

$$Y_{RS} = (X_R - X_L)s \quad (4)$$

where X_L and X_R are the stereo signals, w is a coefficient determining the width of the stereo image, and s is a coefficient adjusting the level of surround channels. Notice that 'phantom mono' (identical signals being played through the stereo speakers) can be computed by using $w = 0$ and $s = 0$, and 'wide' stereo by using $w = 1$ and $s = 0$ while feeding the signals to the outer loudspeaker pairs, LL and RR (see Fig. 1). Six different reproduction modes (phantom mono, weak stereo, stereo, wide stereo, weak surround, and surround) were generated by selecting proper values of w and s, and the loudspeakers to play (see Table 1). This selection of reproduction modes was made in order to create a wide range of spatial impressions, thus making the comparison between the two auralization methods more general.

2.3 Measurements

The three different types of measurements using a microphone, a dummy head and a spherical microphone array were performed in a listening room. The room complies with the IEC 268-13 standard [17], which describes an "average living room" acoustically, and has dimensions of $2.8 \times 4.2 \times 7.8m$ ($H \times W \times L$). Six loudspeakers (Genelec 1031A) were positioned at 2.1 m from the center of the setup, and their positions are shown in Fig. 1. The microphone, the two ears of the dummy head, and the center of the spherical microphone array were all 1.25 m above the floor, aligned with the tweeters of the loudspeakers. Four of the six loudspeakers were arranged in accordance with the ITU-R BS.775-1 standard [18]: two additional speakers (LL and RR) were placed at $\pm 45°$ to generate a wider stereo image than the standard one based on $\pm 30°$ angular separation.

2.4 Procedure

The experiment consisted of two head motility conditions, i.e. fixed and rotating, to investigate the influence of head rotation on the audio quality of the auralization using SHB. Half of the subjects started judging the music samples in the fixed-head condition, and the other half in the rotating-head condition to minimize any order effects.

Quantification of two specific auditory attributes, width and spaciousness, as well as of overall preference was achieved by asking subjects to rate their subjective impression on the rating scales. The attribute to be judged was displayed at the top of the screen, and a set of scales was displayed below. Each scale had two end points, which were "narrow" and "wide" for width, "like a cigarette box" and "like a church" for spaciousness, and "not preferred" and "preferred" for preference. Definitions of the two attributes as given by Choisel and Wickelmaier [4] were presented to the subjects prior to the experiment. The subjects were allowed to choose their own criteria to judge overall preference.

The two processing modes (HATS, SHB) and six reproduction modes resulted in twelve scales being presented to the subjects on a given trial. Next to each scale, there was a corresponding button, which served to activate the selected reproduction mode. The activation of the selected reproduction mode resulted in a cross-fading from the previous BRIR database to the selected one. The three attributes and the two program materials required six trials per session, run either in the fixed or the rotating-head condition. The six trials were divided into three groups within each of which the same attribute was presented in two trials with the two musical excerpts. These three groups of trials as well as the two program materials within a group were pre-

sented in a random order to the subjects. The subjects were allowed to take a short break of 1 minute after each trial, during which they stayed in the listening room. A longer break of 10 minutes was taken outside of the listening room after every other trial. The subjects spent approximately 1.5 hour per day working on each head motility condition, resulting in 3 hours total.

3 Results

The ratings of the three auditory attributes were averaged across the 16 subjects for each reproduction mode in the two processing modes (HATS, SHB) and 95%-confidence intervals were determined. The outcome is shown in Figs. 2 to 5. The results of the dummy head measurements (HATS) are drawn with solid lines, and those of SHB with dashed lines. Notice that the graphical scales presented to the subjects were coded with values from 0 to 100, while the figures display a range between 10 to 80 to emphasize the effects.

When the pop music was presented in the fixed-head condition (see Fig. 2), as in all other conditions (see Figs. 3 - 5), the six reproduction modes differed markedly in preference, and in the ratings of the two spatial auditory attributes. The significance of this effect of the experimental manipulation was confirmed by performing a three-factor analysis of variance (ANOVA) [19] with the 6 reproduction modes, the 2 processing modes (SHB, HATS), and the 3 attributes all constituting within-subjects factor. This analysis indicated a highly significant effect of the reproduction mode [$F(5, 75) = 13.38$, $p < 0.001$], which incidentally was of similar magnitude in all other conditions studied (see Figs. 3 - 5). Furthermore, largely similar curves were obtained for the two processing modes, but the SHB processing produced higher responses than the dummy head synthesis, particularly for width and spaciousness. The statistical significance of this discrepancy shows up as a main effect of processing mode [$F(1, 15) = 6.51$; $p = 0.022$] in the ANOVA. It may be the effect of ghost images generated by sidelobes, which create the percept of additional diffuseness in the reproduced sounds.

As regards overall preference, the wide stereo (WS) and the two multi-channel reproduction modes (snd, SND) were judged quite similarly when comparing the two processing modes, but the subjects preferred the SHB processing over the dummy head synthesis in the three two-channel reproduction modes (PM, s, S). This may be due to the fact that the additional diffuseness created spatial impressions resembling those produced by the surround channels. It can also be seen that the subjects made quite similar responses when asked about width or spaciousness, and thus for this particular material hardly distinguished these two attributes. The participants generally preferred the wide stereo (WS) and the multi-channel reproduction (snd), while they disliked the reproduction mode with a higher level of surround channels (SND).

Judging the classical music excerpt reduced the differences between the two processing modes (HATS, SHB), except for judgments of width (see Fig. 3). Here, the main overall effect of processing mode did not reach

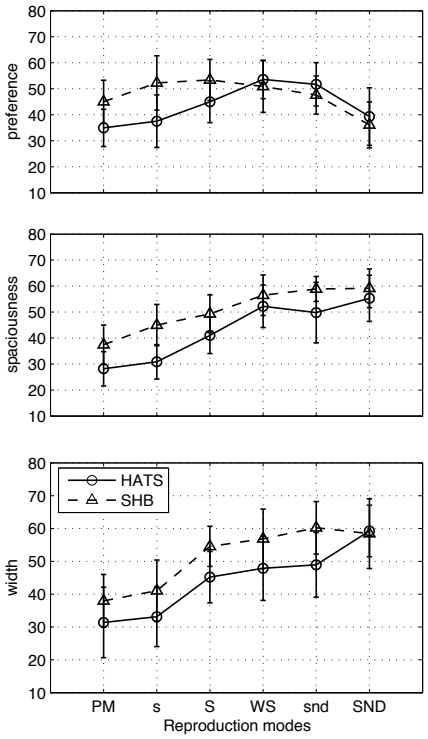

Figure 2: Sound quality ratings of the pop music excerpt in the fixed-head condition. Top: overall preference; center: spaciousness; bottom: width. Dashed line: SHB processing; sold line: HATS synthesis.

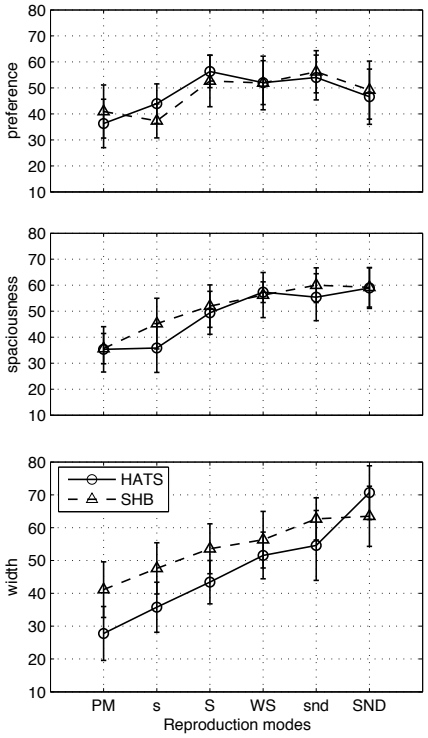

Figure 3: Sound quality ratings of the classical music excerpt in the fixed-head condition. Data arranged as in Fig. 2.

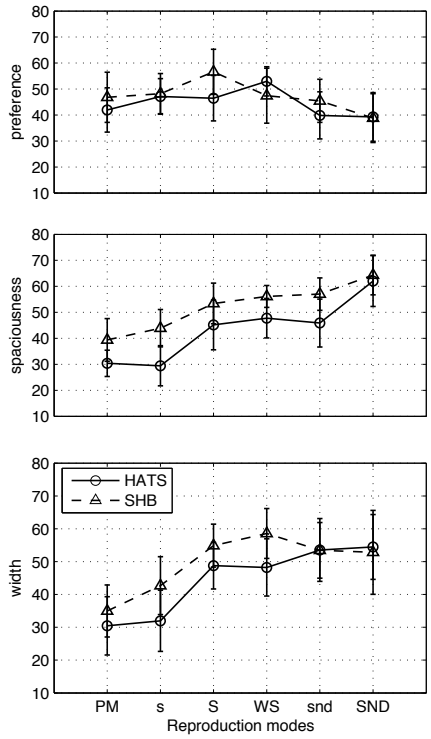

Figure 4: Sound quality ratings of the pop music excerpt in the rotating-head condition. Data arranged as in Fig. 2.

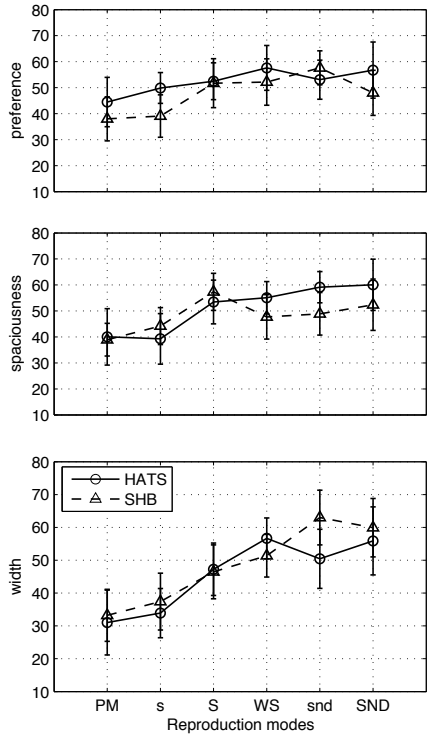

Figure 5: Sound quality ratings of the classical music excerpt in the rotating-head condition. Data arranged as in Fig. 2.

statistical significance [F(1,15) = 1.43; p = 0.25], but the three-way interaction between processing, the reproduction modes, and the attributes did [F(10, 150) = 1.91; p = 0.049], indicating that the divergence seen for the width ratings for the less complex reproduction modes (PM, s, S; bottom panel in Fig. 3) appears to be significant.

This indicates that the SHB processing can approximate listening to the sound fields recorded with a dummy head in terms of spaciousness, overall audio quality, and to some extent, width. For the classical music, the interpretation may be that the effect of ghost images only influences the perception of width, but not of spaciousness. It can still be seen that the two stereo (S, WS) and the two multi-channel reproduction (snd, SND) modes are almost equally preferred while the subjects did not prefer phantom mono (PM) and the narrow reproduction (s).

The results discussed so far imply that auditory attributes of recorded 3D sound fields may be faithfully rendered by measuring the sound field with a spherical microphone array, and reproducing it in a fixed-head condition. Width is the most sensitive attribute and somewhat affected by the beamforming processing, and the perception of the multi-channel reproduction modes (snd, SND) was less affected than that of the simpler reproduction schemes. The results seem to be dependent on the musical excerpts for spaciousness and preference, but not for width. The effect of head rotation will be analyzed in the following.

When the subjects were asked to rotate their head while listening to the pop music excerpt (see Fig. 4), almost identical responses were obtained for width and spaciousness. A four-factor analysis of variance with the two head motility conditions (fixed and rotating) constituting an additional within-subjects factor revealed no significant main effect of head motility condition [F(1, 15) = 0.02, p = 0.89], as well as no significant interactions of head motility with any of the other factors (p > 0.22). Nevertheless, the preference judgments appear to show a smaller effect of processing mode than was evident in the fixed-head condition (Fig. 2). The two multi-channel reproduction modes (snd, SND) are no longer preferred, and the two stereo reproduction modes (S, WS) are slightly preferred over the others.

For the classical music excerpt (see Fig. 5), the two head-motility conditions again yielded quite similar results, except for ratings of width (compare the bottom panels of Figs. 3 and 5). The effect of processing mode on the width ratings became smaller in the rotating-head condition. It is also interesting that in the rotating-head condition spaciousness of wide stereo (WS) and the two multi-channel reproductions was reduced for SHB compared to HATS while preference is quite similar to the fixed-head condition. This was evident in the significant interaction of the attribute judged with the head-rotation condition [F(2, 30) = 7.59, p = 0.002].

These results indicate that allowing for head rotation may modify sound quality judgments to some extent like seen in the rating of width for the classical music and of preference for the pop music, but it certainly does not reveal further differences between the two process-

ing modes (SHB, HATS) when compared to a fixed-head listening test. The results from the present investigation thus show that binaural auralization using SHB can be used for reproducing recorded 3D sound fields while listeners are allowed to rotate their head freely.

4 Conclusion

A binaural auralization method using spherical-harmonics beamforming (SHB) was developed, and it was validated by collecting subjective judgments of auditory attributes, i.e. width, spaciousness, and preference, in a multi-channel loudspeaker setup. When comparing this method with conventional measurements using a head-and-torso simulator, by and large quite similar subjective ratings of the auditory attributes were obtained. The results from the current investigation indicate that the suggested procedure can be applied to situations in which more efficient recording of 3D sound fields is required or where defined operating conditions cannot be repeated for measuring an entire set of head rotation angles, e.g. when auralizing on-road vehicle testing.

References

[1] F. Rumsey, "Spatial quality evaluation for reproduced sound: Terminology, meaning, and a scene-based paradigm", J. Audio Eng. Soc. **50**, 651–666 (2002).

[2] C. Guastavino and B. F. G. Katz, "Perceptual evaluation of multi-dimensional spatial audio reproduction", J. Acoust. Soc. Am. **116**, 1105–1115 (2004).

[3] S. Choisel and F. Wickelmaier, "Extraction of auditory features and elicitation of attributes for the assessment of multichannel reproduced sound", J. Audio Eng. Soc. **54**, 815–826 (2006).

[4] S. Choisel and F. Wickelmaier, "Evaluation of multichannel reproduced sound: Scaling auditory attributes underlying listener preference", J. Acoust. Soc. Am. **121**, 388–400 (2007).

[5] W. R. Thurlow and P. S. Runge, "Effect of induced head movements on localization of direction of sounds", J. Acoust. Soc. Am. **42**, 480–& (1967).

[6] S. Perrett and W. Noble, "The effect of head rotations on vertical plane sound localization", J. Acoust. Soc. Am. **102**, 2325–2332 (1997).

[7] P. Minnaar, S. K. Olesen, F. Christensen, and H. Møller, "The importance of head movements for binaural room synthesis", in *Proceedings of the 2001 International Conference on Auditory Display*, 21–25 (Espoo, Finland) (2001).

[8] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques* (Prentice Hall, London, Great Britain) (1993).

[9] W. Song, W. Ellermeier, and J. Hald, "Using beamforming and binaural synthesis for the psychoacoustical evaluation of target sources in noise", J. Acoust. Soc. Am. **123**, 910–924 (2008).

[10] H. Møller, "Fundamentals of binaural technology", Applied Acoustics **36**, 171–218 (1992).

[11] B. Rafaely, "Plane-wave decomposition of the sound field on a sphere by spherical convolution", J. Acoust. Soc. Am. **116**, 2149–2157 (2004).

[12] B. Rafaely, "Analysis and design of spherical microphone arrays", IEEE Transactions of Speech and Audio Processing **13**, 135–143 (2005).

[13] J. Meyer, "Beamforming for a circular microphone array mounted on spherically shaped objects", J. Acoust. Soc. Am. **109**, 185–193 (2001).

[14] J. Meyer and T. Agnello, "Spherical microphone array for spatial sound recording", in *Audio Engineering Society, 115th Convention*, preprint 5975 (New York, NY, USA) (2003).

[15] S. O. Petersen, "Localization of sound sources using 3D microphone array", Master's thesis, University of Southern Denmark (2004).

[16] W. Song, "Beamforming applied to psychoacoustics - sound source localization based on psychoacoustic attributes and efficient auralization of 3D sound fields", Ph.D. thesis, Aalborg University (2008).

[17] IEC 268-13, "Sound system equipment, part 13: Listening tests on loudspeakers", International Electrotechnical Commission, Geneva, Switzerland (1985).

[18] ITU-R BS.775-1, "Multichannel stereophonic sound system with and without accompanying picture", International Telecommunication Union, Geneva, Switzerland (1994).

[19] D. C. Montgomery, *Design and Analysis of Experiments* (Wiley, New York, USA) (2004).