Assessment of spatial audio signals using a higher order Ambisonics representation

Johann-Markus Batke, Florian Keiler

Technicolor Research and Innovation, Email: {Jan-Mark.Batke\Florian.Keiler}@technicolor.com

Introduction

Spatial audio signals represent information on sound fields in two or three dimensions. Stereo or surround sound are examples of well known formats for music which can recreate the impression of a two dimensional sound field. For three dimensional sound fields (e. g. with height) different extensions of the surround sound format exist, but none of them are well established at the moment. The more recently developed Auro-3D setup [2] appears to be one of the most promising approaches. It is adding four loudspeakers above an existing surround setup resulting in a number of nine loudspeakers.

All of the aforementioned formats carry signals that are directly related to the loudspeaker signals. A more abstract approach of describing three dimensional sound fields is given by the Ambisonics approach [4]. Here an approximation of the desired sound field is described by a number of coefficients that need to be decoded to the used loudspeaker setup. The accuracy of Ambisonics is bounded by the chosen order. In general, the order determines the spatial resolution and the number of coefficients of the obtained sound field description.

In this contribution the effect of Ambisonics encoding and decoding of Auro-3D content is evaluated. For playback a 3D loudspeaker setup is used with loudspeaker positions different to those of the Auro-3D setup. In particular, the impact of the order of the intermediate Ambisonics representation is assessed with regard to localisation errors and colouration effects in a listening test.

Loudspeaker Setups

The Auro-3D loudspeaker setup contains 9 loudspeakers. Five loudspeakers are located according to a 5.1 surround setup. Four elevated positions above the left, right, and surround loudspeakers are added. Figure 1 shows this setup as A1–A5 denoting the 5.1 surround setup and additional elevated positions A6–A9.

We compare the Auro-3D setup to a different 3D loudspeaker setup with 16 loudspeakers. In this setup 8 loudspeakers are regularly distributed on a circle around the listener. Additional 4 speakers are placed below and above this circle as shown in Figure 1 (open balls).

Transcoding of Auro-3D Signals

We aim at playing back content intentionally created for the Auro-3D loudspeaker setup on the 16-loudspeaker setup. The original Auro-3D signals are transcoded to this setup using Ambisonics as a intermediate format.

In a first step the $S = 9$ Auro-3D input signals in vector $\boldsymbol{x}$ are encoded using Ambisonics order N resulting in $O = (N+1)^2$ Ambisonics coefficients. The $O \times S$ encoding matrix $\boldsymbol{\Psi}$ contains the modal information of the Auro-3D loudspeaker positions [4]

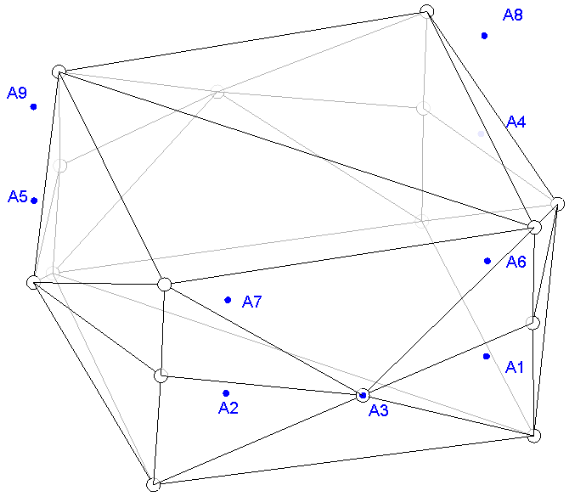

Figure 1: Auro-3D loudspeaker setup (A1–A9) and positions of our 16-loudspeaker setup (○).

$$\boldsymbol{d} = \boldsymbol{\Psi}\boldsymbol{x}. \quad (1)$$

The Ambisonics coefficients are arranged in an O-element vector $\boldsymbol{d}$ that need to be decoded to the addressed loudspeaker setup.

Decoding of vector $\boldsymbol{d}$ is written as

$$\boldsymbol{y} = \boldsymbol{D}\boldsymbol{d} \quad (2)$$

The required decoding matrix $\boldsymbol{D}$ has dimensions $L \times O$ and is used to compute the L loudspeaker signals in vector $\boldsymbol{y}$ for playback. The design of decoding matrix $\boldsymbol{D}$ depends on the positions of the addressed loudspeaker setup. The standard approach “mode matching” [4] has problems with spatial distributions as shown in Figure 1. Other approaches are required to overcome this problem. In our evaluation the Ambisonics decoding is facilitated by a Vector Base Amplitude Panning (VBAP) based approach [1, 5]. Here desired panning functions are defined by VBAP and they are smoothed by means of the spherical harmonics of the chosen order.

Encoding and decoding can be summarised by writing

$$\boldsymbol{y} = \boldsymbol{D}\boldsymbol{\Psi}\boldsymbol{x} = \boldsymbol{T}\boldsymbol{x}. \quad (3)$$

The combination of the encoding and decoding matrices gives the $L \times S$ transcoding matrix $\boldsymbol{T}$. This matrix $\boldsymbol{T}$ in turn can be seen as panning matrix that contains in each column the panning weights for the corresponding source position, which is the position of one of the Auro-3D loudspeakers. The gain values calculated by VBAP can directly be used in matrix $\boldsymbol{T}$. This is equivalent to using the aforementioned Ambisonics encoding and decoding with infinite order.

Assessment

In the listening test different rendering methods for the 16-loudspeaker setup have to be compared to the reference Auro-3D playback.

- mode matching decoding using order $N = 1$ (MM_1)
- VBAP-derived decoding using order $N = 5$ (VBAP_5)
- VBAP-derived decoding using order $N = 10$ (VBAP_{10})
- Direct VBAP panning (VBAP_∞)

The first-order mode matching decoder MM_1 serves as anchor since it leads to much lower localisation accuracy than the higher order decoders. A hidden reference is also included in the listening test. A methodology similar to MUSHRA (MUlti Stimulus test with Hidden Reference and Anchor) is used [3]. The impairments of several attributes taken from [6] have to be graded by the test persons in comparison to the Auro-3D reference on a scale from 1 (very annoying) to 5 (imperceptible). The used attributes include localisation (direction, width), timbre, and envelopment.

Four test signals with durations of 5–15 seconds have been used. Signal ’Speech’ is a short speech sequence with fixed position in the front left created with VBAP for the Auro-3D setup. The ’BumbleBee’ sequence is a synthetic audio scene also created with VBAP and contains a moving bumblebee plus other sounds with fixed positions. The signals ’Part1’ and ’Part4’ are excerpts of classical music recordings using microphones mixed for Auro-3D.

Twelve persons participated in the listening test. These test persons have small to medium experience in spatial hearing. Some of the test persons are expert listeners for evaluating artefacts of stereo codecs. The listening test was conducted in a well prepared listening room with a mean reverberation time of approximately 0.2 s.

Results

Figure 2 shows the results for the rating of the direction of single sources. The hidden reference and the MM_1 anchor were detected correctly. The comparison of the three VBAP variants does not show a significant difference between these methods, although for 2 signals and averaged signals VBAP_{10} is graded slightly better than the other two methods. The attributes’ mean grades of the three VBAP methods are rated approximately in the range 3–4 on the used scale. For a further investigation the difference grades are evaluated. Here two methods are compared and the grades of these methods are subtracted for all test persons. Figure 3 shows difference grades for VBAP_{10} vs. VBAP_∞ and all attributes. For signal ’Part4’ VBAP_{10} is graded significantly better than VBAP_∞. Averaged over all signals VBAP_{10} is graded significantly better for 4 of 7 attributes. The other comparisons using difference grades have the results that VBAP_{10} is graded better than VBAP_5 and that VBAP_∞ is graded better than VBAP_5 except for signal ’Part4’ and for attribute envelopment.

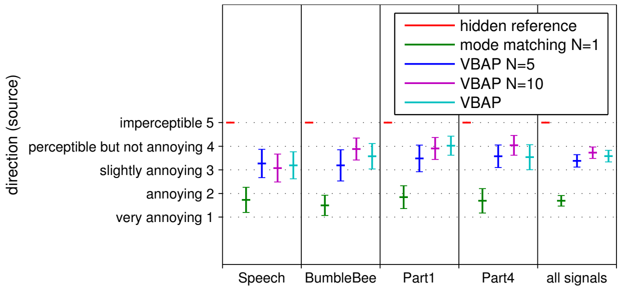

Figure 2: Listening test results for localisation/direction of single source; mean values and 95% confidence intervals are shown.

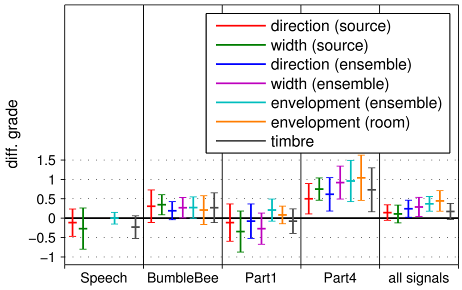

Figure 3: Difference grades for VBAP_{10} compared to VBAP_∞; mean values and 95% confidence intervals are shown.

Summary

The play back of content for the Auro-3D loudspeaker setup with 9 channels on a different 3D loudspeaker setup with 16 loudspeakers is assessed using a listening test. An Ambisonics representation of orders 5 and 10 with a VBAP-derived Ambisonics decoder is compared to the reference and to the direct rendering with VBAP. The smoothed version with Ambisonics order 10 is preferred to the original VBAP panning.

Acknowledgements

We would like to thank Gregor Zielinsky (Sennheiser electronic) for providing us with the Auro-3D music recordings.

References

[1] J.-M. Batke and F. Keiler, “Using VBAP-derived panning functions for 3D ambisonics decoding,” in *Proc. of the 2nd International Symposium on Ambisonics and Spherical Acoustics*, Paris, France, May 6-7 2010.

[2] Galaxy Studios Group, “Homepage Auro-3D,” http://auro-3d.com, visited 2010-07-16.

[3] ITU-R, *Recommendation BS.1534, Method for the Subjective Assessment of Intermediate Quality Level of Coding Systems*, 2001.

[4] M. A. Poletti, “Three-Dimensional Surround Sound Systems Based on Spherical Harmonics,” *J. Audio Eng. Soc.*, vol. 53, no. 11, pp. 1004–1025, Nov. 2005.

[5] V. Pulkki, “Virtual Sound Source Positioning Using Vector Base Amplitude Panning,” *J. Audio Eng. Soc.*, vol. 45, no. 6, pp. 456–466, Jun. 1997.

[6] F. Rumsey, “Spatial Quality Evaluation for Reproduced Sound: Terminology, Meaning, and a Scene-Based Paradigm,” *J. Audio Eng. Soc.*, vol. 50, p. 16, 2002.