DOI: 10.22032/dbt.39950

DOI (proceedings): 10.22032/dbt.39936

vdt Verband Deutscher Tonmeister e.V.

Implementation of Ambisonics Recordings in a Wave Field Synthesis System

M. Fehling, M. Nogalski, E. Wilk
Hamburg University of Applied Sciences, Germany, Email: matthias.fehling@haw-hamburg.de

Abstract

The paper presents different methods of implementation for Ambisoncs recordings in a wave field synthesis system. Using the Ambisonics B-format, virtual microphone signals are extracted from the recording. The signals thus generated can be placed and reproduced at the corresponding positions in the wave field synthesis system. To further improve the performance of Ambisonics recordings in the system the possibility of sound source localisation using a sound field microphone is explored. With the presented methods it is possible to determine the position of individual recorded sound sources and to create a virtual, aligned microphone signal pointing at the source. Reproducing a single sound source gives good results in the system. When playing multiple sound sources, the problems of ”sweet-spot” based systems arise.

1. Introduction

The Hamburg University of Applied Sciences maintains the Interactive Immersive Audiolab ($\mathrm{I}^2\mathrm{A}$)[1], a wave field synthesis system that is used for interactive audio drama productions as well as projects by a wide variety of artists.

In order to make content production more convenient for such systems and even use existing spatial recordings, it would be helpful to be able to utilize a more common and already established audio format, such as the Ambisonics format.

Especially the flexibility of Ambisonics recordings should be made usable in the system. Thus, the production of content can be facilitated and working with the system will become more accessible.

Our WFS system was developed by FourAudio [2] and consists of a rectangular assembly of linear speaker arrays. There are 56 speakers on the long sides of the rectangle and 48 on the short sides resulting in a total of 208 channels with a spacing of 10 cm between each speaker. The size of the rectangle is approximately 5 m by 6 m and the room containing the system is acoustically optimized.

The microphone used is an Ambeo microphone by Sennheiser.

Goal of the study is to play back a B-Format Ambisonics recording in the WFS system in a meaningful way.

In the following we are going to present methods used to prepare the signals for playback. This is followed by the presentation of the measurable results. The results are evaluated and an outlook for future development is given.

2. Extracting the Signal

The first thing that needs to be done to an Ambisonics recording is to extract a mono signal in a defined direction ϕ. This can be done by applying formula (1), [7]:

$$S = W \cdot A + (1 - A) \cdot (cos\phi \cdot X + sin\phi \cdot Y) \quad (1)$$

where *W, X, Y* are the corresponding signals of an Ambisonics B-format:

- W : omnidirectional signal

[1] https://i2audiolab.de/
[2] http://fouraudio.com

- X : figure eight signal on the x-axis
- Y : figure eight signal on the y-axis

Since the WFS system only operates in a horizontal plane, the Z signal, containing height information of the recording, is ignored.

The resulting signal S resembles the signal of a virtually directed microphone oriented to ϕ degrees. The directional characteristic can be set by mixing it with the omnidirectional W signal. The factor $A = 0...1$ thus controls the characteristic of the virtual microphone. ($A = 0$: figure eight; $A = 1$: omnidirectional).

The calculated signal can already be used in the WFS system by reproducing it from a sound source at the chosen angle ϕ. This way everything the microphone recorded from the direction determined by the angle ϕ can be reproduced from the same direction in the WFS system.

When trying to reproduce a full 360° soundfield this can be done by aligning multiple sound sources in a circle around the listening area and playing back signals created with formula (1). When using this method, the recording angles of the virtual microphones should be considered. Overlapping of the recording angles should be held minimal. When using hypercardioid characteristics ($A = 0.25$) for example, a setup with seven microphones with a spacing of $51°$ each would result in a minimal overlap of recording angles.

By using this method, a reproduction of the recording is possible. This way the Wave Field Synthesis is emulating a static positioning of seven loudspeakers. This would create a surround playback situation with a sweet spot and would not use the full potential of a WFS system.

It is possible to extract the spatial information of recorded sound sources from an Ambisonics recording. Thus, a recorded sound source can be located and played back at the correct position and the movement of the source can be reproduced correctly within the playback system.

3. Locating a Sound Source

The location of a recorded sound source can be extracted from the recording of the sound field microphone. By calculating the intensity vector $I = [I_\mathrm{X}\ \ I_\mathrm{Y}]$ the direction of the highest intensity at any moment of the recording can be determined. The angle ϕ can now be calculated from the complex argument of the intensity vector, [7]:

$$\phi = arg(I_\mathrm{X} + i \cdot I_\mathrm{Y}). \tag{2}$$

In their paper "Localization of the Sound Source with the Use of the First-order Ambisonic Microphone" Wierzbicki et al. describe three methods of locating sound sources using a soundfield microphone.

- by using the RMS value,
- by using the phase information of the W signal, or
- by using the product of the sound pressure and velocity values.

Since a RMS value can not be negative, using it would only lead to values between 0°and 90°. This requires correction to 360°, therefore this method is not used.

Using the phase information can be problematic since the capsules of a sound field microphone still have runtime differences. Despite the compact design of the microphone, there may occure inaccuracies with this method.

Therefore, we use the third method presented by [7]. It is assumed that with the W signal the pressure component of a recorded sound source is present and the X and Y signals contain the sound velocity information of the corresponding spatial axes. Since the sound intensity is defined as the product of sound pressure and sound velocity, the intensity can be determined as follows:

$$I_\mathrm{X}(n) = \sum_n X(n) \cdot W(n) \tag{3}$$

$$I_\mathrm{Y}(n) = \sum_n Y(n) \cdot W(n) \tag{4}$$

with (n) representing the sample number.

This method results in angles between 0°and 180°, which makes the correction to 360° much easier.

To localize a sound source from a recording, we use formula (3) and (4) to determine I_X and I_Y. With these values the intensity vector $I = [I_\mathrm{X}\ I_\mathrm{Y}]$ is set up and with formula (2) the angle of the vector is determined. This angle corresponds to the direction of the highest intensity, which matches with the direction of the sound source.

The angles calculated as described are transferred to the WFS system as information for the position of the sound source. During playback, the signal created with formula (1) is placed and played back in the system at the appropriate angle. The position of the sound source corresponds to the position during recording.

4. Implementation

The software for WFS rendering used in the system called *wonder*, was developed by the TU Berlin and under GPL license. To play back a signal in the system the following information is needed: The audio signal and the location that signal should be played from.

Using formula (1) the signal can be produced at the desired angle. The localisation process described in section (3) is applied to reproduce the movement of the recorded sound source. This results in a list of ϕ values for the duration of the signal. By using OSC (Open Sound Control) messages the location of the sound source is communicated to the system. Since only angles can be determined the distance to the center of the system was set to 6 m, just outside the radius of the physical loudspeaker array, to avoid problems that can arise when placing sources inside the physical speaker setup. Synchronicity between the signal and its location is achieved by starting the playback and the movement at the same time.

Several possible use cases were tested.

Firstly a single sound source at a known angle was recorded. The mono signal was generated using formula (1) and the signal is played back from the corresponding position in the WFS system.

Secondly multiple sound sources are created in a full circle in an attempt to generate a full 360° soundscape. The angles of the sources are chosen to keep the overlapping of the recording angles at a minimum.

In order to check the accuracy of the method described in section 3, a single moving sound source is recorded with the Ambeo microphone while it is simultaneously being tracked by an infrared tracking system. Thus a deviation from the actual position can be determined.

From the recording, angle ϕ is determined using formulas (3) and (2) for every 10 ms of the signal. The time interval of 10 ms corresponds to a spatial resolution of approximately 3.34 cm which is suitable for the speed at which the sound source was moved. The faster the sound source is moving, the shorter the chosen time interval should be.

For every angle ϕ, the audio signal is generated using formula (1). As microphone characteristic we chose a hypercardioid ($A = 0.25$) in order to focus the signal on the sound source. This results in a number snippets of 10 ms length which are composed together to the final audio signal of the moving source.

The implementations and their effects on listening impressions were evaluated by listening tests.

5. Results

When playing back a single sound source, the audio signal only contains the information from the desired angle. This is clearly audible in the WFS system when comparing the signal reproduced with a correct ϕ value to a signal that was aligned at a different direction.

This method of playback is suitable when playing back the recording of a single sound source. It is not possible to reproduce movement via this method.

When using multiple sources for playback (each source with its individual audio signal) the movement of the sound source can be heard in the WFS system. With this method the system is simulating a static loudspeaker positioning. When the recorded sound source is moving through an area without a speaker placed, the setup creates a phantom source, similar to conventional stereo and surround setups. The movement can be heard and followed by the listener, but only in a confined area in the middle of the system. If the listening position lies outside of this "sweet spot", locating the signal becomes difficult and the audio signal jumps between the virtual speakers. It is possible to implement Ambisonics

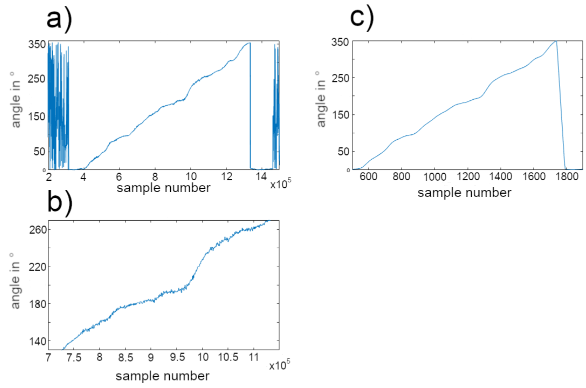

Fig. 1: (a): Calculated angles, (b): enlargement of (a), (c): calculated values after filtering

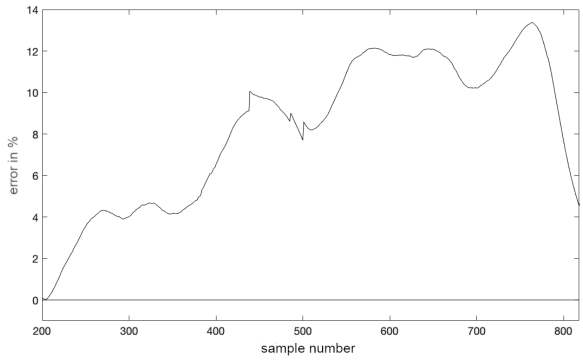

Fig. 2: Percentage error between calculated and tracked values

recordings in this way and it might be suitable to use them for ambience recordings. Though it should be avoided to generate a sweet spot inside a WFS system, as it is a benefit of a WFS system to not have a "sweet spot".

The third method consists in using the location data from the recording to position the sound source correctly in the WFS system.

Fig. 1 shows the calculated values for ϕ. During recording the source was moved around the microphone in a full circle. Therefore values between 0°and 360°are expected. As Fig. 1 a) shows these values can be extracted from the recording. When playing back the signal using these values for ϕ the sound source is not following a clear path, because the position of the source in the WFS system is changing rapidly. These changes in the calculated angles can be seen in fig. 1 b) and lead to audible artifacts during playback. The calculated values are therefore smoothed using a low pass filter. Fig. 1 c) shows the ϕ values after smoothing. Playing back the signal using smoothed angle values yields a much better result. The movement of the source can be followed from any position inside the WFS system.

Fig. 2 shows the percentage error between calculated angle ϕ and the values of the infrared tracking system. Comparing these values results in a maximum error of about 13 %, with a significant increase in the error at low signal amplitude.

The filterlength should be chosen as small as possible while still being long enough to smooth the data to avoid large jumps in the position.

6. Summary

We have shown that it is possible and useful to play back Ambisonics recordings in a Wave Field Synthesis system.

Single audio signals can be extracted from the recording and played back at the desired position. This method, though possible, is not recommended for a WFS system since a soundfield microphone is not required. A simple recording with a conventional microphone in a dry environment would yield the same, if not better results.

Distributing multiple signals at their corresponding positions in a circle leads to a spatial play back utilizing phantom sources. This method is not recommended, since it shows the same restrictions of conventional surround setups: a ”sweet spot” resulting in a confined listening area.

Most promising is the approach of locating the recorded sound source and transferring the location data to the WFS system. Locating the sound source with a soundfield microphone is possible with a small error. When playing back the signal while controlling the movement accordingly, the result is a playback situation where the source can be located at the correct position. Using the system like this, the advantages of WFS systems consisting of a good spatial localisation inside the system at any point, are utilized.

Nevertheless, the method to localize the sound source presented here has its restrictions. Since only the direction of the source can be calculated, it is not possible to determine the distance between the source and the microphone. This could be done by using two microphones and calculating the intensity vector for each microphone, therefore making it possible to determine the intersection of those vectors.

Another restriction is the possibility to process only a single sound source. If the recording contains two or more sources, the resulting intensity vector does not necessarily point to the correct source. This would be a task for future development, since being able to detect multiple sound sources from a single recording and placing them correctly within the WFS system would enhance the effect in the system.

7. References

[1] Bates, Enda and Dooney, Sean and Gorzel, Marcin and O'Dwyer, Hugh and Ferguson, Luke and Boland, Francis M. : Comparing Ambisonic Microphones—Part 2. Audio Engineering Society Convention 142, 2017 May

[2] Benjamin, Eric and Chen, Thomas: The Native B-Format Microphone. Audio Engineering Society Convention 119, 2005

[3] Fohl, Wolfgang: Sound-Perception-Performance. 243–255, Springer, 2013

[4] Fohl, Wolfgang and Wilk, Eva: Enhancements to a Wave Field Synthesis System to Create an Interactive Immersive Audio Environment. Proc. 3rd Int. Conf. on Spatial Audio, 2015.

[5] Frank, Matthias and Zotter, Franz and Sontacchi, Alois: Producing 3D Audio in Ambisonics. Proceedings of the AES International Conference, 2015

[6] Pulkki, Ville: Spatial Sound Reproduction with Directional Audio Coding. J. Audio Eng. Soc. Volume 55, 2007

[7] Wierzbicki, J. and Malecki, P. and Wiciak, J.: Localization of the Sound Source with the Use of the First-order Ambisonic Microphone. Acta Physica Polonica A, Vol. 123 (2013) DOI: 10.12693/APhysPolA.123.1114

[8] Woszczyk et al.: Tetrahedral Microphone: A Versatile ”Spot” and Ambience Receiver for 3D Mixing and Sound Design, 2018