Audio Engineering Society

Convention Paper 7987

Presented at the 128th Convention
2010 May 22–25 London, UK

Linear Simulation of Spaced Microphone Arrays Using B-Format Recordings

Andreas Walther[1] and Christof Faller[1]

[1] *Audiovisual Communications Laboratory, EPFL, 1015 Lausanne, Switzerland*

Correspondence should be addressed to Andreas Walther (andreas.walther@epfl.ch)

ABSTRACT

A novel approach for linear post-processing of B-Format recordings is presented. The goal is to simulate spaced microphone arrays by approximating and virtually recording the sound field at the position of each single microphone. The delays occurring in non-coincident recordings are simulated by translating an approximative plane wave representation of the sound field to the positions of the microphones. The directional responses of the spaced microphones are approximated by linear combination of the corresponding translated B-format channels.

1. INTRODUCTION

There exists a vast amount of literature on the topic of stereo recording. Different methods aim at evoking the perception of a reasonable spatial stereo image and spatial impression during reproduction. Stereo recording principles valuable for two channels can also be applied for multi-channel recording using some additional considerations, such as recording of a center channel. Amongst the many available publications, recording methods are described in [1, 2, 3, 4, 5].

There has been a debate on which microphone technique is optimal for recording spatial audio (two-channel or multi-channel). Some have argued that most precise localization is obtained using coincident microphones [1, 2]. On the other hand, coincident recordings are often criticized for a certain lack of spaciousness. Non-coincident techniques produce phantom sources with higher source width due to the occurring time differences. Therefore, some prefer them for the spacious sound impression they produce. The goal of this paper is not to state an opinion on the merits and drawbacks of the various microphone techniques. The important aspect for our purpose is the notable different sound image produced by coincident and non-coincident techniques.

This paper proposes a technique that processes coincident B-Format [6, 7, 8, 9, 10, 11, 12] recordings to mimic spaced microphone setups. Only linear processing is used. While it is by no means possible to precisely obtain the signals one would obtain from a spaced microphone setup, the goal is to obtain a crude approximation thereof as a tool for alternative B-format to stereo or surround sound conversion.

Given a complete plane wave representation of a sound field, it is possible to translate the measurement point in space. A high-order B-format signal is equivalent to a plane wave representation and can thus also be translated in space. The conversion of a B-format recording to emulate a spaced microphone recording is then, generally speaking, done as follows:

- Translate the B-format to the location of each microphone of the target spaced microphone setup.
- Linearly combine the translated B-format channels to obtain the directivity of each of the spaced microphones.

Given this insight and the necessary simplifications that have to be made due to consideration of only first-order B-format, it is explained how to obtain an estimation of the translated B-format signals and approximate the characteristics of a spaced microphone array.

The paper is organized as follows. Section 2 motivates and details the B-format translation that is used and its application to simulating spaced microphone recordings. Section 3 gives two examples for simulating an ORTF [1] stereo microphone setup based on B-Format. Discussion and conclusions are in Sections 4 and 5, respectively.

2. SIMULATING SPACED MICROPHONES FROM B-FORMAT

2.1. Translation of plane wave representation

One way of describing a plane wave representation of a sound field is by means of a source distribution function $S(t, \phi)$ [13, 14], which for each angle

[1]Named after the former French national broadcaster *Office de Radiodiffusion Télévision Française*.

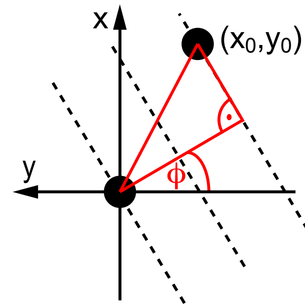

Fig. 1: Translation of observation point.

ϕ corresponds to a signal composed of plane waves arriving from angle ϕ. $S(t, \phi)$ is given relative to a specific point in space. In our case we define $S(t, \phi)$ to be defined relative to the origin ($x = 0$, $y = 0$). If $S(t, \phi)$ is translated to a new point of observation, i.e. $x = x_0$ and $y = y_0$, then the sound (plane waves) from direction ϕ travels

$$d(\phi) = x_0 \cos\phi + y_0 \sin\phi \tag{1}$$

meters less far ($d < 0$ means that the sound has to travel farther). This is illustrated in Figure 1.

Thus, the sound will be delayed by $-d/c$, where c is the speed of sound in air (≈ 340 m/s). The source distribution relative to the new origin, (x_0, y_0), is

$$\tilde{S}(t, \phi) = S(t + \frac{d(\phi)}{c}, \phi)\,. \tag{2}$$

2.2. Translation of B-format

The idea of the first order B-format translation is to convert it to an approximation of the source distribution $S(t, \phi)$. This approximated $S(t, \phi)$ is then translated to a new point in space using (2) and then converted back to B-format.

Let the B-Format recording be defined as

$$\mathbf{B} = \begin{bmatrix} w(t) \\ x(t) \\ y(t) \end{bmatrix}, \tag{3}$$

where $w(t)$, $x(t)$, and $y(t)$ are the B-format omni, x-dipole, and y-dipole signals. Since we only consider horizontal translations in this paper, we do not consider the z-dipole signal. Note that in the following, for simplicity, we assume that $w(t)$, $x(t)$, and $y(t)$ have the same gain.[2]

$S(t, \phi)$ is approximated by considering only a number of discrete directions, i.e.

$$S(t, \phi) = \sum_{i=1}^{N} s_i(t)\delta(\phi - \phi_i)\,, \tag{4}$$

where δ is the Dirac delta function, the considered directions are ϕ_i, and the signals arriving from these directions are $s_i(t)$. We are using $N = 3$ directions, motivated by the fact that three little overlapping directional first order responses fit into the horizontal plane. The signals $s_i(t)$ are computed as first order signals pointing towards the considered directions ϕ_i. In the following, the signals $s_i(t)$ are denoted *intermediate format*.

The three intermediate format signals $s_i(t)$ are obtained using

$$\mathbf{S} = \mathbf{MB} \tag{5}$$

where matrix

$$\mathbf{M} = \begin{bmatrix} \alpha_1 & (1-\alpha_1)\cos\phi_1 & (1-\alpha_1)\sin\phi_1 \\ \alpha_2 & (1-\alpha_2)\cos\phi_2 & (1-\alpha_2)\sin\phi_2 \\ \alpha_3 & (1-\alpha_3)\cos\phi_3 & (1-\alpha_3)\sin\phi_3 \end{bmatrix} \tag{6}$$

defines three microphone responses according to α_i,[3] pointing towards directions ϕ_i.

$\mathbf{S}$ is translated to the new point of observation (x_0, y_0) by applying delays $-d(\phi_i)/c$ to the signals $s_i(t)$, i.e. $\tilde{s}_i(t) = s_i(t + d(\phi_i)/c)$. Given the so-obtained translated intermediate signals $\tilde{\mathbf{S}}$, the translated B-Format is computed as

$$\tilde{\mathbf{B}} = \mathbf{M}^{-1}\tilde{\mathbf{S}}\,. \tag{7}$$

2.3. Simulating spaced microphones

The procedure to simulate a spaced microphone setup is:

- Define the microphone positions such that the origin is symmetrically centered within the target spaced microphone setup.[4]

[2] B-Format otherwise is defined with $x(t)$ and $y(t)$ having 3 dB additional gain compared to $w(t)$.

[3] For example $\alpha_i = 0.5$ corresponds to a cardioid directional response.

[4] The origin is chosen symmetrically centered to ensure similar performance for left and right directions.

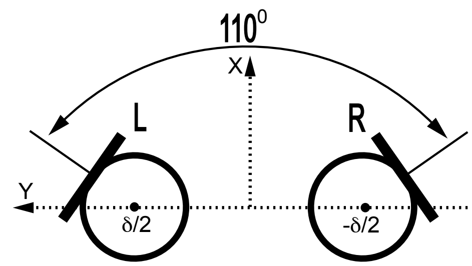

Fig. 2: ORTF stereo microphone setup.

- For each microphone to be simulated do:

- Compute the intermediate format signals $\mathbf{S}$ using (5). Note that for each simulated microphone the same or different intermediate format signals may be used.

- Translate the intermediate format signals to the position of the target microphone and compute the translated B-Format using (7).

- Compute a linear combination of the translated B-Format using direction φ and directivity parameter α of the target microphone,

$$r_\varphi(t) = \alpha\tilde{w}(t) + (1-\alpha)(\tilde{x}(t)\cos\varphi + \tilde{y}(t)\sin\varphi)\,, \tag{8}$$

or, a linear combination considering that the result may not be as expected due to the modified nature of the translated B-Format.

3. EXAMPLE ORTF SIMULATIONS

We arbitrarily chose the well known ORTF setup as an example for demonstrating the proposed technique. ORTF consists of two cardioid microphones spaced $\delta = 17$ cm apart and spanning a 110° angle as shown in Figure 2. The origin in the figure corresponds to the reference position for the B-Format. The positions of the left and right microphones are $(0, \frac{\delta}{2})$ and $(0, -\frac{\delta}{2})$, respectively.

Used for comparison later, Figure 3 depicts the far field impulse response of the left cardioid signal of the ORTF setup as a function of direction of arrival. Front directions between −90 and 90 degrees are shown. The time axis in all subsequent graphs is defined such that time zero corresponds to the time when the impulse would reach the origin.

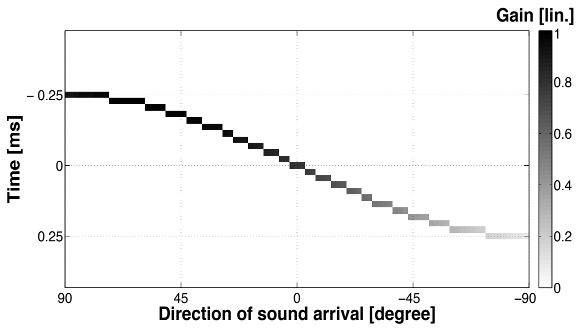

Fig. 3: Impulse responses of the left ORTF microphone as a function of direction of arrival of sound.

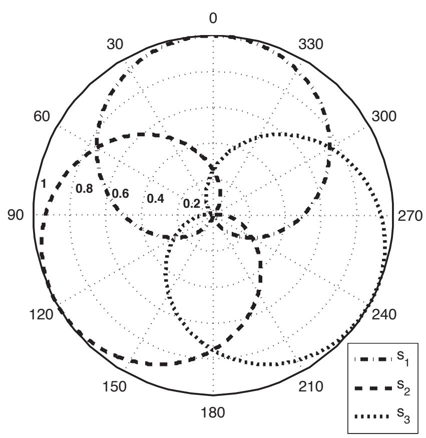

Fig. 4: Example directional responses of intermediate format signals.

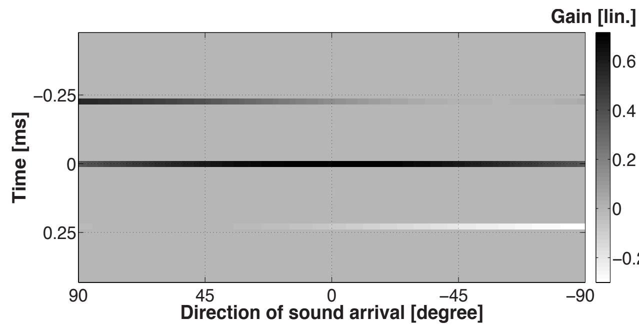

Fig. 5: Impulse responses of the simulated left ORTF microphone as a function of direction of arrival of sound.

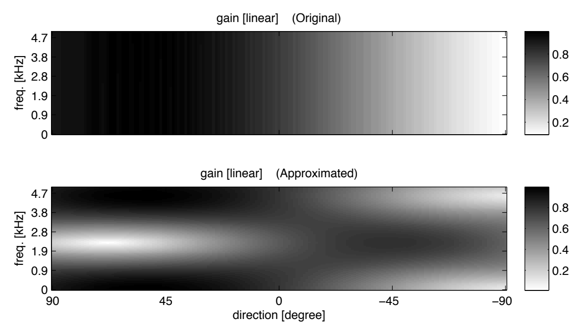

Fig. 6: Magnitude response of the original (top) and simulated (bottom) left ORTF microphone.

3.1. Using one intermediate format

The first example uses one left-right-symmetric intermediate format **S** for simulating both spaced microphones of the ORTF setup to illustrate some traits of the translation procedure. **S** was chosen to consists of three cardioid responses ($\alpha_i = 0.5$) pointing towards $\phi_1 = 0°$, $\phi_2 = 120°$, and $\phi_3 = 240°$, as shown in Figure 4. In the following, data is only shown for the left microphone. Results for the right microphone are corresponding.

S is shifted to position $(0, \frac{\delta}{2})$ to compute the translated B-Format $\tilde{\mathbf{B}}_{\text{left}}$. Subsequently, the approximation of a cardioid microphone pointing towards φ=55° is obtained using (8) with $\alpha = 0.5$ and $\varphi = 55°$. Figure 5 shows the resulting impulse response of the left simulated ORTF cardioid microphone as a function of direction of arrival of sound. While the impulse response of the target ORTF left cardioid (shown in Figure 3) has a dedicated delay and gain for each direction, the simulated microphone has three impulses with fixed delays with direction dependent amplitudes. The number of impulses is equal to the number of channels of the intermediate format.

In the given example, the impulse responses of the original left microphone of the ORTF setup (Figure 3) and the simulated microphone (Figure 5) at first sight look very different. To get more insight on similarity and differences, Figure 6 shows the magnitude response of the original left ORTF microphone

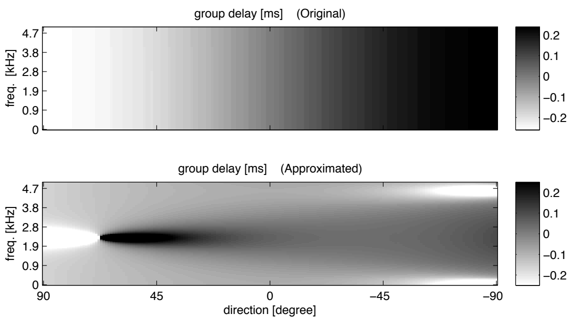

Fig. 7: Group delay of the original (top) and simulated (bottom) left ORTF microphone.

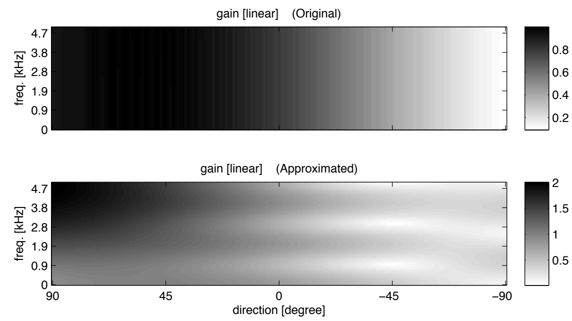

Fig. 8: Magnitude response of the original (top) and simulated (bottom) left ORTF microphone, using tuned intermediate format.

(top panel) and the simulated ORTF microphone (bottom panel). Only frequencies up to 5 kHz are shown, since the data periodically repeats as a function of frequency. The corresponding group delay for the original (top panel) and simulated (bottom panel) microphone are shown in Figure 7. [5] Note that the gain up to about 1.5 kHz is similar between the original and simulated microphone. More generally, around frequencies 0, 4.7, 9.4 kHz, etc. the magnitude response of the original and simulated microphones are similar and between those frequencies they deviate. The group delay does not match well between the original and simulated microphone for any frequency.

Those graphs show that the characteristics of the approximated microphone are not very similar to those of the original. The response characteristics (α_i) and the associated directions (ϕ_i) of the intermediate format strongly influence the overall result. In the following it is shown that a better approximation can be achieved when separate intermediate formats are used for each target microphone.

3.2. Using an intermediate format for each target microphone

The goal was to see if we could improve magnitude response and group delay similarity between the original and simulated ORTF microphones by changing parameters of the intermediate format. A reasonable pointer while looking for good intermediate format parameter values is to consider the magnitude response and the group delay of the imposed filter at a number of reference directions. Plausible choices are for example the limits of the recording angle[6] of the target microphone setup and the center position. The results shown in the following have been experimentally derived by taking into account the reference points 48°,0°, and -48°.

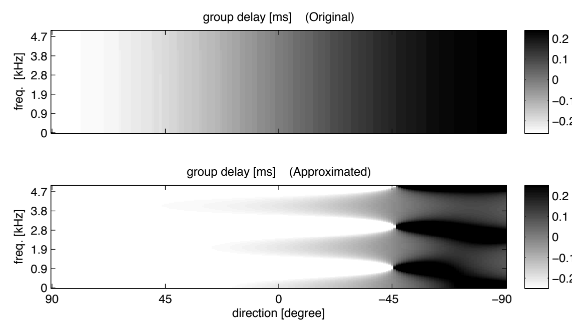

Fig. 9: Group delay of the original (top) and simulated (bottom) left ORTF microphone, using tuned intermediate format.

Keeping the parameters α=0.5 and $\varphi = 55°$ and only manipulating the intermediate format $\mathbf{S}$, a reasonable result for the left microphone of the ORTF setup can be achieved. Manual optimization lead us to use three cardioid responses with directions 90°,

[5] For better clarity, the shown group delay has been limited to the maximum absolute value occurring in the target left ORTF microphone.

[6] The recording angle indicates the useful pick-up sector of a stereo microphone setup. See [15] for a detailed discussion.

145°, and 270° for the intermediate format. The resulting magnitude response and group delay are shown in Figures 8 and 9, [5] respectively. Again, the top panel corresponds to the original left microphone of the ORTF setup and the bottom panel to the simulated left microphone.

It can be seen, that while the target values are again not met, the rough trend of both the magnitude response and the group delay follow the trend of the original. Due to the symmetry of the target microphone setup, we use three cardioid responses with directions 270°, 215°, and 90° for the intermediate format used to simulate the right microphone.

4. DISCUSSION

Analytically accurate simulation of a specific spaced microphone setup is generally not achievable by the proposed linear processing using first order B-format. However, a sound impression mimicking in part the characteristics of spaced microphone arrays can be created. Considering the properties of an exemplary spaced microphone array provides a good basis for optimizing the parameters of the approximation. In the described examples, we restricted the approximation by keeping parameters α and φ, and the translation position as indicated by the target spaced microphone setup. However, better results may be possible by optimizing all parameters and by using frequency dependent parameters.

The preliminary results presented in Section 3.2 show the applicability of the proposed method. The resulting sound image does not resemble that of an original ORTF recording. However, compared to coincident decoding of the B-Format recording, a distinct different, yet reasonable sound image can be achieved. It should be understood, that due to the inherent limitations of the approach, it can by no means substitute recordings made by spaced microphone arrays. Even so, used as a tool for B-Format processing, interesting results can be obtained.

5. CONCLUSIONS

A new method for post-processing of B-Format recordings has been proposed. The characteristics of spaced microphone recordings are imitated based on coincident recordings. The underlying theory of sound field translation was briefly introduced and the limitations imposed by considering only first order B-Format were explained. The application of sound field translation to the simulation of spaced microphone arrays was presented. By means of an example, the limitations of the approach have been demonstrated and an example was shown for optimizing the parameters.

ACKNOWLEDGMENT

Andreas Walther's research is funded by the Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany.

6. REFERENCES

[1] A. Blumlein, "Improvements in and relating to sound transmission, sound recording and sound reproduction systems," *British Patent Specification 394325*, 1931, reprinted in *Stereophonic Techniques*, Aud. Eng. Soc., New York, 1986.

[2] S. P. Lipshitz, "Stereo microphone techniques: Are the purists wrong?" *J. Audio Eng. Soc.*, vol. 34, no. 9, pp. 716–744, Sept. 1986.

[3] M. Williams, "Unified theory of microphone systems for stereophonic sound recording," in *Preprint 82nd Conv. Aud. Eng. Soc.*, Mar. 1987.

[4] G. Theile, "Natural 5.1 music recording based on psychoacoustic principles," in *AES 19th International Conference*, 2001.

[5] J. Eargle, *The Microphone Book*. Focal Press, 2004.

[6] M. A. Gerzon, "Periphony: Width-Height Sound Reproduction," *J. Aud. Eng. Soc.*, vol. 21, no. 1, pp. 2–10, 1973.

[7] ——, "The design of precisely coincident microphone arrays for stereo and surround sound," in *Preprint 50th Conv. Aud. Eng. Soc.*, Mar. 1975.

[8] K. Farrar, "Soundfield microphone: Design and development of microphone and control unit," *Wireless World*, pp. 48–50, Oct. 1979.

[9] ——, "Soundfield microphone - 2: Detailed functioning of control unit," *Wireless World*, pp. 99–103, Nov. 1979.

[10] E. Benjamin and T. Chen, "The native B-format microphone: Part 1," in *AES 119th Convention, New York, USA*, 2005.

[11] H. Wittek, C. Haut, and D. Keinath, "Doppel-MS - eine Surround-Aufnahmetechnik unter der Lupe," in *24. Tonmeistertagung, Leipzig, Germany*, 2006.

[12] D. G. Malham and A. Myatt, "3-D sound spatialization using ambisonic techniques," *Computer Music Journal*, vol. 19, no. 4, pp. 58–70, Winter 1995.

[13] M. A. Poletti, "A unified theory of horizontal holographic sound systems," *J. Aud. Eng. Soc.*, vol. 48, no. 12, pp. 1155–1182, Dec. 2000.

[14] C. Faller, "Signal processing for audio and acoustics," 2006, course Notes, Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland.

[15] H. Wittek and G. Theile, "The recording angle - based on localisation curves," in *AES 112th Convention, Munich, Germany*, 2002.