Transferable Acoustics based on Spatial Analysis and Re-Composition

Klaus Hostniker, Alois Sontacchi

Institute of Electronic Music and Acoustics, University of Music and Performing Arts Graz, A-8010 Graz
sontacchi@iem.at
http://iem.at

Abstract

Within this paper we present a new concept how to transfer the acoustic properties form one to another room, whereby both rooms should be equipped with a 3D audio rendering system. Starting from specific assumptions the transferability will be demonstrated. The basic required procedures will be introduced and deduced. At first, an acoustic measurement procedure is applied, both to the target and the reproduction room to retrieve directional room impulse responses. Secondly, resulting measurement responses are analyzed to obtain metadata, which are used for the spatial re-synthesis process. Based on these metadata the transfer procedure will be derived by using the compact notation of spherical harmonics (SH). This approach will allow optimal and efficient overall results. Besides that, it provides the possibility to examine systematically differences in several measured rooms, and to modify the properties of the observed reproduced room acoustics. Primarily results of the spatial analysis stage are given in this article.

1 Introduction

Room responses are closely related to spatial impressions and perception of sound in a room or hall. In such environments the propagation of sound has a significant influence on the auditory impression. The acoustic properties can be described by many parameters. In [1-3] several important aspects concerning the spatial impression have been reported and various objective descriptive parameters have been developed. To characterize the acoustics of natural rooms standardized measurement procedures e.g. ISO NORM 3382 and ISO NORM 18233 have been defined. However, the measurement of room impulse responses with a single undirected microphone is to the greatest extent insufficient.

The presented approach is motivated by practical reasons. Modern electronic music pieces e.g. operatic productions typically imply spatial sound design and effects. Sound rendering techniques like Ambisonics (cf. [4]) have proven to provide great improvements for such applications. Ambisonics simplifies the adaptation to various loudspeaker layouts at different event locations without considering any further changes to the defined space-time-trajectories of the sound objects. Whereby, several realizations of such perceivable trajectories depend on the concrete available room acoustics. Therefore, special rehearsals for the spatial aspects are commonly required. These additional rehearsals often stress budget or even exhibit at late-stage that the compositional ideas concerning the spatial effects cannot be realized, as considered. To overcome this problem a practical solution is requested.

We aim to transfer the acoustic space from a given event location to our laboratory to provide composers the possibility to examine the target acoustic sound scene. Alternatively, the established spatial sound scene in our labs is transferred to another place. Therefore, we are heading a twofold aim. On the one hand the aim is to measure and analyze directional room impulse responses that are independent of the source signal. On the other hand stands the aim of a reproduction and transfer of natural room acoustics.

To provide a systematic approach, measurements in real rooms are examined. As a concrete example here, the paper describes the analysis and re-synthesis procedure of collected directional room impulse responses from the concert hall in the MUMUTH (House of Music and Music Theatre) and the IEM Cube (Institute of Electronic Music and Acoustics) at the University of Music and Performing Arts in Graz, Austria (cf. Fig. 1). In both rooms an Ambisonics 3D sound rendering system for the upper hemisphere is installed.

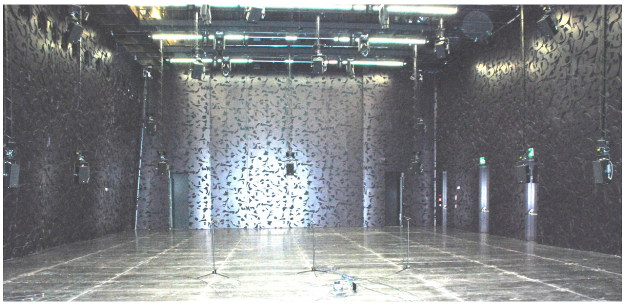

Figure 1: Individual directed room impulse responses measurements of each loudspeaker in B-format in the center of the room (Photo: MUMUTH, H. Pomberger).

Recent studies (cf. [5-8]) consider multi-channel coincident measurements to determine the location of sound source in the three dimensional. Additionally, they provide methods and quantities to derive and describe the spatial properties of sound sources. However, these approaches do not deal with short transient sound events, as it is required while facing with room impulse responses. Hence, we will sketch an algorithm to derive the requested parameters.

In the case of (Higher Order) Ambisonics rendering systems we will show how to combine the obtained results within the reproduction procedure. The combined approach will provide a new possibility to transfer or establish different room acoustics and in addition, it will allow a reduction of interfering and annoying reflections within the reproduction room considering certain reproduction constellations.

The paper is outlined as follows, in section 2 the concept, assumptions and required steps are introduced. The descriptions of the analysis and re-synthesis procedures are given in section 3. Section 4 shows first results and section 5 concludes this contribution.

2 Method

The transfer of acoustic properties can be realized with several rendering techniques. However, the advantage of the proposed schema, using the spherical harmonics (SH) is the straight forward integration into the existing Ambisonics rendering concept. It will be shown that, both the equalization of the room acoustics in the target room as well as the transferred acoustic properties can be easily adjusted to the general conditions, e.g. different loudspeaker setups. Furthermore, the requested effort in spatial accuracy concerning direct sound, early reflections and late diffuse reverberation can be independently scaled.

The proposed acoustic transfer concept is subject to restrictions and specific assumptions. Firstly, both rooms, *Room*$_0$, where the acoustic properties are captured as well as the other *Room*$_1$, to where these properties are transferred have to measured and analyzed. Secondly, at least a loudspeaker array in *Room*$_1$ for the rendering is required. The excitation of *Room*$_0$ can also be realized with single loudspeaker sampling the evaluated enclosing surface at several defined positions (cf. [9]). Thirdly, *Room*$_1$ should be acoustically highly damped otherwise its equalization has to be considered. Finally, only controllable reflections can be modified i.e. using upper hemisphere loudspeaker arrangements cannot alter ground reflections.

2.1 Concept

Our description starts in the *Room*$_0$, which acoustics should be transferred to the target *Room*$_1$. In the following equations the used representatives are frequency dependent variables, whereby the argument is omitted in order to simplify the description. Furthermore, the used vectors are time-varying but the matrices remain invariant.

$$\mathbf{b}_0 = \begin{bmatrix} H_1^{11} & \cdots & H_l^{11} & \cdots & H_L^{11} \\ \vdots & \ddots & & & \\ H_1^{nm} & & H_l^{nm} & & \\ \vdots & & & \ddots & \\ H_1^{NN} & \cdots & & & H_L^{NN} \end{bmatrix} \cdot \mathbf{x} \qquad (1)$$

At the left side of eq. 1 we see the measureable physically resulting spherical harmonics description of the overall room response at the microphone position. Matrix $\mathbf{H}_0$ in eq.1 covers all the transfer functions from each loudspeaker l across the room to the SH, ordered by the super index nm. It should be mentioned that the microphone array properties are already implicit converted into SH. The vector $\mathbf{x}$ represents the sound pressure excitation distribution over the loudspeaker arrangement. This distribution is obtained by a targeted virtual sound field representation $\mathbf{b}$ in SH via the decoding process expressed in Eq. 2.

$$\mathbf{x} = \mathbf{D}_0 \cdot \mathbf{b} \qquad (2)$$

Combining eq.1 and 2, we obtain the transformed representation of the target sound field in SH, which is caused primarily by the room transfer functions.

$$\mathbf{b}_0 = \mathbf{H}_0 \cdot \mathbf{D}_0 \cdot \mathbf{b} \qquad (3)$$

As mentioned above the matrix $\mathbf{H}_0$ is the product of a matrix multiplication given in eq.4. This equation states, that we have to consider the acoustical transfer path between the loudspeakers and the microphones captured in matrix $\mathbf{H}_{ML}$ and the encoding matrix $\mathbf{C}_M$ which converts to the SH representation.

$$\mathbf{H}_0 = \mathbf{C}_M \cdot \mathbf{H}_{ML} \qquad (4)$$

Similar to the eq.3 we can calculate the received SH representation in *Room*$_1$ by applying the targeted sound field representation **b.**

$$\mathbf{b}_1 = \mathbf{H}_1 \cdot \mathbf{D}_1 \cdot \mathbf{b} \qquad (5)$$

In order to transfer the acoustic properties, we have to receive the same SH representation at the origin of the rendering system as stated in eq. 6.

$$\mathbf{b}_1 \overset{!}{=} \mathbf{b}_0 \qquad (6)$$

To obtain the expected matching of the two realizations we have to introduce the modification matrix M in eq.5.

$$\mathbf{b}_1 = \mathbf{H}_1 \cdot \mathbf{D}_1 \cdot \mathbf{M} \cdot \mathbf{b} \qquad (7)$$

By simply rewriting eq.6 using eq.7 and 3 and solve this equation to determine M we get:

$$\mathbf{M} \overset{!}{=} (\mathbf{H}_1 \cdot \mathbf{D}_1)^{-1} \cdot \mathbf{H}_0 \cdot \mathbf{D}_0 \qquad (8)$$

For the acoustic transfer we have to apply this matrix either directly on the sound field representation $\mathbf{b}$ or we obtain a frequency dependent decoder.

2.2 Procedures

Based on the presented concept we have to determine the acoustic transfer functions H_l^{nm} within both rooms to establish the matrix $\mathbf{H}_{ML}$ in eq.4. Using e.g. the Soundfield SPS200 microphone in combination with the A to B-Format converter to measure the directed room impulse responses, we will obtain the SH representation up to order N=1. The converter is represented with matrix $\mathbf{C}_M$ in eq.4 and can be realized either in hard- or software. To improve the spatial resolution in the representation i.e. N>1 we will apply the preliminary steps used in the DirAC (cf. [5]) approach. However, to cope with the present short transient events we have to adapt both the analysis and re-synthesis procedures. Transient detection and transient analysis will be introduced. To design an adequate decoder we refer to [11].

Completing the proposed concept we have to mention, that the inversion of the matrix product in eq.8 is a crucial task and should be considered carefully. As stated in the beginning of this section, the spatial equalization has to keep the physical restrictions in mind to prevent non-causal and non-minimal phase realizations. If this step has to be considered we suggest using the SVD approach combined with some sort of regularization.

3 Spatial Analysis and Re-Synthesis

Within this contribution the analysis is based on coincident measurements using the Soundfield SPS200 microphone, placed at the origin of the examined arrangement. It delivers a recording of the spatial impulse responses in first-order Ambisonics B-Format. Anyway, several other arrangements providing the possibility to determine the SH representation at the origin might be adequate.

In both examined rooms, an Ambisonics 3D sound rendering system equipped with 29 and 24 loudspeakers, respectively is installed. The spatial impulse response, from every single loudspeaker to each microphone capsule, is measured using the exponential sweep-sine method. Using the four measured impulse responses for each loudspeaker the SH representation is received by applying a conversion matrix.

In the following the analysis and re-synthesis procedure are visited.

3.1 Analysis of spatial impulse responses

First, short transient events are detected in the SH representation responses for each loudspeaker. Therefore, an onset-analysis (cf. [10]) is applied to the undirected first SH channel. In the onset detection procedure each instant of time, where relevant energy alterations occurs, will be stored in an onset list (OL). These instants relate to the direct sound and observable discrete reflections. We have evaluated several onset approaches and finally decided to use the High-Frequency Content (HFC) approach to detect these events.

At the detected onset instants a special analysis to determine the direction of the energy flow is used. This analysis method is related to measure the instant intensity vector, which can be calculated using the first four SHs (cf. [5]).

$$\mathbf{I}(n) = [I_x(n), I_y(n), I_z(n)]^T = p(n) \cdot \mathbf{u}(n) \quad (9)$$

whereby $p(n) = b_1(n)$, and $\mathbf{u}(n) = [b_2(n), b_3(n), b_4(n)]^T$

To draw attention to physical relations e.g. low and high frequency components of transient events evolve different in time, and neighboring onsets can overlap in time, the calculation of intensity vectors is split up into several frequency bands.

The analysis procedure captures the following steps:

- General band limitation from 50 to 5000Hz
- Analyze transient regions starting 1ms before OL (i) until at the maximum 1ms before OL (i+1)
- The maximum overall duration of these regions is restricted to 10ms.
- Evaluated regions are examined in several critical frequency bands.
- The observed region is adjusted to n-times periods of the examined center frequency.
- Windowing function reduces artifacts at the beginning and at the end of the selected region.
- Sub-regions within the selected regions are defined that meet an expected energy criteria i.e. a threshold is defined by the RMS of the undirected SH component within this region.
- Intensity vectors for the remaining reduced regions according to eq.9 are calculated.
- Determine the θ (azimuth) und φ (elevation) directions based on the intensity vector (eq.10 and eq.11).

$$\theta(n, f_c) = -\arctan\left(\frac{I_y(n, f_c)}{I_x(n, f_c)}\right) \quad (10)$$

$$\varphi(n, f_c) = -\arctan\left(\frac{I_z(n, f_c)}{\sqrt{I_x^2(n, f_c) + I_y^2(n, f_c)}}\right) \quad (11)$$

- Calculate the median values of θ and φ using circular statistics within one frequency band.
- Based on the energy relations between frequency bands the above obtain directions are weighted according to trust and averaged again.

Additionally, a diffuseness parameter (see eq.12) is required to describe not localized regions.

$$\psi(n, f_c) = 1 - \frac{\left\| \langle \mathbf{I}(n, f_c) \rangle \right\|}{c \langle E(n, f_c) \rangle} \quad (12)$$

where $\| \;\|$ denotes the norm of the vector and $\langle \;\rangle$ denotes time averaging and E is the overall Energy in the SH representation.

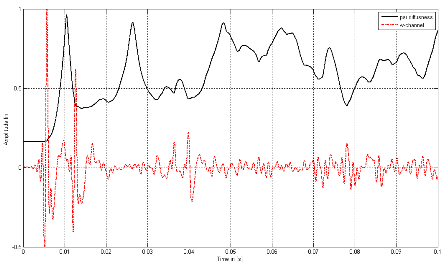

Figure 2: Diffuseness (black solid line) vs. evolving impulse response (red dashed line) in the band filtered undirected SH channel (W-channel of the SH representation of loudspeaker nr1, MUMUTH).

The tail of the spatial impulse responses, where no further discrete reflections can be distinguished we apply is simply functionally modeled. Therefore, we use frequency dependent exponential decaying band limited noise bursts.

3.2 Re-synthesis of higher spatial resolution

In the re-synthesis stage we improve the spatial resolution of the measured impulse responses. With the aid of the above analyzed parameters θ and φ at the related instants in the OL and the given undirected SH signal, we can determine any arbitrary SH order signal representation.

At periods where no transients occur we can observe that a certain degree of diffuseness appear (cf. Fig. 2). Within these crossover regions the spatial resolution is decreased according to the diffuseness parameter ψ (see eq.12). We suggest feeding more energy towards lower spatial resolution. Therefore, higher orders of the SH representation fade out earlier, whereby, their energy is recovered towards the lower order channels to retain the overall energy.

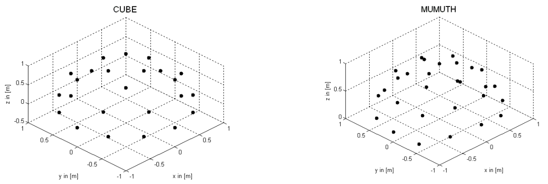

Figure 3: Real loudspeaker layout projected on a unit sphere with radius 1 (left: CUBE and right: MUMUTH).

4 Preliminary Results

Table 1 provides some room-acoustics facts of the examined rooms. Results of the determined positions according to presented analysis procedure are summarized in table 2.

	CUBE	MUMUTH
Volume, size	10x11x5 [m^3]	32x17x11[m^3]
RT$_{60}$ (overall)	0,7 [s]	1,6 [s]
# loudspeakers	24	29
critical distance	3,5 [m]	10 [m]

Table 1: Parameters related to room acoustics.

	Elevation				Azimuth			
			abs dev.				abs dev.	
	μ	std	μ	std	μ	std	μ	std
MUMUTH	***0,7***	***5,2***	***4,8***	***4,5***	***7,3***	***8,3***	***5,8***	***5,6***
rig1@ 8°	-0,7	5,9	6,6	5,5	8	4,9	3,4	3,3
rig2@ 32-36°	1	2,7	2,1	1,6	7,7	7	5,8	3,3
rig3@ 59-63°	2,4	7,2	5,8	3,9	5,5	14,7	10,5	9,3
CUBE	***-0,9***	***4,1***	***3,1***	***2,5***	***-2,6***	***8***	***6,3***	***4,8***
rig1@ 2°	-1,6	2,8	2,2	1,8	-2,2	4,2	3,5	2,1
rig2@ 30°	-0,4	3,3	2,8	1,5	-2,6	8,2	6,9	3,6
rig3@ 58°	0,2	8,3	6,7	3,2	-3,3	16,4	13,6	4,9

Table 2: Deviation and absolute deviation comparing measured and real loudspeaker positions in the two rooms are shown. Rigs are specified by elevation.

In table 2, a bias related to the azimuth adjustment of the microphone can be identified in both cases. The differentiated examination of the rigs shows an increased deviation in determined elevation with increasing height of the rigs. Furthermore, the deviation dependencies are similar in both rooms. Compared to the realizable spatial resolution (cf. [9]) with these two rendering systems (CUBE up to fifth and MUMUTH up to sixth Ambisonics order) the obtained results are basic lower. Anyhow, the retrieved accuracy should be improved.

5 Conclusion

We have presented a new concept to transfer the acoustical characteristics of rooms. Preliminary investigations of our measurements using a Soundfield microphone show first practicable results. Imprecision can be addicted to the microphone arrangement itself (cf. [9]), to interfering grinding reflection related to the lower rigs and positioning dependent reflections from the close wall. In future, we will examine the Soundfield microphone concerning direction dependent accuracy, and will use the technique presented in [12] to improve our measurements.

References

[1] L. L. Beranek, Concert hall acoustics, In J. Acoustic Soc. A m., 92.1-39, 1992.

[2] D. Griesinger, General Overview of Spatial Impression, Envelopment, Localization, and Externalization, *In Proceedings of the Audio Eng. Soc.*, 15th Int. AES Conference, Copenhagen, 1998.

[3] G.A. Souloudre et al., Objective Measures of Listener Envelopment in Multichannel Surround Systems. *In Journal of the Audio Eng. Soc*, 1995.

[4] Franz Zotter, et al., An Alternative Ambisonics Formulation: Modal Source Strength Matching and the Effect of Spatial Aliasing, *In AES 126th Convention*, Munich, Germany, May, 2009.

[5] V. Pulkki, Ch. Faller, Directional Audio Coding: Filterbank and STFT-based Design, *In AES 120th Convention*, Paris, France, May, 2006

[6] M. Kallinger, et al., Analysis and adjustment of planar microphone arrays for application in Directional Audio Coding, *In AES 124th Convention*, Amsterdam, The Netherlands, 2008.

[7] J. Vilkamo, et al., Directional Audio Coding: Virtual Microphone-Based Synthesis and Subjective Evaluation. *In J. Audio Eng. Soc.,* Vol. 57, No. 9, September, 2009.

[8] G. Del Galdo et al., Optimized Parameter Estimation in Directional Audio Coding Using Nested Microphone Arrays, In *Proceedings of the 127th AES Convention*, New York, USA, October 9–12, 2009.

[9] F. Zotter, Analysis and Synthesis of Sound-Radiation with Spherical Arrays", *Doctoral Thesis*, Institute of Electronic Music and Acoustics, University of Music and Performing Arts Graz, 2009.

[10] J.P. Bello et al, A Tutorial on Onset Detection in Music Signals, *In IEEE Transactions on speech and audio processing*, 2005.

[11] F. Zotter et al., The Virtual T-Design Ambisonics-Rig Using VBAP. *In Proceedings of the 1st EAA-EuroRegio*, Ljubljana, Slovenia, September, 2010.

[12] K. Freiberger, A. Sontacchi, Sound Source Localization with a coincident microphone array utilizing similarities, *submitted to the 129 AES Convention*, San Francisco, USA, November, 2010.