Array-based HRTF pattern emulation for auralization of 3D outdoor sound environments with direction based muffling of sources

Pieter Thomas, Timothy Van Renterghem and Dick Botteldooren

Department of Information Technology (INTEC), Ghent University,
St.-Pietersnieuwstraat 41, B-9000 Gent, Belgium
pieter.thomas@intec.ugent.be

Abstract

The use of spatial audio reproduction techniques is widely employed for the subjective analysis of concert halls and, more recently, complex outdoor sound environments. In this work, a binaural reproduction technique is developed based on a 32-channel spherical microphone array, optimized for the simulation of a virtual microphone with directional characteristics that approximate the directivity of the human head. A set of weights is calculated for each microphone of the constituting array based on a regularized least-square solution. This technique allows for adaptation of the auditory scene based on source direction. The performance of the technique has been evaluated by means of listening tests, and its use for the auralization of outdoor soundscapes has been illustrated.

1 Introduction

Different techniques exist to capture a given soundfield for binaural reproduction, each with different degrees of complexity. The most straightforward method is to record a soundfield with a Head-and-Torso Simulator (HATS). Although its simplicity is attractive, there are some inherent drawbacks. Firstly, the HATS is constructed based on average human dimensions, meaning that individual physical properties of the listener's ears, head and torso are not taken into account. Furthermore, the position of the HATS is fixed and determines the orientation of the listener in the reproduction phase.

To allow rotation of the listener's head, Algazi et al. [1] developed the Motion-Tracked Binaural sound (MTB) technique. Basically, the soundfield is sampled by multiple microphones, located on a sphere with dimensions comparable to a human head (a first order spherical-head model). A binaural signal is extracted from the microphones closest to the listener's ears.

In this paper, a more complex technique is proposed based on the use of a microphone array to emulate the directivity pattern of the human listener (section 2). In this way, individual physiological characteristics can be taken into account, while head rotation can be implemented by updating the microphone weights. The proposed reproduction technique is validated by means of listening tests in section 3. Furthermore, it will be illustrated how the pattern emulation technique can be used to auralize the effect of an environmental noise reduction measure (see [2] for a more extensive study). In contrast to HATS and MTB based recordings, the effect of such measures can be easily included by spatial selective muffling. This technique is of specific interest for urban planners, as it provides a tool to simulate the auditory effect of their designs [2].

2 Methodology for directivity pattern emulation

2.1 Microphone array design

The microphone array is specifically designed for HRTF pattern emulation and consists of a hard plastic sphere with a 7-cm radius as a substitute for the human head, facilitating the calculation of the microphone filters needed to approximate the HRTFs, in particular at high frequencies. 32 omnidirectional microphones (Knowles Acoustics type FG-23329) are distributed on the sphere allowing focus on azimuthal rather than elevational pattern reconstruction, corresponding to the larger sensitivity of humans regarding azimuthal source discriminating resolution. Apart from the two poles, the remaining 30 microphones are uniformly distributed along three circles at elevations of $-45°$, $0°$ and $45°$ (Fig. 1). For the data-acquisition (DAQ), a National Instruments PXIe-1082 chassis with three NI-4498 DAQ cards, counting 16 channels each, is used.

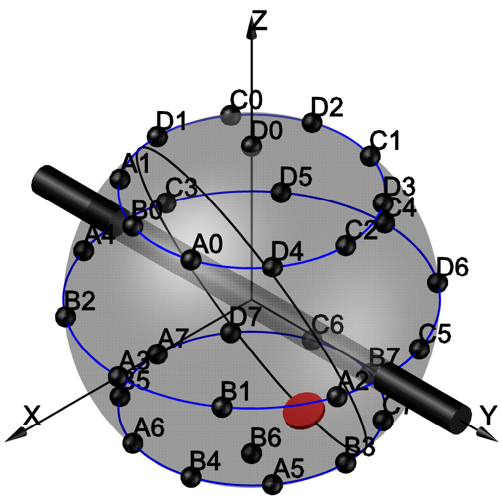

Figure 1: Spherical microphone array geometry. The black bar is part of the array fixation. The red circle indicates the outlet of microphone cables.

2.2 Calculation of microphone array weights

In a first step, the target directivity pattern has to be defined. The set of HRTFs used here is taken from measurements by Gardner employing a HATS (KEMAR model DB-4004) at 710 different orientations (ϕ_l, θ_l) [3]. Compensation filters have been designed to account for the frequency response of the measurement equipment, ear canal and reproduction headphones used in the listening test (Sennheiser HD-280 pro).

When applied to auralization, direction-based sound-source muffling is introduced by attenuating the impulse responses $\mathrm{HRIR}(t, \phi_l, \theta_l)$ with the direction-dependent insertion loss $\mathrm{IL}(\phi_l, \theta_l)$.

The set of attenuated HRTFs is only available for 710 discrete directions and elevations higher than $-40°$ (polar gap). Therefore, an inter/extrapolation strategy based on a spherical harmonic decomposition of the HRTF dataset has been developed to estimate the target directivity pattern at intermediate orientations [4].

In a second step, the directivity patterns of each individual microphone on the sphere have been measured with a dedicated setup at 614 discrete directions. In this way, deviations in microphone placement, level and phase differences between microphones, influences of windshield, cable outlet and fixation mechanisms are accounted for. To provide a continuous representation and to smooth out pattern irregularities, a spherical harmonic decomposition has been conducted on the magnitude and spectrally unwrapped phase. For frequencies lower than 200Hz, only the spherical harmonic coefficients until $N = 3$ are taken into account to eliminate fast varying spatial fluctuations from loudspeaker instabilities and phase noise.

In the final step, the appropriate filter coefficients $W_i(f)$ for each microphone signal can be estimated so that the resulting array directivity pattern approaches the target (muffled) HRTF directivity pattern [2]:

$$\mathrm{HRTF}(f, \phi, \theta) = \sum_{i=1}^{32} W_i(f)\mathrm{MIC}_i(f, \phi, \theta), \quad (1)$$

with $\mathrm{HRTF}(f, \phi, \theta)$ and $\mathrm{MIC}_i(f, \phi, \theta)$ the complex-valued directivity pattern of the (muffled) HRTF, resp. i-th array microphone. The microphone weights $W_i(f)$ are calculated as a regularized least-square solution of the discretized version of Eq. 1:

$$\mathbf{W}(f) = \left[\mathbf{MIC}^H\mathbf{MIC} + \mu\mathbb{I}_{32}\right]^{-1} \mathbf{MIC}^H\,\mathbf{HRTF}, \quad (2)$$

with **HRTF** and **MIC** the smoothed directivity patterns evaluated at each direction (ϕ_m, θ_m) of an M-point uniformly sampled grid. **HRTF** is of size $M \times 1$ and **MIC** is of size $M \times 32$, with $M = 614$ points. A regularization of $\mu = 0.01$ is included to avoid over-fitting and decrease the sensitivity of the fit to sensor noise and errors in e.g. microphone characteristics.

The pattern emulation technique typically shows good results below a certain cross-over frequency f_X. At higher frequencies, the HRTF pattern fitting performance decreases due to spatial aliasing and inaccuracies in the target and measured directivity patterns.

In order to provide the listener with a full-spectrum binaural signal, the pattern emulation technique is replaced with the MTB technique at high frequencies, by taking the signal from the microphone closest to the location of the ear. Basic sound source muffling is then included by attenuating the left resp. right ear signal with the overall IL of the complete left resp. right hemisphere.

3 Validation by listening tests

The quality of the pattern emulation technique is evaluated by means of listening tests. As the microphone array is mainly designed for binaural reconstruction of outdoor sound environments, and more specific for the auralization of traffic noise reduction measures, a location near the E17 highway in Belgium is chosen where a mound was going to be installed [2].
In total, 14 female listeners and 23 male listeners took part in the test, with ages ranging between 21 and 57. Several samples were selected from binaural recordings, made with a HATS (B&K Type 4128C), and binaurally converted array recordings. A static set of array coefficients is used here, related to a head orientation with interaural axis parallel with the highway. For frequencies higher than f_X, only the high-pass filtered signal from microphones A2 (left ear) and A4 (right ear) is kept.

In a first part of the listening test, the quality of the binaural reproduction methodology is evaluated. Firstly, listeners were asked to categorize samples as true binaural recordings or reproductions, see Table 1. The rightmost column indicates the probability that the null-hypothesis, assuming random choice, is accepted, resulting from a two-sided binomial test.
HATS recordings are clearly identified as true binaural recordings, which proves that the listeners indeed have a good understanding of what is meant by a 'real' recording. For array reproductions, including the reproduction solely based on microphone A2-A4 ('L-R array rec.'), listeners could not identify whether or not array extracts were true recordings or reproductions.

Secondly, listeners were asked to rank five reproductions with different f_X according to their similarity with the original HATS recording. To determine the mean ranking and check whether or not mean rankings are significantly different, a multiple comparison test is performed by applying Tukey's least significant difference procedure on the outcome of an ANOVA test (Fig. 2). Clearly, the samples with $f_X = 4490$Hz are ranked worst. The L-R reproduction and reproduction with $f_X = 1122$Hz have the best ranking, with a significant difference from samples with $f_X = 2245$Hz and 4490Hz, but not from samples with $f_X = 1782$Hz.

In a second part of the listening test, the ability of the proposed methodology to auralize the effect of outdoor noise reduction measures has been tested by considering the effect of the construction of a 6m-high complex L-shaped mound (84.5m x 65.3m). Auralized samples are extracted from a priori recordings by using array weights corresponding to the HRTF directivity patterns muffled with the direction dependent insertion loss of the mound, simulated with the ISO 9613-2 model, which is a widely used engineering model to predict attenuation of sound outdoors.

Table 1: Test 1: responses given by the listeners, presented with samples recorded with the HATS or binaurally synthesized from array recordings. ($\chi^2 = 40.13$, $p < 0.01$)

	Indicated recording	Indicated reproduc.	p-value
HATS rec.	130	55	$< 0.01^{**}$
L-R array rec.	19	18	1
Array, $f_{1122\text{Hz}}$	15	22	0.324
Array, $f_{1782\text{Hz}}$	16	21	0.511

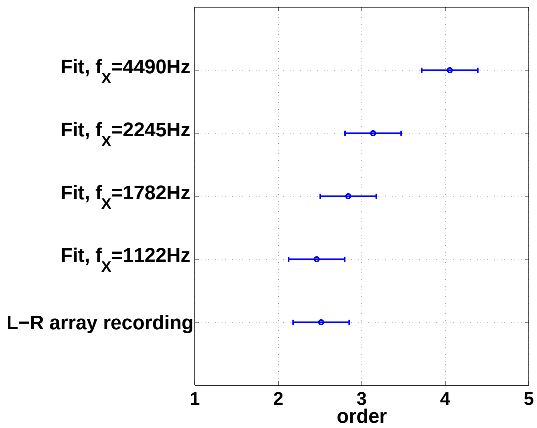

Figure 2: Test 2: results of the ranking for different f_X. Mean ranking and 95% confidence intervals are indicated (1 means most similar).

Table 2: Test 3: responses given by the listeners, presented with auralized samples and samples recorded before or after installation of the mound. (χ^2 = 303.81, $p < 0.01$)

	Indicated before	Indicated after	p-value
HATS rec. bef.	74	0	$< 0.01^{**}$
Array rec. bef.	73	1	$< 0.01^{**}$
HATS rec. aft.	5	32	$< 0.01^{**}$
Array rec. aft.	1	36	$< 0.01^{**}$
Array sim. ISO	0	37	$< 0.01^{**}$

Table 3: Test 4: responses given by the listeners, presented with auralized samples and samples recorded after installation of the mound. (χ^2 = 17.31, $p < 0.01$)

	Indicated recording	Indicated simulat.	p-value
HATS rec. aft.	84	64	0.118
Array sim. ISO	13	24	0.099
L-R sim. ISO	10	27	$< 0.01^{**}$

Firstly, listeners are asked to indicate whether or not the fragments were recorded before or after construction of the mound to check if the mound has an effect on the soundfield (Table 2). HATS recordings, samples from array recordings with and without the mound and auralized samples are included. Results of this test clearly show that listeners can distinguish between fragments recorded before and after construction, whether or not fragments have been binaurally synthesized from an array recording or recorded with a HATS. Furthermore, almost all auralizations made from array recordings are categorized as recorded after construction.

In the final test, the listener was asked to categorize samples as a true recording or a simulation (Table 3). Samples included true HATS recordings made after installation of the mound and auralizations based on the pattern emulation technique and the MTB technique with hemispherical IL model for the full frequency range ('L-R sim.'). Results show that listeners have great difficulty to distinguish whether samples have been auralized or recorded after installation of the mound: 43% of the a posteriori HATS recordings have been categorized as simulations. For auralizations based on the pattern emulation technique the assumption of random choice remains valid (p-values ≈ 0.1), while for samples auralized using the full range MTB technique this assumption can be rejected, as almost 73% of the samples were indicated as simulated. This is to be explained by the lack of directional resolution of the source muffling.

4 Conclusion

A binaural reproduction technique has been proposed based on a 32-channel spherical microphone array. The technique is based on a combination of the HRTF directivity pattern emulation technique for low and mid frequencies and the MTB technique at high frequencies. Based on listening tests, a cross-over frequency of 1782Hz is proposed. When used for binaural reproduction, the pattern emulation technique delivers similar performance as the full range MTB technique. However, the pattern emulation technique shows its value when auralizing the effect of complex-shaped noise reduction measures, as it allows to include directional dependent source muffling. The necessity of this is illustrated by the listening tests: while listeners could not categorize HATS recordings or auralizations as recordings or reproductions, a significant majority was still able to identify auralizations based on the full range MTB technique.

References

[1] Algazi, R., Duda, R., and Thompson, D., "Motion-tracked binaural sound," J. Audio Eng. Soc., 52, pp. 1142–1156, 2004.

[2] Thomas, P., Tools for urban sound quality assessment, Ph.D. thesis, Ghent University, 2014.

[3] Gardner, W. G. and Martin, K. D., "HRTF measurements of a KEMAR," J. Acoust. Soc. Am., 97, p. 3907, 1995.

[4] Ahrens, J., Thomas, M., and Tashev, I., "HRTF magnitude modeling using a non-regularized least-squares fit of spherical harmonics coefficients on incomplete data," in Proceedings of APSIPA, 2012.