A comparative study of spherical microphone arrays based on subjective assessment of recordings reproduced over different audio systems

Simeon Delikaris-Manias, Vincent Koehl, Mathieu Paquier, Rozenn Nicol, Jérôme Daniel

Forum Acusticum 2011 - Aalborg - Denmark

A comparative study of spherical microphone arrays based on subjective assessment of recordings reproduced over different audio systems

Simeon Delikaris-Manias, Vincent Koehl, Mathieu Paquier
European Center of Virtual Reality, UEB - UBO - ENIB - LISyC EA 3883, 29200 Brest, France,
delikaris@enib.fr
{vincent.koehl, mathieu.paquier}@univ-brest.fr.

Rozenn Nicol, Jerome Daniel
Orange Labs, TECH/OPERA, 22307 Lannion, France,
{rozenn.nicol,jerome.daniel}@orange-ftgroup.com.

Summary
This study presents an evaluation of sound recordings acquired with two rigid spherical microphone arrays of the exact same size fitted with different type of capsules. These two sets of capsules were of small and large membrane sizes respectively. Objective evaluation has been previously performed by the authors by analyzing the reconstructed spherical harmonic components. Listening tests are conducted to evaluate the perceived quality of the signals acquired with the two arrays and indicate the possible degradation due to the signal processing techniques used to calculate the spherical harmonic components. The subjective evaluation has indicated a slight but significant preference on the array with the large membrane sensors.

PACS no. 43.60.+d, 43.90.+v

1. Introduction

Three dimensional audio aims at reconstructing a sound field and conveying the listener to a virtual sound space. This can be achieved either by using monophonic signals and encoding-decoding them into a three dimensional render or use microphone arrays and then process the acquired multichannel signals. Using microphone arrays requires that the acquired signals are encoded into a specific format which in this case is in the spherical harmonics domain. Producing this type of encoding provides enough information to then perform a decoding to a number of speakers for spatial reproduction.

In this paper a comparison between two spherical microphone arrays using different capsules is presented and analyzed. The aim of this paper is to assess the performance of two spherical microphone arrays with different capsules and to determine whether the choice of the capsule affects the overall perceived quality of the spatial encoding. Previous work has objectively evaluated the differences between the two arrays by characterizing the spherical harmonic components [1]. Differences in the capsules responses were revealed but little difference was observed in the calculated spherical harmonic components. In the present study each spherical array is characterized by evaluating music excerpts with listening tests.

The objective is to investigate whether it is beneficial to use higher quality capsules and whether this benefit can be perceived throughout various reproduction systems. In addition, it is examined to what extent the signal processing techniques, for encoding and decoding the microphone signals, and the reproduction setups affect the intrinsic characteristics of microphones. Three reproduction systems were chosen for the purpose of the listening tests consisting of a monophonic, a stereophonic and a five channel setup . Raw capsule outputs and multichannel decodings obtained from each of the two arrays were fed into these three setups. The comparison between raw and decoded signals aims at investigating if the preference between capsules is preserved in the decodings from arrays made of the same capsules. The different setups were used to determine the influence of the reproduction system over the perceptual assessment.

Previous studies on the assessment of timbre have shown that multichannel reproduction setups are acknowledged to be less discriminant than monophonic ones in loudspeaker measurements and room equalization [2, 3]. The choice of a monophonic reproduction system was based on the fact that previous research on quality of reproduced sound focuses on the timbral aspects of monophonic reproduction [4]. As the multichannel reproduction systems become more popular there are studies that address the problems of identifying and evaluating sound quality [5, 6]. Subjective evaluation of surround sound microphones showed that the choice of the musical pieces affects the preference of the listener which is also expected in this study [7]. Spatial sound attributes are not being investigated as this study focuses on the preference and similarities between recordings. According to the signal to noise ratios and frequency responses of both sets of capsules, it is assumed that the large diaphragm capsules are of a higher quality, meaning a flatter frequency response and higher signal to noise ratio.

2. FOURIER-BESSEL SERIES

Closed spherical microphone arrays consist of microphones distributed uniformly around a rigid sphere. A sound-field can be described in the spherical harmonics domain as a Fourier-Bessel series. This is derived from the Helmholtz equation in spherical coordinates in a source free region. This series is described as:

$$p(kr,\theta,\phi) = \sum_{m=0}^{\infty} i^m j_m(kr) \sum_{n=0}^{m} \sum_{\sigma=\pm 1} A_{mn}^{\sigma}, \quad (1)$$

where $A_{mn}^{\sigma} = B_{mn}^{\sigma} Y_{mn}^{\sigma}(\theta,\phi)$, k is the wavenumber, θ and ϕ are the angular dependencies (azimuthal and elevation respectively) and $j_m(kr)$ the spherical Bessel functions.

The spherical harmonic components are given by:

$$B_{mn}^{\sigma} = \frac{1}{i^m j_m(kr)} \iint_S p(kr,\theta,\phi) Y_{mn}^{\sigma}(\theta,\phi) dS, \quad (2)$$

where $Ymn^{\sigma}(\theta,\phi)$ are the spherical harmonics which represent the two angular dependencies θ and ϕ:

$$Y_{mn}^{\sigma}(\theta,\phi) = \sqrt{C(m,n)(2-\delta_{0,n}) P_{mn}(sin\delta)}, \quad (3)$$

where $C = (2m-1)\frac{(m-n)!}{(m+n)!}$ and P_{mn} are the associated Legendre functions.

When spherical microphone arrays are used for sound acquisition the signals are captured by the capsules and then matrixed and filtered. The spherical harmonic components Bmn^{σ} can be estimated from Equation 1 and 2. In matrix form this can be expressed as:

$$B_{mn}^{\sigma} = J^{-1} Y p_q, \quad (4)$$

where J^{-1} is the radial dependence matrix, Y the spherical harmonics matrix and p_q the signal acquired from each sensor $q = 1 \ldots 8$.

A detailed analysis on the spherical harmonic analysis can be found in [8] and [9]. This study focuses on horizontal spherical microphone arrays fitted with eight omni-directional sensors on the same plane. The spherical harmonic components are calculated according to Equation 4 were p consisted of the eight sensor input signals.

3. EXPERIMENTS AND IMPLEMENTATION

The experimental arrangement and procedure consist in recording a quartet with the two spherical microphone arrays, processing the microphone signals and decoding the signals to single and multichannel reproduction setups.

3.1. Acquisition setup

A performance of a quartet has been recorded and short (4-7 seconds) musical excerpts have been selected. The recording took place in a recording room of the University of Brest as shown in Figure 1. The sound scene consists of four musicians. The instruments from left to right are a flute, a clarinet, a double bass and an oboe. The recording position was chosen as the position where a conductor would be placed.

The two spherical microphone arrays have been placed one on top of the other as shown in Figure 1, the large diaphragm array (L) is the one on the top and the small diaphragm array (S) directly below. As this experiment is performed to determine perceptual differences between recordings from two arrays, it should be remarked that this configuration result to similar diffraction properties. The process of obtaining the spherical harmonics components which are fed into the reproduction setup is described by previous work of the authors [1]. An analysis of the theory and design of high order spherical microphone arrays is presented in [10]. These recordings were used for monophonic and multichannel comparisons.

3.2. Reproduction setup

The reproduction chain consists of five loudspeakers arranged at 0, 30, −30, 120, −120 degrees according to Figure2 and all the loudspeakers were matched in term of loudness. The recorded monophonic signal and the calculated omni-directional component (B_{00}^{+1}) were reproduced from the frontal speaker, two signals from anti-diametric sides of the each sphere were also

Figure 1. Microphone setup for the acquisition of the music excerpts.

Setup	Source positions	Microphone signals
monophonic	0^o	frontal sensor, omni-directional component B_{00}^{+1}
stereophonic	$+30^o$, -30^o	anti diametric sensors, 1^{st} order decoding
five channel	0, $+30^o$, -30^o $+120^o$, -120^o	3^{rd} order decoding, 4^{th} order decoding

Table I. Source positions and the corresponding input signals.

captured and in addition with a 1^{st} order stereophonic decoding calculated from the B_{11}^{+1} and B_{11}^{-1} components were reproduced from the speakers at 30 and -30 degrees. Finally two 5 channel decodings calculated by using all the eight microphone signals from each array, were fed to all the loudspeakers. The various microphone signals and the corresponding reproduction setups and are shown in Table I.

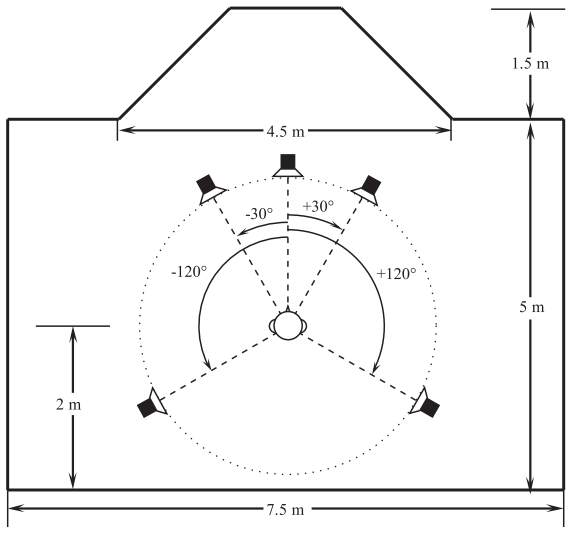

Figure 2. Reproduction setup for the assessment of the musical excerpts.

Frequency (Hz)	125	250	500	1000	2000	4000
RT_{60} (sec)	0.5	0.4	0.6	0.5	0.4	0.4

Table II.

Reverberation time of the room where the recordings and listening tests took place was also measured. Results are shown in Table II.

3.3. Perceptual Evaluation

Thirteen subjects were positioned at the center of the setup facing the frontal loudspeaker at 0 degrees as shown in Figure 2. Two different types of tests were performed: a preference test aiming at evaluating an overall preference between recordings performed with the two arrays and a similarity test. Both tests were conducted based on a MATLAB graphical user interface, controlled through a notebook with a touch screen. During the test the subject could listen to four different excerpts recorded with the two different microphone array systems and compare them in terms of similarity and preference. The reproduction of the excerpts could be repeated. When the choice of the subject was determined and reviewed, he could choose the appropriate answer by moving a slider to the appropriate position and then clicking respond. Concerning the similarity test the slider ranged from 0 to 1 where 0 means identical and 1 completely different. In the preference test the slider ranged from -1 to 1 where -1 meant that the sound A was preferred and 1the sound B. Additional adjective labels were added to the two sliders: for the preference test (Sound A strongly preferred, Sound A preferred, No preference, Sound B preferred, Sound B strongly preferred) and for the similarity test (Identical, Slightly Different, Different, Very Different, Extremely Different). A training program consisting of five random trials from the preference and similarity test was used to familiarize the listener to the procedure. Both the tests and the trials of each test were randomized.

4. RESULTS

The mean ratings for the preference and similarity tests are shown in Figure 3. A t-test showed that the average preference was higher than 0 ($t(311) = 1$; $p = 0.0034$), which indicates that the large membrane array was significantly preferred (Figure 3). The mean similarity lies between Slightly Different and Different and is significant different than zero. That indicates a clearly perceived difference between the two arrays by the subjects. A 2-way analysis of variance is carried out to study the effects of the restitution system, the excerpt and their possible interaction over the preference (Table III) and similarity ratings (Table IV).

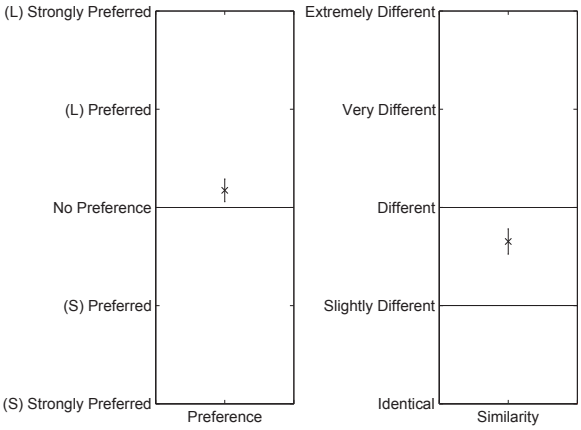

Figure 3. Overall mean values for preference and similarity within their 95% confidence interval.

Source	SS	dof	MS	F	p
System	0.9319	5	1.18638	0.75	0.5976
Excerpt	1.3676	3	0.45586	1.84	0.1834
S*E	3.7188	15	0.24792	0.92	0.5418

Table III. ANOVA preference test results.

Source	SS	dof	MS	F	p
System	0.5635	5	0.1127	0.77	0.584
Excerpt	0.5078	3	0.16927	1/16	0.3574
S*E	2.1876	15	0.14584	1.87	0.0255*

Table IV. ANOVA similarity test results.

According to the ANOVA results there is no effect of the system or excerpt because the mean values are equal. The only significant effect was found for the System*Excerpt (S*E) interaction on similarity ratings, which indicates that some combinations of System and Excerpt were more discriminative than others. In contrary to previous studies on loudspeakers or room response compensation [3], the stereo and multichannel system are as discriminative as in the monophonic system. Independently of the system or excerpt, the large membrane array was slightly but significantly preferred. Although that limited preference, the benefits of the higher quality capsules are preserved for all the reproduction system under study.

5. CONCLUSION

In this paper we present a comparison between two identical spherical arrays of microphones fitted with different capsules.Previous study [1] revealed that the objective comparison of the two arrays indicated little difference in the calculated spherical harmonic components. A subjective evaluation is performed by using listening tests which measure the perceptive differences between recordings acquired with the two microphones arrays over different loudspeaker setups. The differences between the two restitution systems were perceived but only a slight but significant preference has been identified for the large membrane sensor array. Future work will investigate whether these differences and the overall preference for the assumed higher quality microphone array will be enhanced with expert listeners.

Acknowledgement

The authors wish to thank Schoeps Mikrofone GmbH and especially Dr Helmut Wittek for the loan of the large membrane microphones and the musicians (“les quatre vents”) for participating in the recordings..

References

[1] S. Delikaris-Manias and V. Koeh. and M. Paquier and R. Nicol and J. Daniel: Does Capsule quality matter? A comparison study between spherical microphone arrays using different types of omnidirectional capsules, Second Ambisonic Symposium, 2010.

[2] F. E. Toole: Subjective Measurements of Loudspeakers: A Comparison of Stereo and Mono Listening, Audio Engineering Society Convention 74, 1983

[3] A. Devantier and H. Sean and O. Sean: Comparison of Loudspeaker-Room Equalization Preferences for Multichannel, Stereo, and Mono Reproductions: Are Listeners More Discriminating in Mono?, Audio Engineering Society Convention 124, 2008

[4] A. Gabrielsson and Håkan Sjögren: Perceived sound quality of sound-reproducing systems, Journal of the Acoustical Society of America, Vol. 65, No. 4, 1979

[5] C. Guastavino and B. F. G. Katz: Perceptual evaluation of multi-dimensional spatial audio reproduction, The Journal of the Acoustical Society of America, Vol. 116, No. 2, 2004

[6] N. Zacharov and K. Koivuniemi: Unravelling the Perception of Spatial Sound Reproduction: Analysis and External Preference Mapping, Audio Engineering Society Convention 111, 2001.

[7] A. Marui and T. Kamekawa and H. Irimaji: Subjective Impression of Surround Sound Microphone Arrays,Audio Engineering Society Surround Experiments Project Report, 2006.

[8] T.D Abhayapala: Generalized framework for spherical microphone arrays: Spatial and frequency decomposition, Proc. ICASSP 2008. IEEE International Conference on Acoustics, Speech and Signal Processing, 2008.

[9] S. Bertet and J. Daniel and S. Moreau:3D Sound Field Recording with Higher Order Ambisonics - Objective Measurements and Validation of Spherical Microphone, Audio Engineering Society Convention 120, 2006.

[10] T.D Abhayapala: Generalized framework for spherical microphone arrays: Spatial and frequency decomposition, Proc. ICASSP ’02. IEEE International Conference on Acoustics, Speech, and Signal Processing, 2002.