Spherical Array Systems - On the Effect of Measurement Errors in Terms of Perceived Auralization Quality

F. Melchior [1,3], Z. Kuang[2,4], D. de Vries [3], S. Brix[4]
[1] *IOSONO GmbH, Germany, Email: frank.melchior@iosono-sound.com*
[2] *Institute of Acoustics, Chinese Academy of Sciences, China*
[3] *Delft University of Technology, The Netherlands*
[4] *Fraunhofer IDMT, Germany*

Introduction

The properties of diffuse acoustic fields can be analyzed using the spatial correlation coefficients as given for single frequency diffuse fields by Cook et al. in [3]. Jacobson has extended the analysis to velocity as well as to a broadband case using the spatial coherence function in [5]. Rafaely verified and extended the results to broadband diffuse fields by simulations in [6]. Elko analyzed the spatial correlation coefficients for first-order microphone setups in [2]. Their work deliver an analysis framework to study the properties of diffuse acoustic fields in simulations and measurements in terms of objective parameters. It also delivers the analytic reference description used in this paper. Array measurements are a flexible way to simulate various virtual sensor configurations based on a single measurement. For auralization applications, a room can be measured once with a high spatio-temporal resolution. Afterward, the microphone configuration used to get the desired impulse responses can be flexibly chosen. In this work, the properties of reproducing these virtual two-channel sensor configurations in a diffuse field are investigated. Based on diffuse field simulations of ideal and non-ideal spherical microphone arrays, virtual sensor configurations are generated by wave field extrapolation. As an alternative approach, a frequency dependent decorrelation technique is applied to generate a new impulse response based on a single measurement. The properties of these three configurations are analyzed in terms of perceptual quality by a listening experiment.

Signal processing

The theory of spherical harmonic decomposition (SHD), wave field extrapolation (WFE), and spatial coherence is briefly reviewed in the following. These delivers the signal processing used to calculated the different items of the listening experiment.

Spherical Harmonic Decomposition

Let $S(r, \phi, \theta, k)$ be a function measured with a spherical microphone array, whereas r is the radius, ϕ is the azimuth, θ is the co-elevation, and k is the wavenumber. The spherical harmonic coefficients $A_{nm}(k)$ of order n and mode m with $-n \leq m \leq n$ can be determined using [10]:

$$A_{nm}(k) = \frac{1}{b_n(kr)} \int_0^{2\pi} \int_0^{\pi} S(r, \phi, \theta, k)\, \overline{Y_n^m(\phi, \theta)}\, d\phi \sin\theta\, d\theta, \quad (1)$$

where $\overline{Y_n^m(\cdot)}$ is the complex conjugate spherical harmonic function. Assuming an open sphere array with cardioid microphones, the mode strength $b_n(kr)$ is given by [10]

$$b_n(kr) = \frac{1}{2} \cdot j_n(kr) - i \cdot \frac{1}{2} \cdot j_n'(kr), \quad (2)$$

where $j_n(\cdot)$ is the spherical Bessel function of the first kind, $j_n'(\cdot)$ is the derivation with respect to the argument, and $i = \sqrt{-1}$. In real world applications, the sound field $S(r, \phi, \theta, k)$ must be sampled at discrete positions which turns the integral in Eq. (1) into a sum over Q grid points

$$A_{nm}(k) = \frac{1}{b_n(kr)} \sum_{q=1}^{Q} w_q\, S_q(r, \phi, \theta, k)\, \overline{Y_n^m(\phi_q, \theta_q)}, \quad (3)$$

where w_q are the quadrature weights. Spatial sampling requires a limited bandwidth to avoid spatial aliasing, i. e., the coefficients $A_{nm}(k)$ must become zero above a certain maximum order N_{field} before sampling the sphere. This is not the case for sound fields composed of plane waves and the reader is referred to [8] for studies on the effects of sampling order-unlimited fields. For single plane wave sound fields, A_{nm} can be computed theoretically using [10]:

$$A_{nm} = 4\pi\, i^n\, \overline{Y_n^m(\phi_0, \theta_0)}, \quad (4)$$

where (ϕ_0, θ_0) is the direction of arrival (DOA) of the wave.

Wave Field Extrapolation

The results of the SHD can be used to extrapolate the measured sound field for virtual sensors with arbitrary first-order characteristics. Since the coefficients $A_{nm}(k)$ are independent from any position (r, ϕ, θ), they can be used to compute the sound pressure $p(R, \varphi, \vartheta, k)$ in an arbitrary point (R, φ, ϑ), [10]

$$p(R, \varphi, \vartheta, k) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} A_{nm}(k)\, j_n(kR)\, Y_n^m(\varphi, \vartheta). \quad (5)$$

A straight-forward approach to determine the sound velocity $v_\Omega(R,\varphi,\vartheta,k)$ in direction $\Omega = (\Phi,\Theta)$ can be derived using the Fourier transform of the Euler's equation, given by

$$i\,k\,c\,\rho_0\ \mathbf{v}(R,\varphi,\vartheta,k) = \nabla p(R,\varphi,\vartheta,k), \tag{6}$$

where $\mathbf{v}(\cdot)$ is the sound velocity vector, c is the sound velocity, and ρ_0 is the density of air. The pressure gradient $\nabla p(\cdot)$ is defined in spherical coordinates as [10]

$$\nabla p \equiv \frac{\partial p}{\partial R}\mathbf{e}_R + \frac{1}{R\sin\vartheta}\cdot\frac{\partial p}{\partial \varphi}\mathbf{e}_\varphi + \frac{1}{R}\cdot\frac{\partial p}{\partial \vartheta}\mathbf{e}_\vartheta, \tag{7}$$

where $\mathbf{e}_{(\cdot)}$ are the basis unit vectors and the dependencies of $p(R,\varphi,\vartheta,k)$ have been omitted. The radial derivative of the pressure $p(R,\varphi,\vartheta)$ in Eq. (5) is given by

$$\frac{\partial p}{\partial R} = k\sum_{n=0}^{\infty}\sum_{m=-n}^{n} A_{nm}(k)\, j_n'(kR)\, Y_n^m(\varphi,\vartheta). \tag{8}$$

The derivative of Eq. (5) with respect to the azimuth φ can be computed using

$$\frac{\partial p}{\partial \varphi} = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} A_{nm}(k)\, j_n(kR)\,\frac{\partial}{\partial \varphi}\, Y_n^m(\varphi,\vartheta) \tag{9}$$

with [9]

$$\frac{\partial}{\partial \varphi}\, Y_n^m(\varphi,\vartheta) = i\,m\, Y_n^m(\varphi,\vartheta). \tag{10}$$

The derivative of Eq. (5) with respect to the co-elevation ϑ can be obtained by

$$\frac{\partial p}{\partial \vartheta} = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} A_{nm}(k)\, j_n(kR)\,\frac{\partial}{\partial \vartheta}\, Y_n^m(\varphi,\vartheta) \tag{11}$$

with [9]

$$\frac{\partial}{\partial \vartheta}\, Y_n^m(\varphi,\vartheta) = \frac{1}{2}\cdot K_1 \cdot Y_n^{m+1}(\varphi,\vartheta)\cdot e^{-i\varphi} - \frac{1}{2}\cdot K_2 \cdot Y_n^{m-1}(\varphi,\vartheta)\cdot e^{i\varphi} \tag{12}$$

and $K_{1/2} = \sqrt{n(n+1) - m(m\pm 1)}$. The sound velocity $v_\Omega(R,\varphi,\vartheta,k)$ is found by computing the directional derivative of the sound velocity vector $\mathbf{v}(\cdot)$ in Eq. (6) with respect to the direction Ω, i. e.,

$$v_\Omega(R,\varphi,\vartheta,k) = \frac{1}{i\,k\,c\,\rho_0}\nabla p \circ \begin{pmatrix} \sin\Theta\cos\Phi \\ \sin\Theta\sin\Phi \\ \cos\Theta \end{pmatrix}. \tag{13}$$

Superposing (5) and (13) yields the response $S_\Omega(R,\vartheta,\varphi,k)$ of a virtual sensor with steering direction Ω,

$$S_\Omega(R,\varphi,\vartheta,k) = \beta\cdot p + (1-\beta)\,\rho_0\,c\cdot v_\Omega, \tag{14}$$

where the first-order sensor parameter is $\beta = 0.5$ in case of cardioid sensors. In practice, the first sum in Eq. (5) must be limited to a specific maximum order N. A large N allows a more accurate WFE over larger areas but can lower the array robustness against microphone noise and positioning errors in Eq. (3), especially at low frequencies and on small array radii. Rafaely pointed out in [7] that microphone arrays provide the highest robustness if the maximum order is set to $N \approx kr$, whereas r is the array radius. It can be shown that in this case, the WFE can be computed accurately within the area enclosed by the microphone array.

Spatial Coherence

The coherence $\gamma_{xy}^2(k)$ between two arbitrary signals $x(t)$ and $y(t)$ is defined as [1]

$$\gamma_{xy}^2(k) \equiv \frac{|S_{xy}(k)|^2}{S_{xx}(k)\,S_{yy}(k)}, \tag{15}$$

where $S_{xy}(k)$ is the cross power spectral density and $S_{xx}(k)$ and $S_{yy}(k)$ are the auto power spectral densities, respectively. In the following, $x(t)$ and $y(t)$ are assumed as the output of two spatially spaced microphones and thus $\gamma_{xy}^2(k)$ is referred to as *spatial coherence*. It can be easily related to frequency using $k = \omega/c$. In an ideal broadband diffuse sound field, the spatial coherence between two omni-directional sensors with distance d can be computed theoretically with [5]

$$\gamma_{pp}^2(k,d) = \left(\frac{\sin(kd)}{kd}\right)^2. \tag{16}$$

The spatial coherence between two particle velocities perpendicular to each other and both positioned in the same plane at a distance d can be calculated using [5]

$$\gamma_{v_\perp v_\perp}^2(k,d) = 9\left(\frac{\sin(kd) - (kd)\cos(kd)}{(kd)^3}\right)^2. \tag{17}$$

The reader is referred to Elko [2] for an analytic description of the spatial coherence functions of arbitrary two-dimensional sensor layouts.

Frequency Dependent De-correlation

A time domain signal or impulse response $p(t)$ can be analyzed using a short time Fourier transform (STFT) given by

$$P(\tau,\omega) = \int_{-\infty}^{\infty} p(t)w(t-\tau)e^{-i\omega t}dt, \tag{18}$$

where $w(t-\tau)$ denotes the shifted version of the window function $w(t)$. The phase spectrum of the time variant spectra $P(\tau,\omega)$ are denoted as $\Theta(\tau,\omega) = \angle P(\tau,\omega)$. The phase spectra can be used to achieve a desired coherence $\widetilde{\Theta}(\tau,\omega)$ by applying:

$$\widetilde{\Theta}(\tau,\omega) = \Theta(\tau,\omega)\cdot\gamma(\omega) + \Theta_N(\tau,\omega)\cdot(1-\gamma(\omega)), \tag{19}$$

where $\Theta_N(\tau,\omega)$ denotes a random phase spectrum. The desired signal in the time domain can be achieved by applying the inverse short time Fourier transform after changing the original phase spectrum to the modified version $\widetilde{\Theta}(\tau,\omega)$, resulting in the new representation $\widetilde{P}(\tau,\omega)$:

$$\widetilde{p}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\widetilde{P}(\tau,\omega)e^{i\omega t}d\tau d\omega. \tag{20}$$

In the implementation used for the listening experiment, a discrete version was applied. The parameters for the discrete short time Fourier transform were a window length of 512 samples and a FFT-size of 1024 samples. To avoid artifacts due to the phase modification, an additional cosine-tapered window was used in the reconstruction process.

Listening Experiment

The aim of the experiment was to compare the virtual microphones which were generated based on spherical array measurements with an ideal simulation of the desired stereo setup. Both were calculated based on the same diffuse field. Furthermore, the simulation of the desired coherence function as described in the previous section was included. For all configurations, impulse responses were calculated. These were convolved with four different dry audio signal, consisting of a 50 ms pink noise bursts, continuous pink noise, a sentence of male speech and a short loop of drums with accompanying electric bass. The listening experiment design was a quality grading experiment based on [4]. The stereo configurations used were a coincident figure-of-eight setup (BL) with a angle difference between the microphones of 90° and a configuration of two pressure sensors with a distance of 0.5 m (AB). During a training phase, the subjects had the opportunity to listen to samples of the test items, to become familiar with the audio material and the expected artifacts. 15 subjects in the age range of 22 to 43 years participated in this listening experiment. No hearing problems were reported by the 3 female and 12 male subjects.

Test Items

The diffuse sound field was simulated by calculating a plane wave sound field with 1202 nearly equal distributed directions (Lebedev grid). For each plane wave a decorrelated white noise signal was generated. This was temporally shaped using an envelope. The parameters used for the envelope are an attack time of 30 ms, a sustain time of 15 ms and a release time of 500 ms. An air damping filter was applied using a STFT based time variant filter according to the mean distance.

As a reference the impulse responses are calculated by a superposition all plane waves according to the directivity and direction of the microphones to simulate. The same principle was used to calculate the impulse responses of a single microphone of the spherical arrays. From the output of these arrays, the impulse responses of two stereophonic setups are computed using the spherical harmonic extrapolation. Two different spherical array set-ups with a radius of 0.28 m are simulated (Open sphere cardioid configuration). *Array 1* is an ideal simulation of the diffuse field without any measurement errors. *Array 2* includes the following measurement errors:

- 80 dB signal-to- noise ratio
- 1° normal distributed random positioning error
- 4° offset error in co-elevation
- realistic cardioid characteristic

Both arrays consist of 2030 sampling positions on a nearly uniformly spaced grid (Lebedev). The maximum level used in the calculations was set to $N = 38$. The used level for a specific frequency was set to $N = \lceil kr \rceil$ to achieve the highest robustness against measurement errors. For the item named as *simulation* a second impulse response was derived based on one reference impulse response using frequency dependent de-correlation. The coherence function required was calculated using Eqs. (16) and (17).

A hidden reference and an anchor consisting of a mono version of the reference audio signal was also included. The test was performed using electrostatic headphones and high quality A/D converters.

Experiment Results

In the overall rating, the AB configuration (see Figure 1) based on the array without measurement error were graded significantly better than the array including measurement errors. The simulation is in the same range as the spherical array configurations. The same holds for the figure-of-eight configuration. In view of the results for the different audio signals it was found that when using the speech signal, nearly all items yield excellent results, only the simulation was lightly worse. The reason is, that the spectral content of the speech signal excludes the critical frequency range of the arrays. In case of the noise stimulus, the ratings tend to be worse because spectral differences were perceived more easily by all of the subjects. For the noise bursts, the spherical array set-ups were graded significantly worse than the simulation, especially in case of the AB configuration. In this constellation, the extrapolation errors of the array are much stronger than in case of the figure-of-eight configuration. Furthermore, the extrapolation error can be perceived much better in transient signals than in static noise. In the BL set-up, the difference between the array including errors and setup without can be distinguished. The simulation is in the same range as the error free set-ups. One can conclude again that the audio signal has very strong influence on the perception of different artifacts of the array processing. In case of diffuse sound fields for virtual microphone set-ups, the simulation of the coherence function delivers results comparable to those of the array processing.

Conclusions

In this paper the perceptual influence of measurement errors of spherical open cardioid microphone arrays was analyzed for diffuse field measurements. It was shown that the non-ideal simulation are not graded significantly worth that an ideal simulation in case of stereophonic auralization using an AB and a coincident figure-of-eight set-up. Depending on the virtual sensor configuration, the extrapolation error is an important problem, which needs further investigation. The proposed simulation of

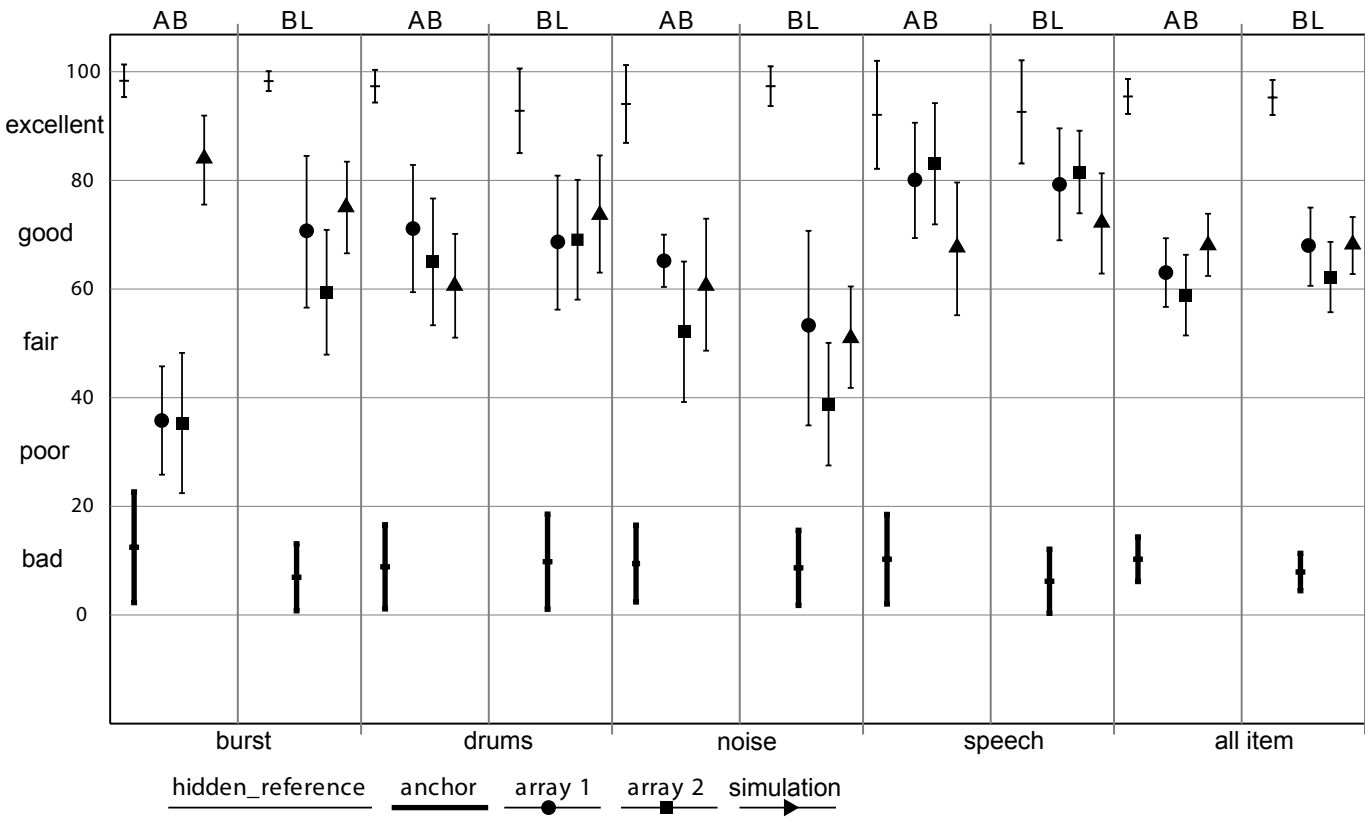

Figure 1: Results of the listening experiment: Mean values and 95% confidence intervalls.

the coherence function of virtual sensors is perceptual equivalent to the used array configuration. This is an important result because the computational effort in array processing is very high for complete impulse responses. A possible solution is to use the array processing only for the early part of the impulse response and calculate the late/diffuse part using the frequency dependent de-correlation method.

Acknowledgment

The authors like to thank Oliver Thiergart for his support preparing the listening test items and all subjects for participating in the listening experiment.

References

[1] J.S. Bendat and A.G. Piersol. *Engineering Applications of Correlation and Spectral Analysis*. Wiley, 1980.

[2] Michael Brandstein and Darren Ward, editors. *Microphone Arrays*. Springer, 2001.

[3] Richard K. Cook, R. V. Waterhouse, R. D. Berendt, Seymour Edelman, and M. C. Thompson JR. Measurement of correlation coefficients in reverberant sound fields. *Journal Acoustic Society of America*, 27(6):1072–1077, November 1955.

[4] International Telecommunication Union (ITU). ITU-R BS.1534-1 Method for the subjective assessment of intermediate quality level of coding systems. Technical report, ITU Radiocommuncation Assembly, 2003.

[5] Finn Jacobson. The diffuse sound field. Technical Report 27, The Acoustic Labroratory, Technical University of Denmark, 1979.

[6] Boaz Rafaely. Spatial-temporal correlation of a diffuse sound field. *Journal Acoustic Society of America*, 107(6):3254–3258, June 2000.

[7] Boaz Rafaely. Analysis and design of spherical microphone arrays. *IEEE Transactions on speech and audio processing*, 13(1):135–143, January 2005.

[8] Heinz Teutsch. *Modal Array Signal Processing: Principles and Applications of Acoustic Wavefield Decomposition*. Springer, 2007.

[9] D.A. Varshalovich, A.N. Moskalev, and V.K. Khersonskii. *Quantum Theory of Angular Momentum*. World Scientific, 1988.

[10] E.G. Williams. *Fourier Acoustics*. Academic Press, 1999.