

Beamforming using a spherical microphone array based on legacy microphone characteristics

Nils Peters^{1,2} and Andrew W. Schmeder¹

¹*Center for New Music and Audio Technologies (CNMAT), UC Berkeley, USA*

²*International Computer Science Institute (ICSI), Berkeley, USA*

Email: nils@icsi.berkeley.edu, andy@cnmat.berkeley.edu

Abstract

This paper presents a numerical approach that projects frequency-dependent directivity patterns of classic recording microphones onto steerable beams created with a spherical microphone array. In an anechoic chamber, the spatial and timbral characteristics of a legacy recording microphone and the characteristics of a 120-channel spherical microphone array were measured. Using a least-square matching approach, the measured frequency responses were used to calculate the set of filters that synthesize the desired legacy recording microphone characteristic from the 120-channel spherical microphone array. Synthesized microphone-beams are shown and compared with the measured characteristics of the original legacy microphones.

Introduction

Spherical microphone array technology allows for sound-field analysis and beamforming, and the creation of dynamic beams in any desired direction. Although flexible beamforming features may be desired in audio production, Tonmeisters tend to prefer first-order legacy recording microphones over novel microphone arrays. This preference is probably related to the familiarity in working with recording microphones. Contrary to engineering measurement applications that require neutral microphones with a flat frequency response, in sound recording, Tonmeister often desire microphones with “character”, expressed through parameters such as sensitivity, nonlinear distortions and off-axis frequency response. Consequently, the selection and placement of the “right” microphone becomes a highly subjective and irreversible decision.

Because beamforming technology can capture a sound-field with a flexible directivity and steering orientation, these kinds of preproduction decisions can be moved to the production process.

Unfortunately, beamforming technology has a reputation for not capturing the sound with the same character as recording microphones. This might be due to the fact that beamforming algorithms often prioritize the creation of highly directed beams while timbral features are of secondary importance. Although the authors are not aware of studies on the importance of specific microphone parameters on overall preference, research in sound reproduction quality suggests that timbral attributes are of higher importance than spatial attributes [1]. Also,

an array’s performance and sound character may be affected by manufacturing differences of a large number of material components, such as capsules, microphone-preamps, or AD-converters.

In this paper we use a spherical microphone array (the SAM-array) to synthesize the frequency-dependent directivity response of recording microphones. We first measured the frequency response of a recording microphone and the frequency response of our spherical microphone array. Then, using a numerical approach, we compute the best-fit filters for a filter-and-sum beamforming algorithm that matches the frequency response of the recording microphone. Theoretical results are compared with real-measurements. We also show how this virtual recording microphone can dynamically be re-oriented and how the desired microphone’s directivity pattern can be modified at run time.

One might wonder why we aim to simulate a recording microphone using a microphone array. As with many old synthesizers, tube amplifiers and other electro-acoustical instruments, there is a desire to simulate the complex behavior of rare, expensive and sensitive musical devices to enable their use in today’s digital audio workstation environments.

A decade ago, the company Antares¹ released the *Microphone Modeler*. This audio effect strived to virtually change the brand and type of the microphone used in a mono-channel recording. The user defines a) what microphone model was used within a recording and b) what microphone model is desired. Then the algorithm aimed to match the frequency response of both microphones to create the desired effect. Since the Microphone Modeler is discontinued, it was probably commercially not very successful. The algorithm only modeled the on-axis frequency response of the desired microphones using a one-channel input signal, neglecting other crucial microphone parameter such as the off-axis frequency response.

¹<http://www.antarestech.com/products/amm.shtml>

A recording microphone

Microphones are categorized according to their ideal directivity Γ in classes as listed in Table 1. The directivity can be computed with Equation 1 with the angle of incidence δ .

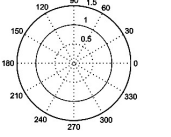
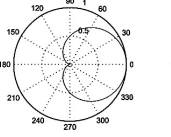
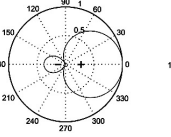
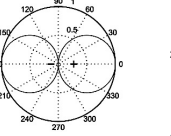
Omni-directional	Cardioid	Hyper-cardioid	Figure-8
			
$a = 1.0$	0.5	0.3	0.0

Table 1: Common first-order microphone directivity pattern

$$\Gamma_i = a + (1 - a) \cdot \cos \delta_i \quad 0 \leq a \leq 1. \quad (1)$$

Manufacturers usually supply octave-smoothed, two-dimensional cross-sections of the directivity pattern as a reference, because the directivity depends on frequency and microphone model. We measured the directivity of a popular large-diaphragm cardioid condenser microphone (2200 US\$) from 576 directions in an anechoic chamber. The measurement procedure is described later in this paper. The horizontal polar patterns of the measurement compared with the specification sheet by the manufacturer shows a lot of resemblance across the entire frequency range (see Figure 1). However, these horizontal polar pattern do not reveal common axial asymmetries in the directivity, which are only visible in the spherical balloon plots (Figure 2).

A transformation of this directivity pattern into spherical harmonics is visualized in Figure 3. One can see that most prominent contributions are of zero and first order spherical harmonics. This is expected, since the recording microphone has a cardioid characteristic which can be expressed with zero and first order spherical harmonics. Also clearly visible are the omnidirectional characteristic in low frequency bands due to the high contribution of the zero order harmonics, and an

increasing contribution of the first order harmonics in the mid-frequencies around 1 kHz. Second order harmonics start to contribute at around 500 Hz to the the measured directivity pattern. It can be observed in Figure 2 how the recording microphone becomes increasingly directed for higher frequencies; the increase in directivity is expressed through additional contributions of higher spherical harmonics.

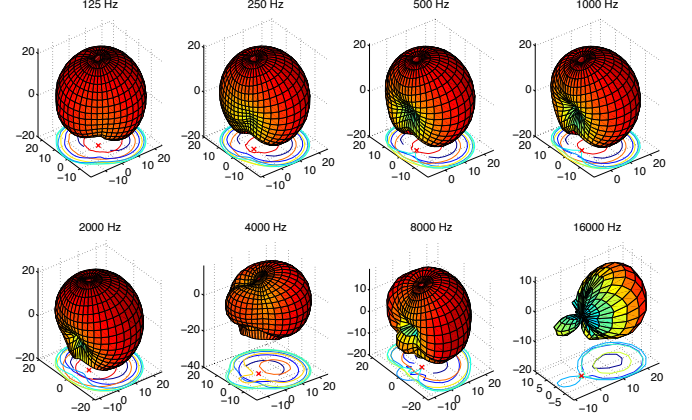


Figure 2: Measured spherical directivity of the recording microphone at different frequencies. Projection contour plots at the floor show directivity at different cross-sections. Red cross indicates projected center of origin.

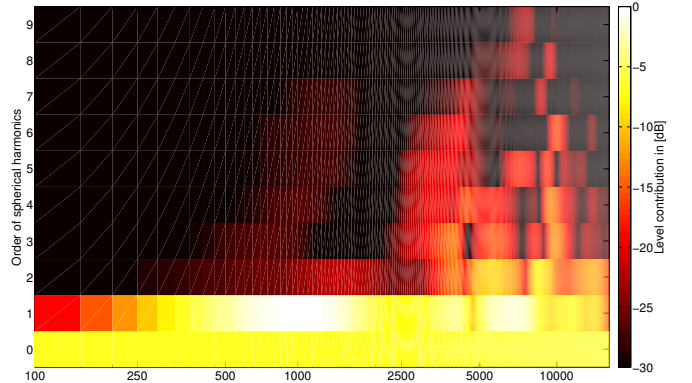


Figure 3: Frequency dependent contribution of different orders of spherical harmonics to the overall directivity of the measured recording microphone.

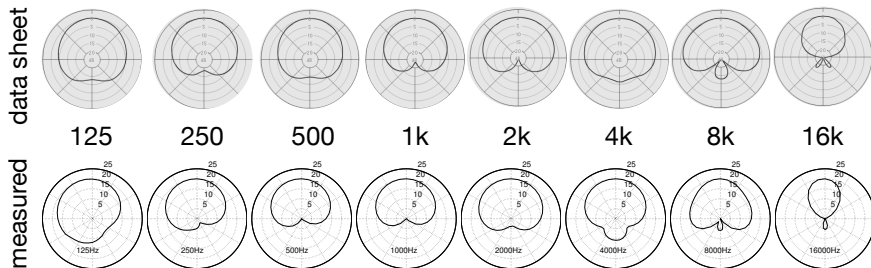


Figure 1: Comparison of manufacturer's directivity specification and measured directivity in horizontal plane (normalized and 3rd octave-smoothed)

The SAM-array

The SphericAl Microphone array (SAM), a co-development of CNMAT and Meyer Sound Laboratories, consists of 144 microphones. The majority (120) are cardioid microphones (DPA 4080), normally placed on a rigid sphere whose radius is about 4.5 cm. The distance between the rigid sphere and the actual microphone diaphragms is about 2.5 cm, resulting in an array radius of about 7 cm. The spherical array order is about $N = 9$, estimated by the number of cardioid microphones ($N = \sqrt{120 - 1}$). With that array order, the spatial aliasing frequency is about 7 kHz, defined by $kr \approx N$, where k is the wavenumber and r the array radius. The Rayleigh resolution limit, which estimates the smallest angle where two proximate plane waves can be separated, is about $\pi/N = 20^\circ$. Because accurate capturing and processing of low frequencies is challenging with a small array diameter, additional 24 omnidirectional microphones (DPA 4060) are mounted as outriggers around the inner sphere in 3 different radii, ranging from 200 cm to 1 m.

All microphone pre-amps and AD-converters are embedded inside SAM, permitting distribution of all 144 audio channels over an ethernet connection in 24 Bit/96 kHz [2]. Additional field-programmable gate arrays (FPGA) for onboard signal processing are also available.

The matching approach

To synthesize the directivity of a recording microphone, we adopted a filter-and-sum beamformer approach by Farina et al. [3]. This approach is based on computing a set of FIR filters (h_n) for the M microphones of a microphone array. These filters modify the microphone signals of the SAM-array so that its sum results in the output signal y (see Figure 5). The output signal y is a microphone beam whose directivity approximates the directivity pattern q of the desired recording microphone.

The frequency responses of the best-fit FIR filters (h_n) are estimated in the spectral domain by solving the least-square equation 2 for each spectral bin k in a half-sided FFT spectrum. This approach requires that the frequency responses of all M microphones from the SAM-array are measured from D source positions around a sphere. These frequency responses are stored in the matrix C , and C^* indicates its complex conjugate. The sensitivity of the recording microphone with respect to the same D source positions is represented by the vector Q . If L different microphone beams are synthesized, Q become a matrix of size $M \times L$. In contrast to [3], Q can vary across spectral bins k to account for the frequency-dependent directivity pattern of the recording microphone. One advantage of this numerical approach is that C comprises the effects of sensor misplacements and individual component irregularities of the microphone array. Therefore the resulting FIR filters h_n ought to compensate these irregularities.



Figure 4: The SAM-array with outriggers (upper picture), and close-ups of the onboard electronics (left) and of the inner sphere (right).

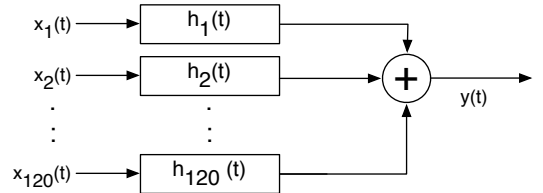


Figure 5: Filter-and-Sum beamforming principle

$$[H_k]_M = \frac{[C_k]_{M \times D}^* \cdot W_{M \times M} \cdot [Q_k]_D \cdot e^{-j\pi k}}{[C_k]_{M \times D}^* \cdot W_{M \times M} \cdot [C_k]_{D \times M} + \beta_k \cdot [I]_{M \times M}} \quad (2)$$

Weighting

We added the weighting matrix W to Equation 2 to account for an unevenly spherical distribution of the measurement positions D . The measurements (see next section) employed an equi-angle sampling grid characterized by a relative increase in sampling density from the equator to the poles. In estimating the best-fit filter response without weighting, the least-square solution would favor the oversampled poles compared to the undersampled equator region. Therefore, a sin weighting function was used. Other sampling schemes require different weighting functions (see [4] for an overview).

Regularization

The regularization parameter β_k controls the tradeoff between accuracy and robustness of an ill-conditioned inversion of $(C_k^* \cdot W \cdot C_k)$. Setting $\beta_k = 0$ produces the best accuracy but has the potential that small sampling errors cause instabilities and excessive signal amplification in H_k . This is especially common on a small array for low frequencies due to sampling errors. Several approaches to estimate an optimal regularization parameter β are known [5]. We implemented an algorithm that selects β based on the condition number κ_k (Equation 3). The condition number κ_k is known to be a measure of a matrix's ease of invertibility, and is 1 for an orthogonal matrix and $\gg 1$ for an ill-conditioned matrix. For each spectral bin k and starting at $\beta_k = 0$, our iterative algorithm increases β_k until the condition number κ reaches a predefined value. This predefined value can depend on frequency-dependent, similarly to what is suggested in 2.

$$\kappa_k = \text{cond}([C_k]_{M \times D}^* \cdot W_{M \times M} \cdot [C_k]_{D \times M}) \quad (3)$$

Steering of the synthesized microphone

Changing the orientation of the synthesized microphone beam can be expressed as modifying Q with regards to the steering angle. The correct modification can be performed via coordinate transforms of the microphone's spherical base solutions (Figure 3) as described in [6] or [7]. Another less efficient strategy is to compute Q by rotating and interpolating the sampled polar pattern.

Recomputing entirely Equation 2 every time Q changes would be too expensive for real-time applications. Because a majority of this computation is independent from Q , Equation 2 can be simplified to:

$$[H_k]_M = [A_k]_{M \times D} \cdot [Q_k]_D \cdot e^{-j\pi k} \quad (4)$$

with

$$[A_k]_{M \times D} = \frac{[C_k]_{M \times D}^* \cdot W_{M \times M}}{[C_k]_{M \times D}^* \cdot W_{M \times M} \cdot [C_k]_{D \times M} + \beta_k \cdot [I]_{M \times M}} \quad (5)$$

The matrix A_k in Equation 5 is independent from the desired beam directivity Q and can therefore be precomputed. Thus, when Q changes, the computational cost of H_k can be reduced to one matrix-vector multiplication per spectral bin k . The directivity vector Q can also be manipulated in different ways: by raising Q to the power of a number larger than 1 increases the directivity of the beam. As a side note, this algorithm scales easily on many-core processor architectures because computing H_k is independent for each k .

Measurements

To apply the matching approach discussed in the previous section, impulse responses from the SAM-array and the recording microphone were measured from 576 positions in an anechoic chamber. Both measurements were facilitated by using a robot arm that rotates the microphone device under test in azimuth and elevation angle with regards to a fixed reference loudspeaker, a Meyer Sound HD-1. The distance of the loudspeaker to the microphone device was 4 m, sufficiently satisfying the far-field assumption. The measurement setup is depicted in Figure 6. For measuring the recording microphone, the mic-preamp and AD converter of the RME Fireface were used. The 120 audio channels of the SAM-array were digitally streamed directly over an ethernet connection to the computer.

For both the SAM-array and the recording microphone, an equi-angle sampling scheme was used. The spatial sampling resolution was 10 degrees in both azimuth and elevation angle. Unfortunately, both devices could not be measured from all spherical directions, because at 2 elevation angles, the robot arm would have interfered with the propagation path. Therefore, only 36×16 instead of 36×18 were collected. In the case of the SAM-array, for each of these 576 measurements, the response of all 120 cardioid microphones were captured. None of the 24 outriggers were measured. For both devices, the impulse responses were computed in 24 Bit/96 kHz from a 1 second exponential sine sweep [8] ranging from 100 Hz to 16 kHz. The high sampling rate of 96 kHz is beneficial for accurately measuring the relative time-of-arrival differences between the array microphones. We did not measure frequency components lower than 100 Hz because beamforming processing with a small array diameter is prone to low frequency measurement errors. A future measurement session will include the 24 outrigger microphones and address these frequencies.

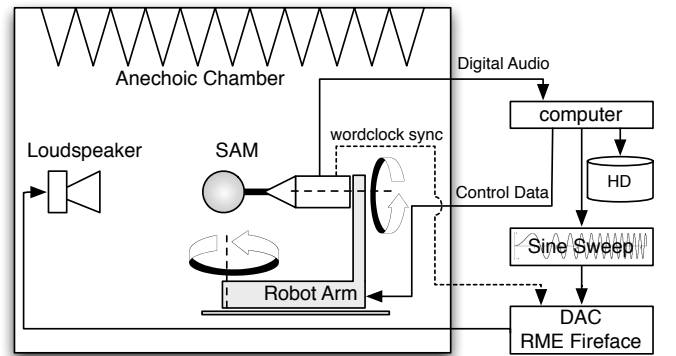


Figure 6: Sketch of the measurement setup

Results

This section presents four results using the matching approach. Two different microphones are re-created: first, an ideal cardioid microphone whose frequency-independent directivity pattern was computed with Equation 1; second, the real-world recording microphone whose measured frequency-dependent directivity pattern is shown in Figure 2. We first study the matching algorithm under ideal conditions, using a simulated SAM-array. Afterward, the frequency responses of the real SAM-array are used. For objective comparisons, all results are computed with a regularization parameter $\beta = 0.0$. The FFT size is 2048 samples at 96 kHz.

Simulated SAM

We simulate the 120 cardioid capsules of the SAM-array using the ViMiC system [9]. ViMiC is a 3D auditory virtual environment where sound sources and sinks in the form of microphones can be freely placed in a virtual room. According to parameters for characterizing sources, room, and microphones, the virtual microphone signals are computed with the proper time-of-flight delays and level attenuation of direct sound and optional early reflections. The simulation approximates the SAM-array with frequency-independent cardioids microphones and without the inner rigid sphere. The latter approximation is reasonable because the modal magnitude responses between an open cardioid array and a cardioid array with a rigid sphere are reasonably similar [10].

Figure 7 shows the matching results of an ideal cardioid microphone onto the simulated microphone array. At first, one can see in Figure 7(b) that the ideal cardioid microphone is very well re-created up to the 4 kHz. Above this frequency, the computed beam continues to be directional, but matching errors start to distort the smooth cardioid directivity pattern. Figure 7(a) shows the normalized magnitude response of the 120 array filters h computed with Equation 2. These filters have their largest amplification around 10 kHz - a region where the matching approach according to Figure 7(b) fails to work. Also, for low frequencies, a few filters attenuate the input signal by about 80 dB and can be attributed to the high condition number of the ill-conditioned matrix (see Equation 3). A frequency-dependent regularization as shown in Equation 2 will prevent those extreme attenuation and amplification values at the expense of a less accurate directivity matching at those frequencies.

Figure 8 shows the matching results of the measured directivity response of the legacy recording microphone onto the simulated microphone array. A visual comparison of the measured directivity pattern (Figure 2 with Figure 8(b)) indicates that the character of the recording microphone is very well re-created up to the 4 kHz. Beginning with 8 kHz, the directivity pattern is getting more and more distorted, also observable in the floor projections of the directivity cross-sections in Figure 8(b).

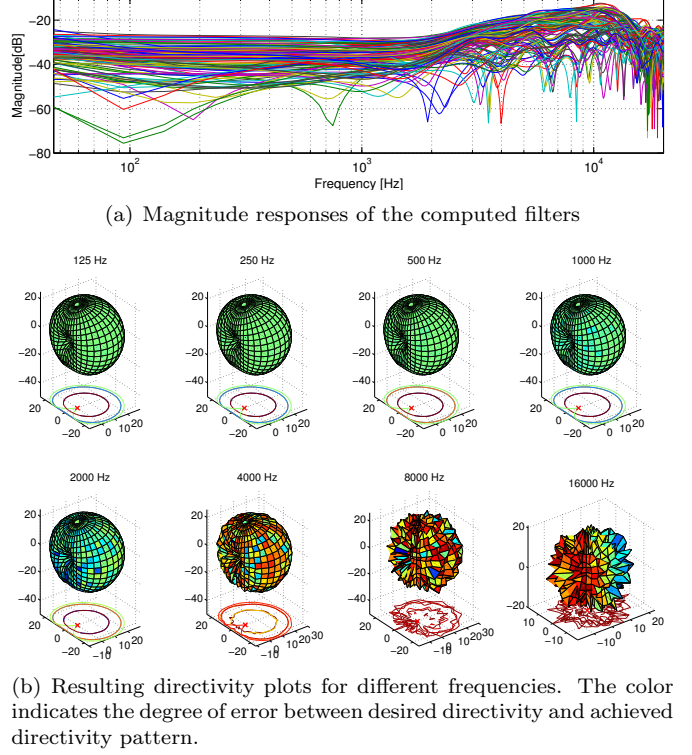


Figure 7: Synthesizing an ideal cardioid microphone with the simulated SAM-array

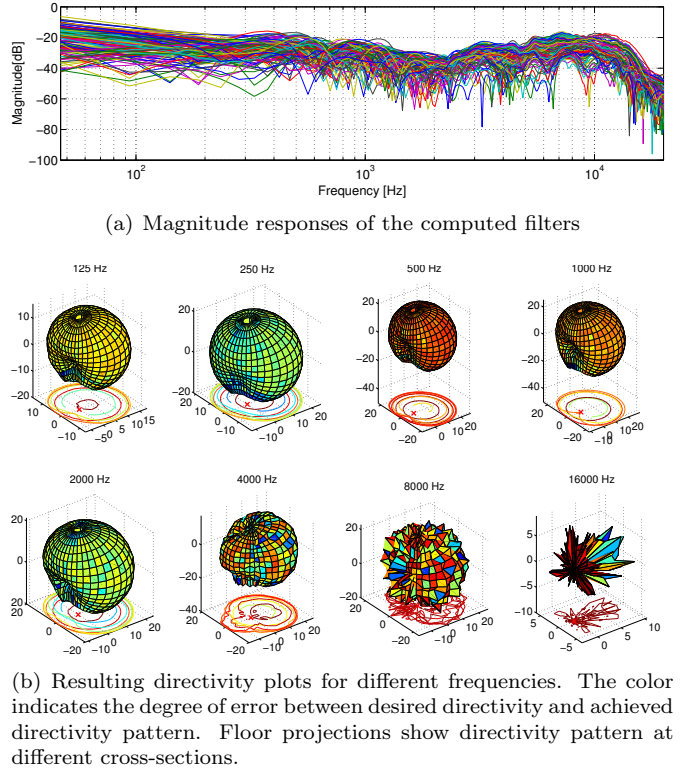


Figure 8: Synthesizing the measured recording microphone with the simulated SAM-array

Measured SAM

In this section we create microphone beams based on the measured frequency responses of the SAM-array.

Figure 9 depicts the results of creating a beam with an ideal cardioid characteristic. A visual evaluation of Figure 9(b) shows that the matching algorithm works well for frequencies up to 2 kHz. Above this frequency, the results degrade in the same way as described in the previous section: the beam is still directed and steers into the right direction, but the smoothness of the cardioid directivity pattern increasingly distorts with higher frequencies. Figure 9(a) shows the normalized frequency response of the 120 generated array filters. Up to about 1.5 kHz, the magnitude responses are between -40 dB and 0 dB (despite a few filters which have a larger attenuation at low frequencies). From 1.5 kHz to 4 kHz, the magnitude responses steadily decrease from -40 dB to -80 dB. Remarkably, this magnitude drop happens in the frequency range where the matching errors start to affect the re-created directivity pattern (see Figure 9(b)).

Figure 10 depicts the results of re-creating the legacy recording microphone. Up to 2 kHz, the legacy recording microphone is reasonably well re-created with the SAM-array, see Figure 10(b). Again, at higher frequencies, the matching errors increase and the directivity pattern becomes distorted. The magnitude responses of the computed filters, shown in Figure 10(a), are similar, yet not equal, compared to those for re-creating the ideal cardioid microphone in Figure 9(a). Compared to those of the simulated SAM-array in Figure 8(a), the filters look different, which could be due to the frequency-dependent sensitivity of the SAM-array sensors.

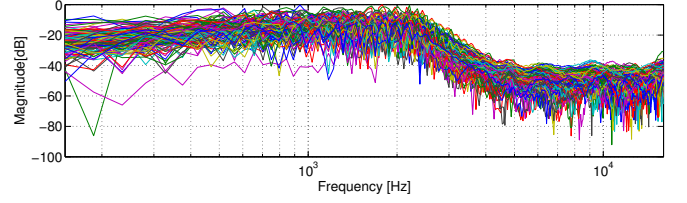
Discussion

The matching results of the simulated SAM-array (Figure 7 and 8) compared to those achieved with the measured SAM-array (Figure 9 and 10) are similar in accuracy for low and mid frequency up to about 2 kHz. Above 2 kHz, the matching accuracy using the measured SAM-array data decreases. With the simulated SAM-array, such decrease in matching accuracy starts one octave later at about 4 kHz.

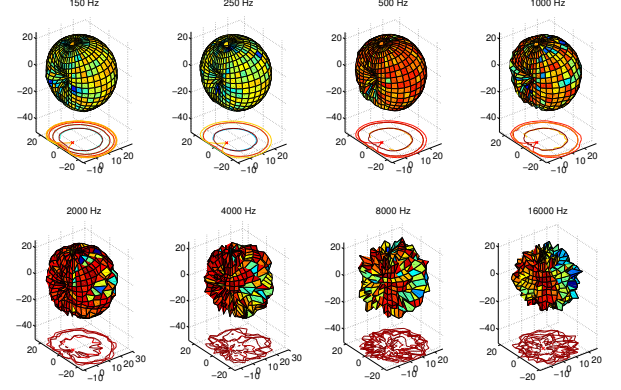
As a measure of overall matching accuracy, we introduce the mean error ϵ_k , which we define as the difference between the desired directivity Q and the computed directivity of the beam, averaged over all D sampling positions (Equation 6). This mean error ϵ is frequency dependent as denoted through the index k .

$$\epsilon_k = \frac{1}{D} \left[\sum_{d=1}^{D=576} \left| Q_k(d) - \sum_{i=1}^{M=120} C_k(d, i) \cdot H_k(i) \right| \right] \quad (6)$$

Figure 11 shows ϵ as a function of the frequency for all four matching results. Both simulated results have a mean error below 1 dB up to 4 kHz. For higher frequencies, the errors increase fast. The mean errors of the results using the measured SAM-array data are similar to those of the simulated array up to about 1 kHz.

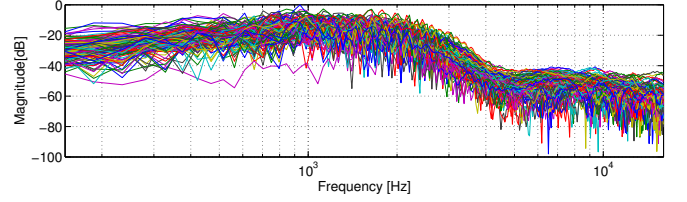


(a) Magnitude response of the 120 computed filters h

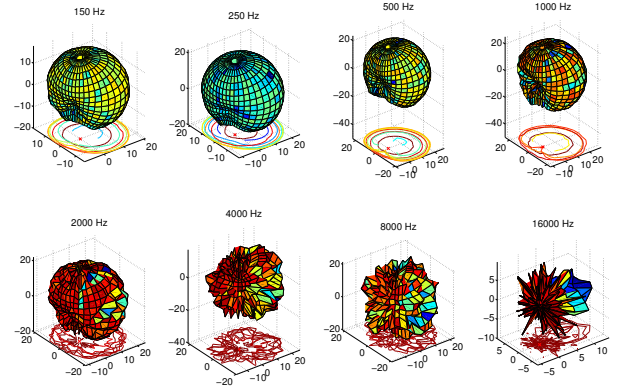


(b) Resulting directivity plots for different frequencies. The color indicates the degree of error between desired directivity and achieved directivity pattern.

Figure 9: Synthesizing an ideal cardioid microphone with the SAM-array



(a) Magnitude responses of the 120 computed filters h_n



(b) Resulting directivity plots for different frequencies. The color indicates the degree of error between desired directivity and achieved directivity pattern.

Figure 10: Synthesizing the measured recording microphone with the SAM-array

Then, the matching error of the results based on the measured SAM-array increases to about 4 dB at 4 kHz.

Besides sampling errors due to spatial aliasing, which are expected for frequencies above 8 kHz, it is unclear what the cause is for the divergence of the mean

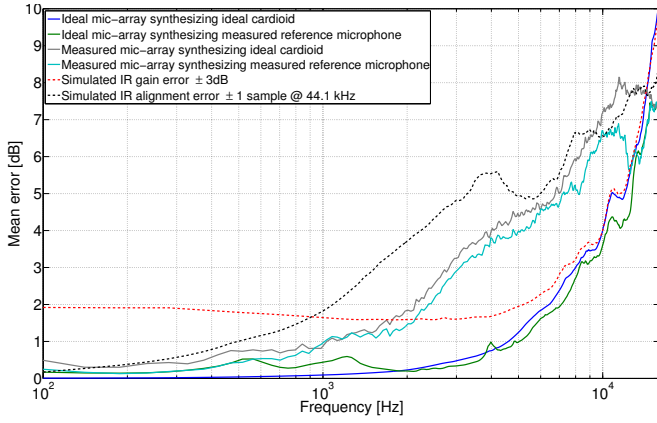


Figure 11: Mean error comparison of the matching results using simulated and real measured data

error. To find an answer to that question, we simulated two different errors which could have occurred in the process of measuring SAM’s frequency responses: first, a ± 3 dB sensitivity error randomly added to the frequency responses in c ; second, a temporal jitter of 1 sample randomly introduced to c which corresponds to about 0.8 cm flight distance at the speed of sound. The effect of these simulated errors on ϵ are shown with dashed lines in Figure 11. Compared to the mean error of the simulated SAM-array, the random sensitivity error constantly increases the mean error by about 2 dB from low to high frequencies. At 5 kHz both error curves start to converge again. A temporal jitter of ± 1 sample can significantly affect the matching accuracy starting at mid frequencies of about 500 Hz and rises to 5.5 dB at 4 kHz. This error simulation suggests that the time alignment of the measured frequency responses is important and needs to be carefully controlled to achieve the desired beam directivity for higher frequencies. The mean error of matching results using the measured SAM-array has a similar progression, but starts one octave later. This could indicate a small jitter error in the measurements.

Conclusion and future work

This paper described a numerical matching approach to synthesize legacy recording microphones with a spherical microphone array. We explained the motivation for this work and described the mathematical concepts to find the best filters for driving the microphone array. Using a 120-channel microphone array, results of this approach show that it is possible to synthesize the frequency-dependent directivity pattern of a recording microphone for a wide frequency band. For high frequencies, this approach starts to fail. Simulated measurement errors suggest that for best matching results, it is crucial to capture the frequency responses of the spherical microphone array with best temporal accuracy.

Future work will focus on increasing the matching accuracy for higher frequencies and to synthesize other microphone types. The discussed results assume that the recording microphone is used as a main microphone with sources in the far-field. Because many recording microphones are used as spot microphones with sound

source in close proximity, it would be interesting to synthesize microphones for this use case too.

Clearly, the flexible beamforming and beamsteering capabilities of microphone arrays can be useful in sound recording applications. When evolving from first-order microphones to the new generation of higher-order spherical microphone arrays for sound recording purposes, we need to understand what the salient parameters of recording microphones are. Is it the frequency-dependent directivity, the non-linearities, or is it the wooden box inside which an expensive microphone is packaged? Maintaining or even improving any of these salient parameters in the design of microphone arrays and beamforming algorithms will help to increase the use of microphone arrays in musical applications.

Acknowledgement

Thanks to Meyer Sound Laboratories for access to their anechoic chamber and technical support, and to Lorenzo Chiesi for clarifying [3]. Nils Peters is supported by the German Academic Exchange Service (DAAD).

References

- [1] F. Rumsey, S. Zieliński, R. Kassier, and S. Bech, “On the relative importance of spatial and timbral fidelities in judgments of degraded multichannel audio quality,” *J. Audio Eng. Soc.*, vol. 118, no. 2, pp. 968–976, 2005.
- [2] R. Avizienis, A. Freed, T. Suzuki, and D. Wessel, “Scalable connectivity processor for computer music performance systems,” in *International Computer Music Conference*, pp. 523–526, 2000.
- [3] A. Farina, A. Capra, L. Chiesi, and L. Scopece, “A spherical microphone array for synthesizing virtual directive microphones in live broadcasting and in post production,” in *Proc. of the AES 40th International Conference*, (Tokyo, Japan), 2010.
- [4] F. Zotter, “Sampling strategies for acoustic holography/holography on the sphere,” in *Fortschritte der Akustik: DAGA Tagungsband*, (Rotterdam, Netherlands), 2009.
- [5] P. Hansen, *Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion*. No. 4, Society for Industrial Mathematics, 1998.
- [6] N. Gumerov and R. Duraiswami, “Fast, exact, and stable computation of multipole translation and rotation coefficients for the 3-d helmholtz equation,” 2001.
- [7] F. Zotter, *Analysis and Synthesis of Sound-Radiation with Spherical Arrays*. PhD thesis, University for Music and Dramatic Arts, Graz, Austria, 2009.
- [8] A. Farina, “Simultaneous measurement of impulse response and distortion with a swept-sine technique,” in *108th AES Convention, Preprint 5093*, (Paris, France), 2000.
- [9] J. Braasch, N. Peters, and D. L. Valente, “A loudspeaker-based projection technique for spatial music applications using virtual microphone control,” *Computer Music Journal*, vol. 32, no. 3, pp. 55–71, 2008.
- [10] P. Plessas and F. Zotter, “Microphone arrays around rigid spheres for spatial recording and holography,” in *Fortschritte der Akustik: DAGA Tagungsband*, (Berlin, Germany), 2010.