EuroRegio2016
June 13-15, 2016
Porto ▪ Portugal

USING MICROPHONE ARRAYS TO RECONSTRUCT MOVING SOUND SOURCES FOR AURALIZATION

Fanyu Meng, Michael Vorlaender

Institute of Technical Acoustics, RWTH Aachen University, Germany
{fanyu.meng@akustik.rwth-aachen.de)

Abstract
Microphone arrays are widely used for sound source characterization as well as for moving sound sources. Beamforming is one of the post processing methods to localize sound sources based on microphone array in order to create a color map (the so-called "acoustic camera"). The beamformer response lies on the array pattern, which is influenced by the array shape. Irregular arrays are able to avoid the spatial aliasing which causes grating lobes and degrades array performance to find the spatial positions of sources. With precise characteristics from the beamformer output, the sources can be reconstructed regarding not only spatial distribution but also spectra. Therefore, spectral modeling methods, e.g. spectral modeling synthesis (SMS) can be combined to the previous results to obtain source signals for auralization.
In this paper, we design a spiral microphone array to obtain a specific frequency range and resolution. Besides, an unequal-spacing rectangular array is developed as well to compare the performance with the spiral array. Since the second array is separable, Kronecker Array Transform (KAT) can be used to accelerate the beamforming calculation. The beamforming output can be optimized by using deconvolution approach to remove the array response function which is convolved with source signals. With the reconstructed source spectrum generated from the deconvolved beamforming output, the source signal is synthesized separately from tonal and broadband components.

Keywords: auralization, synthesizer, microphone arrays, beamforming, SMS

PACS no. 43.60.Fg

1 Introduction

Moving sound source auralization finds its application in environmental noise issues caused by vehicles to achieve better noise control and subsequent traffic planning in densely populated urban and rural areas. Auralization is an effective modern technique, which makes it possible to perceive simulated sound intuitively instead of describing the acoustic properties with abstract numerical quantities. Simply put, auralization converts numerical acoustic data to audible sound files, through three procedures: sound source generation, sound propagation and reproduction [1].
The sound source generation consists of forward and backward approaches. The forward method is based on the *a priori* knowledge of the sources, such as physical generation mechanism and spectral data to obtain the source signal; while the backward method acquires the signal by inverting the propagation procedure (e.g., directivity, Doppler Effect, spherical spreading) from the recording [2]. For multiple sources propagating waves simultaneously, especially moving sources, sound source

signals can't be obtained by near field recordings in an anechoic chamber. In this case, the inverse backward method can be utilized in moving sound source synthesis for auralization.

Beamforming is a popular sound source localization method based on the array technique [3]. However, the beamforming output in the frequency domain cannot be directly considered as the source spectrum. It is the convolution of the spectrum and the array's point spread function (PSF). Therefore, DAMAS (Deconvolution Approach for the Mapping of Acoustic Sources) was applied to remove the PSF from the output [4]. Despite improvement of the acoustic image, DAMAS has the drawback of high computational cost in handling large matrices and iterations. In order to accelerate the beamforming process, the Kronecker Array Transform (KAT) as a fast separable transform is possible [5], where "separable" suggests that the microphones and the reconstruction grids are distributed rectangularly and nonuniformly [6]. By using these methods, sound sources can be precisely localized in a much reduced computational duration. In this research, beamforming is extended to auralization as well. The positions of the moving sound sources are identified, and the beamforming output is then used to reconstruct the source signals.

Despite the advantages mentioned above, the beamforming output spectrum cannot be directly taken as the source signal. It is known that DAMAS and other deconvolution methods can be used to remove the PSF. However, the PSF is rather unpredictable. For example, as sound from moving vehicles propagates in an outdoor environment, the measurement conditions are under poorer control than in a free-field condition in an anechoic chamber. Thus, some non-predictable acoustic effects can occur in the meanwhile, leading to uncertain PSF. Even if the measurement conditions are well controlled and the beamforming's characteristics are perfectly removed, the reconstructed signal can only be used for specific cases, in which the measurement for the particular sound sources takes place. Besides, not all information in the source needs to be reflected in the auralization, since the human hearing system is not sensitive enough to perceive every detail. Under these considerations, parameterization overcomes the drawbacks mentioned above. The beamforming output spectrum can be parameterized with excluding some unnecessary features. Spectral modelling synthesis (SMS) is a way to parameterize and synthesize a spectrum separately in deterministic and stochastic components on the fact that most sounds consist of these two components [7]. Parameters are generated to represent the source using SMS. Another benefit of parameterization is that it enables variable source signals' generation out of one sample. By changing parameters, the sound samples are generated dynamically according to different acoustic scenes. This provides more possibilities for auralization in the real-time virtual reality system without having to conduct repeated psychoacoustic measurements [8].

Even though some recordings are blurred by the poorly conditioned measurement environment, the frequencies of the tones in the spectra after de-Dopplerization can still remain correct [2]. The frequency, amplitude and phase are obtained by peak detection and continuation. The deterministic component is synthesized by generating sinusoidal signals with the previously mentioned parameters, and subsequently the broadband component is represented by the subtraction of the synthesized tonal signals from the original beamforming output spectrum.

The objective of this paper is to develop an efficient synthesizer for moving sound sources based on microphone arrays. The sound field produced by a moving sound source is described and a de-Dopplerization technique is introduced to prepare for beamforming. Using an array of 32 microphones, spiral and separable arrays are designed with similar resolution. The moving sound sources are localized by applying beamforming with de-Dopplerized input signals. Furthermore, beamforming is extended to source signal synthesis. Parameterization based on SMS utilizes beamforming output spectrum as a sample, with which different sound samples are able to be generated to adapt to different acoustic scenarios.

2 Moving sound source and de-Dopplerization

According to [9], the sound field generated by a moving sound source denotes as:

$$p(t) = \frac{1}{4\pi}\frac{q'(t-\frac{R(t)}{c})}{R(1-M\cos(\theta(t)))^2} + \frac{q(t)}{4\pi}\frac{v(\cos\theta(t)-M)}{R(t)^2(1-M\cos(\theta(t)))^2} \quad (1)$$

where $q(t)$ is the source strength, $q'(t)$ is the derivative of $q(t)$, $R(t)$ is the distance between source and receiver, c is the sound speed, v is the source moving speed, $M = v/c$ is the Mach number and $\theta(t)$ is the angle between source moving direction and source-receiver direction.
When the receiver is far away from the moving source with at a relatively low speed (normally M < 0.2), the previous equation is rewritten by omitting the second term:

$$p(t) = \frac{1}{4\pi}\frac{q'(t-\frac{R(t)}{c})}{R(1-M\cos(\theta(t)))^2} \quad (2)$$

To eliminate of the Doppler Effect, the recordings need to be interpolated and re-sampled. The reception time is calculated by $t = t_e + r/c$ by taking emission time as the reference time. Then the recorded signal is interpolated and re-sampled according to the equally-spaced reception time. This procedure is called de-Dopplerization. The de-Dopplerized signal is denoted as $\tilde{p}(t)$.

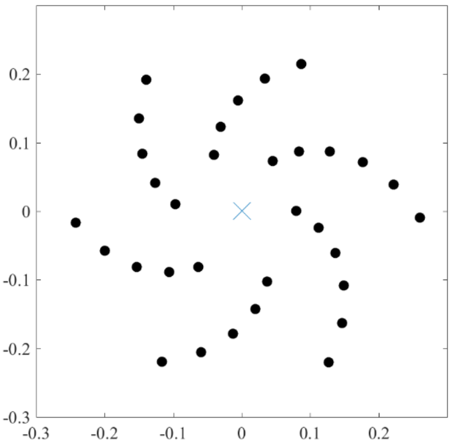
3 Microphone array design

3.1 Spiral array

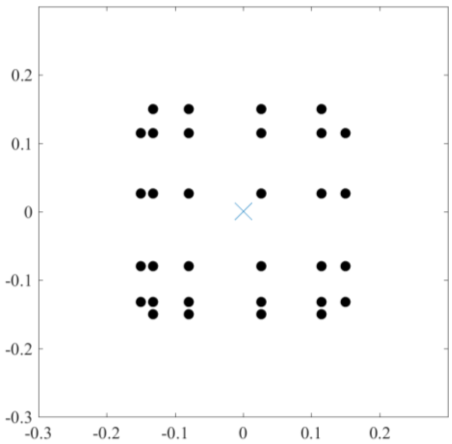
Spiral array has the advantages of decreasing MSL (maximum sidelobe level) and avoiding grating lobes over regular array [10]. In this research, an Archimedean spiral array is applied. The basic parameters are given in Table 1. Figure 1(a) shows its layout.

Table 1 – The basic parameters of the spiral microphone array

Microphone number	Spacing	Diameter	Resolution		
			Frequency	Steering angle	Distance
			3 kHz	30°	1.5 m
32	0.04-0.06 m	0.50 m	0.64 m		



(a) Spiral array



(b) separable array

Figure 1 – Spiral and separable array layout

3.2 Separable array

To accelerate the beamforming and deconvolution process using KAT, separable array geometry is necessary. To achieve comparable results, the resolutions of the spiral and separable arrays should be similar, and the microphone numbers remain the same. Therefore, the diameter of the separable array is set to 0.3 m, namely half of the spiral array's size. Non-redundant array [11] is able to keep higher resolution capability using a small number of microphones compared to a longer uniform array. In this research, 6 microphones are taken and aligned linearly, with the spacing set to 0.02 m, 0.07 m, 0.12 m, 0.26 m and 0.30 m between microphones, as illustrated in Figure 1(b).
Extending the linear non-redundant array to two-dimensional, a 6x6 array is obtained. After eliminating microphones in the four corners, the reduced number of the microphones remains identical with the spiral array. Figure 2 gives the beam patterns comparison at 3 kHz. First of all, after removing the microphones in the corners, the beam pattern remains similar. Secondly, the spiral and reduced separable arrays share similar beam width, and so is the resolution. It is obvious that the sidelobe levels of the separable arrays are almost 10 dB higher than that of spiral array. However, the sidelobes are not relevant any more by applying appropriate deconvolution methods.

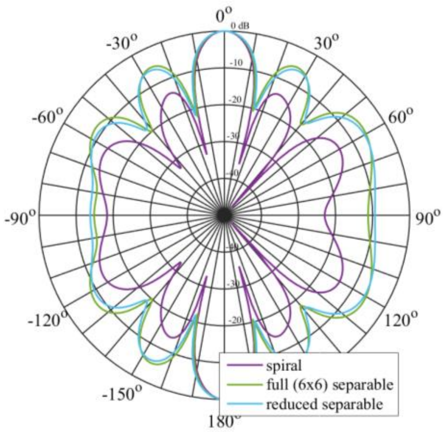


Figure 2 – The beam pattern of the three arrays

4 Beamforming and output spectrum

4.1 Delay-and-sum beamforming

Beamforming is a general way to localize sound sources based on temporal and spatial filtering using microphone array [12]. The most conventional beamforming is delay-and-sum beamforming (DAS), which reinforces the signal by delaying at each microphone and adding all the signals. The output of DAS denotes

$$y(t) = \sum_{m=1}^{M} w_m \tilde{p}_m(t+\tau_m) \tag{3}$$

where M is the number of microphones (all variables with lower case m represents the m th microphone), w_m is the weight of the output signal, $\tilde{p}_m$ is the de-Dopplerized signal, $\tau_m = (r_m - r_0)/c$ is the time delay, r_m is the distance between source and the microphone, and r_0 is the distance between source and array origin.

4.2 DAS output spectra

A plane (1.5 m x 5 m) is moving in the x-direction, carrying two point sources at the speed of 40 m/s. Two point sources are placed on the plane with 2 m spacing. A microphone array is set at 1.5 m away from the moving direction. The array origin is on the z-axis (Figure 3). The plane is meshed into grids, with 5 cm spacing between each other. Each grid represents a potential sound source, so that the array can steer its angle to "scan" the plane to search for the sources. The left source consists of a 2 kHz tone and noise, and the right source signal contains the same noise as in the left one. In both cases, additive white Gaussian noise (AWGN) with SNR = 20 dB is added. The sound pressure RMS of the tone and noise are both 1 Pa.

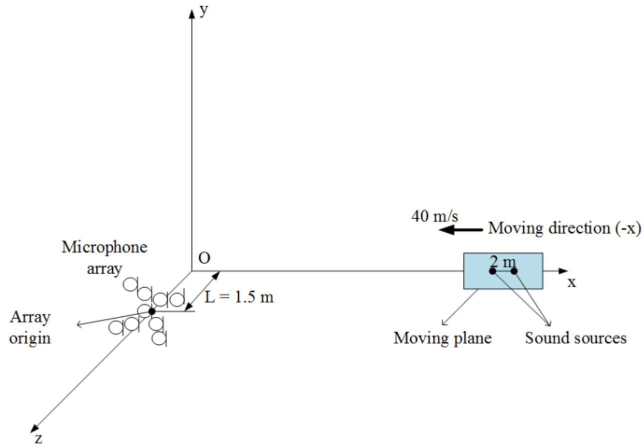


Figure 3 - Moving sound source measured by microphone array

The good localization ability of the deconvolution method using spiral and separable arrays has been confirmed to be comparable [13]. Therefore, this paper only shows the color map generated by DAS. The localization results of the two arrays at 2 kHz are shown in Figure 4. In these figures, the positions of the two sources are (1.5, 0.75) and (3.5, 0.75).
As previously examined, the spiral array has lower sidelobe levels and the resolutions of the two arrays are similar. The separable array has comparatively higher sidelobe levels. However, DAMAS has the capability to remove strong sidelobes. In this sense, separable array combining KAT can be used for beamforming. In addition, both arrays cannot resolve well at lower frequency (e.g., 1.6 kHz). For some frequency bands, the localization deviation reaches 5 – 10 cm.
In terms of localization capability, the separable array can replace the spiral array with DAMAS reducing the sidelobe levels and KAT accelerating the whole procedure.

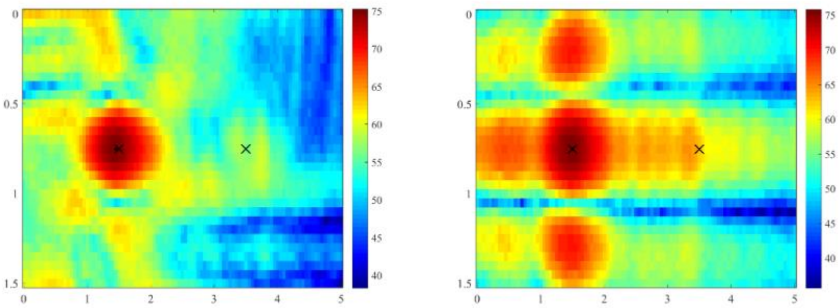


Figure 4 – Localization results of the spiral array and the separable array at 2 kHz ("×" represents the real source position and "·" represents the maximum value in the color map)

5 Sound source synthesis

5.1 SMS analysis and parameterization

The beamforming output spectrum representing the source is obtained by steering the array angle to the detected source position. Short-time Fourier Transform (STFT) is conducted on these spectra representing sources. The prominent peaks are detected in each magnitude spectrum, and then peak continuation is tracked along all the frames in the time domain. The deterministic component is synthesized by the sum of all the detected tones in all trajectories and frames. Afterwards, the broadband component is modeled by the original spectrum subtracted by the synthesized tonal component.

In SMS, there are several parameters are involved, such as window size, type, maximum peak amplitude (MPA), maximum guide number (MGN), maximum sleeping time (MST) and maximum peak deviation (MPD) during continuation detection [7]. Since the frequency resolution is limited when the window size is small, interpolation is conducted in the spectra before using SMS to increase the resolution to 1 Hz.

Figure 5 shows the beamforming output spectra of the two sources with spiral and separable arrays. The peak in each figure is obvious due to the removal of the Doppler Effect, and the amplitudes of the peaks are almost the same. Thus, in the sense of this synthesis step, the separable array is also comparable to the spiral one.

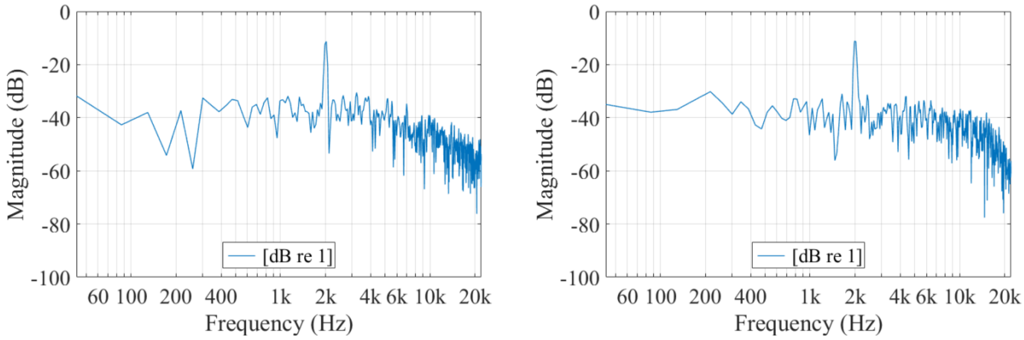


Figure 5 – Beamforming output spectra with array angle steering to the left source position (left: spiral array, right: separable array)

Taking MPA as a variable, the other parameters are given in Table 2. In this paper, the MGN is determined dynamically according to the peaks in the first frame.

Table 2 – Parameters of SMS

Window size	Hop size	Window type	MST	MPD
512	16	Kaiser	3 frame	150 Hz

The results with varying MPA are shown in Figure 6. As can be seen, when MPA = 0.025 Pa, a clear 2 kHz tone can be tracked along all the frames; while for the other cases, incorrect trajectories are found. With the source information already given in this example, the result can be verified directly. If the source is unknown, there is no *a priori* knowledge. Under this condition, the parameters need to be determined deliberately with proper simulation and verification.

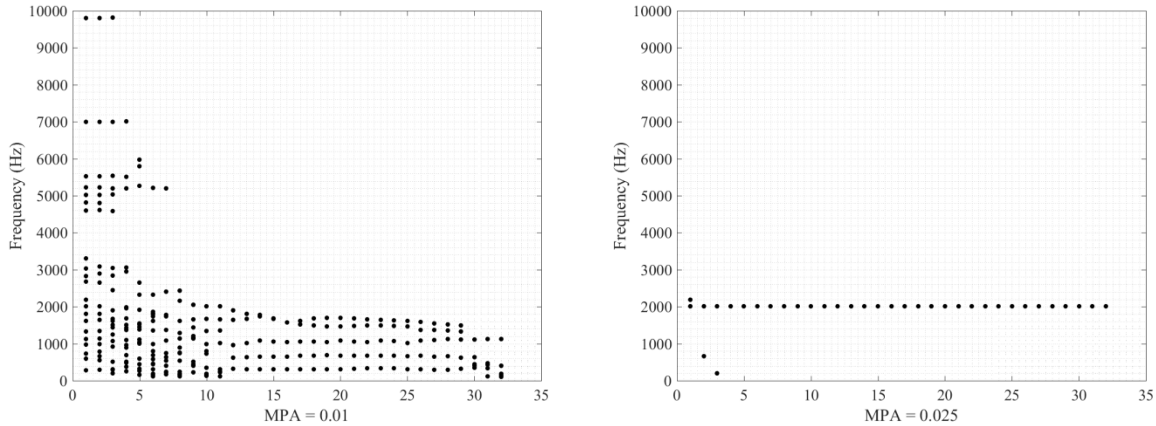


Figure 6 – Peak detection results with varying MPA

5.2 SMS Synthesis

After the determination of the peaks in each trajectory, the frequency, amplitude and phase (f_{ft}, A_{ft} and φ_{ft} respectively) are determined at the same time. Where $f = 1, \dots, F$ and $t = 1, \dots, T$ are the frame and trajectory numbers. The synthesized tonal signal at each frame is denoted as

$$s(n) = \sum_{1}^{t=T} A_{ft} \cos\left(2\pi f_{ft} n + \varphi_{ft}\right) \tag{4}$$

where $n = 1, \dots, N$ is the sample in each frame. Adding the synthesized signals at each frame and trajectory gives the final representation of the sinusoidal (deterministic) signal.
Subsequent residual signal is the subtraction of the deterministic component from the original beamforming spectrum. There is noise included during the recording, thus the level of the residual broadband component can be reduced accordingly to approach the original source signal. Adjusting noise is another benefit of parameterization for dynamic auralization to compensate for noise uncertainty in the measurements.

6 Conclusions and outlook

This paper describes a synthesizer of moving sound sources based on microphone arrays and parameterization for auralization. A de-Dopplerization technique in the time domain is introduced. DAS uses the de-Dopplerized signals as the inputs. The spiral and separable arrays with similar resolutions are compared in the localization and signal reconstruction parts. The aforementioned abilities are not significantly different if deconvolution method is applied to reduce sidelobe levels. In this regard, separable arrays can be applied instead of irregular arrays. On the one hand, separable array allows KAT to accelerate the beamforming and deconvolution procedure; on the other hand, the rectangular geometry of the separable array is easier to be established. DAS is then extended to source signal reconstruction. SMS parameterized the DAS output spectrum, and the signals can be then generated separately in deterministic and broadband components.
This research introduces the possibility to combine beamforming and spectral modeling to synthesize moving sound sources. The results suggest that such synthesizer is validated for the specific case discussed. The optimized parameterization procedure needs further investigation with more simulations. Additionally, deconvolved beamforming output as the source spectrum is necessary to be included in this synthesizer. Since sound field produced by real moving sources is more complicated,

on-site measurements are necessary to verify the simulation results presented in this work. In addition, since the target sound source is auralized, it is essential to further the understanding using psychoacoustic analysis and listening test with the synthesized signals.

Acknowledgements

The authors acknowledge Bruno Masiero from University of Campinas for the cooperation design of the separable array.

References

[1] Vorlaender, M. *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2007.

[2] Meng, F; Wefers, F; Vorlaender, M. Moving Sound Source Simulation Using Beamforming and Spectral Modelling for Auralization. *DAGA, 42. JAHRESTAGUNG FÜR AKUSTIK*, Aachen, 14.-17. March 2016, In CD-ROM.

[3] Van Trees, H. L.; Harry, L. *Detection, estimation, and modulation theory, optimum array processing*. John Wiley & Sons, New York, USA, 2002.

[4] Brooks, T. F.; Humphreys, W. M. A deconvolution approach for the mapping of acoustic sources (DAMAS) determined from phased microphone arrays. *Journal of Sound and Vibration*, 294 (4), 2006, pp 856–879.

[5] Ribeiro, F. P.; Nascimento, V. H. Fast Transforms for Acoustic Imaging—Part I: Theory. *Image Processing, IEEE Transactions*. 20 (8), 2011, pp 2229–2240.

[6] Coelho, R.F., Nascimento, V.H., de Queiroz, R.L., Romano, J.M.T. and Cavalcante, C.C. *Signals and Images: Advances and Results in Speech, Estimation, Compression, Recognition, Filtering, and Processing*. CRC Press, 2015.

[7] Serra, X; Smith, J. Spectral modeling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition. *Computer Music Journal*, 1990, pp 12–24.

[8] Miner, N. E. *A Wavelet Approach to Synthesizing Perceptually Convincing Sounds for Virtual Environments and Multi-Media*, PhD dissertation. University of New Mexico, 1998.

[9] Morse, P. M.; Ingard, K. Uno. *Theoretical acoustics*. Princeton: Princeton University Press, 1968.

[10] Christensen, J.J; Hald, J. *Technical Review: Beamforming*. Bruel & Kjaer, 2004.

[11] Vertatschitsch, E.; Haykin, S. Nonredundant arrays. *Proceedings of the IEEE*, 74 (1), 1986, pp 217–218.

[12] Johnson, D. H.; Dudgeon, D. E. *Array signal processing: concepts and techniques*. Simon & Schuster, 1992.

[13] Ribeiro, F. P.; Nascimento, V. H. Fast Transforms for Acoustic Imaging—Part II: Applications. *Image Processing, IEEE Transactions*, 20 (8), 2011, pp 2241–2247.