On the Design of a MEMS Microphone Array for a Mobile Beamforming Application

Alexander Poets[1], Roman Schlieper[1], Stephan Preihs[1], Jürgen Peissig[1]

[1] *Leibniz Universität Hannover, Institut für Kommunikationstechnik, 30167 Hannover, Deutschland*

Email: alexander.poets@gmx.com, schlieper@ikt.uni-hannover.de, stephan.preihs@ikt.uni-hannover.de, peissig@ikt.uni-hannover.de

Abstract

Considering the extensive use of smartphones and their outstanding camera capabilities these days, it's desirable to record a high-quality audio signal in tandem with capturing video, e.g. at a rock concert. For a rather immersive sound experience when playing back smartphone-recorded material it's mandatory to preserve the acoustic environment's spatial properties. On that note, a design of a miniature microphone array to be rear-mounted on a smartphone is proposed. In the course of this work, we'll focus on the development of a proper microphone arrangement that meets application-specific requirements. Therefore, several geometries are evaluated in terms of directivity index, half-power beam width and side lobe attenuation using a simulation-driven approach. Once a suitable geometry has been found, the actual sensor placement is determined by thinning the array using a genetic algorithm. Then, a prototype of the design is implemented utilizing digital MEMS microphones. Finally, a spatial filter (beamformer) is applied to the raw sensor data in order to allow the capture of multi-channel audio and the suppression of interfering background noise.

Foundations of Array Modeling

Before being able to discuss the suitability of selected sensor arrangements, a framework for simulation-driven microphone array assessment needs to be proposed first. To start with, the far-field radiation pattern of an arbitrary homogenous transducer array can be obtained utilizing the principle of pattern multiplication, which is known from but not limited to the field of antenna engineering. Accordingly, an array's radiation pattern $B(\theta, \phi)$ is equal to that of one of its identical, same-oriented transducers $R(\theta, \phi)$ multiplied by the so-called array factor $AF(\theta, \phi)$:

$$B(\theta, \phi) = R(\theta, \phi) \cdot AF(\theta, \phi) \tag{1}$$

The array factor corresponds to the radiation pattern of a geometrically equivalent array, where the transducers are assumed to be isotropic, and is defined as follows:

$$AF(\theta, \phi) = \sum_{n=1}^{N} w_n^* \cdot \exp\left(-jk^{\mathsf{T}} p_n\right) \tag{2}$$

At each sensor position p_n, a signal caused by an impinging monochromatic plane wave with propagation vector

$$k = -\frac{2\pi}{\lambda} \begin{pmatrix} \cos\theta \cos\phi \\ \cos\theta \sin\phi \\ \sin\theta \end{pmatrix} \tag{3}$$

and wavelength λ is received and, before being summed, optionally phase-shifted using a complex weight w_n^*, which allows the array's main response axis to be electronically steered in a desired direction [1]. Due to reciprocity, the radiation pattern doesn't depend on whether the array is transmitting or receiving. The following convention for the specification of spherical coordinates has been used:

- The elevation coordinate $\theta \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ measures the angle between the vector and its orthogonal projection onto the xy-plane.
- The azimuth coordinate $\phi \in [-\pi, \pi]$ measures the angle between the positive x-axis and the vector's orthogonal projection onto the xy-plane.

From the radiation pattern, some properties that are suitable to quantify an array's performance may be derived. The directivity index (expressed in dBi) compares the radiant intensity on the main response axis with the average radiant intensity of an isotropic source that radiates the same total power:

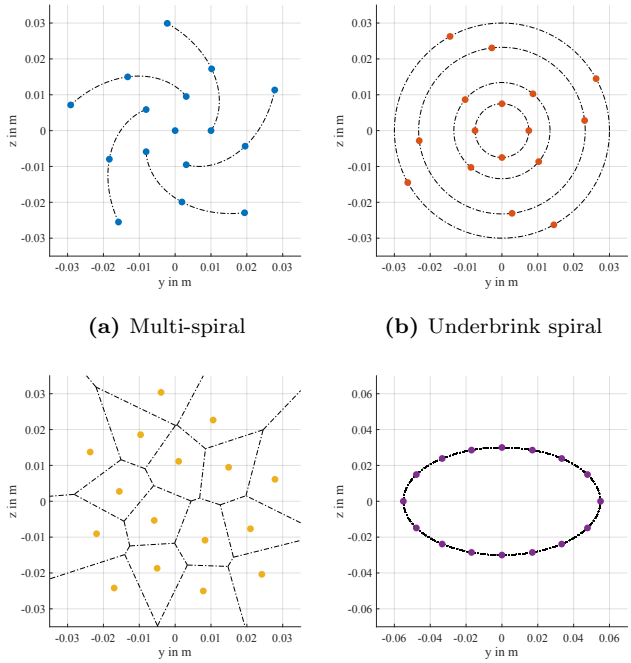
$$DI = 10 \cdot \log_{10}\left(\frac{|B(\theta_0, \phi_0)|^2}{\frac{1}{4\pi}\iint_{4\pi} |B(\theta, \phi)|^2 \cos\theta \, d\theta \, d\phi}\right) \tag{4}$$

As the directivity index indicates how much energy is directed towards the main response axis, a high value is preferable [2]. Another quantity that may be used for microphone array assessment is the half-power beam width, which measures the spatial extent of the main lobe at the approx. $-3\,\mathrm{dB}$ half-power point. Therefore, usually the angular distance between the azimuth coordinates at which the radiant intensity has dropped to half of its peak value is computed. A narrow main beam not only improves the array resolution but also increases the SNR of the array output [3]. Equally important to that effect is the side lobe attenuation, which expresses the difference in radiated power between the main beam and the peak side lobe. Since the side lobe-emitted energy is drawn from the main beam, it's desired to keep the power density across the side lobe region low [4]. With respect to above quantities, the performance of certain geometries is assessed and compared in the next section.

Evaluation of Array Designs

When designing any type of transducer array, there are a few characteristics to consider: the aperture resp. spatial extent of the array, the type and count of transducers as well as their physical orientation and arrangement. For this application, the array is to be rear-mounted on an iPhone 8/X, which results in a maximum aperture size of approx. 11 x 6 centimeters. Regarding the sensors, 16 of Infineon's IM69D130 digital MEMS microphones have been used, since more PDM devices cannot be interfaced with the xCORE-200 explorer kit from XMOS, which

served as USB audio adapter. Mounting the microphones on a finite baffle (the PCB, in this case) allows for either linear or planar array designs. With linear arrays, the resolution capability is limited to one axis, which is a major drawback considering the beam cannot be steered in both azimuth and elevation plane [1]. Hence, the following focuses on planar arrangements only. In microphone array design, ordering the sensors in a spiral-like manner has established. This is mainly due to two reasons: First, spirals naturally exhibit inter-element spacings void of redundancy and are therefore less prone to spatial aliasing. Second, their symmetric property ensures that for different steering directions a similar beam pattern is obtained [5]. There are several popular designs known from literature, among which the three most promising ones have been selected for further evaluation in the context of this application, namely the multi-, Underbrink and Vogel spiral. The multi-spiral (see Figure 1a) is constructed by taking the arm of a logarithmic spiral, on which the elements are spaced in uniform arc lengths, and rotating it equally about the origin. Placing the sensors equidistantly on concentric circles creates an Underbrink spiral as shown in Figure 1b [6]. Being inspired by a sunflower head's seed pattern, the Vogel spiral (see Figure 1c) constitutes a sophisticated design that yields Pareto optimal results in terms of half-power beam width and side lobe attenuation [4]. In addition to the spirals, a design that features a wider aperture and elliptically arranged elements is considered (see Figure 1d). Note that in all designs the sensors were placed in the yz-plane, so that the main response is obtained along the x-axis. Assessing the array performance now requires the computation of the respective reception patterns. The beam



(a) Multi-spiral **(b)** Underbrink spiral

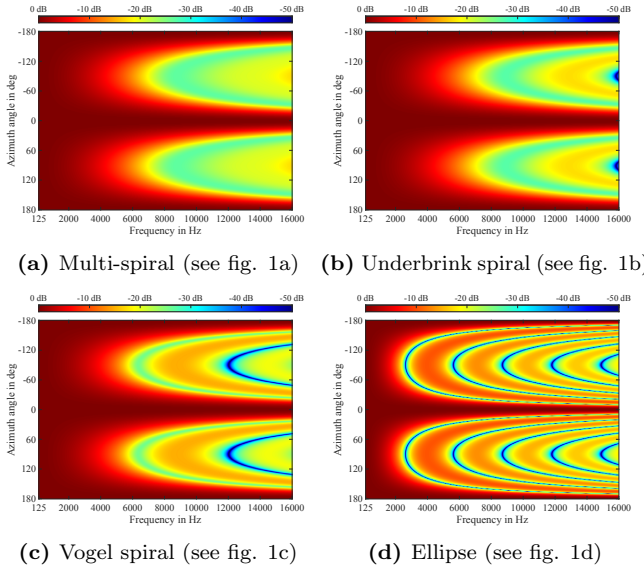
(c) Vogel spiral **(d)** Ellipse

Figure 1: These are the microphone arrangements taken into account for evaluation.

pattern of an array equipped with omnidirectional microphones like the IM69D130s just equals the array factor, as $R(\theta, \phi) = 1\ \forall \theta, \phi$ applies in that case (see Equation 1). For the sake of limiting complexity, the arrays were simulated under free field conditions, i.e. the finite baffle has not been taken into account. Further on, only the azimuthal reception pattern served as basis for the assessment: When filming with a smartphone, it is mostly held in landscape orientation, so that steering the beam in the horizontal (xy-)plane is of primary interest. Only on the rare occasion of filming in portrait orientation, the array is rotated by 90 degrees and thus the beam must be steered in the vertical (xz-)plane.

Simulation Results

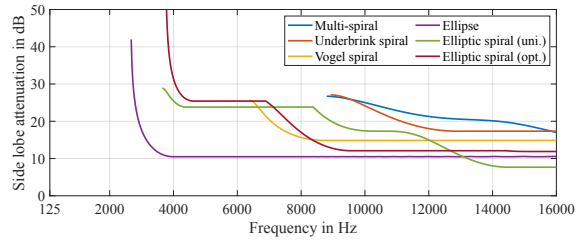
In the simulation, the speed of sound was assumed to equal $343.2\,\mathrm{m} \cdot \mathrm{s}^{-1}$, which holds true in dry air at 20 °C. Figure 2 shows the beam patterns of the respective microphone arrangements for the azimuth plane at zero degrees elevation and frequencies between 125 Hz and 16 000 Hz. Apparently, of the considered designs, the multi-spiral performs best in terms of side lobe attenuation. The elliptically shaped array yields by far the worst results in that respect, having a side lobe attenuation of only approx. 10 dB at 4 kHz. On the other hand, the ellipse has a much narrower main beam and a significantly better low frequency response than any of the spirals, which is due to its wider aperture. The spirals don't differ substantially from one another in beam width and show an almost omnidirectional response below 800 Hz. Since a wide main lobe is more sensitive to sound not coming directly from the steering direction, beamforming might not have as much of a directional effect as desired. Accordingly, the spirals might not be considered the best option in the context of this application. Figure 3 confirms the spirals' superiority in side lobe attenuation but also their rather poor performance in half-power beam width and directivity index, especially in the low frequency range.



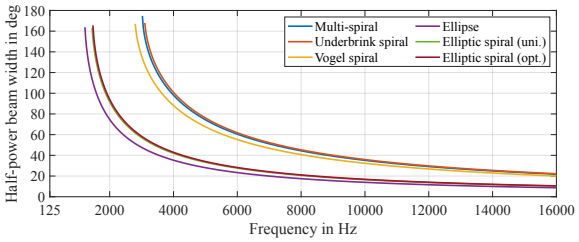
(a) Multi-spiral (see fig. 1a) **(b)** Underbrink spiral (see fig. 1b)

(c) Vogel spiral (see fig. 1c) **(d)** Ellipse (see fig. 1d)

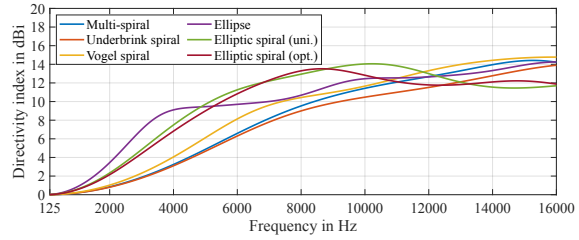
Figure 2: Azimuthal reception patterns of the designs shown in Figure 1 at zero degrees elevation plotted versus frequency.



(a) Azimuthal side lobe attenuation at 0° elevation



(b) Azimuthal half-power beam width at 0° elevation



(c) Directivity index

Figure 3: Measures to quantify the performance of the various array geometries considered throughout this work.

This may be explained using the Rayleigh criterion which states that an array's beam width is proportional to its angular resolution:

$$HPBW \propto \frac{\lambda}{B} = \frac{c}{f \cdot B} \tag{5}$$

It can be seen that with increasing wavelength λ and decreasing baseline B (maximum distance between two array elements), the beam widens and the angular resolution gets worse. In other words, the larger the array, the better the performance at low frequencies [3]. As the ellipse has a longer baseline of 11 cm compared to the spirals' short 6 cm, it outperforms them in half-power beam width. Considering the benefits of each geometry, the spirals' high side lobe attenuation and the ellipse's narrow beam resp. high directivity index, it seems reasonable to propose a combination of them in the next section.

The Elliptic Spiral

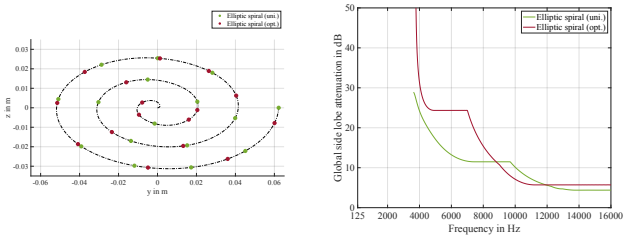
Deriving an elliptic spiral is accomplished by multiplying the polar equation of an ellipse by that of an Archimedes' spiral:

$$r = \frac{v\varphi \cdot ab}{\sqrt{b^2 \cos^2 \varphi + a^2 \sin^2 \varphi}} \tag{6}$$

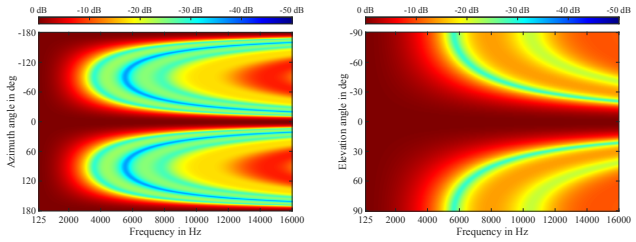
Choosing $v = 1/t2\pi$ and $\varphi \in [0, t2\pi]$ then yields an elliptic spiral of t turns with width $2a$ and height $2b$. In order to place N elements on its arm that are spaced a uniform arc length apart, it is required to know of the arm's total length. The arc length of a polar curve over the interval $0 \leq \varphi \leq \xi$ is given by [7]

$$L(\xi) = \int_0^{\xi} \sqrt{r^2 + \left(\frac{dr}{d\varphi}\right)^2}\, d\varphi \tag{7}$$

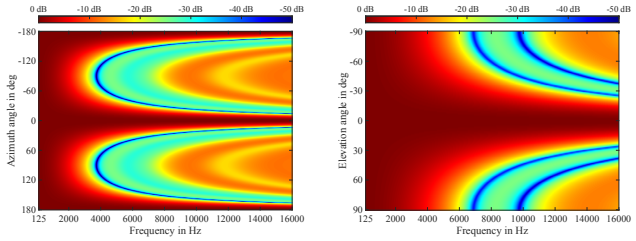
The next step is to determine the arc lengths at which each of the N elements is placed: $l_n = n \cdot L(t2\pi)/N$. Finally, the angles at which $L(\xi)$ takes the values of l_n are computed, i.e. solving $L(\xi_n) - l_n = 0$ for ξ_n. Using Equation 6, the radial coordinates corresponding to ξ_n may be calculated. An elliptic spiral that complies with the application's requirements ($N = 16$, $a = 0.055$, $b = 0.03$, $t = 3$ and $v = 0.06$) is shown in Figure 4a along with the reception patterns for both the azimuth and elevation plane (see Figure 4c/4d). As can be seen from Figure 3,



(a) Elliptic spirals (uni. & opt.) **(b)** Global side lobe attenuation



(c) Uni. spacing (az. plane) **(d)** Uni. spacing (el. plane)



(e) Opt. spacing (az. plane) **(f)** Opt. spacing (el. plane)

Figure 4: The elliptic spiral along with its reception patterns.

the elliptic spiral achieves values similar to those of the ellipse in directivity index and half-power beam width but without compromising as much on side lobe attenuation. Looking at the reception pattern in Figure 4d, it is evident that this performance doesn't apply to the elevation plane. In order to improve the design in terms of elevational side lobe attenuation while retaining that of the azimuth plane, an array thinning optimization method has been utilized, which is usually used to reduce the operating expenses of large antenna arrays. The underlying idea of the procedure is to systematically eliminate elements from the array without sacrificing the system's performance considerably [8]. In this scenario, however, the application differs a little: The intention was not to

reduce the element count but rather to seek for a proper arrangement of the microphones on the elliptic spiral's arm, so that an optimal trade-off between azimuthal and elevational side lobe attenuation is obtained. Instead of optimizing both objectives directly (and thereby doing a Pareto optimization), it is convenient to have only a single objective function, i.e. the global side lobe attenuation, which corresponds to the peak value of the largest side lobe in an array's three-dimensional reception pattern. Since this is a function of frequency, a value for which the array is optimized had to be set. The best results were yielded when minimizing the global side lobe attenuation for a frequency of 7 kHz. The overall approach can be summarized as follows: As a starting point, the positions of 1600 sensors that are equally spaced on an elliptic spiral's arm were calculated. Among these (which may be considered as a discretization of the search space) a subset of 16 positions that minimizes the global side lobe attenuation and violates none of the provided constraints was then identified. This may be classified as an integer programming problem with 16 variables, where each variable can take a value from 1 to 1600 denoting a possible sensor position's index. The problem was constrained by the following conditions:

- A solution may not include the same sensor position multiple times.
- The Euclidean distance between adjacent sensors must be at least 5 mm in order to ensure that a solution can be implemented in hardware.
- The azimuthal half-power beam width of a solution may not exceed 25° at 7 kHz so as not to sacrifice a narrow beam for low side lobes.

The optimization was performed using an evolutionary algorithm, namely the genetic algorithm (GA) solver that MATLAB provides with the Global Optimization Toolbox. A simple genetic representation to encode the candidate solutions is given by a bit string of length 1600, where each bit is associated with a possible sensor position. If a sensor position is included in a candidate solution, its corresponding bit is set to 1, and to 0 otherwise. The fitness of the candidate solutions was evaluated using the objective function and therefore matches the global side lobe attenuation. The solver provided the optimized sensor arrangement shown in Figure 4a. Considering the before/after comparison of the global side lobe attenuation (see Figure 4b), it is apparent that the optimized version of the elliptic spiral yields better results than the version with evenly distributed sensors, at least in the frequency range up to 9 kHz. The improvement in elevational side lobe attenuation can also be deduced from viewing the respective reception patterns in Figure 4d/4f. Further on, the optimization didn't cause much deterioration in half-power beam width and directivity index, as can be seen from Figure 3b and 3c.

Implementation

A prototype implementation of the optimized elliptic spiral array is shown in Figure 5. The device has been attached to the smartphone using two 3D-printed holding clamps. Being connected to the XMOS development board, the array can be accessed as 16-channel USB audio device in MATLAB. In order to utilize the array for stereo recordings, two beams are formed, one for each output channel. Therefore, a STFT is applied to the sensor data first. Then, in frequency domain, each DFT bin is multiplied by the complex weight of a narrowband beamformer (while retaining the Hermitian symmetry). Finally, the output signal is synthesized by taking the inverse STFT of the filtered frequency data. This procedure is commonly referred to as subband beamforming [1].

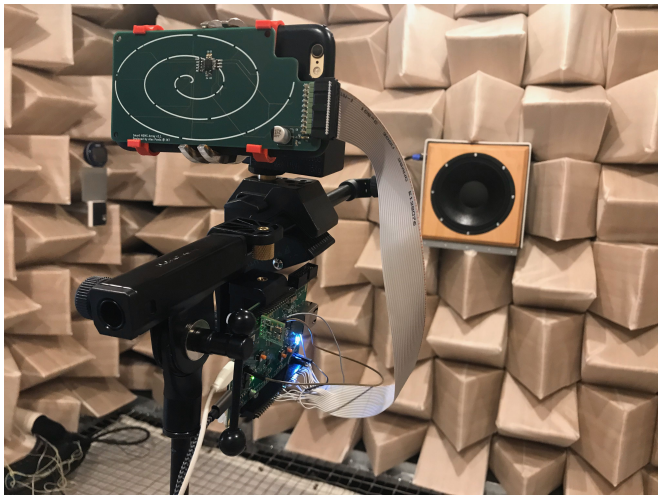


Figure 5: The prototype in the anechoic chamber of the IKT.

References

[1] Van Trees, H. L.: Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory. Wiley-Interscience, New York, 2002

[2] Beranek, L. and Mellow, T.: Acoustics: Sound Fields, Transducers and Vibration (Second Edition). Academic Press, London, 2019

[3] Grythe, J.: Evaluating array resolution. Norsonic technical note, Oslo, 2015

[4] Sarradj, E.: Optimal Planar Microphone Array Arrangements. Tagungsband DAGA 2015 – 41. Deutsche Jahrestagung für Akustik, 2015, 220-223

[5] Mortsiefer, C. and Peissig, J.: Design of a Ceiling-Microphone Array for Speech Applications with Focus on Transducer Arrangements and Beamforming Techniques. Tagungsband DAGA 2017 – 43. Deutsche Jahrestagung für Akustik, 2017, 544-547

[6] Prime, Z. and Doolan, C.: A Comparison of Popular Beamforming Arrays. Proceedings of the Annual Conference of the Australian Acoustical Society, 2013

[7] Arc Length – from Wolfram MathWorld, URL: `http://mathworld.wolfram.com/ArcLength.html`

[8] Jain, R. and Mani, G. S.: Solving "Antenna Array Thinning Problem" Using Genetic Algorithm. Applied Computational Intelligence and Soft Computing – Volume 2012, Hindawi Publishing Corporation, 2012