

# DESIGN OF A MICROPHONE ARRAY FOR NEAR-FIELD CONFERENCING APPLICATIONS

Pieter Thomas\*, Dick Botteldooren

Reinout Verburch, Michael Catrysse

research group WAVES  
Ghent University  
Technologiepark-Zwijnaarde 15  
B-9052 Gent, Belgium

Televic Conference  
Leo Bekaertlaan 1  
B-8870 Kachtem, Belgium

## ABSTRACT

Microphone arrays are becoming increasingly popular for conferencing applications and near-field speech recording. In this work, a 16-element cylindrical microphone array is designed for beamforming towards a nearby speaker, while reducing the influence of competing talkers. A two-stage approach is used to obtain the desired array directivity pattern, optimizing both microphone locations and filter weights. The positions of the microphones are optimized by using a hybrid optimization technique, taking into account the influence of the nearby acoustic environment. FIR filter coefficients for each microphone are derived from a regularized least-squares (LSQR) solution, combined with null-steering. The performance of the design is evaluated experimentally and compared with a classically used goose-neck microphone.

*Index Terms*— Microphone array design, sensor placement optimization, conferencing

## 1. INTRODUCTION

To date, gooseneck microphones are commonly used for speech enhancement and recording in conferencing environments. However, in order to minimize the influence of competing sources and maximize speech intelligibility, they need to be positioned close to the speaker's mouth. This is visually distracting the attention, while the microphone is frequently pulled about by the speaker.

With the development of low-cost high-quality digital MEMS microphones, the use of microphone arrays becomes a more feasible alternative [1]. Broadband microphone arrays allow beamforming towards a target speaker, while the signal from unwanted noise sources or competing talkers can be reduced, improving speech intelligibility [2]. Source localization techniques allow a speaker to be actively tracked, and when combined with adaptive beamsteering, source movements can be accounted for without the need to move the

microphone. Furthermore, the array can be positioned out of sight, while still guaranteeing good performance.

Generally, in microphone array design, focus is put on calculating the set of filter coefficients for each microphone, while microphone positions have been fixed a priori based on well-known geometries. It has been shown however that the array geometry has the main influence on array performance [3]. Regular microphone positions (e.g. linear, circular or planar arrays) have been studied extensively [2, 4] and their performance can be directly related to the number of microphones and some form factors. However, as stated by Yu et al. [3], irregular arrays with more diverse microphone positioning potentially achieve better performance over a broadband (speech) frequency range.

This paper discusses the design of a dedicated 16-element cylindrical microphone array for conferencing applications with focus on both microphone filter weights and microphone positioning. The optimization technique to find the most suited microphone positions is based on the hybrid descent optimization technique proposed by Feng et al. [5, 6]. In their work, however, only plane microphone configurations were optimized, while no influence of the acoustic environment was taken into account (free-field propagation). As for near-field beamforming, the influence of the environment cannot be neglected, an improved design methodology is proposed here. The methodology is based on a simplified acoustical model to reduce calculation cost. Once the microphone positions are known, suitable microphone FIR coefficients are derived from a least-squares solution based on more detailed numerical simulations of the acoustics [7, 8, 9], combined with an improved null-steering technique [4] to reduce the influence of nearby speakers even further.

## 2. MICROPHONE ARRAY DESIGN

The microphone array consists of a half-cylindrical shape on which the microphones are distributed. The array is placed in front of the speaker on a solid surface (conferencing desk).

\*Corresponding author: pieter.thomas@intec.ugent.be

The authors are grateful for the support by the Flanders Innovation and Entrepreneurship (VLAIO) agency, Project number IWT 150955.

Opting for a cylindrical microphone array increases spatial diversity, due to shielding and diffraction by the cylinder and the 3D location of the microphones, which are no longer confined to a single plane.

The response  $G(\mathbf{r}, f)$  of a microphone array with  $N$  microphones to a source at location  $\mathbf{r}$  can be written in the frequency domain as a linear combination of the received microphone signals:

$$G(\mathbf{r}, f) = \sum_{i=1}^N H_i(f) A_i(\mathbf{r}, f) = \mathbf{H}(f)^T \mathbf{A}(\mathbf{r}, f), \quad (1)$$

with  $H_i(f)$  the frequency dependent array microphone filters,  $A_i(\mathbf{r}, f)$  the transfer functions from a source located at position  $\mathbf{r}$  to the  $i$ -th microphone of the array,  $\mathbf{H}(f) = [H_1(f) \cdots H_N(f)]^T$  and  $\mathbf{A}(\mathbf{r}, f) = [A_1(\mathbf{r}, f) \cdots A_N(\mathbf{r}, f)]^T$ .  $\mathbf{A}(\mathbf{r}, f)$  is dependent on the microphone positions, the array shape and (nearby) acoustic environment.

The aim is to optimize both the microphone positions and array filters so that the resulting array response  $G(\mathbf{r}, f)$  approximates a desired response  $G_d(\mathbf{r}, f)$ . The desired response is defined over a discrete evaluation range  $(\mathbf{r}, f) \in \Omega = \Omega_p \cup \Omega_s$ . In this case, we write  $G_d(\mathbf{r}, f) = A \cdot e^{-j2\pi f \frac{\|\mathbf{r}-\mathbf{r}_c\|}{c}}$ , a combination of a pure delay with attenuation  $A = 1$  in the pass-band  $\Omega_p$  and  $A = 10^{-3}$  in the stop-band  $\Omega_s$ .  $\mathbf{r}_c$  is the centre of gravity of the microphone array positions.

In total, a set of  $M = 443$  discrete evaluation points  $\mathbf{r}$  are randomly chosen on a half spherical surface, centred around the centre of the array, with radius 0.8 m, the distance between the speaker and the array centre. A higher point density is chosen around the location of the speaker ( $A = 1$ ; 38 points  $\mathbf{r} \in \Omega_p$ ). The remaining 405 points are part of  $\Omega_s$  ( $A = 10^{-3}$ ) with a slightly smaller inter-point distance between 20 cm and 40 cm around the speaker to ensure a fast decay of the sensitivity pattern for noise sources close to the speaker. At each of the evaluation points 60 frequencies  $f$  are chosen, logarithmically spaced between 30 Hz and 8 kHz to cover the frequency range of speech (sample frequency  $f_s = 16$  kHz is assumed).

## 2.1. Microphone position optimization

In a first stage, the microphone positions are optimized by minimizing the error between the resulting array response and the desired response as a function of the  $N$  microphone locations  $\mathbf{r}_{\text{mic}} = [\mathbf{r}_1 \cdots \mathbf{r}_N]$ :

$$E(\mathbf{r}_{\text{mic}}) = \sum_{(\mathbf{r}, f) \in \Omega} \|G(\mathbf{r}, f) - G_d(\mathbf{r}, f)\|^2. \quad (2)$$

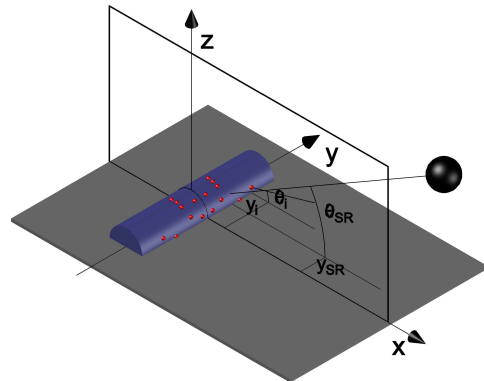
For each iteration of the optimization algorithm, this error function needs to be evaluated. Consequently the transfer functions  $A_i(\mathbf{r}, f)$  need to be recalculated for the specific microphone positions under investigation. The cost of this calculation strongly depends on the complexity of the acoustic

environment. To reduce the calculation time in the first stage, it is therefore decided to use a simplified analytical model of the nearby environment as a substitute for extensive numerical simulations. Here, the half-cylindrical microphone array on the desk is modelled as an infinite half-cylinder, positioned on a hard infinite plane. This model only accounts for the near-field acoustics, but neglects the influence of diffractions at the edges of the desk and array, the room and speaker. However, the advantage of this approximation is that the acoustic response to a spherical sound source located at distance  $d$  from the cylinder axis can be expressed analytically [10]. In cylindrical coordinates, the pressure at a point  $(\rho, \delta\theta, \delta y)$ , with  $\rho$  the distance to the cylinder axis,  $\delta\theta$  the difference in elevation between source and receiver and  $\delta y$  the distance between source and receiver projected on the cylinder axis, can be written as:

$$\begin{aligned} p_{tot}(\rho, \delta\theta, \delta y) &= p_{inc}(\rho, \delta\theta, \delta y) + p_{scat}(\rho, \delta\theta, \delta y) \\ &= \int_{-\infty}^{+\infty} e^{-j\lambda\delta y} \sum_{n=0}^{\infty} \frac{j(-1)^n \epsilon_n H_n^{(2)}(\sqrt{k^2 - \lambda^2}d)}{2} \cos n\delta\theta \\ &\quad \times \left( J_n(\sqrt{k^2 - \lambda^2}\rho) + H_n^{(2)}(\sqrt{k^2 - \lambda^2}\rho) C_n(\lambda) \right) d\lambda, \end{aligned} \quad (3)$$

with  $J_n(\cdot)$  the Bessel function of the first kind of order  $n$ ,  $H_n^{(2)}(\cdot)$  the Hankel function of the second kind of order  $n$  and  $C_n(\lambda) = -\frac{nJ_n(\sqrt{k^2 - \lambda^2}a) - \sqrt{k^2 - \lambda^2}aJ_{n+1}(\sqrt{k^2 - \lambda^2}a)}{nH_n^{(2)}(\sqrt{k^2 - \lambda^2}a) - \sqrt{k^2 - \lambda^2}aH_{n+1}^{(2)}(\sqrt{k^2 - \lambda^2}a)}$ , with  $a$  the radius of the hard reflecting cylindrical surface,  $k = 2\pi f/c$  the wavenumber and  $\epsilon_n$  the Neumann factor.

The total transfer function  $A_i(\mathbf{r}, f)$  between source (speaker) and receiver ( $i$ -th microphone) on the half-cylindrical surface located on the desk is then written as  $A_i((a, \theta_i, y_i), f) = p_{tot}(a, |\theta_i - \theta_{SR}|, |y_i - y_{SR}|) + p_{tot}(a, |\theta_i + \theta_{SR}|, |y_i - y_{SR}|)$ , with  $\theta_i$  and  $\theta_{SR}$  the elevation of the  $i$ -th microphone and source with respect to the infinite reflecting plane and  $y_i$  and  $y_{SR}$  the distance along the cylinder axis, referred to the centre of the cylinder (see Figure 1).



**Fig. 1.** Cylindrical microphone array with optimized microphone positions indicated in red ( $N = 16$ ).

The optimization of microphone positions is equal to finding the  $2N + 1$  unknowns  $\theta_i$ ,  $y_i$  and  $a$  so that Equation 2 is minimized. However, the error between target and wanted sensitivity pattern is also function of the chosen microphone weights  $H_i(f)$ . Here, the weights are not explicitly included into the microphone positioning optimization process, but calculated by using the analytical solution of a regularized least-squares problem, based on the calculated transfer functions  $A_i(\mathbf{r}, f)$  between microphones and points  $(\mathbf{r}_m, f)$  of the evaluation range  $\Omega$  [8]:

$$\mathbf{H}(f) = \frac{\mathbf{A}^H(f)}{\mathbf{A}^H(f)\mathbf{A}(f) + \mu\mathbf{I}_N} \mathbf{G}_d(f), \quad (4)$$

with  $\mathbf{A}(f) = [\mathbf{A}(\mathbf{r}_1, f) \cdots \mathbf{A}(\mathbf{r}_M, f)]^T$  of size  $M \times N$ ,  $\mathbf{G}_d(f) = [\mathbf{G}_d(\mathbf{r}_1, f) \cdots \mathbf{G}_d(\mathbf{r}_M, f)]^T$  of size  $M \times 1$  and  $\mu$  a regularization parameter. Here,  $\mu = 10^{-5}$  is taken.

A hybrid descent optimization methodology is chosen to minimize Equation 2 [5]. This is an iterative technique, where in a first step a genetic algorithm is applied. The resulting 7 best unique solutions are then used as starting points for a gradient-based method to reach the nearby (local) minimum. In a next iteration, the unique solutions are included in the start population of the genetic algorithm, and the process is restarted until convergence is reached. By using a genetic algorithm, points close to the global minimum are reached, while the application of a gradient-based method assures that the absolute minimum is found. Additional constraints are put on the microphone positions to make sure they are restricted within the physical dimensions of the cylinder (cylinder length is 40 cm), while distances between microphones should be greater than 1.5 cm for ease of construction. To decrease convergence time, the number of unknowns is reduced by choosing a symmetrical microphone positioning to pursue a sensitivity pattern symmetrical around the speaker in front of the array i.e.  $y_i = -y_{N-i+1}$  and  $\theta_i = \theta_{N-i+1}$ . Furthermore, the radius  $a$  was also fixed to the highest allowable value (5 cm), maximally exploiting the effect of spatial diversity.

The result of the optimization process is given in Figure 1, where the locations of the microphones are indicated on the cylindrical surface.

## 2.2. Microphone filter weights optimization

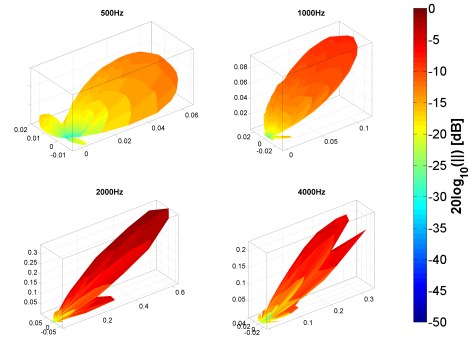
In the second stage, the finiteness of the design (diffraction at edges) and non-simplified acoustic environment (shape of desk) is taken into account by optimizing the design of the microphone FIR filters  $\mathbf{H}(f)$ . Calculation of the FIR coefficients is based on the least-squares approach, Equation 4, where the different transfer functions  $\mathbf{A}(f)$  are calculated by using a 3D numerical solver (Actran - Free Field Technologies), modelling the real environment. Here,  $\mu = 2$  is chosen to reduce the weight of the filter coefficients and shorten the FIR filter length.

In order to further reduce the influence of competing speakers, which are uniformly spaced (distance 0.9 m) in this specific conferencing environment, the solution of the least-squares problem is combined with a null steering algorithm [4]. The idea is to minimize the error between the sensitivity pattern  $G(\mathbf{r}, f) = \mathbf{H}(f)^T \mathbf{A}(\mathbf{r}, f)$  and an approximation  $G'(\mathbf{r}, f) = \mathbf{H}'(f)^T \mathbf{A}(\mathbf{r}, f)$ , subject to null constraints for certain directions  $\mathbf{r}_j$ :  $\mathbf{H}'(f)^T \mathbf{A}(\mathbf{r}_j, f) = 0$  for  $j = 1 \cdots K$ . The solution of this constrained optimization problem can be written as

$$\mathbf{H}'(f) = \mathbf{H}(f) \left( \mathbf{I}_N - \mathbf{C} [\mathbf{C}^H \mathbf{C}]^{-1} \mathbf{C}^H \right), \quad (5)$$

with  $\mathbf{C} = [\mathbf{A}(\mathbf{r}_1, f) \cdots \mathbf{A}(\mathbf{r}_K, f)]$  the constraint matrix. This solution can be interpreted as the component of the filter coefficients  $\mathbf{H}(f)$  in the subspace orthogonal to the constraint subspace. In this paper, 9 constraints are used: six nulls are steered towards the first, second and third speaker position on each side of the target speaker, and three nulls are steered to the back, to competing speakers on rows in front of the main speaker. Solving Equation 5 requires inversion of  $\mathbf{C}^H \mathbf{C}$ . However, for low frequencies, columns of  $\mathbf{C}^H \mathbf{C}$  can become highly correlated, making inversion ill-defined. Therefore, a regularization parameter  $\mu_0$  is introduced and the inversion is calculated as  $[\mathbf{C}^H \mathbf{C} + \mu_0 \mathbf{I}_K]^{-1}$ . Here, we put  $\mu_0 = 1$ .

In a last step, the  $L$ -taps FIR filters  $h_i(t)$  are estimated by taking the inverse FFT of the  $L$  filter coefficients, calculated at linearly spaced frequencies. Values at these frequencies are extracted by interpolating the amplitude and unwrapped phase of  $\mathbf{H}'(f)$ , calculated at the 60 logarithmically spaced frequency points. In the following, a filter length of  $L = 128$  taps is seen to be sufficiently long to allow good performance, while keeping time-delay to a minimum.



**Fig. 2.** Microphone array directivity pattern at 500 Hz, 1 kHz, 2 kHz and 4 kHz. The axes of the balloon plots give the  $x$ ,  $y$  and  $z$  component of  $|G'(\mathbf{r}, f)|^2$ , while the color scales with  $10 \log_{10} |G'(\mathbf{r}, f)|^2$ .

Figure 2 shows the resulting sensitivity patterns for the 16-element microphone array, evaluated at 80 cm from the array centre. As can be seen, the beam is aimed at the speaker

location, but the magnitude at the target location decreases towards the lower frequencies (high-pass filtering). At frequencies below 1 kHz, spatial selectivity is hard to achieve given the restricted dimensions of the array. The optimization algorithm tries to minimize the error (Equation 2) at as many evaluation points as possible - which are chosen more numerous in the stopband than in the passband region. As a consequence, at lower frequencies, more effort is put into optimizing the stopband conditions than the passband conditions.

Additionally, the null steering algorithm introduces zeros towards competing talkers. At high frequencies, the constraint matrix is highly uncorrelated, and all null directions can be accounted for. At the lowest frequencies (500 Hz and lower), columns of the constraint matrix become correlated, and some nulls cannot be achieved, while the introduction of nulls at certain directions gives rise to higher sidelobe levels.

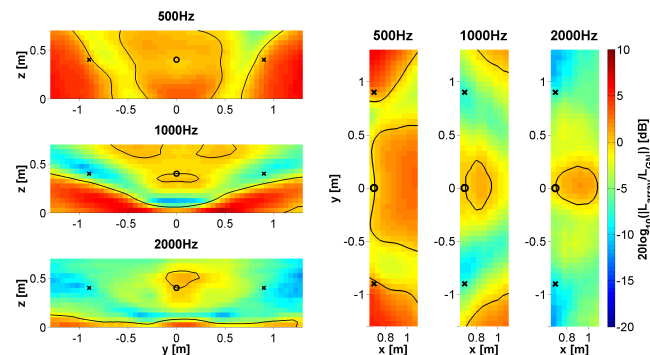
### 3. ARRAY - GOOSE-NECK PERFORMANCE

The aim of the research was to develop a microphone array with better or similar directional sensitivity than a classically used goose-neck microphone. To check the design performance, a prototype of the array was created with 16 Knowles Acoustics microphones (type FG-23329) and high-quality data-acquisition system (NI PXIe-1082 chassis with three NI-4498DAQ cards). The sensitivity of the prototype and the goose-neck microphone was measured by using a dedicated measurement set-up, which allowed us to automatically move a source over a 3D grid behind the conference desk, placed in an anechoic chamber (cf. Figure 3). A B&K HATS Type 4128C with mouth simulator was used as source. With this set-up, the sensitivity of both the array and goose-neck microphone was measured for a 5 cm grid with lower corner (0.65 m, -1.3 m, 0 m) and (1.1 m, 1.3 m, 0.7 m) as upper corner.



**Fig. 3.** Automated measurement set-up for the sensitivity comparison between microphone array and goose-neck microphone.

Figure 4 indicates the level difference (in octave band levels) between the microphone array and goose-neck microphone. Both have been normalized with the level at the target direction. Localized around the target location, a positive level difference is seen, indicating that the array induces an amplification of the source, compared to the goose-neck microphone. At larger distances from the target source and especially near competing talkers, a level reduction is seen (negative difference). For low frequency octave bands however, the band of level reduction around competing talkers narrows, and some competing sources are no longer attenuated (or even amplified), compared to the goose-neck microphone.



(a) YZ section at  $X = 0.7$  m. (b) XY section at  $Z = 0.4$  m.

**Fig. 4.** YZ and XY cross-sections of the level difference between the microphone array and the goose-neck microphone levels. Results are given for the 500 Hz, 1 kHz and 2 kHz octave bands. The black line indicates the contour with 0 dB difference. The circle indicates the target position, the cross gives the position of the competing talker.

### 4. CONCLUSION

In this work a half-cylindrical 16-element microphone array has been designed for conferencing applications, taking the nearby acoustic environment into account. An array design technique has been proposed, optimizing both positions and FIR filters, by using a hybrid descent optimization algorithm combined with a regularized least-squares solution and null-steering technique to further reduce competing sources. In a first stage, positions have been optimized by using a model of reduced complexity of the nearby environment, while in a second stage, optimization of the FIR coefficients is based on more accurate numerical simulations of the source-microphone transfer functions. Results have been compared with a classically used goose-neck microphone and an improvement in directional characteristics is seen for frequencies in the speech range. Spectral characteristics (flatness) of the array however still require equalization.

Future work may focus on how the technique can be used for the design of more complex array shapes. The use of array thinning techniques might be an option.

## 5. REFERENCES

- [1] I. Hafizovic, C. Nilsen, M. Kjolerbakken, and V. Jahr, "Design and implementation of a MEMS microphone array system for real-time speech acquisition," *Applied Acoustics*, vol. 73, pp. 132–143, 2012.
- [2] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*, Springer-Verlag, 2008.
- [3] J. Yu and K. Donohue, "Geometry descriptors of irregular microphone array related to beamforming performance," *EURASIP Journal on Advances in Signal Processing*, vol. 2012, no. 249, 2012.
- [4] H. Van Trees, *Optimum Array Processing*, Wiley-Interscience, 2002.
- [5] Z. Li, K. Yiu, and Z. Feng, "A hybrid descent method with genetic algorithm for microphone array placement design," *Applied Soft Computing*, vol. 13, pp. 1486–1490, 2013.
- [6] Z. Feng, K. Yiu, and S. Nordholm, "Placement design of microphone array in near-field broadband beamformers," in *IEEE transactions on signal processing*, 2012, vol. 60.
- [7] P. Thomas, W. Wei, T. Van Renterghem, and D. Botteldooren, "Measurement-based auralization method for the assessment of noise mitigation measures," *Journal of Sound and Vibration*, vol. 379, pp. 232–244, 2016.
- [8] P. Thomas, T. Van Renterghem, and D. Botteldooren, "Array-based HRTF pattern emulation for auralization of 3D outdoor sound environments with direction based muffling of sources," in *Proceedings of the 139th International AES Convention*, 2015.
- [9] E. Rasumow, M. Blau, S. Doclo, M. Hansen, S. Van de Par, D. Puschel, and V. Mellert, "Least squares versus non-linear cost functions for a virtual artificial head," in *Proceedings of Meetings on Acoustics*, 2013.
- [10] T. Li and M. Ueda, "Sound scattering of a spherical wave incident on a cylinder," *Journal of the Acoustical Society of America*, vol. 87, no. 5, pp. 1871–1879, 1990.