

Robustness of virtual artificial head topologies with respect to microphone positioning

Eugen Rasumow, Matthias Blau, Martin Hansen
Institut für Hörtechnik und Audiologie, Jade Hochschule, 26121 Oldenburg, Germany

Simon Doclo, Steven van de Par, Volker Mellert
Carl von Ossietzky Universität Oldenburg, 26121 Oldenburg, Germany

Dirk Püschel
Akustik Technologie Göttingen, 37073 Göttingen, Germany

Summary

A multimicrophone array is presented that can be used to approximate the frequency dependent directional characteristics of an artificial head. The desired HRTFs can be realised by a set of appropriate filters. Such a setup may be referred to as a virtual artificial head. Virtual artificial heads are much more flexible than real artificial heads, since, e.g., the filters can be adjusted to match an individual set of HRTFs. However, virtual artificial heads are sensitive to small errors in the characteristics and the position of the individual microphones. In the present work, the relevance of different microphone topologies and the robustness with respect to positioning errors of the microphones is investigated. First, a method for optimizing the microphone positions based on a Golomb array topology, which originates from number-theoretic considerations, is introduced. The method successively computes a set of microphone positions with the possibility to vary the number of microphones without changing the general topology. Second, the robustness against positioning errors is improved by applying a white noise gain regularisation constraint for the computation of the filter coefficients. It is shown by numerical simulations, for a two dimensional array, that both procedures considerably improve the robustness.

PACS no. 43.60.Fg, 43.66.Pn

Introduction

The use of so-called artificial heads, which are a replica of real human heads, is common practice today. Alternatively, the desired frequency-dependent beam pattern of human head related transfer functions (HRTFs) can also be approximated by a microphone array with appropriate filters (cf. for instance [5]). Such a setup may be referred to as a virtual artificial head (VAH). The resulting directivity pattern of the VAH does not only depend on the filter coefficients but also on the number and the topology of the microphones used in the array. Thus, microphone positioning is a crucial step in the setup of a VAH. In this paper a method which originates from number-theoretic considerations is introduced for a planar array and compared to traditional methods with respect to accuracy and robustness.

Calculation of filter coefficients $\underline{\mathbf{w}}(\omega)$

A given (two-dimensional) directivity pattern $D(\omega, \Theta_i)$ depends on the frequency ω and direction of arrival Θ_i . The analytical derivation of the N-dimensional filter coefficients $\underline{\mathbf{w}}(\omega)$, with N the number of microphones, can e.g. be performed by minimising a least-squares cost function that connects the desired directivity $D(\omega, \Theta_i)$ to the resulting directivity $H(\omega, \Theta_i)$. The resulting directivity $H(\omega, \Theta_i)$ in turn depends on the steering vector $\underline{\mathbf{d}}(\omega, \Theta_i)$ of the microphones and the filter coefficients $\underline{\mathbf{w}}(\omega)$,

$$H(\omega, \Theta_i) = \underline{\mathbf{w}}^H(\omega) \underline{\mathbf{d}}(\omega, \Theta_i). \quad (1)$$

The steering vector $\underline{\mathbf{d}}(\omega, \Theta_i)$ was derived analytically assuming omnidirectional microphones (i.e. the microphones do not alter the directivity of adjacent microphones) and far-field sound propagation.

In general, a chosen cost function can either be minimised for all frequencies simultaneously (broadband optimisation) or separately for each frequency bin

(narrowband optimisation). In the following computations the narrowband least-squares cost function

$$J(\underline{\mathbf{w}}(\omega)) = \sum_{i=1}^P |H(\omega, \Theta_i) - D(\omega, \Theta_i)|^2 \quad (2)$$

was used, where the parameter P represents the number of discrete directions Θ_i . In [1] it was shown that the minimisation of equation (2) with regard to $\underline{\mathbf{w}}(\omega)$ can be computed by

$$\underline{\mathbf{w}}(\omega) = \underline{\mathbf{Q}}^{-1}(\omega) \underline{\mathbf{a}}(\omega), \text{ with} \quad (3)$$

$$\underline{\mathbf{Q}}(\omega) = \sum_{i=1}^P \underline{\mathbf{d}}(\omega, \Theta_i) \underline{\mathbf{d}}^H(\omega, \Theta_i) \quad (4)$$

$$\underline{\mathbf{a}}(\omega) = \sum_{i=1}^P \underline{\mathbf{d}}(\omega, \Theta_i) D^*(\omega, \Theta_i). \quad (5)$$

Random Sampling method

The resulting directivity $H(\omega, \Theta_i)$ of the VAH depends on the desired directivity $D(\omega, \Theta_i)$ and the number and the steering vector of the microphones. If one assumes independent omnidirectional microphones (which is a fair assumption for an array with small electret microphones), the steering vector $\underline{\mathbf{d}}(\omega, \Theta_i)$ depends only on the relative position of the microphones.

Assuming further there is a predetermined grid of possible microphone positions, the best microphone topology could be determined using a brute force approach, by computing a set of filter coefficients and the resulting $H(\omega, \Theta_i)$ for each possible combination of microphone positions and then choosing the one with the lowest value of the cost function. This eventually would lead to the optimal microphone topology, but even with a moderate grid density, the time required to probe all possible combinations of microphone positions is prohibitive.

Alternatively, one could generate topologies at random instead of systematically probing each possible topology. This will be referred to as the ‘‘Random Sampling’’ method. In this study, $K = 10000$ randomly chosen topologies on a $x \times y$ plane within a $10\text{cm} \times 10\text{cm}$ plane were compared, and the one with the lowest least-squares error was finally selected. The number of $K = 10000$ different topologies was chosen as a reasonable compromise between all possible topologies and computational cost.

Golomb method

The microphone topology or, more specifically, the inter-microphone distance determines the frequency-dependent beamwidth. A good base for an optimal broadband fit to the frequency- and angle dependent directivity pattern $D(\omega, \Theta_i)$ would be a topology with

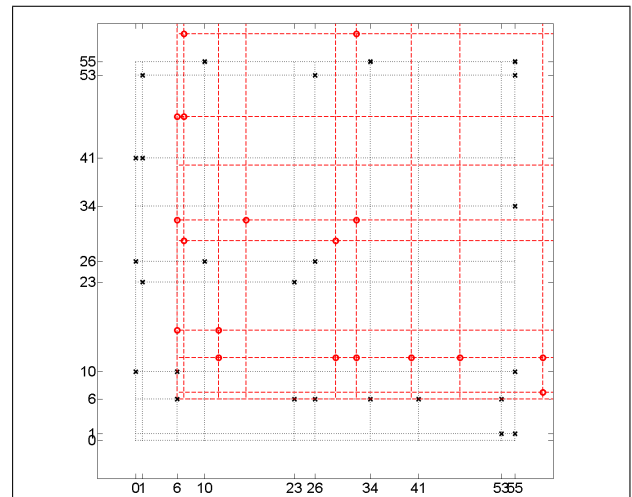


Figure 1. Procedure to find a set of legitimate nodes according to [2]. The dotted black lines indicate a two-dimensional grid originating from a Golomb-ruler with order $M = 10$. Black crosses indicate chosen nodes and the red lines show a shifted copy of the chosen nodes.

as many as possible different inter-microphone distances in all possible directions. Since the number of microphones is limited, the topology of choice must present a compromise between frequency- and directional accuracy. A topology that fulfills the requirement of as many as possible inter-microphone distances in one dimension is the so-called Golomb ruler. A Golomb-ruler is a numerical series that does not contain any inter-mark distance more than once. A possible method to expand this scale to a two- or three-dimensional grid was proposed in [2]. The procedure for determining a set of valid nodes according to [2] can be subdivided into seven steps:

1. Create a grid using the same one-dimensional Golomb-ruler in both dimensions (see gray structure in figure 1)
2. Select a first reference-node on the grid arbitrarily
3. Select a second test-node on the grid randomly
4. Copy and shift the topology along both dimensions by all possible distances (see red structure in figure 1)
5. Check condition: Shifted nodes must not coincide with more than one of the previously selected nodes within each shift
6. Keep the test-node as an additional reference node if (5) is true. Repeat from (3) if (5) is false.
7. Repeat this procedure until all additional nodes were tested or a desired number of valid nodes has been reached.

Since the reference- and test-nodes are chosen randomly, the position and number of valid nodes may vary with each new run of the method.

In general, the Golomb method is less computationally expensive compared to the Random Sampling method.

Regularisation

When the filter coefficients $\mathbf{w}(\omega)$ are computed using eq. 3, it is known that very small changes of the microphone characteristics (gain/phase) and the microphone positions can drastically deteriorate the resulting directivity pattern $H(\omega, \Theta_i)$ (cf. [1]). Even if the microphone characteristics are known very accurately (e.g. using a proper calibration), there could still be changes in the position of the microphones, e.g. due to environmental parameters. In order to make the procedure more robust, regularisation techniques can be employed. A well-known regularisation technique is the so-called diagonal loading (a special case of Tikhonov regularisation, see for instance [4]), in which the cost function in 2 is replaced by

$$J_r(\mathbf{w}(\omega)) = J(\mathbf{w}(\omega)) + \mu \underbrace{\mathbf{w}^H(\omega, \mu) \cdot \mathbf{w}(\omega, \mu)}_{WNG(\omega, \mu)} \quad (6)$$

where $WNG(\omega, \mu)$ represent the so-called white noise gain. The filter minimising the cost function in 6 is given by

$$\mathbf{w}(\omega, \mu) = \left(\underline{\underline{\mathbf{Q}}}(\omega) + \mu \mathbf{I}_N \right)^{-1} \cdot \mathbf{a}(\omega), \quad (7)$$

where \mathbf{I}_N represents the $N \times N$ -dimensional identity matrix. The regularisation parameter μ has to be carefully chosen, such that on the one hand it provides robustness against microphone position errors and on the other hand the resulting directivity pattern does not strongly deviate from the desired directivity pattern. A common way to choose an appropriate μ is to impose a constraint onto the white noise gain, i.e. $WNG(\omega, \mu) \leq \beta$. The parameter μ is derived by minimising $(WNG(\omega, \mu) - \beta)^2$, see for instance [3]. β , in contrast, has to be chosen manually according to the expected error of the steering vector. In general, β should be proportional to the desired accuracy and reciprocal to the resulting robustness. This means that lower β enhance the robustness at the expense of accuracy and vice versa.

Numerical Simulation

The HRTF of subject 1005 from the IRCAM database¹ was used as the desired directivity $D(\omega, \Theta_i)$ pattern in the horizontal plane to a hypothetical microphone array with $N = 4, 6, 8, 10, 12, 14, 16, 18, 19, 20, 22, 23, 24$ microphones using both introduced methods for microphone positioning. In order to compare the

accuracy of the positioning methods, the error parameter $E(\text{CB})$ was chosen, with

$$E(\text{CB}) = \frac{1}{P} \sum_{i=1}^P \left| 20 \lg (|\delta(\Theta_i)|) \text{ dB} \right| \quad (8)$$

$$\delta(\Theta_i) = \frac{1}{M} \sum_{\omega=\omega_1}^{\omega_M} \frac{H(\omega, \Theta_i)}{D(\omega, \Theta_i)}. \quad (9)$$

Here, $\delta(\Theta_i)$ (eq. 9) is the ratio of the resulting $H(\omega, \Theta_i)$ to the desired $D(\omega, \Theta_i)$ directivity, averaged in critical bands (CB) with 50% overlap. ω_1 and ω_M indicate the lowest and highest frequency bin within each CB. To further simplify matters, the absolute dB-value of $\delta(\Theta_i)$ is averaged over all P discrete directions of arrival, resulting in $E(\text{CB})$.

As described before, even small positioning errors can cause huge deviations in the resulting directivity (cf. for instance [1]). In order to evaluate the effect of positioning errors, a uniformly distributed random shift in x - and in y - direction, both with maximal absolute values of $|\Delta x_{max}| = |\Delta y_{max}| = 0.1\text{mm}$, was superimposed onto the true microphone positions. The resulting errors $E(\text{CB})$ with and without positioning errors are shown in fig. 2.

Within the application of a VAH, it seems advantageous to vary the number of microphones with frequencies since the directivity pattern is rather smooth at lower and rather peaky at higher frequencies. Thus it is of interest to investigate the minimal error $E(\text{CB})$ per critical band while the microphone number is assumed variable. This characterisation will be referred to as the best-case scenario, shown in fig. 3.

Results and discussion

Effect of positioning errors

The resulting errors when using the Random Sampling method both with unbiased and randomly corrupted steering vectors are shown in the left column of fig. 2. Using the unbiased steering vector (top row in fig. 2), the resulting error is relatively small (≈ 1 dB), except at higher frequencies ($f \geq 3$ kHz) and smaller microphone numbers ($N \leq 12$). This phenomenon is reasonable, since the directivity pattern is peakier at higher frequencies, requiring more microphones to adapt to it.

When a randomised positioning error with an amplitude of $|\Delta x_{max}|, |\Delta y_{max}| \leq 0.1\text{mm}$ is introduced, the error deteriorates drastically (middle row in fig. 2), especially for frequencies $500 \text{ Hz} \leq f \leq 5000 \text{ Hz}$ and microphone numbers $N \geq 12$. These errors clearly show that robustness needs to be improved more when many microphones are used.

The deterioration of accuracy due to positioning errors is primarily limited to lower frequencies, since a bias in the phase response (which is the main effect of

¹ The HRTF-database is available on <http://recherche.ircam.fr>

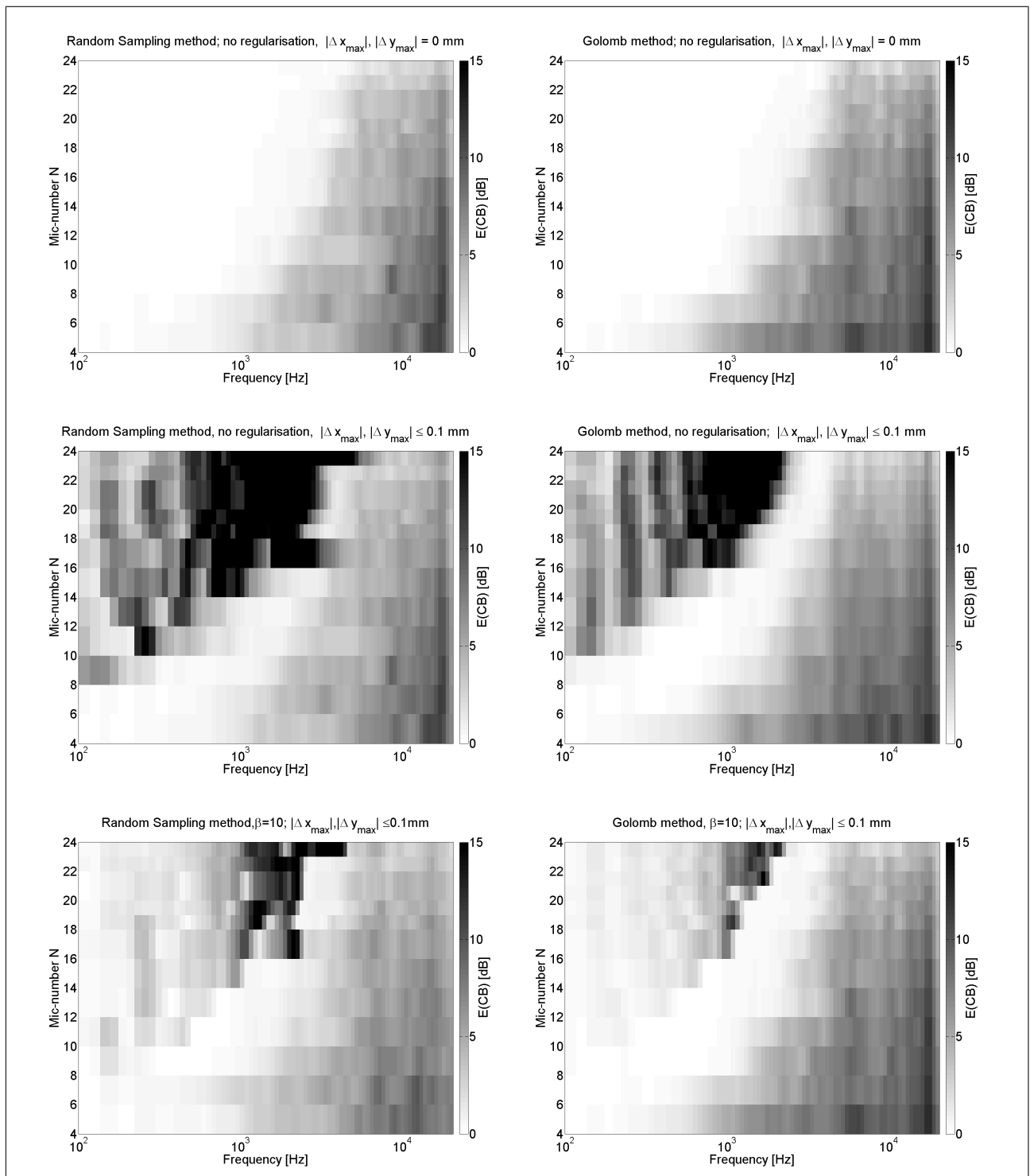


Figure 2. Simulated accuracy $E(CB)$ (eq. 8) of the virtual artificial head, as a function of the number of microphones in the array. The left column shows results for the Random Sampling method, the right column for the Golomb method. **Top row:** no artificial positioning error, no regularisation, **middle row:** random positioning error $|\Delta x_{max}|, |\Delta y_{max}| \leq 0.1\text{mm}$, no regularisation, **bottom row:** random positioning error $|\Delta x_{max}|, |\Delta y_{max}| \leq 0.1\text{mm}$, regularisation with $\beta = 10$.

positioning errors) is more critical at lower frequencies.

The resulting errors using the Golomb method with unbiased and corrupted steering vectors are depicted

in the right column of fig. 2. The error obtained with the unbiased steering vector is approximately equal to that obtained with the Random Sampling method (top row in fig. 2). Thus, the Golomb method only achieves a small advantage (lower computing costs)

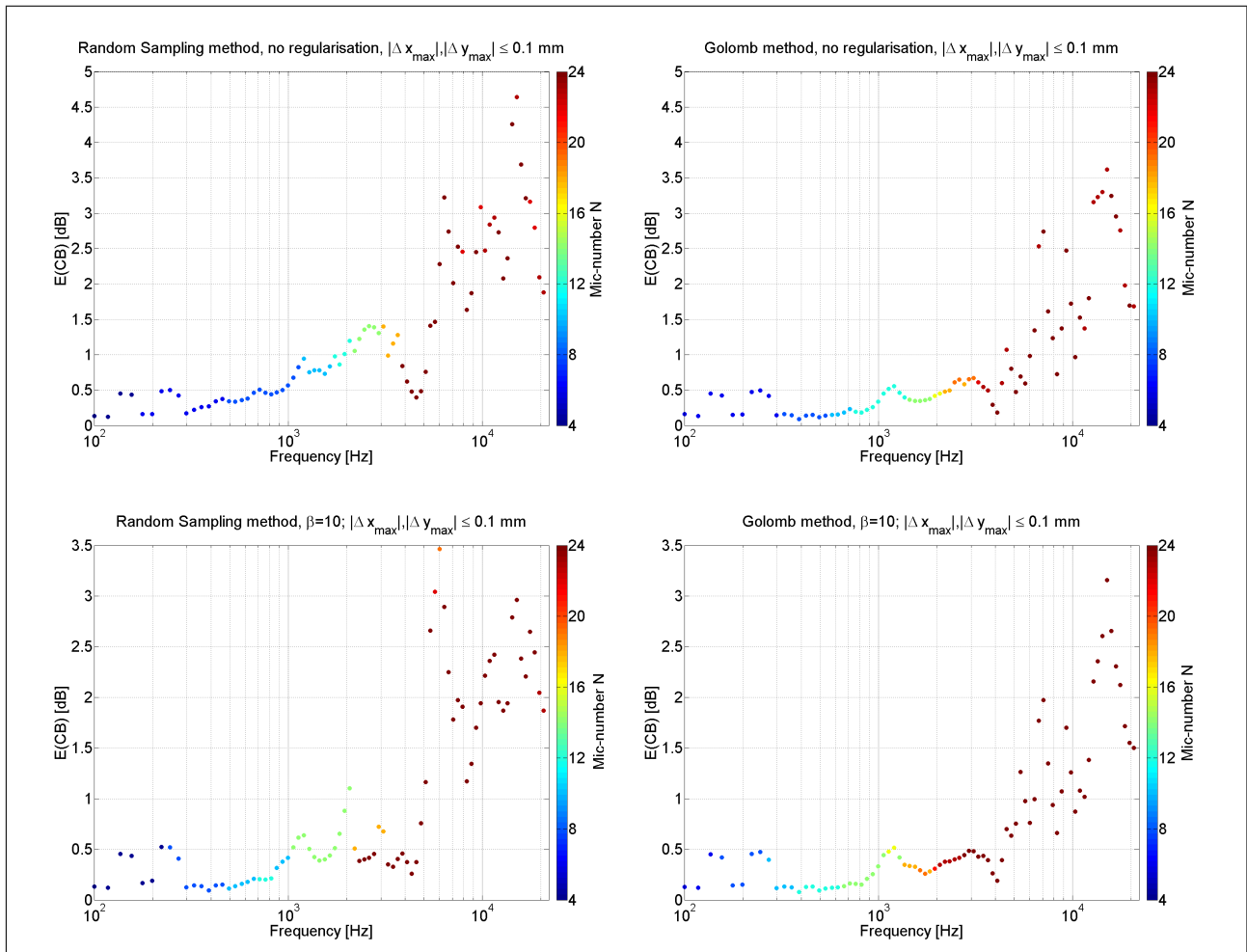


Figure 3. Simulated accuracy $E(\text{CB})$ (eq. 8) of the virtual artificial head, for optimal numbers of microphones per critical band. The left column shows results for the Random Sampling method, the right column for the Golomb method. **Top row:** random positioning error $|\Delta x_{\max}|, |\Delta y_{\max}| \leq 0.1\text{ mm}$, no regularisation, **bottom row:** random positioning error $|\Delta x_{\max}|, |\Delta y_{\max}| \leq 0.1\text{ mm}$, regularisation with $\beta = 10$. Note that the ordinate scale is smaller than in the top row. The microphone-number that leads to each minimal error is indicated by the color.

compared to the Random Sampling method when it is reasonable to assume a perfectly unbiased steering vector.

However, the errors of both positioning methods become much larger when the steering vector is assumed to have a random positioning error of $|\Delta x_{\max}|, |\Delta y_{\max}| \leq 0.1\text{ mm}$ (middle row in fig. 2). In this case, the directivity pattern using the Golomb method seems to be more robust compared to the directivity pattern using the Random Sampling method, according to the error metric $E(\text{CB})$. More specifically, the error is smaller with the Golomb method at frequencies $500\text{ Hz} \leq f \leq 5000\text{ Hz}$ and microphone numbers $N \geq 12$. These regions appear to be particularly affected by positioning errors. Moreover, the Golomb method enables a wider frequency range with smaller error.

However, the Golomb method seems to be disadvantageous for frequencies $f \geq 5\text{ kHz}$ when only a few microphones ($N \leq 12$) are used. This points to the

fact that the Golomb topology with $N \leq 12$ has an unfavourable spatial distribution. In fact the spatial distribution of the Golomb topology was only monitored for $N = 24$. On that score the error using the Golomb method probably would have been smaller if the spatial distribution of the topology would have been monitored for each single microphone number successively.

Effect of regularisation

A further improvement can be achieved by applying the regularisation as in equation 6 and 7, see bottom row in fig. 2. It can be seen that regularisation reduces the error, especially for microphone numbers $N \geq 12$. Relating to a positioning-error of $|\Delta x_{\max}|, |\Delta y_{\max}| \leq 0.1\text{ mm}$, the parameter $\beta = 10$ seems to be appropriate since this regularisation improves robustness and does not deteriorate accuracy for higher frequencies considerably. With greater positioning errors, robustness may be improved by using

smaller β . Therefore, β should be adjusted for the expected errors, to optimise the trade-off between robustness and accuracy.

Best-case scenarios: optimal number of microphones per critical band

As could be seen above, the optimal number of microphones depends on the frequency band: it is lower at low frequencies and higher at high frequencies. This feature can be exploited in a VAH, by choosing the optimal number of microphones in each frequency band. This best-case scenario is illustrated in fig. 3.

Again, errors differ only slightly between both positioning methods when no positioning error is assumed (not shown here).

When a positioning error of $|\Delta x_{max}|, |\Delta y_{max}| \leq 0.1\text{mm}$ is applied, the best-case error is smaller using the Golomb method for the entire frequency range. Especially at frequencies $f \geq 1000$ Hz, the Golomb-method seems to yield topologies that enable smaller errors within the fitting process. As could be expected, the best-case error becomes smaller when the regularisation with $\beta = 10$ is applied using the same positioning error of $|\Delta x_{max}|, |\Delta y_{max}| \leq 0.1\text{mm}$.

It should also be noted that only the topologies derived with the Golomb method enable to achieve the depicted minimal error with a fixed setup (by setting corresponding filter coefficients to zero), because the small-number topologies are a subset of the large-number topologies.

Conclusions

A method for optimising the positioning of microphones for a virtual artificial head (VAH), based on a two-dimensional expansion of a Golomb ruler, is presented. This method is less computationally expensive compared with heuristic methods. Topologies derived using the Golomb method in general enable a smaller error compared to the Random Sampling method in general. A further advantage of the Golomb method for the depicted application within a VAH is the variable microphone number for a fixed setup.

The robustness of the performance can be considerably improved applying a white noise gain constraint when positioning errors occur. The crucial parameter for this regularisation β needs to be optimised for the expected positioning error individually.

Acknowledgement

This project was partially funded by Bundesministerium für Bildung und Forschung under grant no. 17080X10.

References

- [1] **Doclo S. & Moonen M.**: Design of broadband beamformers robust against gain and phase errors in the microphone array characteristics. *IEEE Transactions on signal processing*, 51(10):2511-2526, October 2003.
- [2] **DE 10 2010 012 388 A1** Sender- bzw. Sensoranordnung zur Erzielung optimierter Richtcharakteristik für Sende- oder Empfangsvorrichtungen in zwei und drei Dimensionen
- [3] **Dörbecker M.** Mehrkanalige Signalverarbeitung zur Verbesserung akustisch gestörter Sprachsignale am Beispiel elektronischer Hörhilfen, PhD thesis, no. 10 in *Aachener Beiträge zu Digitalen Nachrichtensystemen (ABDN)* (Vary, P., ed.), Verlag Mainz in Aachen, 1998.
- [4] **Cox H., Zeskind R., & Kooij T.** Practical supergain. *IEEE Trans. Acoust., Speech, Signal Processing*, 34(3):393-398, June 1986.
- [5] **Mellert, V. & Tohtuyeva, N.**, 1997. Multi-microphone arrangement as substitute for dummyhead recording technique. In *Proc. 137th ASA Meeting*. S. 3117.