TRIGONOMETRIC INTERPOLATION BEAMFORMING FOR A CIRCULAR MICROPHONE ARRAY

Christian Schüldt

Google Inc., Kungsbron 2, Stockholm, Sweden

ABSTRACT

Polynomial beamforming has previously been proposed for addressing the non-trivial problem of integrating acoustic echo cancellation with adaptive microphone beamforming. This paper demonstrates a design example for a circular array where traditional polynomial beamforming approaches exhibit severe (over 10 dB) directivity index (DI) oscillations at the edges of the design interval, leading to severe DI degradation for certain look directions. A solution, based on trigonometric interpolation, is proposed that stabilizes the oscillations significantly, resulting in a DI that deviates only about 1 dB from that of a fixed beamformer over all look directions.

Index Terms— Polynomial beamforming, trigonometric interpolation, circular microphone array, acoustic echo cancellation

1. INTRODUCTION

Systems for hands-free speech communication typically comprise multiple signal processing components, such as acoustic echo cancellation (AEC), microphone beamforming, non-linear processing, noise reduction, comfort noise, etc. [1, 2] that have to function together in order to ensure excellent audio quality. Combining AEC and adaptive microphone beamforming traditionally means either beamforming first, followed by echo cancellation, or echo cancellation first followed by beamforming [3]. Both approaches have pros and cons; beamforming first (followed by one AEC operating on the beamformed signal) is advantageous from a computational complexity perspective, but requires the AEC to re-adapt whenever the beamformer changes focus. AEC-first does on the other hand mean that the AEC operates independently of the beamformer (i.e., no need to re-adapt the AEC when the beamformer changes look-direction), but also requires high computational load (one AEC per microphone), and possibly mitigated beamformer performance during double-talk and/or echo-path change if the different AEC filters are not equally converged, as uncorrelated mismatch between sensors are amplified by the white noise gain (WNG) [4]. (A method to reduce the computational complexity of multi-microphone AECs based on relative transfer functions has recently been presented [5], but has the implication of requiring small inter-microphone distance.)

A way of solving the beamforming-/AEC-first problem was introduced by Hamalainen & Myllyla [6], by integrating AEC with polynomial beamforming [7], which essentially splits up the beamforming into two parts; a fixed filter-and-sum part, followed by a polynomial postfilter. As the beamsteering is performed by the polynomial postfilter, an AEC can be placed after the filter-and-sum-part, but before the postfilter, and will not be affected by a changed look-direction. This method has been extended by Mabande et al. [8, 9] by re-formulating the beamforming filter design as a constrained convex optimization problem, based on an approach originally presented in [10]. The method has also been extended to 2-D [11]. One of the main advantages of the design approach in [10] is that the WNG, which essentially is a measure of beamformer robustness, is added as a design parameter through a quadratic constraint in the optimization. Since the WNG is such an important parameter in practical microphone array design, this is an advantage over other similar design methods such as e.g., differential microphone arrays [12], where the WNG must be controlled indirectly by other parameters such as model order and array size. Also, a general advantage is that the approach requires less computational complexity compared to data-dependent beamformers such as e.g., the MVDR beamformer, which typically involves a matrix inverse or eigenvalue decomposition [13] (of the estimated noise pseudo-coherence matrix), as well as not having to rely on estimated signal statistics.

However, one well known issue with polynomial interpolation in general (not directly tied to beamforming) is Runge's phenomenon; a problem of oscillation at the edges of an interval that occurs for polynomials of high degree over a set of equispaced interpolation points. In this paper, it is shown that the phenomenon is also present in polynomial beamforming, and that it can severely degrade the performance. A way of handling this in some sense for symmetrical arrays is presented in [9], where array symmetry is used to allow reduction of the interpolation range, while mirroring the beamforming filters around some symmetry plane. (This technique of mirroring filters has been used for other beamformer designs as well [14].) Unfortunately, mirroring the beamforming filters does not mean fewer number of signals for the AECs to process. (See section 2.1 in this paper for a more in-depth discussion on this.)

This paper introduces the concept of trigonometric interpolation as a solution to the previously described problems of polynomial beamforming. It is demonstrated that in a design example where the interpolation based beamforming approach in [8] exhibits multiple issues and strong directivity index (DI) oscillations, the proposed trigonometric interpolation performs much better in terms of stable DI over all look directions.

2. TRIGONOMETRIC INTERPOLATION BEAMFORMING

Consider an array of M microphones with a planar wave arriving from an angle ϕ. The output of a P-th order trigonometric interpolation beamformer, for frequency ω in look-direction θ, can be expressed as

$$B_\theta(\omega, \phi) = \sum_{p=0}^{P} t_p(\theta) \sum_{m=1}^{M} W_{m,p}(\omega) g_m(\omega, \phi), \qquad (1)$$

where $t_p(\theta)$ is a trigonometric basis function according to

$$t_p(\theta) = \begin{cases} 1, & \text{for } p = 0 \\ \cos(\lceil \frac{p}{2} \rceil \theta), & \text{for odd values of } p > 0 \\ \sin(\lceil \frac{p}{2} \rceil \theta), & \text{for even values of } p > 0, \end{cases} \qquad (2)$$

where $\lceil \cdot \rceil$ denotes the ceiling, $W_{m,p}(\omega)$ denotes the fixed beamforming coefficients, and $g_m(\omega, \phi)$ is the m-th sensor response to the incoming planar wave.

It should be noted that the beamformer formulated in (1) bears much resemblance to the polynomial beamforming described in e.g., [8], with the difference being the basis functions. Considering a circular array with equidistant microphones, the geometric symmetry and periodicity suggests that the choice of a trigonometric basis is suitable. Moreover, regardless of array geometry, steering angle periodicity is captured by the trigonometric basis functions since $t_p(\theta) = t_p(\theta + 2\pi n) \quad \forall n \in \mathbb{Z}$.

Formulating (1) as a convex optimization problem can be done in the same way as for the polynomial beamformer [8]. This analysis is very briefly described below for the sake of completeness. (The interested reader is referred to [8] for details.)

The goal is to jointly optimize, in a least square sense, the beamformer responses for I different look-directions θ_i, $i = 1, \ldots, I$. In contrast to polynomial beamforming [8, 10, 11], where an interpolation factor between -1 and 1 (representing some angular range) is used to freely steer the beam in any direction, here the angle θ is used for interpolation "as-is". Hence, the fundamental idea is that even though the beamformer is optimized for a fixed set of look directions θ_i, the beam will be steerable in all directions.

Numerical optimization is used, meaning that both the frequency range, and the angular range are discretized into $q = 1, \ldots, Q$ frequencies and $k = 1, \cdots, K$ angles, respectively. The optimization problem for frequency ω_q can then be written as [8]

$$\underset{\boldsymbol{w}_{\mathrm{f}}(\omega_q)}{\operatorname{argmin}} \sum_{i=1}^{I} ||\boldsymbol{G}(\omega_q)\boldsymbol{P}_i \boldsymbol{w}_{\mathrm{f}}(\omega_q) - \mathbf{b}_{\mathrm{des},i}||_2^2, \qquad (3)$$

subject to the constraints on the distortionless response and WNG according to

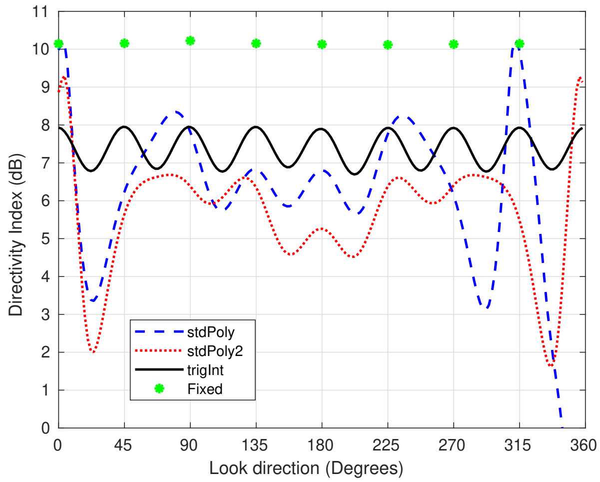
$$\boldsymbol{\alpha}_i^\top(\omega_q)\boldsymbol{P}_i \boldsymbol{w}_{\mathrm{f}}(\omega_q) = 1, \ \frac{|\boldsymbol{\alpha}_i^\top(\omega_q)\boldsymbol{P}_i \boldsymbol{w}_{\mathrm{f}}(\omega_q)|^2}{||\boldsymbol{P}_i \boldsymbol{w}_{\mathrm{f}}(\omega_q)||_2^2} \geq \gamma \qquad (4)$$

$\forall i = 1, \ldots, I$. In (3) and (4), the PM length vector $\boldsymbol{w}_{\mathrm{f}}(\omega_q)$ contains all fixed beamforming filter coefficients $W_{m,p}$, the $K \times M$ matrix $[\boldsymbol{G}(\omega_q)]_{km} = g_m(\omega_q, \phi_k)$ contains the sensor responses over the K angles and across the M sensors, the $M \times M(P{+}1)$ matrix $\boldsymbol{P}_i = \boldsymbol{I}_M \otimes [t_0(\theta_i), t_1(\theta_i), \cdots, t_P(\theta_i)]$ consists of the trigonometric basis function factors, the K length vector $\mathbf{b}_{\mathrm{des},i}$ consists of the desired responses for the different angles, while the M length vector $\boldsymbol{\alpha}_i = [g_1(\omega_q, \phi_{\mathrm{des},i}), g_2(\omega_q, \phi_{\mathrm{des},i}), \cdots, g_M(\omega_q, \phi_{\mathrm{des},i})]$ holds the sensor responses for the i-th look-direction. The Matrix $\boldsymbol{I}_M$ is an $M \times M$ identity matrix, $\otimes$ denotes the Kronecker product, and the design parameter γ is the WNG limit. The problem as given in (3) and (4) is convex [8], and can be solved with a convex optimization software package such as e.g., CVX [15].

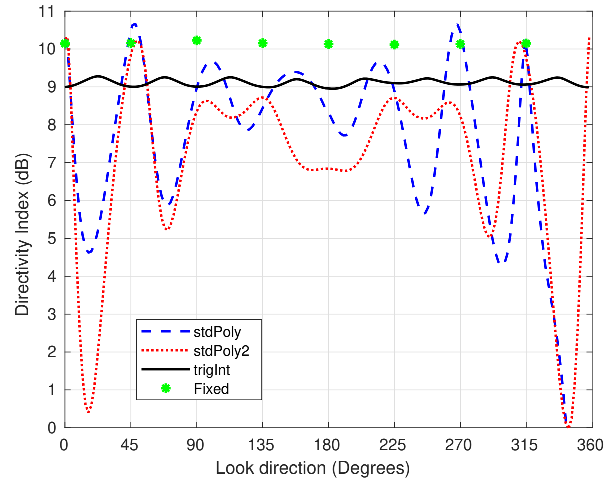
2.1. Integrating echo cancellation

The beamformer given in (1) can be seen as consisting of two parts; one fixed filter-and-sum part (the inner sum over M), and one polynomial postfilter (the outer sum over P). The fixed-filter-and-sum part does not depend on the look direction θ, which means that echo cancellation can be done after this stage and still be unaffected by a changed look direction. As a contrast to having one AEC per microphone (i.e., M AECs), it can be seen that if integrating the AECs after the filter-and-sum part of the beamforming, but before the polynomial postfiltering, the number of required AECs are $P + 1$. Hence, if $P + 1 < M$, a reduction in computational complexity is obtained. (See e.g., [6] for more details.)

As a contrast, the symmetry exploiting method presented in [9] (and [14] for another beamformer design), which basically constitutes a mirroring of the beamforming filters about some symmetry plane(s), does not reduce the number of required AECs. The reason for this is that even though a smaller set of beamforming filters can be used, the filters still need to be mirrored and applied to *different* microphone signals, producing in total a larger set of output signals that *each require their own* AEC. For example, consider the case of a circular



(a) $P = 4$



(b) $P = 6$

Fig. 1. Comparison of the resulting directivity index for the methods stdPoly, stdPoly2 and trigInt (proposed) for two different settings of P and the $M = 8$ microphone circular array described in the text.

array with $M = 8$ uniformly spaced microphones, and assume that symmetry around two axes are used, meaning that interpolation is performed only between 0 and 90 (and the set of beamforming filters are mirrored around both axes to allow 360 degree pick-up). If choosing $P = 2$, i.e., 2-nd degree polynomial interpolation, a total of 3 basis filters are required, meaning 3 AECs for this range (i.e., 0 - 90 degrees). Mirroring the filters effectively means that the same filter coefficients are applied to microphone signals with different index. However, since this produces another set of output signals, another set of associated AECs are required in order to allow continuous echo cancellation for all 360 degrees. Hence, for $P = 2$ (meaning 3 AECs) and symmetry around two axes, this means a total of 12 AECs, which is more than the number of microphones in the array in this example. Another design example, taken from [9]; using $M = 6$, $P = 3$ and exploiting symmetry to reduce the interpolation range by a factor of $2M$ turns out to require $P + 1 \times 2M = 48$ AECs.

3. SIMULATIONS

An $M = 8$ microphone circular equidistant array with 38 mm radius was simulated. The optimization problem given in (3) and (4) was solved for $I = 8$ desired look-directions $\{0, 45, 90, \cdots, 315\}$ degrees, with the frequency range discretized into $Q = 128$ frequency bins between 0 and 8 kHz and the angles discretized into $K = 360$ bins. The main lobe of the desired frequency response was defined with a beamwidth of 16 degrees. Two settings of P, i.e., the number of trigonometric basis functions, were considered $P = 4$ and $P = 6$. (Note that P should be even due to how $t_p(\theta)$ is defined, see (2).) This approach is denoted *trigInt* in the following text.

As a comparison, the same simulation was done for the polynomial interpolation approach as described in [8], here denoted *stdPoly*. Due to the non-periodic nature of the polynomial interpolation, i.e., the inability to model that $\theta_i = 0$ degrees and $\theta_i = 360$ degrees represent the same direction-of-arrival, simulations was also conducted for $I = 9$, where the first 8 desired look directions were as described previously, and the 9-th look direction was 360 degrees. This approach is denoted *stdPoly2* in the following text.

For all beamformers, the WNG constraint was set to -10 dB, and the sampling frequency was set to 16 kHz.

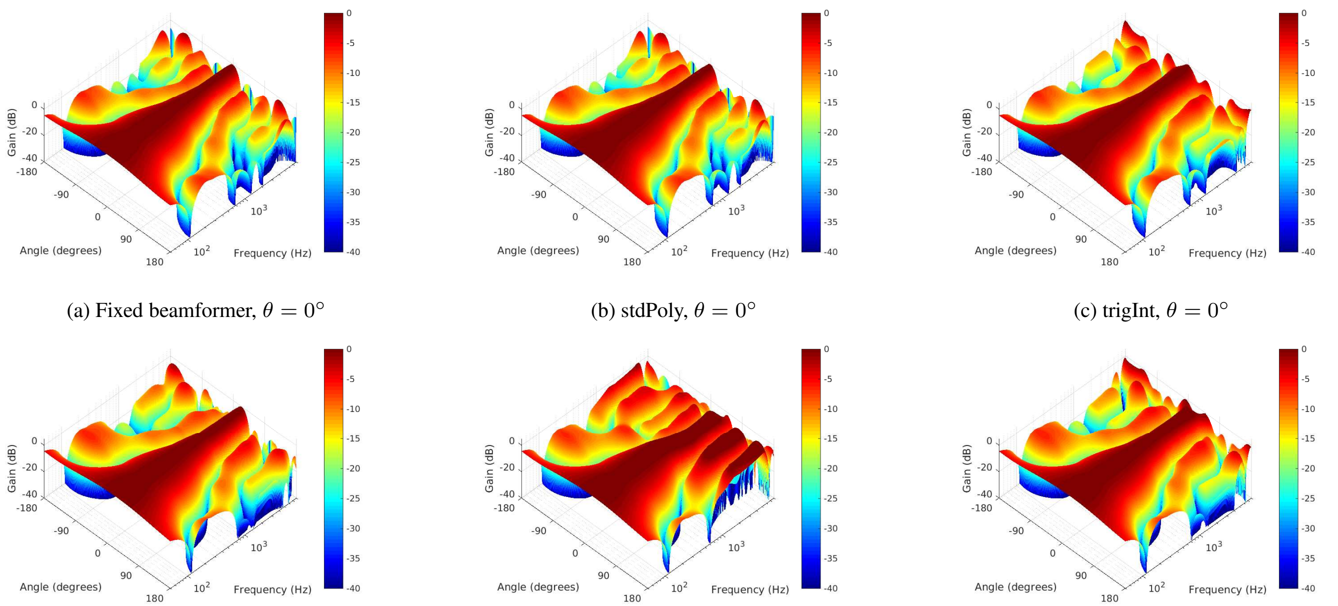
3.1. Directivity index

To evaluate the performance, the directivity index (DI), here calculated as

$$DI(\theta) = 20 \log_{10} \frac{K - L}{QL} \sum_{\omega} \frac{\sum_l |B_\theta(\omega, \theta_l)|}{\sum_{k \neq l} |B_\theta(\omega, \phi_k)|}, \qquad (5)$$

where $l = 1, \cdots, L$ and L is the discrete number of look direction bins, was evaluated for a uniform set of look-directions θ between 0 and 360 degrees, with 2 degrees angular resolution. It should be noted that the mismatch between the interpolated beamformer frequency response and the fixed beamformer frequency response is *not* quantified in this paper. The motivation for this is that e.g., variations in stop-band response is irrelevant as long as a similar level of attenuation is maintained, which is something that is captured by the DI.

Fig. 1 shows the resulting DI for the three compared methods; trigInt (proposed), stdPoly and stdPoly2, for design parameter P set to 4 (Fig. 1a) and 6 (Fig. 1b), respectively. Also



(a) Fixed beamformer, $\theta = 0°$

(b) stdPoly, $\theta = 0°$

(c) trigInt, $\theta = 0°$

(d) Fixed beamformer, $\theta = 25°$

(e) stdPoly, $\theta = 25°$

(f) trigInt, $\theta = 25°$

Fig. 2. Beamformer responses of the three compared methods; stdPoly, stdPoly2 and trigInt (proposed) for two different look-directions: 0 degrees (one of the look-directions used in the optimization) and 25 degrees (*not* used in the optimization). Polynomial order $P = 6$ is used in all cases.

shown in the figure, for reference, are the resulting DIs of a fixed beamformer optimized for one specific look-direction according to the approach described in [10], for all θ_i. It can clearly be seen that the proposed method achieves a much more stable DI over all look directions, while both stdPoly and stdPoly2 exhibit severe oscillation at the edges of the interval, i.e., Runge's phenomenon, which is a well-known problem for high order polynomial fitting. Also worth noting is that the DI of stdPoly totally breaks down for look directions close to 360 degrees due to the interpolation method not modeling the periodicity, as explained previously. This issue is avoided by adding a 9-th desired look direction at 360 degrees, i.e., stdPoly2, but at the cost of even stronger DI oscillation close to 0 degrees. Also worth noting is the oscillations being stronger for trigInt2 with higher order ($P = 6$ compared to $P = 4$), which is consistent with what can be expected from standard polynomial interpolation. The DI of the trigInt approach on the other hand is significantly more stable and much closer to the DI of the fixed beamformer optimized for one specific look direction. In fact, for $P = 6$, the DI of trigInt is only about 1 dB lower than that of the fixed beamformer for all look-directions.

3.2. Beamformer responses

In order to better visualize the beamformer performance, a number of responses for different look directions θ are shown in Fig. 2. Fig. 2a), b) and c) show the beamformer responses of the fixed beamformer, stdPoly and trigInt, respectively, for look direction $\theta = 0$ degrees and $P = 6$. It can be seen that the beamformer response of stdPoly looks almost identical to that of the fixed beamformer, while the response of trigInt appears slightly different at higher frequencies ($>$ 1 - 2 kHz). The DI is however not significantly different (as shown in Fig. 1).

Fig. 2d), e) and f) show the beamformer responses of look direction $\theta = 25$ degrees. For this look direction, stdPoly exhibits significant performance degradation at higher frequencies, which results in a significant DI reduction (again, as shown in Fig. 1). The proposed method trigInt, on the other hand, has a response that is much more similar to that of the fixed beamformer, with the exception of some degradation very close to the Nyquist frequency.

4. CONCLUSIONS

This paper has presented a novel approach to interpolation for a circular array, based on trigonometric interpolation. It was shown that while the conventional polynomial beamforming suffers from multiple issues; not modeling the steering angle periodicity and the strong DI oscillation at the edges of the design interval, the proposed approach does not. For a specific design with 8 microphones and a 6-th order trigonometric interpolation beamformer, it was shown that the DI deviates only about 1 dB from that of a fixed beamformer over all look directions. Future work will focus on the problem of source localization in the context of trigonometric interpolation beamforming.

5. REFERENCES

[1] E. Hänsler and G. Schmidt, *Acoustic echo and noise control: a practical approach*, vol. 40, John Wiley & Sons, 2005.

[2] C. Schüldt, F. Lindstrom, and I. Claesson, "A combined implementation of echo suppression, noise reduction and comfort noise in a speaker phone application," in *Proc. IEEE Int. Conference on Consumer Electronics (ICCE)*, Jan 2007, pp. 1–2.

[3] W. Kellermann, "Strategies for combining acoustic echo cancellation and adaptive beamforming microphone arrays," in *Proc. IEEE Int. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 1997, vol. 1, pp. 219–222.

[4] M. Brandstein and D. Ward, *Microphone arrays: signal processing techniques and applications*, Springer Science & Business Media, 2013.

[5] M. L. Valero and E. A. P. Habets, "Multi-microphone acoustic echo cancellation using relative echo transfer functions," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2017, pp. 229–233.

[6] M. Hamalainen and V. Myllyla, "Acoustic echo cancellation for dynamically steered microphone array systems," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2007, pp. 58–61.

[7] M. Kajala and M. Hamalainen, "Filter-and-sum beamformer with adjustable filter characteristics," in *Proc. IEEE Int. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2001, vol. 5, pp. 2917–2920.

[8] E. Mabande and W. Kellermann, "Design of robust polynomial beamformers as a convex optimization problem," in *Proc. IEEE Int. Workshop Acoustic Echo, Noise Control (IWAENC)*, 2010, pp. 1–4.

[9] E. Mabande, M. Buerger, and W. Kellermann, "Design of robust polynomial beamformers for symmetric arrays," in *Proc. IEEE Intl. Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 1–4.

[10] E. Mabande, A. Schad, and W. Kellermann, "Design of robust superdirective beamformers as a convex optimization problem," in *Proc. IEEE Intl. Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2009, pp. 77–80.

[11] H. Barfuss, M. Bachmann, M. Buerger, M. Schneider, and W. Kellermann, "Design of robust two-dimensional polynomial beamformers as a convex optimization problem with application to robot audition," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2017, pp. 106–110.

[12] J. Benesty, J. Chen, and I. Cohen, *Design of Circular Differential Microphone Arrays*, Berlin, Germany: Springer-Verlag, 2015.

[13] C. Pan, J. Chen, and J. Benesty, "Performance study of the MVDR beamformer as a function of the source incidence angle," *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, vol. 22, no. 1, pp. 67–79, 2014.

[14] C. Lai, S. Nordholm, and Y.H. Leung, "Design of robust steerable broadband beamformers with spiral arrays and the farrow filter structure," in *Proc. Intl. Workshop Acoust. Echo Noise Control (IWAENC), Tel-Aviv, Israel*, 2010.

[15] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," `http://cvxr.com/cvx`, Mar. 2014.