Robust Frequency Invariant Beamforming with Low Sidelobe for Speech Enhancement

Yiting Zhu\textsuperscript{1,*} and Xiang Pan\textsuperscript{1}

\textsuperscript{1}College of Information Science and Electronic Engineering, Zhejiang University, Zheda Road, No.38,310000, Hangzhou, People’s Republic of China

*Email: zhuyiting@zju.edu.cn

Abstract. Frequency invariant beamformers (FIBs) are widely used in speech enhancement and source localization. There are two traditional optimization methods for FIB design. The first one is convex optimization, which is simple but the frequency invariant characteristic of the beam pattern is poor with respect to frequency band of five octaves. The least squares (LS) approach using spatial response variation (SRV) constraint is another optimization method. Although, it can provide good frequency invariant property, it usually couldn’t be used in speech enhancement for its lack of weight norm constraint which is related to the robustness of a beamformer. In this paper, a robust wideband beamforming method with a constant beamwidth is proposed. The frequency invariant beam pattern is achieved by resolving an optimization problem of the SRV constraint to cover speech frequency band. With the control of sidelobe level, it is available for the frequency invariant beamformer (FIB) to prevent distortion of interference from the undesirable direction. The approach is completed in time-domain by placing tapped delay lines (TDL) and finite impulse response (FIR) filter at the output of each sensor which is more convenient than the Frost processor. By invoking the weight norm constraint, the robustness of the beamformer is further improved against random errors. Experiment results show that the proposed method has a constant beamwidth and almost the same white noise gain as traditional delay-and-sum (DAS) beamformer.

1. Introduction

Microphone arrays combined with FIBs are widely utilized in speech enhancement [1-3]. Compared to the frequency-domain broadband beamformer, the FIB in time-domain is attractive for its low computational complexity and constant frequency response with uniform beamwidth to avoid the effect of low pass filtering.

Harmonic nesting based on discrete Fourier transform is a traditional approach which achieves the frequency-invariant property between octaves [4]. However this method operates different subarrays at different frequency bands so that the beam pattern within one octave is still frequency dependent. In [5], harmonic nesting and filter-and-sum beamforming are combined to improve the frequency-invariant property within an octave. Instead of focusing on linear arrays like the above works, a simple FIB design utilizing multi-dimensional inverse Fourier transform is investigated in [6-7], which can be used for two or three-dimensional arrays. Recently, the convex optimization is utilized in the FIB design as a direct optimization approach [8]. However such beamformer using Frost processor needs pre-steering delays which are not an integer multiple of the sampling period, leading to some difficulties in implementation. In [9], the pre-delays are avoided by tapped delay lines (TDL) structure...
and complex weighted method. Another drawback of many convex optimization approaches is that they cover only one or two octaves. To get the closed-form array weights of the FIB[10] and further improve the frequency invariant, the least squares (LS) cost function combined with the SRV constraint is proposed, whose frequency band covers five octaves (speech frequencies). However, direct implementation of the least square cost function neglects the robustness of the beamformer so that the output signal has bad intelligibility and can only be used in source localization.

In this paper, we combine the spatial response constraint and the sidelobe control together as the cost function, and use the convex optimization method to solve the beam pattern synthesis problem to achieve a frequency-invariant response. To improve the robustness of the FIB, norm constraint of the steering weights are imposed, so that the proposed beamformer has good speech intelligibility just like conventional beamformer. Moreover, we use TDL to implement integer pre-delays [11] before FIR filter with real coefficient to avoid the complex structure of Frost processor [12].

2. Improved FIB

Consider an M-element array with a known arbitrary geometry, the sensors spatially sample the signal field at the locations \( p_m, m = 1, 2, ..., M \). The input is a plane wave with frequency region \([F_l, F_u]\) propagating in the direction \( \theta \), \( u(\theta) \) represents a unit vector that can be expressed as

\[
\mathbf{u}(\theta) = \begin{bmatrix} -\cos \theta \\ -\sin \theta \end{bmatrix}
\]

The time delay based on coordinate origin of every sensor is defined as

\[
r_m(\theta) = \frac{u(\theta)^T p_m}{c}
\]

where \( c \) is the velocity of propagation in the medium, and the superscript \( T \) denotes the transpose.

Figure 1 shows the modified time-domain beamformer structure. The processor has M sensors and each sensor is followed by a TDL and a FIR filter of length \( L \). The beamforming output is the sum of the ML weighted tap signals.

Let \( T_s \) represents the sampling period. \( T_m = \text{int}[ -r_m(\theta_m) / T_s ] \times T_s, m = 1, 2, ..., M \) is the integer pre-delay attached after each sensor, where \( \text{int}[\cdot] \) denotes the nearest integer so that the pre-delays are more convenient for practical implementation by using TDL.

The sampled data from the \( m \)th sensor is \( x_m(n) \). Thus, the data after integer pre-delay is defined as \( x_m(n) = \bar{x}_m(n-T_m) \). And the data of of every FIR filter tap is given by

\[
x_{ml}(n) = x_m(n-(l-1)\times T_s), l = 1, ..., L
\]
Thus, complex frequency response of the mth sensor FIR filter at frequency $f_k$ is

$$H_m(f_k) = \sum_{l=1}^{L} h_m e^{-j\phi_l(\theta) f_k T} = e^{i\phi_m(f_k)} h_m \quad \text{\textit{MERGEFORMAT}} (4)$$

where $e(f_k) = \left[ 1, e^{-j 2\pi f_k T}, \ldots, e^{-j(L-1) 2\pi f_k T} \right]^T$, and $h_m = \left[ h_{m1}, h_{m2}, \ldots, h_{ml} \right]^T$ are the adjustable weights which follow behind the mth sensor, $f_k$ is the frequency grid of the frequency region $[F_f, F_u]$. The weight vector of mth sensor at frequency $f_k$ is caused by the combined effects of pre-delay and FIR filter, which can be expressed as

$$w_m(f_k) = e^{i\phi_m(f_k)} h_m g_m(f_k), m = 1, 2, \ldots, M \quad \text{\textit{MERGEFORMAT}} (5)$$

where $g_m(f_k) = e^{-j 2\pi f_k T}$ is the phase shift caused by pre-delay. Finally, the weight vector of frequency $f_k$ is given by

$$w(f_k) = \left[ e^{i\phi_m(f_k)} h_m g_m(f_k), e^{i\phi_m(f_k)} h_m g_m(f_k), \ldots, e^{i\phi_m(f_k)} h_m g_m(f_k) \right]^T \quad \text{\textit{MERGEFORMAT}} (6)$$

For every frequency grid $f_k$, the two-dimensional array response of the beamformer is

$$p(\theta, f_k) = w^T(\theta, f_k) v(\theta, f_k), k = 1, 2, \ldots, K \quad \text{\textit{MERGEFORMAT}} (7)$$

where $v(\theta, f_k)$ is the array manifold vector given by

$$v(\theta, f_k) = \left[ e^{-j 2\pi f_k T}, e^{-j 2\pi f_k T}, \ldots, e^{-j 2\pi f_k T} \right]^T \quad \text{\textit{MERGEFORMAT}} (8)$$

is the propagation delay of the mth array element form direction $\theta$. Combining Eqs.(6) and Eqs.(8) into Eqs.(7) gives

$$p(\theta, f_k) = H^T \left[ v(\theta, f_k), g(f_k) \otimes e(f_k) \right], k = 1, 2, \ldots, K \quad \text{\textit{MERGEFORMAT}} (9)$$

$g(f_k) = \left[ g_1(f_k), g_2(f_k), \ldots, g_m(f_k) \right]^T$, $\otimes$ represents the Hadamard product operator, and $\otimes$ represents Kronecker product operator, $h = [h_{11}, h_{12}, \ldots, h_{ml}, h_{m1}, h_{m2}, \ldots, h_{ml}]^T$ is the $ML \times 1$ real-valued weight vector.

To get frequency invariant beam pattern, the SRV constraint is defined as:

$$SRV = \sum_{\phi_m \in \Theta_m, f_k \in [F_f, F_u]} \left| H^T a(\theta_m, f_k) - H^T a(\theta_m, f_0) \right|^2 \quad \text{\textit{MERGEFORMAT}} (10)$$

where $f_0$ is reference frequency, $a(\theta_m, f_k) = v(\theta, f_k) g(f_k) \otimes e(f_k)$, and $\Theta_m$ is the angular grid of the whole angle region $\Theta$, which can be same for different frequency. SRV measures the Euclidean distance of beam response between the reference frequency and other operating frequencies. The smaller the value of SRV is, the better performance of the beamformer has. Obviously, when SRV is fixed to zero, the beamformer has a frequency-invariant response.

In order to suppress interference from undesirable direction, the control of sidelobe level is obtained by adding a constraint for the sidelobe response at reference frequency $f_0$, which is defined as

$$J_{SL} = \sum_{\phi_m \in \Theta_{sl}} \left| H^T a(\theta_m, f_0) \right|^2 \quad \text{\textit{MERGEFORMAT}} (11)$$

where $\Theta_{sl}$ is the sidelobe region. Combining Eq.(10) and Eq.(11), the cost function can finally be expressed as

$$J = \sum_{\phi_m \in \Theta_m, f_k \in [F_f, F_u]} \left| H^T a(\theta_m, f_k) - H^T a(\theta_m, f_0) \right|^2 + \alpha \sum_{\phi_m \in \Theta_{sl}} \left| H^T a(\theta_m, f_0) \right|^2 = h^T Ch \quad \text{\textit{MERGEFORMAT}} (12)$$

where $\alpha (\alpha > 0)$ is a compromise parameter between the sidelobe level and the frequency-invariant property, and

$$C = \sum_{\phi_m \in \Theta_m, f_k \in [F_f, F_u]} \left[ a(\theta_m, f_k) - a(\theta_m, f_0) \right] \times \left[ a(\theta_m, f_k) - a(\theta_m, f_0) \right] + \alpha \sum_{\phi_m \in \Theta_{sl}} a(\theta, f_k) a^*(\theta, f_0) \quad \text{\textit{MERGEFORMAT}} (13)$$

Let $C = U^T U$ be the Cholesky factorization, we have
Thus, minimizing $h^* Ch$ is equivalent to minimizing $\|U^* h\|$. Consequently, the FIB design problem is formulated as follows:

$$\min_{h} \epsilon \quad \text{subject to } h^* a(\theta_0, f_0) = 1 \quad \|U^* h\| \leq \epsilon \quad \|h\| \leq \Delta$$

\* MERGEFORMAT (15)

where $h^* a(\theta_0, f_0) = 1$ denotes that the beamformer response at the reference frequency in the look direction achieves the unity response. $\epsilon$ is an auxiliary variable to control the cost function. Eq(15) is different from the constraints in Ref.[10], because $\Delta$ is used to improve the robustness of the FIB in the improved method.

3. Simulation and experimental result

A speech enhancement experiment with a 16-element linear microphone array with element spacing of 0.05m is used for evaluating the performance of the improved FIB. Table.1 shows the specific experimental parameters. A speech source is placed at a distance of 10 meters and at the normal direction of the microphone array.

| Table 1. Parameters of improved FIB. |
|--------------------------------------|
| experimental parameters | value |
| Steering direction | $0^\circ$ |
| Number of microphones | 16 |
| Working band | [300,3400]Hz |
| Sampling frequency | 8000Hz |
| Element spacing | 0.05m |
| Reference frequency | 2400Hz |
| The length of FIR filters | 30 |
| Compromise parameter | 0.5 |
| The norm of FIR filter tap weights | 0.5 |
| Subband number | 100 |
| The uniform grid of $[-90^\circ,90^\circ]$ | $2^\circ$ |
| The side-lobe region | $[-90^\circ,-50^\circ] \cup [50^\circ,90^\circ]$ |

Figure 2 depicts the beam pattern of the improved FIB over the frequency range of 300 to 3400 Hz in simulation. As can be observed, the resulting beam pattern has approximately constant mainlobe and the response in the look direction is close to the unity response. Moreover, $2.5^\circ$ difference between the maximum beamwidth and the minimum beamwidth at all frequencies further displays the frequency-invariant property of the broadband beamformer. Due to using the sidelobe control design, the sidelobe level of the improved FIB is under -17 dB.

Speech signals collected in the outdoor environment with the noise level of 53-55 dB are shown in Figure.3. The 6.5 s speech is chosen from the listening materials of CET-6 (College English Test, Level 6). The corresponding spectrograms are achieved by the short-time Fourier transform (STFT).
To prove the improved FIB has the same white noise gain as the traditional DAS beamformer, the results provided by the DAS beamformer are also shown in Figure 3(c). Obviously, after a large part of high frequency noise has been filtered out, the resulting speech signal's clarity and intelligibility in Figure 3(d) have been significantly improved in comparing with the polluted speech signal in a single channel as shown in Figure 3(b).

Furthermore, the performance of the improved FIB is evaluated by the output signal-to-noise ratio (SNR) and the perceptual evaluation of speech quality (PESQ) index [3]. Note that the PESQ value is calculated using the whole data of 6.5 s and the SNR is calculated with the data of 1.5 s. The results are shown in Table 2. For the speech signal with SNR of 4.6136 dB (single channel), the output SNRs of the DAS-beamformer and the improved FIB are 12.5324 dB and 12.8445 dB, respectively. Correspondingly, the PESQ values for speech in the single channel, the DAS-based beamformer output and the FIB output are 1.287, 1.965 and 1.904, respectively. Therefore, the proposed FIB has almost the same white noise gain as the DAS-based beamformer, which has been proved to be robust in practical implementation.

![Figure 2. Beam pattern of the improved FIB in simulation for a look direction of $0^\circ$ over the bandwidth of 300 to 3400Hz using a uniform linear array consisting of 16 microphones.](image)

![Figure 3. Waveforms and Spectrograms of speech signals (a)Original speech (b)Speech received in channel1 (c)DAS-based beamformer output (d)Improved FIB output.](image)
Table 2. The improvements in terms of output SNR and the PESQ index.

| Method        | SNR   | PESQ |
|---------------|-------|------|
| DAS-beamformer| 12.5324 | 1.965 |
| Improved-FIB  | 12.8445 | 1.904 |

4. Conclusion
A Robust frequency invariant beamforming with low sidelobe is proposed for speech enhancement. Our method minimizes the SRV constraint while maintaining the robust of beamformer using norm constraint on tap weights and providing the sidelobes of array pattern under a given threshold value. Furthermore, the convex optimization-based FIB is suited for wideband signals covering five octaves, and the proposed beamformer is convenient and cheap to implement for using integral tapped precisions. This approach can be used in arbitrary geometry array, and does not require significant computational complexity. The outdoor speech trials have verified the effectiveness of the proposed approach in terms of the output SNR and the PESQ measurement.

Acknowledgments
The authors would like to thank the financial support from the National Natural Science Foundation of China (No.6151397 and No.61171148)

References
[1] Mars R, Reju V G, Khong A W H 2014 A frequency-invariant fixed beamformer for speech enhancement *Signal and Information Processing Association Annual Summit and Conference (APSIPA)* pp 1-6
[2] Zou Y X, Li B, Liu W 2010 A broadband speech enhancement technique based on frequency invariant beamforming and GSC 2010 *International Conference on Green Circuits and Systems (ICGCS 2010)* pp 175-179
[3] Cho K, Nam K W, Lee J C, Hong S H, Kwon S Y, Han J, Kim D, Lee S, Kim I N 2014 A comparison of frequency-invariant beamforming algorithms for hearing aids-differential microphone-based beamformers and the broadband beamformer *Biomedical Engineering Letters* 4 pp 166-175
[4] Smith R 1970 Constant beamwidth receiving arrays for broadband sonar systems *Acustica* 48 pp 21-26
[5] Chou T 1995 Frequency-independent beamformer with low response error in Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing 5 pp 2995-2998
[6] Liu W, Weiss S, McWhirter J G, Proudler I K 2007 Frequency invariant beamforming for two-dimensional and three-dimensional arrays *Signal Process* 87 pp 2535-2543
[7] Liu W, Weiss S 2008 Design of frequency invariant beamformers for broadband arrays *IEEE Trans. Signal Processing* 56 pp 855-860
[8] Yan S F, Ma Y L, Hou C H 2007 Optimal array pattern synthesis for broadband arrays *Journal of Acoustic Society of America* 122 pp 2686-2996
[9] Peng, Chen, Yihui, 2008 SRV constraint based FIB design for wideband linear array *Journal of Systems Engineering and Electronics* 21 pp 941-947
[10] Zhao Y, Liu W, Langley R 2011 Application of the least squares approach to fixed beamformer design with frequency-invariant constraints*EIT Signal Processing* 5 pp 281-291
[11] Yan S, Hou C, Ma X 2007 Convex optimization based time-domain broadband beamforming with sidelobe control *Journal of the Acoustical Society of America* 121 pp 46-49
[12] Frost O L An algorithm for linearly constrained adaptive array processing 1972 *Proceedings of the IEEE* 60 pp 926-935