Microphone Array Directivity Improvement in Low-Frequency Band for Speech Processing

Mikhail Stolbov[1,2] and Sergei Aleinik[2,3(✉)]

[1] Speech Technology Center, Krasutskogo-4, St. Petersburg 196084, Russia
stolbov@speechpro.com
[2] ITMO University, 49 Kronverkskiy pr., St. Petersburg 197101, Russia
sergealeinik@gmail.com
[3] Alango Technologies Ltd., St. Petersburg, Russia

Abstract. This paper presents a new method of improving microphone array directivity in the low-frequency band. The method is based on a sub-band processing technique. We also evaluate the parameters and characteristics of the method and consider some of its practical implementations.

Keywords: Microphone array · Sub-band processing · Directivity pattern · Directivity index

1 Introduction

It is well known that conventional microphone arrays (MA) for speech processing have poor directivity in the low-frequency band (lower than 1000 Hz) [1–4]. At the same time, it is known that this frequency band is important for speech processing systems, since a large part of both speech signal and noise is concentrated in this band. There are two basic ways to solve this problem: (1) applying different superdirectivity algorithms [2–4] and (2) increasing the size of the MA by using the harmonically nested microphone subarrays technique with sub-band processing [3, 5].

However, superdirectivity algorithms perform badly in non-stationary noise environments, are very sensitive to noise (or signal + noise) covariance matrix estimation errors [6], as well as to microphone gain mismatch [7], etc., which limits their use in speech signal processing.

On the other hand, non-adaptive nested microphone subarrays with sub-band processing show good results, as noted in [3, 5]. A typical scheme of microphone layout for a MA with 9 microphones and 3 subarrays is described in detail in [5]. Here we note that in such schemes all subarrays have the same numbers of microphones but different distances between microphones, and improvement of directivity in the low-frequency (LF) band is achieved by increasing the length of the MA while keeping the total number of microphones constant. In the present paper we propose a different approach.

2 The Proposed Method

2.1 The Basic Idea

Typically, the distance d between microphones in a discrete MA is chosen based on the absence of side lobes with a large amplitude [1]:

$$d < \frac{c}{2f_{\max}}, \tag{1}$$

where: c is speed of sound in the air (343.1 m/s for 20 ºC) and $f_{\max}$ is the maximum frequency of the operating range of the MA in Hz. Let us denote the total number of microphones in the MA as N, and the aperture length of the MA as L: $L = d(N - 1)$.

It is known that if (1) is satisfied, further increasing N with a fixed L (i.e. decreasing d) does not lead to directivity improvement. On the other hand, (1) shows that for the low frequency range it is not necessarily to use all MA microphones. We can take, for example, only the edge microphones (the 0-th and $N - 1$). It turns out that for the low-frequency range, we obtain improved directivity even in comparison with the case where all MA microphones are used.

Let us consider a standard equidistant linear MA with the number of microphones $N = 9$ and the distance between microphones $d = 0.05$ m. Consequently, the aperture length L of the MA is equal to: $L = d(N - 1) = 0.4$ m. The normalized horizontal amplitude directivity pattern $D(f, \varphi_d, \varphi)$ of such an MA (for far-field assumption) can be obtained as [1]:

$$D(f, \varphi_d, \varphi) = \left| \frac{1}{N} \sum_{n=0}^{N-1} \exp\left(j \frac{2\pi f}{c} nd(\sin(\varphi) - \sin(\varphi_d)) \right) \right|, \tag{2}$$

where: f is the signal frequency in Hz; φ is the direction of arrival; φ_d is the desired look direction and '$|\cdot|$' denotes the "magnitude of complex value" operator.

The directivity pattern (DP), calculated using (2), for MA with $N = 9$, $d = 0.05$, $\varphi_d = 0$ and $f = 428.87$ Hz is shown in Fig. 1, solid line.

It is clear that the MA has poor directivity for the given f. At the same time, if we take only the edge microphones (0-th and 8-th, $d_{0,8} = L = 0.4$) and calculate the DP for the same f for this 2-microphone sub-array, we get the dashed curve shown in Fig. 1, i.e. we get better DP.

It seems strange, but when signal frequency is lower than some threshold level, it is better to use the edge microphones rather than all microphones in the MA.

2.2 Detailed Study of the Method

In fact there is nothing strange about the above mentioned result. It is known that for a dual-microphone array with a 0.4 m distance between microphones, the first zeroes in the DP for $\varphi = \pm 90$ can be calculated as: $\lambda/2 = 0.4$ (where λ is the wavelength); which implies that $f = c/0.8 = 343.1/0.8 = 428.875$ Hz. At the same time, for an MA with a continuous aperture (and the same length) we will have the condition of first
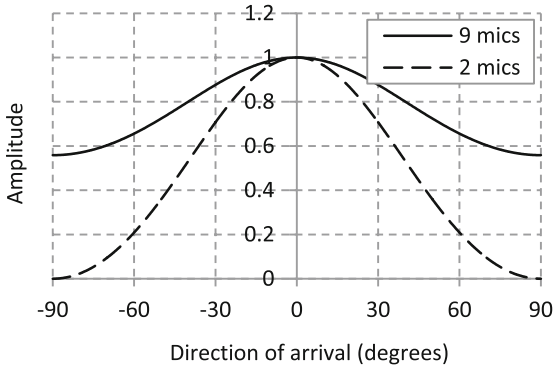
Fig. 1. Directivity patterns, calculated using all 9 (solid line) and 2 edge (dashed line) microphones of the MA for $f = 428.87$ Hz.

zeroes at $\varphi = \pm 90$ as: $\lambda = 0.4$ [1] i.e. the frequency in this case is two times higher. Frequency for a discrete MA with $N = 9$ occupies an intermediate position.

Now consider the directivity index (DI), $G(f, \varphi_d)$ of a linear MA. $G(f, \varphi_d)$ characterizes the directivity of the MA: the greater $G(f, \varphi_d)$, the better the directivity. Without loss of generality we can calculate it in Cartesian coordinates, using a single-dimensional horizontal DP (as in Fig. 1) and $-90 \leq \varphi \leq 90$ in degrees. In this case $G(f, \varphi_d)$ can be obtained as:

$$G(f, \varphi_d) = \frac{180}{\int_{-90}^{90} |D(f, \varphi_d, \varphi)|^2 \cos(\varphi) d\varphi} \tag{3}$$

Figure 2 demonstrates the directivity index $G(f, \varphi_d)$ for $\varphi_d = 0$ and for the initial MA ($N = 9$, $d = 0.05$ m); for the MA created using two microphones on the edges
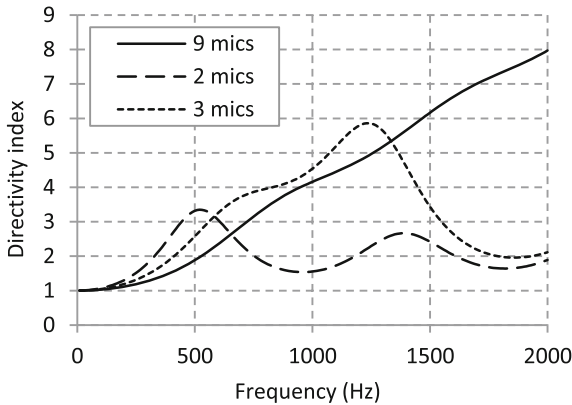

Fig. 2. Directivity index as function of signal frequency for three MA's with the same length and different number of microphones.

$(N = 2, d = 0.4$ m) and for the MA created using three (the central and the edges) microphones, i.e. $N = 3, d = 0.2$ m.

It can be seen that up to 600 Hz the MA with $N = 2$ has the best directivity. The MA with $N = 3$ has the best directivity in the interval 600–1300 Hz. The MA with $N = 9$ works best only when $f > 1300$ Hz. It should also be noted that the difference in $G(f, \phi_d)$ for $N = 3$ and $N = 9$ is not so large, so we will further focus only on the MA with $N = 2$.

Maximizing (3) for $N = 2, d = 0.4$ and $\varphi_d = 0$ we find that $G(f, \varphi_d)$ has a maximum when $f = 523$ Hz. The corresponding DP as well as the DP of the initial MA with $N = 9$ are shown in Fig. 3.
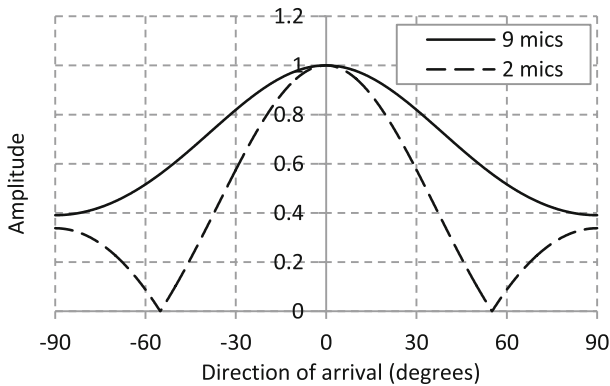

Fig. 3. Directivity patterns, calculated using all 9 (solid line) and 2 edge (dashed line) microphones of the MA for $\phi_d = 0, f = 523$ Hz.
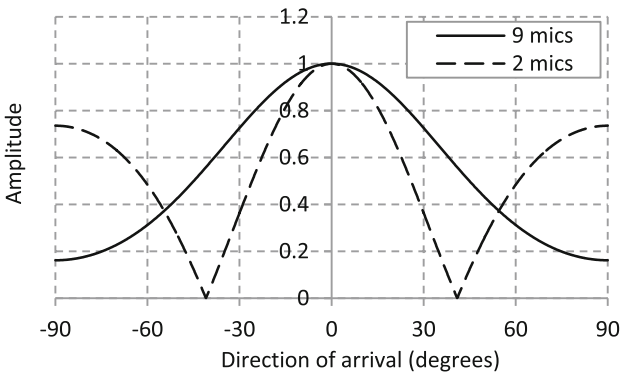

Fig. 4. Directivity patterns, calculated using all 9 (solid line) and 2 edge (dashed line) microphones of the MA for $\phi_d = 0, f = 654.54$ Hz.

Figure 3 demonstrates the advantages of using the two microphone scheme instead of the total number of microphones in the LF band. Further increase of signal frequency leads to the point of intersection of $G(f, \varphi_d)$ for $N = 2$ and $N = 9$: $f = 654.54$ Hz. The corresponding DPs are shown in Fig. 4.

In our opinion, Fig. 4. shows performance deterioration for $N = 2$, as the side-lobe level of the DP increases rapidly. Figures 3 and 4 suggest that the boundary frequency between using 2 and 9 microphones should be chosen in the interval [523, 654] Hz.

2.3   Dependence on the Look Direction

The above suggestions were obtained when the look direction was perpendicular to the line of the microphones, i.e. for $\varphi_d = 0$. In real life we often have $\varphi_d \neq 0$ and it is clear that directivity patterns, directivity indexes and, consequently, boundary frequencies depend on the φ_d. These dependences for the MA with $L = 0.4$ m are shown in Fig. 5.
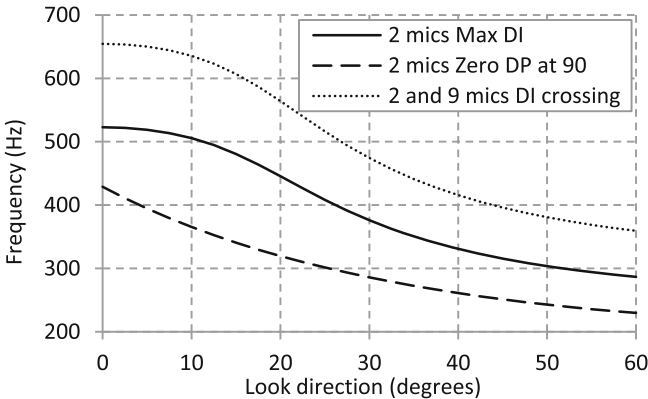

Fig. 5. Three basic frequencies as functions of the look direction ϕ_d.
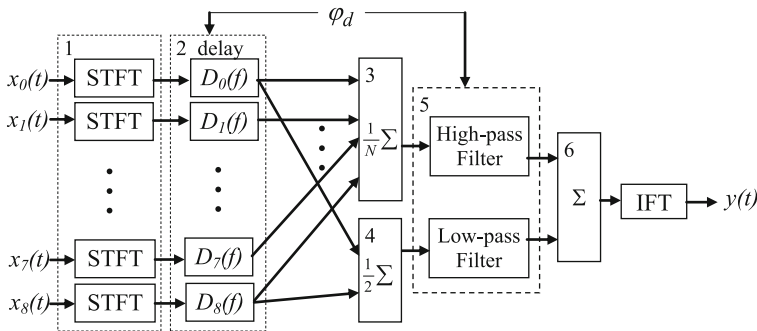

Fig. 6. Flow-chart of the MA with the proposed method.

It can be seen that all frequencies decrease when φ_d increases. So the boundary frequency of the filters should be controlled depending on the φ_d. We suggest choosing Max DI frequency (solid curve on Fig. 5) or the arithmetical mean between Max DI frequency and the frequency at which the 2 and 9 mics DI cross.

The structure of the MA with the proposed method is shown in Fig. 6.

The first two blocks of the chart are well-known: they are the short time Fourier transform (Block 1) that transforms input signals $x_i(t)$, $i = 0, N - 1$ into the frequency domain, and the frequency-domain delay (Block 2) where all signals are delayed by multiplication by the complex steering vector $D_m(f, \varphi_d)$ [9] to move the MA beam to the look direction. After the delay, all signals $X_i(f, \varphi_d)$, $i = 0, N - 1$ are summed and normalized to $1/N$ in Block 3. At the same time the two signals $X_0(f, \varphi_d)$ and $X_{N-1}(f, \varphi_d)$ are summed and normalized to $1/2$ in Block 4. Then both sums are filtered in Block 5 where the cut-off frequency is calculated according to the look direction φ_d (Fig. 5). Finally, both filtered sums are added together, after which inverse Fourier transform is performed for the resulting signal.

As a result, we have the following: high-frequency components of the input signal are passed to the MA output through Block 3 and the high-pass filter (all MA microphones work); at the same time, the LF components of the input signal are passed to the MA input through Block 4 and the low-pass filter, i.e. only $X_0(f, \varphi_d)$ and $X_{N-1}(f, \varphi_d)$ are used, which gives us better MA directivity in the LF band.

3   Conclusion

In this paper we describe a new method for improving MA directivity in the low-frequency band. This method can be useful for speech processing in MAs in different areas, for example in speaker verification [8], in multimodal systems [9], etc.

Note that it is easy to implement a 3-microphones scheme (two microphones on the edges and the central one) in the flowchart in Fig. 6. In order to do this, it is necessary to add the third sum block and the third band-pass filter to Block 5. However, the improvement will not be significant (see Fig. 2).

It is clear that the proposed method can be used for LF direction improvement not only in equidistant MAs but in all MAs that include a set of microphones, even for nested microphone subarrays described in [5].

The method can also be used in planar MAs. For circular planar MAs, for example, the microphone ring with the highest diameter is the analog of two microphones on the edges for linear MAs. The proposed scheme may be improved by using adaptive superdirectivity algorithms, for example, those presented in [10]. This will further strengthen the MA directivity in the LF band, but will cause a shift in boundary frequency. This modification of the method remains to be studied in future work.

References

1. McCowan, I.: Microphone arrays: a tutorial (2001). URL: http://www.idiap.ch/~mccowan/arrays/tutorial.pdf

2. Bitzer, J., Simmer, K.U.: Superdirective microphone arrays. In: Brandstein, M., Ward, D. (eds.) Microphone Arrays: Signal Processing Techniques and Applications, pp. 19–38. Springer, Heidelberg (2001)

3. Aleinik, S., Stolbov, M.: A comparative study of speech processing in microphone arrays with multichannel alignment and zelinski post-filtering. In: Ronzhin, A., Potapova, R., Fakotakis, N. (eds.) SPECOM 2015. LNCS, vol. 9319, pp. 34–41. Springer, Heidelberg (2015)

4. Löllmann, H.W., Vary, P.: Post-filter design for superdirective beamformers with closely spaced microphones. In: Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA 2007), pp. 291–294 (2007)

5. Fischer, S., Kammeyer, K.D.: Broadband beamforming with adaptive postfiltering for speech acquisition in noisy environment. In: Proceedings of International Conference on Acoustics, Speech, and Signal Processing, pp. 359–362 (1997)

6. Carlson, B.: Covariance matrix estimation errors and diagonal loading in adaptive arrays. IEEE Trans. Aerosp. Electron. Syst. **24**(3), 397–401 (1988)

7. Ba, D.E., Dinei, F., Cha, Z.: Enhanced MVDR beamforming for arrays of directional microphones. In: IEEE Conference on Multimedia and Expo, pp. 1307–131 (2007)

8. Kozlov, A., Kudashev, O., Matveev, Y., Pekhovsky, T., Simonchik, K., Shulipa, A.: SVID speaker recognition system for NIST SRE 2012. In: Železný, M., Habernal, I., Ronzhin, A. (eds.) SPECOM 2013. LNCS, vol. 8113, pp. 278–285. Springer, Heidelberg (2013)

9. Karpov, A., Akarun, L., Yalcin, H., Ronzhin, A., Demiroz, B., Coban, A., Zelezny, M.: Audio-visual signal processing in a multimodal assisted living environment. In: Proceedings of INTERSPEECH 2014, pp. 1023–1027 (2014)

10. Buck, M., Roessler, M.: First order differential microphone arrays for automotive applications. In: Seventh International Workshop on Acoustic Echo and Noise Control (IWAENC 2001), Darmstadt, pp. 19–22 (2001)