COMPLEX SPECTRUM CIRCLE CENTROID METHOD FOR NOISE REDUCTION IN ARRAY SIGNAL PROCESSING

Kazuhito Inoue, *Takashi Okajima*, *Yutaka Kamamoto*, *Takuya Nishimoto* and *Shigeki Sagayama*

Graduate School of Information Science and Technology
The University of Tokyo
Hongo, Bunkyo-ku, Tokyo, Japan
Email: {inoue,t-okajima,kamamoto,nishi,sagayama}@hil.t.u-tokyo.ac.jp

Introduction. Microphone array signal processing is actively studied for various purposes such as speech recognition and hands-free telecommunication systems. The main idea is utilization of differences in direction of arrival between sources of target and noise signals to multiple microphones. The simplest technique is widely-known Delay-and-Sum (DS) that adjusts delays added to microphone inputs so that the target signal from a particular direction synchronizes across multiple microphones while noises from different directions do not. This technique has advantage that no training is required, though performance of noise reduction is not sufficient. On the other hand, adaptive types of microphone array signal processing such as AMNOR[1] and other adaptive beamforming methods commonly require time segments of noise signal in order to train filter coefficients. These methods often fail to track rapid changes of environmental characteristics such as moving noise sources. Other methods such as blind source separation based on independent component analysis[2] are known to give high performance without knowledge of the direction of target signal, but the nonlinear optimization often faces serious local convergence difficulties.

This paper discusses an approach to microphone array signal processing based on geometrical manipulation on complex spectrum plane, characterized by nonlinearity and framewise operation without training. We previously introduced this idea to speech recognition[3]. In this paper, we evaluated the further potential of this method for noise reduction use.

CSCC Method. Primarily, we assume that acoustic characteristics (gains, directivities, etc.) of microphones are identical (or can be equalized by adjusting gains and delays at each frequency). If target signal $s(t)$ propagates and arrives at K microphones simultaneously at time t while a noise signal $n(t)$ in each microphone arrives with different time delay $(\tau_1, \tau_2, \cdots, \tau_K)$, the observed signal $m_i(t)$ at i-th microphone and its Fourier transform $M_i(\omega)$ are given by:

$$m_i(t) = s(t) + n(t - \tau_i), \quad M_i(\omega) = S(\omega) + N(\omega)e^{-j\omega\tau_i}, \quad i = 1, 2, \cdots, K. \tag{1}$$

according to the basic properties of Fourier transform, where ω denotes angular frequency and $S(\omega)$ and $N(\omega)$ denote Fourier transforms of $s(t)$ and $n(t)$, respectively. From microphone signals, we easily obtain framewise complex spectrum (typically multiplied by a short-time window). Geometrically, Eq. (1) implies that $M_i(\omega)$ lies on a circle of radius $\|N(\omega)\|$ centered at $S(\omega)$ on the complex spectrum plane as shown in Figure 1. The complex spectrum of target signal $S(\omega)$ can be restored by estimating the center of the circle that all complex points $M_i(\omega)$ lie on. We call this method "Complex Spectrum Circle Centroid (CSCC) method." Meanwhile, Delay-and-Sum (DS) method uses the center of gravity (arithmetic mean) of microphone inputs: $\bar{M} = \frac{1}{K}\sum_{i=1}^{K} M_i(\omega)$. The circle centroid is obviously a point placed in equal distance from the circumferential points. We estimate the centroid as a point $\tilde{S}(\omega) = X + jY$ by minimizing the variance of K squared distances from $M_i(\omega) = x_i + jy_i$, i.e.,

$$\tilde{S}(\omega) = \underset{X+jY}{\operatorname{argmin}} \left\{ \frac{1}{K}\sum_{i=1}^{K}\Big((X-x_i)^2+(Y-y_i)^2\Big)^2 - \Big(\frac{1}{K}\sum_{i=1}^{K}\Big((X-x_i)^2+(Y-y_i)^2\Big)\Big)^2 \right\}. \tag{2}$$

Differentiating the above equation with respect to X and Y and setting the gradients equal to zero, the optimal $\tilde{S}(\omega)$ is obtained by:

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \frac{1}{2}\begin{pmatrix} \mathrm{Var}[x_i] & \mathrm{Cov}[x_i, y_i] \\ \mathrm{Cov}[x_i, y_i] & \mathrm{Var}[y_i] \end{pmatrix}^{-1} \begin{pmatrix} \mathrm{Cov}[x_i, x_i^2] + \mathrm{Cov}[x_i, y_i^2] \\ \mathrm{Cov}[y_i, y_i^2] + \mathrm{Cov}[y_i, x_i^2] \end{pmatrix} \tag{3}$$

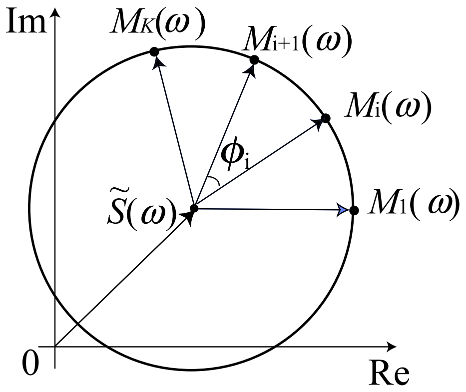

Figure 1: Microphone signals $M_i(\omega)$ located on a circle in the complex spectrum plane with the target signal $S(\omega)$ being the circle centroid

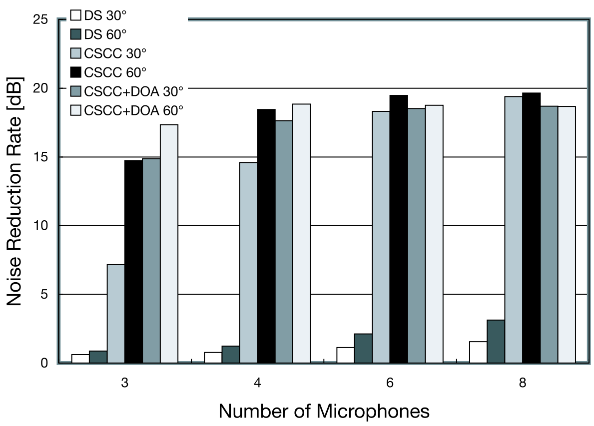

Figure 2: Noise reduction rates of CSCC and DS methods

denoting the variance of a by $\mathrm{Var}[a]$ and the covariance of b and c by $Cov[b,c]$. Eq.(3) shows that parameters X and Y are obtained analytically without iterative calculation. The solution of Eq.(3) is guaranteed to exist unless the correlation coefficient between x and y, $r_{xy} = 1$, i.e., all spectrum points $M_i(\omega),\ i = 1, 2, \ldots, K$ lie on a straight line in the complex plane. Even though r_{xy} is always guaranteed to be no greater than 1, in numerically bad conditions such as $r_{xy} > 0.99$, we use the center of gravity of K points $M_i(\omega)$, i.e., the delay-and-sum solution here, instead of the circle centroid. Its solutions X and Y give the estimated complex spectrum centroid for each frequency-bin. Signal waveform is restored through inverse Fourier transform of the estimated $S(\omega)$.
We can further combine this method with direction of arrival (DOA) estimation of noise. After obtaining $S(\omega)$ by Eq.(3), we can easily evaluate the angle between $M_i(\omega) - S(\omega)$ and $M_{i+1}(\omega) - S(\omega)$, $(i = 1, \cdots, K-1)$, which should be proportional to the time lags τ_i $(i = 1, \cdots, K)$. Under consideration that this angle must be also proportional to frequency (i.e., phase-linear property), we can apply regression line approximation to the estimated angles along frequency-bins. $S(\omega)$ can be reestimated with the help of the approximated regression line and thus excessive errors in the first estimation step of $S(\omega)$ can be reduced.

Experiments and results. For the evaluation of the proposed method, simulated experiments were performed. The array was assumed to consist of 3, 4, 6, 8 and 12 microphones spaced at 5cm. The directions of target and noise were assumed to be 0[deg] and 30 or 60[deg], respectively. Signal data of target and interfering speech were excerpted from ASJ-JNAS corpus of read newspaper articles. 8 testsets were artificially created by adding speech data of male or female speakers. All signals were sampled at 16 kHz, and analyzed with Hamming window where frame length and shift were 32ms and 16ms, respectively. To quantify the performance of our method, we measured the Noise Reduction Rate (NRR = output SNR − input SNR). The NRR results shown in Figure2 demonstrates the superiority of the CSCC method. Even if the number of microphones was fewer, proposed method gave high performance by estimating DOA though computing cost was more expensive.

Conclusion. In this paper, we proposed the Complex Spectrum Circle Centroid (CSCC) method for restoring the target signal from multiple microphone input signals in noisy environment without training filter coefficients. The noise reduction process is non-linear to the input and framewise operation. The proposed method was evaluated in simulation experiments and shown to be significantly effective in noise reduction. This new method is still an on-going work to be further explored, both theoretically and experimentally.

References

[1] Y. Kaneda and J. Ohga, "Adaptive microphone-array system for noise reduction," *IEEE Trans. Acoust. Speech Signal Process.*, vol. ASSP-34, no. 6, pp. 1391–1400, 1986.
[2] A. Hyvärinen, J. Karhunen and E. Oja, *Independent Component Analysis*, John Wiley & Sons, 2001.
[3] S. Sagayama, T. Okajima, Y. Kamamoto and T. Nishimoto, "Complex Spectrum Circle Centroid for Microphone-Array-Based Noisy Speech Recognition," *Proc. ICSLP2004*, vol.4, pp. 825-828, 2004.