A DIRECTIONAL NOISE SUPPRESSOR WITH A SPECIFIED BEAMWIDTH

Akihiko Sugiyama and Ryoji Miyahara†

Information and Media Processing Laboratories
NEC Corporation
† Internet Terminal Division, NEC Engineering
1753 Shimonumabe, Nakahara-ku, Kawasaki 211-8666, Japan
aks@ak.jp.nec.com

ABSTRACT

This paper proposes a directional noise suppressor with a specified constant beamwidth. A directional gain is calculated based on interchannel phase difference and combined with a spectral gain commonly used in single-channelnoise suppressors (NSs). The beamwidth can be specified as passband edges of the directional gain. In order to implement frequency-independent constant beamwidth, frequency-proportionate band-edge phase differences are determined for the passband. Evaluation with signals recorded by a commercial PC demonstrates good agreement between the theoretical and the measured directivity. The signal-to-noise ratio improvement and the PESQ score for the enhanced signal are improved by 24.4 dB and 0.3 over a conventional NS.

***Index Terms*—** Beamformer, Noise suppressor, Phase difference, DOA, Directional gain, Beamwidth

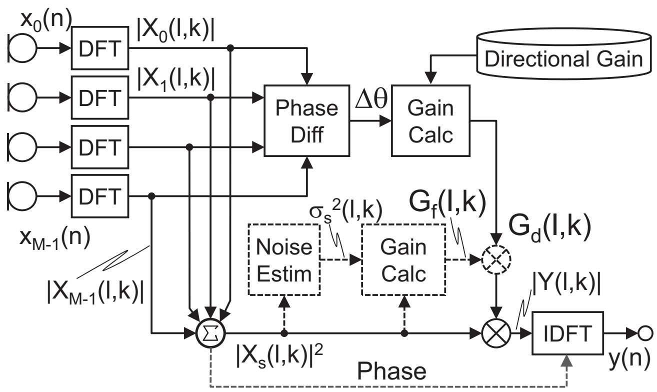

Fig. 1. Directional noise suppressor with a specified beamwidth.

1. INTRODUCTION

Signal enhancement is an indispensable technology for communications and human-computer interaction in noisy environment. A variety of algorithms have been developed with a single [1]-[4], dual [5]-[18], and multiple microphones [19]-[23]. Single-microphone and some of two-microphone algorithms [5]-[13] use only the magnitude information and performance is saturated as the signal-to-noise ratio (SNR) becomes low [24]. Other two-microphone algorithms [14]-[18] are based on the adaptive noise canceller (ANC) structure, where both magnitude and phase information are used. ANCs have demonstrated superior performance for some applications [15, 18]. However, neither single-microphone nor two-microphone algorithms described so far do not form directivity. They apply suppression in the spectral domain and are useful for non-directional noise.

For directional noise, acoustic beamformers, also known as microphone arrays (MAs), are effective [19]–[23]. MAs, different from antenna arrays, require a large number of sensors (microphones) to form a sharp beam because of a long wave length of acoustic signals. It is a potential drawback for consumer applications which may not have sufficient space to accomodate many microphones. In addition, MAs have a limitation from a viewpoint of constant beamwidth across frequency. The mainlobe and a null in a low frequency are wider than those in a high frequency due to a longer wavelength, leading to poor spatial selectivity.

Solutions to this problem are a combinations of arrays of different size dedicated to different frequency ranges [25, 26]. A most common example is a harmonically-nested array [27]–[33]. However, as far as a nested technique is employed, increase in the array size and the number of microphones is inevitable. A recently proposed solution with an auxiliary microphone performs well with a compact microphone array [34]. An auxiliary input forms an ANC structure that is integrated into an MA and cancels low-frequency noise and interference. Nevertheless, increase of a microphone may not be acceptable for cost-and-space-conscious products. Without increase of microphones and array size, it is desirable to suppress both directional and non-directional noise in a wide frequency range.

Phase-based time-frequency (T-F) masking can be another solution to this selectivity degradation with a small number of microphones. Aarabi et al. [35] uses a phase-difference error between two microphones in each T-F block to calculate a directional gain. Qazi et al. [36] presents a wider variety of directional gains with different characteristics. The obtained directivities have a sharp beam in a look direction even with two microphones. However, the beamwidth is not considered in the design of the directional gain and only a directional interference is assumed.

This paper proposes a directional noise suppressor with a specified constant beamwidth. The following section discusses calculation of a spectral and a directional gain with beamwidth specification. In Section 3, evaluation results are presented with respect to the directivity and enhanced signal quality compared with a conventional NS to demonstrate how much the added directivity contributes to the superior performance.

2. DIRECTIONAL NOISE SUPPRESSOR

Figure 1 illustrates a blockdiagram of the proposed directional noise suppressor. It calculates two gains, namely, a spectral gain $G_f(l,k)$

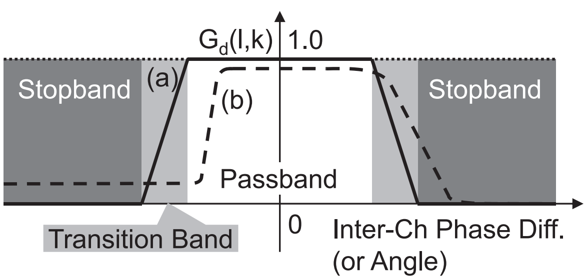

Fig. 2. Directional gain $G_d(l, k)$.

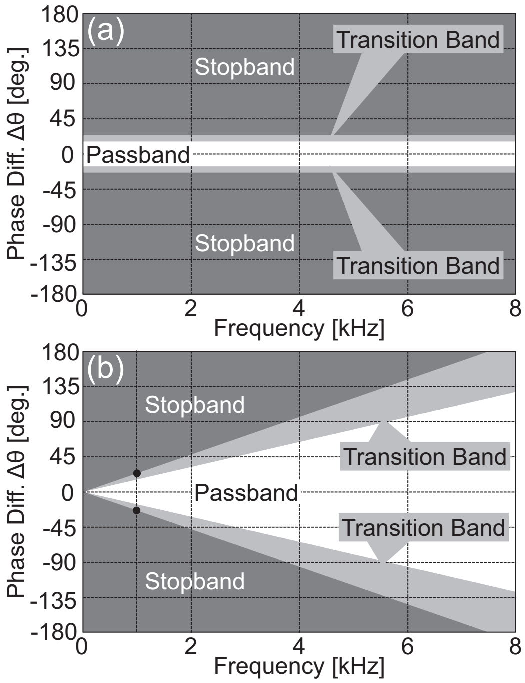

Fig. 3. Frequency dependency of directional gain with respect to phase difference. (a) Constant with frequency, (b) Proportional to frequency. Passband edge: ± 20 deg., Stopband edge: ± 30 deg., both at 1 kHz.

and a directional gain $G_d(l, k)$, where l and k are the frame and the frequency index. The spectral gain $G_f(l, k)$ is calculated in the traditional noise suppression framework (*e.g.* [1, 2]). The directional gain $G_d(l, k)$ is designed in advance and stored in memory. It suppresses all signal components other than the target signal. The target signal components are identified by direction-of-arrival (DOA) represented by interchannel phase difference $\Delta\theta(l, k)$.

2.1. Calculation of the spectral gain $G_f(l, k)$

The input signal $x_m(n)$ in channel m ($0 \leq m \leq M - 1$) is segmented into frames and applied DFT (discrete Fourier transform) to generate a corresponding frequency-domain signal $X_m(l, k)$. In order to calculate a spectral gain, it is necessary to estimate the noise power. Assuming that the target signal is located on the line perpendicular to the array surface, the sum-beamformer output power $|X_s(l, k)|^2$ is used for noise estimation.

$$|X_s(l, k)|^2 = \left| \sum_{m=0}^{M-1} X_m(l, k) \right|^2, \tag{1}$$

where M is the number of channels. When the target signal is located off the above line, widely known beam steering can be applied before (1). Once $|X_s(l, k)|^2$ is calculated, any noise estimation algorithm [3, 4] can be used to obtain a noise power estimate $\sigma_s^2(l, k)$. It is also possible to form a null beamformer with $X_m(l, k)$ and use its output power as an estimated noise power. With $\sigma_s^2(l, k)$ and $|X_s(l, k)|^2$, a spectral gain $G_f(l, k)$ can be calculated by a traditional noise suppression algorithm [1, 2].

2.2. Design of the directional gain $G_d(l, k)$

A directional gain is designed in advance such that the signal components coming from the target-signal direction are passed and all others are sufficiently suppressed. An example is given as a solid line (a) in Fig. 2. Because the target direction is assumed to be 0 degrees, $G_d(l, k)$ takes a value of unity around 0 degrees and a value 0 otherwise. In addition to the passband around 0 degrees and the stopbands, there may be transition bands where the directional gain $G_d(l, k)$ gradually decreases from the passband to the stopband. The passband, transition bands, and stopbands can be arbitrary specified as design issues. The shape of the directional gain $G_d(l, k)$ is flexible, too. A dissymmetric shape such as a dashed line (b) in Fig. 2 is possible.

The directional gain can be designed in the following way. Let us first assume that the passband edge angles are $\pm\phi$ radian. The time lag τ between signals from adjacent microphones is given by

$$\tau = d \cdot \sin\phi / c, \tag{2}$$

with the signal DOA ϕ, a microphone spacing d, and a sound velocity c. The time lag τ corresponds to a phase difference $\Delta\theta$ as

$$\exp\{-j\Delta\theta\} = exp\{-j\omega\tau\} = exp\{-j2\pi f\tau\}, \tag{3}$$

where f is the signal frequency. (2) and (3) leads to

$$\Delta\theta = 2\pi f d \sin\phi / c. \tag{4}$$

It is understood from (4) that the phase difference $\Delta\theta$ representing a passband edge of the phase difference should be proportional to the frequency.

Figure 3 illustrates frequency dependency of the directional gain $G_d(l, k)$. (a) and (b) represent the constant band-edge phase difference and frequency-proportionate band-edge phase difference, respectively. The passband edge DOAs of $\phi = \pm 20$ degrees and the stopband edge DOAs of $\phi = \pm 30$ degrees are assumed at 1 kHz in (b). As an example, stopband edges are calculated for $f = 1$ kHz, $d = 4.5$ cm, $\phi = \pm 30$ degrees, $c = 346.3$ m/s. (4) gives $\Delta\theta = 0.41$ radian which corresponds to ± 24 degrees and are marked by bullets in Fig. 3 (b).

Figure 4 compares the frequency dependency of the directional gain with respect to the DOA for two cases in Fig. 3. For a constant $\Delta\theta$ over frequency in Fig. 3 (a), the edge DOA ϕ should be smaller for a higher frequency as in (4), leading to a frequency dependent beamwidth in Fig. 4 (a). On the contrary, a frequency proportionate $\Delta\theta$ cancels out f on the right-hand side of (4), resulting in a frequency independent beamwidth in Fig. 4 (b). A phase difference $exp\{-j\Delta\theta\}$ may take the same values at a high and a low frequency as in (3) and (4). This comes from the periodicity of the exponential function. It appears as aliasing in Fig. 4 (a) and (b).

2.3. Calculation of the directional gain $G_d(l, k)$

The directional gain $G_d(l, k)$ is determined based on the angle $\Delta\theta$ representing interchannel phase difference. As the simplest case, let us assume $M = 2$. An angle $\Delta\theta(l, k)$ is given by

$$\Delta\theta(l, k) = \angle\{X_0(l, k)/X_1^*(l, k)\} = \theta_0(l, k) - \theta_1(l, k), \tag{5}$$

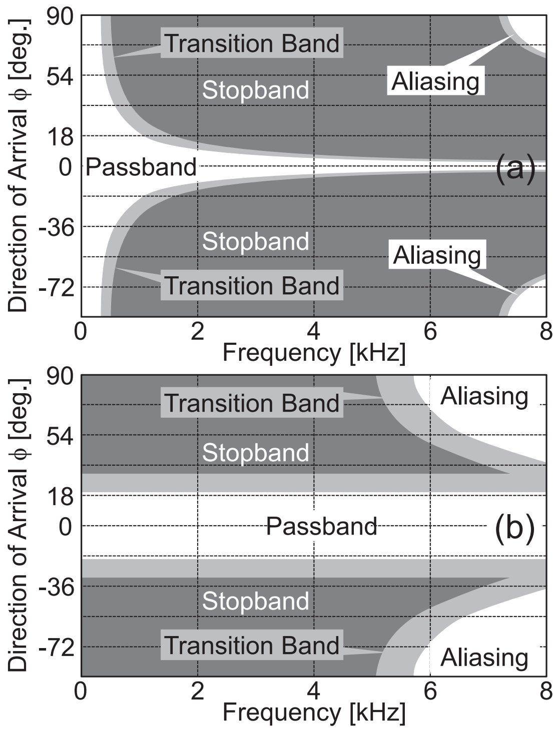

Fig. 4. Frequency dependency of directional gain with respect to DOA. (a) Constant $\Delta\theta$ over frequency, (b) Frequency proportionate $\Delta\theta$.

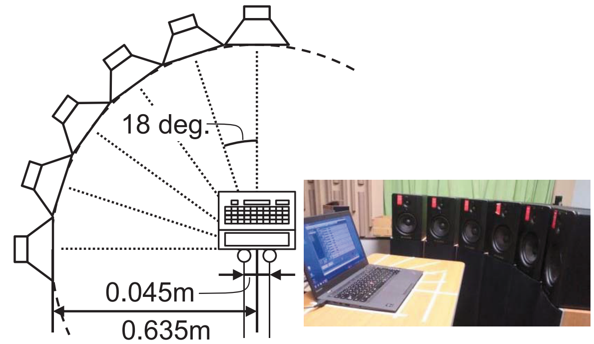

Fig. 5. Evaluation set-up with a PC and six loudspeakers.

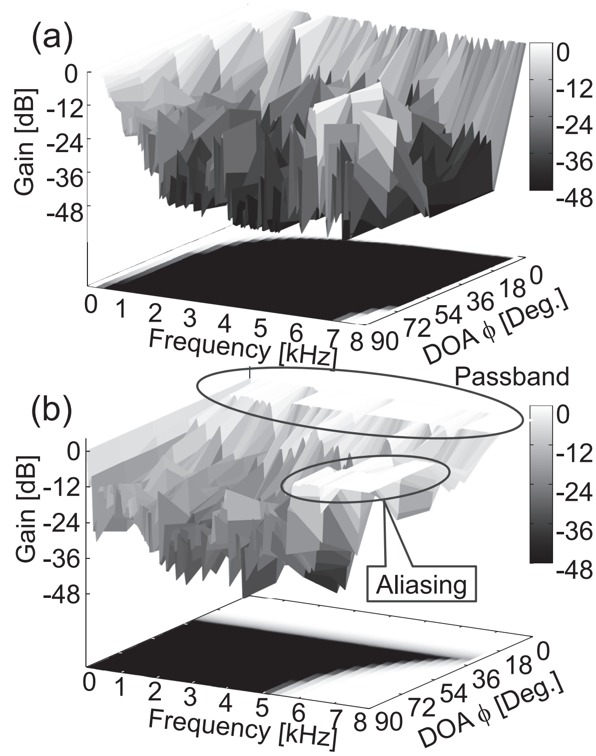

Fig. 6. Measured directivity for constant and variable passband widths. (a) Constant passband, (b) Frequency proportionate variable passband.

where θ_0 and θ_1 are the phase of $X_0(l,k)$ and $X_1(l,k)$. When there are more than two channels, interchannel phase difference of multiple adjacent channels can be used to obtain more accurate phase difference by averaging for example. Referring to Fig. 2, $\Delta\theta(l,k)$ in a specified passband returns $G_d(l,k) = 1$ and no suppression is performed by the directional gain. For other values of $\Delta\theta(l,k)$, $G_d(l,k) < 1$ is returned and suppression is performed accordingly. The directional gain $G_d(l,k)$, as depicted in Fig. 4 (b), is stored in memory and used for gain calculation based on interchannel phase difference $\Delta\theta(l,k)$.

2.4. Overall suppression with a directional and a spectral gain

The final enhanced signal power in each frequency is obtained by multiplying the power sum of all microphone signals $|X_s(l,k)|^2$ by two gain values as

$$|Y_s(l,k)|^2 = G_f(l,k)G_d(l,k)|X_s(l,k)|^2. \tag{6}$$

$|Y_s(l,k)|$ is combined with the phase of the sum-beamformer output and applied an inverse DFT to obtain a time-domain enhanced signal.

3. EVALUATIONS

A laptop PC equipped with two built-in microphones was placed on a table in a $5 \times 5 \times 2.5$ m room. The microphone spacing was 4.5 cm.

3.1. Measurement of the beam pattern

Six loudspeakers were placed around the PC with an 18 degree spacing and a 63.5 cm distance as shown in Fig. 5. A chirp signal to cover a frequency range of $0 - 8$ kHz was radiated from each loudspeaker one after another. The recorded 2-channel signals were processed by constant passband over frequency in Fig. 3 (a) and frequency proportionate passband in (b) with $G_f(l,k) = 1$. A passband beamwidth was set to $20\,\mathrm{deg}$. A ratio of the input to the output signal power is depicted in Fig. 6. In case of frequency proportionate passband in (b), almost constant beamwidth across frequency is obtained. Its projection on the ground is similar to Fig. 4 (b). The same projection of (a) representing constant passband over frequency shows good agreement with Fig. 4 (a). It is clearly demonstrated that frequency proportionate passband is effective to form constant beamwidth along frequency.

3.2. Signal enhancement

A female speech sampled at 16 kHz was played back in front of the PC (0 degrees) at a distance of 61 cm. A babble noise was played back at 91.4 cm away and 60 degrees off the speech direction at an

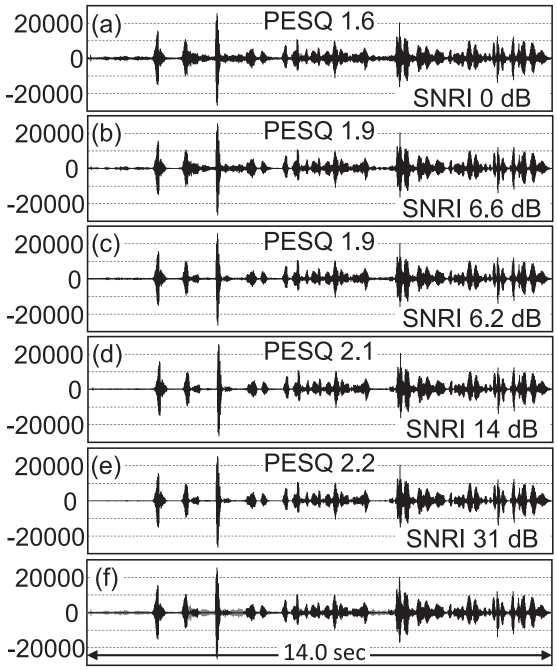

Fig. 7. Enhanced signal for babble noise. (a) Noisy signal (speech + babble noise), (b) Conventional NS [$G_d(l,k) = 1$], (c) Directional NS with constant passband [$G_f(l,k) = 1$], (d) Directional NS with frequency proportionate variable passband [$G_f(l,k) = 1$], (e) Directional NS [$G_f(l,k) \neq 1, G_d(l,k) \neq 1$], (f) Clean speech (black) and babble noise (gray).

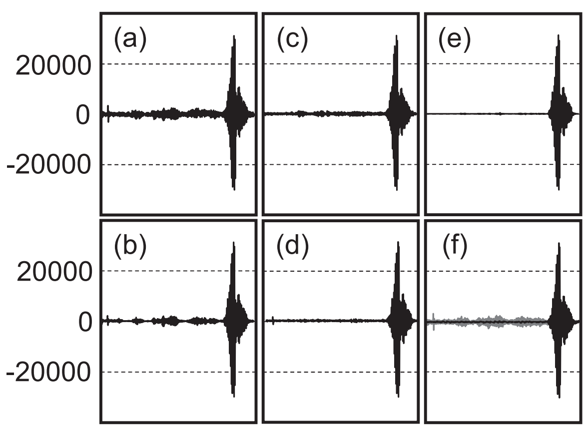

Fig. 8. Close-up comparison of Fig. 7 for the first 4 sec.

SNR of 15.6 dB. The recorded signal looks like Fig. 7 (a). A spectral gain $G_f(l,k)$ was calculated by [3]. Enhanced signals are compared in Fig. 7 (b)-(f) with PESQ and SNR improvement (SNRI) [37].

The conventional spectral NS output with $G_d(l,k) = 1$ in (b) is comparable to the proposed directional NS output with $G_f(l,k) = 1$ in Fig. 7 (c). This is because the babble noise played back by a single loudspeaker can be considered directional. A better PESQ and an SNRI in the enhanced signal of (d) over (c) exhibits the effect of frequency proportionate variable passband. (d) has much smaller residual noise than (c), demonstrating that the passband should be designed to be frequency proportionate. (e) is the enhanced signal by both directional and spectral gains, where the effective gain for the noisy signal is $G_d(l,k)G_s(l,k)$. (e) provides the best quality of

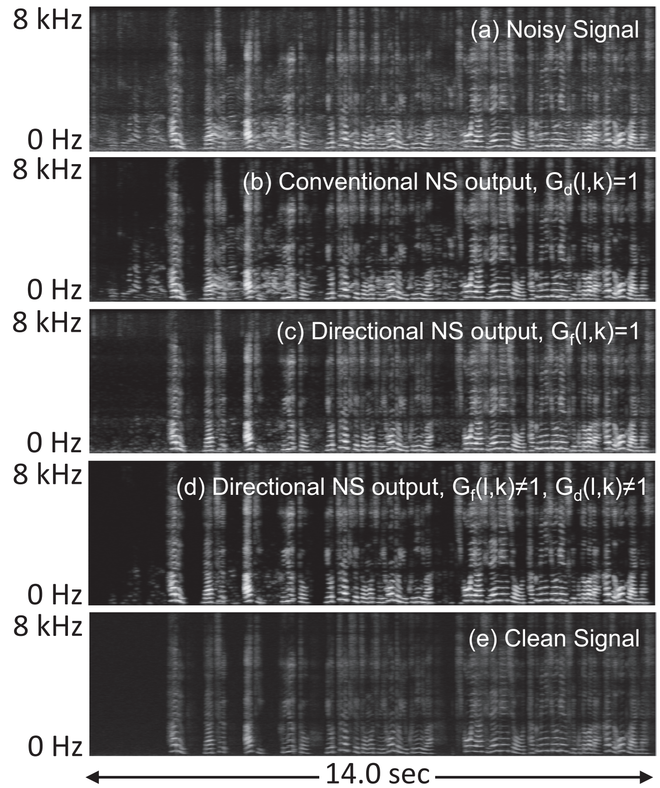

Fig. 9. Enhanced signal for babble noise in spectrogram. (a) Noisy signal (speech + babble noise), (b) Conventional NS [$G_d(l,k) = 1$], (c) Directional NS with variable passband width [$G_f(l,k) = 1$], (d) Directional NS [$G_f(l,k) \neq 1, G_d(l,k) \neq 1$], (e) Clean speech.

the enhanced signal in Fig. 7. It should be noted that the proposed directional NS in (e) provides much more significant improvement on SNRI than PESQ over conventional spectral NS in (b). Directional gain contributes more to suppression of directional noise. The proposed directional NS improves PESQ by 0.3 and increases SNRI by 24.4 dB over the conventional spectral NS. For more detailed comparison, the first non-speech and speech sections are highlighted in Fig. 8, (a) through (f).

Figure 9 compares output spectrogram by different settings of the proposed directional NS. A directional NS output in (c) clearly has smaller residual noise in non-speech sections than the conventional spectral NS output in (b). The proposed directional NS with combined spectral and directional gains provides the best result that is comparable to the clean signal in (e) .

4. CONCLUSION

A directional noise suppressor with a specified constant beamwidth has been proposed. A directional gain has been obtained from DOA of the target signal or, equivalently, interchannel phase difference. The directional gain has been designed in advance and multiplied by the noisy signal with a spectral gain. A frequency proportionate variable passband has been incorporated in the directional gain to provide constant beamwidth over frequency. Recorded signals by a commercial PC with two microphones has demonstrated good agreement between the theoretical and the measured directivity. It has been shown that signal-to-noise ratio improvement is 24.4 dB higher and PESQ is 0.3 improved compared to conventional single-channel noise suppressor with a babble noise.

5. REFERENCES

[1] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. ASSP, vol.27, no. 2, pp. 113–120, Apr. 1979.

[2] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-32, no. 6, pp. 1109–1121, Dec. 1984.

[3] M. Kato, A. Sugiyama and M. Serizawa, "Noise suppression with high speech quality based on weighted noise estimation and MMSE STSA," Proc. IWAENC2001, pp. 183–186, Sep. 2001.

[4] J. Taghia, J. Taghia, N. Mohammadiha, J. Sang, V. Bouse, and R. Martin, "An evaluation of noise power spectral density estimation algorithms in adverse acoustic environments," Proc. ICASSP2011, pp.4640–4643, May 2011.

[5] M.-S. Choi and H.-G. Kang, "A two-channel minimum mean-square error log-spectral amplitude estimator for speech enhancement," Proc. HSCMA2008, pp.152–155, May 2008.

[6] J. Freudenberger, S. Stenzel and B. Venditti, "A noise PSD and cross-PSD estimation for two-microphone speech ehancement systems," Proc. SSP2009, pp.709–712, Aug. 2009.

[7] S. -Y. Jeong, K. Kim, J. -H. Jeong, K. -C. Oh, and J. Kim, "Adaptive noise power spectrum estimation for compact dual channel speech enhancement," Proc. ICASSP2010, pp.1630–1633, Apr. 2010.

[8] K. Kim, S. -Y. Jeong, J. -H. Jeong, K. -C. Oh, and J. Kim, "Dual channel noise reduction method using phase difference-based spectral amplitude estimation," Proc. ICASSP2010, pp.217–220, Apr. 2010.

[9] N. Yousefian and P. C. Loizou, "A dual-microphone speech enhancement algorithm based on the coherence function," IEEE Trans. ASLP, Vol. 20, No. 2, pp.599–609, Feb. 2012.

[10] M. Jeub, C. Herglotz, C. Nelke, C. Beaugeant, and P. Vary, "Noise reduction for dual-microphone mobile phones exploiting power level differences," Proc. ICASSP2012, pp.217–220, Mar. 2012.

[11] J. Zhang, R. Xia, Z. Fu, J. Li, and Y. Yan, "A fast two-microphone noise reduction algorithm based on power level ratio for mobile phone," Proc. ICSLP2012, pp.206–209, Dec. 2012.

[12] Z.-H. Fu, F. Fan and J. -D. Huang, "Dual-microphone noise reduction for mobile phone application," Proc. ICASSP2013, pp.7239–7243, May 2013.

[13] J. Taghia, R. Martin, J. Taghia and A. Leijon, "Dual-channel noise reduction based on a mixture of circular-symmetric complex Gaussians on unit hypersphere," Proc. ICASSP2013, pp.7289–7293, May 2013.

[14] B. Widrow, J. R. Glover, Jr., J. M. McCool, J. Kaunitz, C. S. Williams, R. H. Hearn, J. R. Zeidler, E. Dong, Jr., R. C. Goodlin: Adaptive noise cancelling: principles and applications, *Proc. IEEE*, **63**, (12), pp.1692–1716, 1975.

[15] A. Sugiyama, "Low-distortion noise cancellers – Revival of a classical technique," *Speech and audio processing in adverse environment*, Chap. 7, Hänsler and Schmidt, ed. Springer, 2008.

[16] X. Zhang, H. Zeng, and A. Lunardhi, "Noise estimation based on an adaptive smoothing factor for improving speech quality in a dual-microphone noise suppression system," Proc. ICSPCS2011, pp.1–5, Dec. 2011.

[17] A. Sugiyama, M. Kato, and M. Serizawa, "A low-distortion noise canceller with an SNR-modified partitioned power-normalized PNLMS algorithm," Proc. APSIPA ASC 2009, pp.222–225, Oct. 2009.

[18] A. Sugiyama, and R. Miyahara, "A low-distortion noise canceller with a novel stepsize control and conditional cancellation," Proc. EUSIPCO2014, Sep. 2014.

[19] M. Brandstein and D. Ward (Eds), "Microphone arrays," Springer, 2001.

[20] W. Herbordt and W. Kellermann, "Adaptive beamforming for audio signal acquisition," *Adaptive signal processing, Applications to real-world problems*, Chap. 6, Benesty and Huang, ed. Springer, 2003.

[21] G. W. Elko and J. Meyer, "Microphone arrays," *Handbook of speech processing*, Chap. 50, Benesty, Sondhi, and Huang, ed. Springer, 2008.

[22] J. Benesty, Y. Huang, and J. Chen (Eds), "Microphone array signal processing," Springer, 2008.

[23] L. J. Griffiths and C. W. Jim, "An alternative approach to linear constrained adaptive beamforming," IEEE Trans. Antennas and Propagations, vol. AP-30, no. 1, pp.27–34, Jan. 1982.

[24] A. Sugiyama and R. Miyahara, "Phase randomization, - A new paradigm for single-channel signal enhancement," Proc. ICASSP2013, pp.7487–7491, May 2013.

[25] R. P. Smith, "Constant beamwidth receiving arrays for broadband sonar systems," Acustica 23:21–26, 1970.

[26] E. L. Hixon and K. T. Au, "Wide-bandwidth constant bandwidth acoustic array," J. of the Acoust. Soc. of America, 48(1):117, July 1970.

[27] M. M. Goodwin and G. W. Elko, "Beam dithering," Proc. Audio Eng. Soc., Oct. 1992.

[28] M. M. Goodwin and G. W. Elko, "Constant beamwidth beamforming," Proc. ICASSP'93, pp.169–172, Apr. 1993.

[29] D. Tucker, "Arrays with constant beam-width over a wide frequency range," Nature, 180:496–497, Sep. 1957.

[30] T. Chou, "Frequency independent beamformer with low response error," Proc. ICASSP'93, pp.169–172, Apr. 1993.

[31] J. L. Flanagan, D. A. Berkeley, G. W. Elko, J. E. West, and M. M. Sondhi, "Autodirective microphone systems," Acustica, Vol. 73, pp.58–71, 1991.

[32] W. Kellermann, "A self-steering digital microphone array," Proc. ICASSP'91, pp.3581–3584, Apr. 1991.

[33] F. Khalil, J. P. Jullien, and A. Gilloire, "Microphone array for sound pickup in teleconference systems," J. Audio Eng. Soc., Vol. 42, No. 9, pp.691–700, Sep. 1994.

[34] A. Sugiyama and R. Miyahara, "A new generalized sidelobe canceller with a compact array of microphones suitable for mobile terminals," Proc. ICASSP2014, pp.820–824, May 2014.

[35] P. Aarabi and G. Shi, "Phase-based dual-microphone tobust speech enhancement," IEEE Trans. Syst., Man, and Cyber., Vol. 34, No. 4, pp.1763–1773, Aug. 2004.

[36] O. u. R. Qazi, B. v. Dijk, M. Moonen and J. Wouters, "Speech understanding performance of cochlear implant subjects using time-frequency masking-based noise reduction," IEEE Trans. Bio. Eng., Vol. 59, No. 5, pp.1364–1373, May 2012.

[37] "Voice enhancement devices," ITU-T G.160, Edition 2.0, Jun. 2012.