A DIRECTIONAL NOISE SUPPRESSOR WITH AN ADJUSTABLE CONSTANT BEAMWIDTH FOR MULTICHANNEL SIGNAL ENHANCEMENT

Akihiko Sugiyama and Ryoji Miyahara†

Information and Media Processing Labs., NEC Corporation
†Embedded System Division, NEC Engineering
1753, Shimonumabe, Nakahara-ku, Kawasaki-shi, Kanagawa 211–8666, JAPAN

ABSTRACT

This paper proposes a directional noise suppressor with an adjustable constant beamwidth for multichannel signal enhancement. A directional gain based on inter-channel phase difference is combined with a spectral gain commonly used in noise suppressors (NS). 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. Stereo perception is preserved by weighting stereo input with the common directional and spectral gain. Evaluation with signals recorded by a commercial PC demonstrates that the signal-to-noise ratio improvement and the PESQ score for the enhanced signal are equally improved in two channels by 26.1 dB and 0.2 over a conventional NS. ILD difference between the input and the output is small when the target-signal dominates the input signal.

Index Terms— Multichannel, Beamformer, Noise suppressor, Phase difference, Directional gain, Constant beamwidth

1. INTRODUCTION

A wide variety of applications which capture audio signals are exposed to different types of noise and interference. For any moderate, relatively stationary noise, single channel signal enhancement [1]-[6], or noise suppression (NS), is useful as a simple method. When the environment is more adverse such as lower signal-to-noise ratio (SNR) and nonstationary noise, dual-microphone solutions [7]-[20] are more suitable. Some types of noise or interference consist of point signal sources. In such a case, it is known that acoustic beamformers, also known as microphone arrays (MAs), are effective [21]–[25].

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 accommodate many microphones. In addition, MAs have a limitation from a viewpoint of constant beamwidth across frequency. Beams and nulls in a low frequency range are wider than those in a high frequency range due to a longer wavelength, leading to poor spatial selectivity. A solution to this problem is a combination of arrays of different sizes dedicated to different frequency ranges [26, 27]. A most common example is a harmonically-nested array [28]–[34]. Nevertheless, increase of the array size and the number of microphones imposed by a nested technique may not be acceptable for cost-and-space-conscious consumer products such as mobile phone handsets and personal computers (PCs).

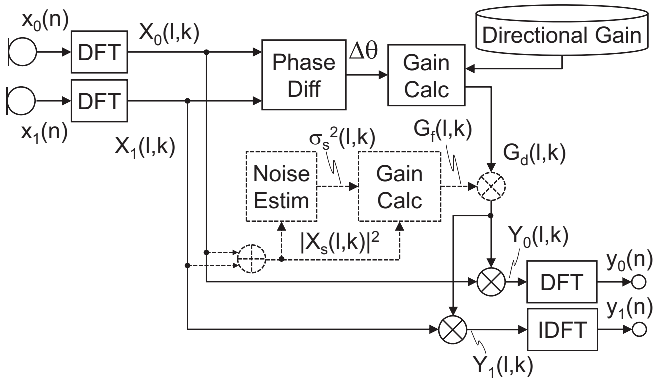

Fig. 1. Blockdiagram of proposed multichannel directional noise suppressor (a stereo example).

Moreover, the principle of MAs is delay adjustment and addition or subtraction. The output signal is always monaural independent of the number of microphones. Therefore, signal enhancement based on directivity formed by a conventional MA cannot be applied to multichannel signal acquisition. One of sharply increasing applications of multichannel signal acquisition is audio visual (AV) recording in personal mobile devices such as smartphones.

Phase-based time-frequency (T-F) filtering [35]-[37] with a small number of microphones, typically two, can be a possible solution to this selectivity degradation and multichannel beamforming. It calculates a directional gain based on some phase information obtained from multichannel input signals. Aarabi et al. [35] uses a phase-difference error between two microphone signals 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. Sugiyama et al. [37] presents a directional gain design method which guarantees a constant beamwidth along frequency. In any phase-based T-F filtering method, a directional gain is calculated from some phase information with no addition/subtraction of multichannel input signals. The obtained gain can be applied to all multichannel signals for signal selectivity to achieve multichannel beamforming.

This paper proposes a directional noise suppressor with an adjustable constant beamwidth for multichannel signal enhancement. The following section presents a multichannel beamforming framework followed by a constant-beamwidth directional gain design. In Section 3, evaluation results are presented with respect to the signal quality for speech recognition compared to a conventional NS and preservation of multichannel sound image.

2. MULTIHANNEL DIRECTIONAL NOISE SUPPRESSOR

2.1. Overall Structure

Figure 1 illustrates a blockdiagram of the proposed multichannel directional noise suppressor. Although the number of channels can be determined arbitrary, this figure assumes the simplest case with two channels. It calculates both a spectral gain $G_f(l,k)$ and a directional gain $G_d(l,k)$, where l and k are the frame and the frequency index. Each input signal $x_m(n)$ in channel m is transformed into a frequency-domain signal $X_m(l,k)$ by discrete Fourier transform (DFT). 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 calculated with M being the number of channels as

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

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]-[6] or a null beamformer can be used to obtain a noise power estimate $\sigma_s^2(l,k)$. 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]. The directional gain $G_d(l,k)$, which is large in the vicinity of a look direction and small otherwise, is designed in advance and stored in memory. It suppresses all signal components other than the target signal. Examples are given as a solid line (a) and a dashed line (b) in Fig. 2.

Target signal components are identified by direction-of-arrival (DOA) represented by interchannel phase difference $\Delta\theta(l,k)$. Assuming a simplest case with $M=2$, $\Delta\theta(l,k)$ is given by

$$\Delta\theta(l,k) = \angle\{X_0(l,k)\cdot X_1^*(l,k)\} = \theta_0(l,k) - \theta_1(l,k), \quad (2)$$

where θ_0 and θ_1 are the phase of $X_0(l,k)$ and $X_1(l,k)$ and $*$ represents complex conjugate. Interchannel phase difference of multiple adjacent channels can be used to obtain more accurate phase difference for $M > 2$. $\Delta\theta(l,k)$ in a specified passband returns $G_d(l,k) = 1$ which performs no directional suppression. For other values of $\Delta\theta(l,k)$, $G_d(l,k) < 1$ is returned to suppress the interference accordingly.

The final enhanced signal in each frequency is obtained by multiplying each microphone signals $X_m(l,k)$ by two gain values as

$$Y_m(l,k) = G_f(l,k)G_d(l,k)X_m(l,k). \quad (3)$$

$Y_m(l,k)$ is applied an inverse DFT to obtain a time-domain enhanced signal in chnnel m.

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

A directional gain $G_d(l,k)$ is determined for each value of l and k based on the phase difference $\Delta\theta(l,k)$. Therefore, it will be expressed as $G_d(\Delta\theta(l,k))$. $G_d(\Delta\theta(l,k))$ is designed in advance such that the signal components coming from the look direction are passed and all others are sufficiently suppressed. Assuming that the look direction is perpendicular to the array surface, *i.e.* 0 degrees, $G_d(\Delta\theta(l,k))$ takes a value of unity around 0 degrees and a value 0 otherwise. A transition band may be applied for smooth connection between the passband and the stopband. The passband, transition bands, and stopbands can be arbitrary specified as design issues.

First, a directional gain $G_d(l,k_0)$ at a fundamental frequency k_0 is desined as shown in Fig. 2. With a passband edge DOA angle $\pm\phi$,

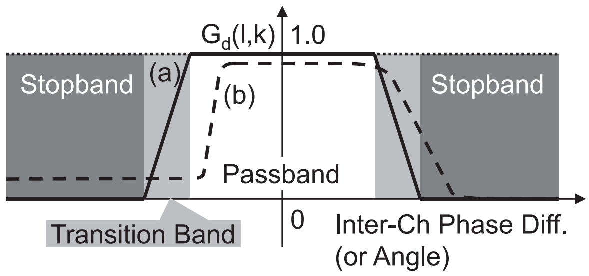

Fig. 2. Directional gain $G_d(l,k)$ examples.

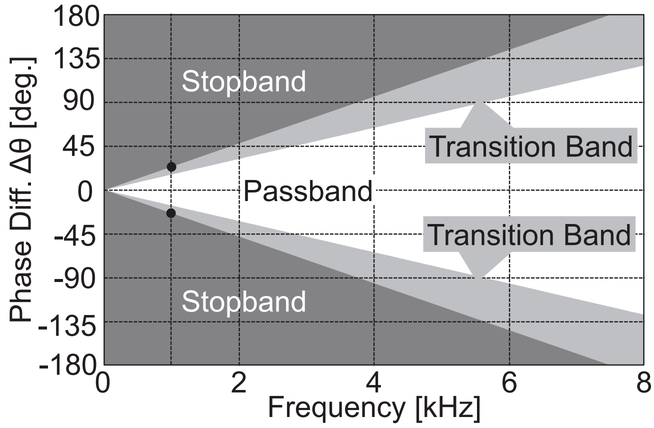

Fig. 3. Directional gain for frequency vs. phase difference $\Delta\theta$. Bandwidth of passband, transition band, and stopband are proportional to frequency. Passband edge: ± 20 deg., Stopband edge: ± 30 deg., both at 1 kHz.

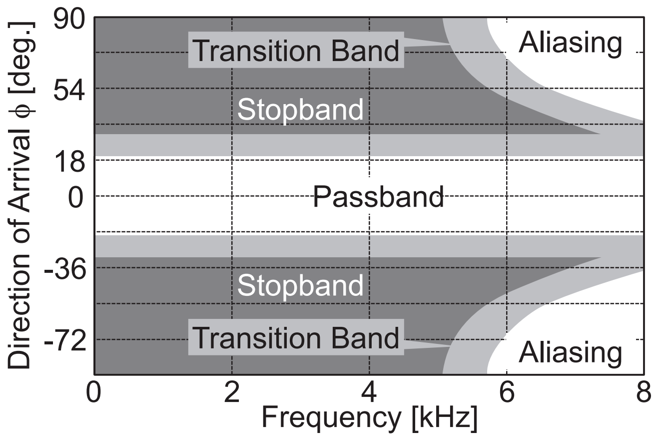

Fig. 4. Directional gain for frequency vs. DOA ϕ. Constant beamwidth is implemented by frequency-proportionate phase difference.

a phase difference $\Delta\theta(l,k_0)$ is given by

$$\Delta\theta(l,k_0) = 2\pi\frac{k_0}{N}\frac{d\sin\phi}{c}, \quad (4)$$

where N, d, and c are a DFT block size, a microphone spacing, and the sound velocity. Because the phase difference should be propor-

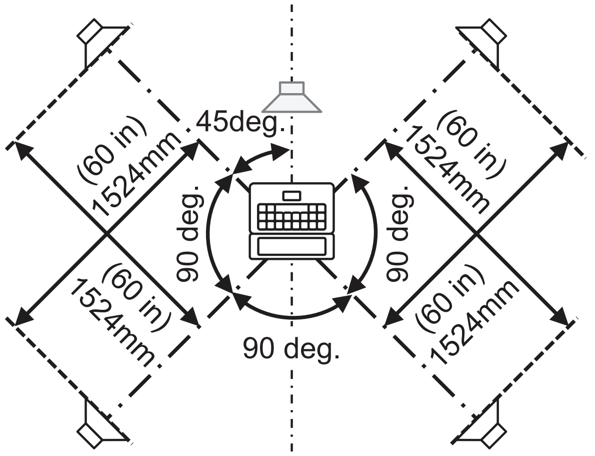

Fig. 5. Layout of four loudspeakers for noise source.

tional to the frequency for a constant beamwidth, the phase difference $\Delta\theta(l,k)$ for an arbitrary frequency k should be obtained by

$$\Delta\theta(l,k) = 2\pi\frac{k}{N}\frac{d\sin\phi}{c} = \frac{k}{k_0}\Delta\theta(l,k_0). \tag{5}$$

Substituting (5) in the original definition, the directional gain $G_d(\Delta\theta(l,k))$ at an arbitrary frequency k is given by

$$G_d(\Delta\theta(l,k)) \quad = \quad G_d(\frac{k}{k_0}\Delta\theta(l,k_0)). \tag{6}$$

With $G_d(l,k_0)$ and (6), a set of directional gains for $0 \leq k \leq N-1$ can be designed.

Figure 3 illustrates the directiona gain $G_d(l,k)$ for frequency vs. phase difference $\Delta\theta(l,k_0)$. A constant beamwidth makes the relationship between the frequency and the phase difference linear. In Fig. 3, the passband edge DOAs of $\phi = \pm 20$ degrees and the stopband edge DOAs of $\phi = \pm 30$ degrees are assumed at a fundamental frequency $k_0 = 1$ kHz. As an example, the phase difference $\Delta\theta(l,k_0)$ at stopband edges are calculated for $d = 4.5$ cm and $c = 346.3$ m/s. Equation (4) gives $\Delta\theta = 0.41$ radian for the stopband edges which corresponds to ± 24 degrees as marked by bullets in Fig. 3. A corresponding plot to Fig. 3 representing a constant beamwidth is shown in Fig. 4. Passband edges of ± 20 degrees and stopband edges of ± 30 degrees are observed as they were set in the design process. A phase difference $exp\{-j\Delta\theta\}$ may take the same value at multiple frequencies. This comes from the periodicity of the exponential function and appears as aliasing in Fig. 4.

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 with a reverberation time of 320 ms. The microphone spacing was 4.5 cm. The screen face was fixed with an angle of 110 degrees to its keyboard and the distance from the center of its screen hinges to a loudspeaker for target-speech radiation was set to 609.6 mm (24 in). Four loudspeakers were arranged for noise sources as illustrated in Fig. 5. An interfering speech signal was located 914.4 mm (36 in) away from the center of the screen hinges with an angle of 60 degrees to the look direction. The target signals consisted of 10 male and 10 female native English speakers. The target-signal-to-noise ratio was adjusted to 16 dB for the noise and 5 dB for the interfering speaker. A commercially available speech recognition engine was used.

Table 1. Parameters

Passband @ 1 kHz	± 30 degrees
Stopband @ 1 kHz	± 45 degrees
Passband gain	1.0
Stopband gain	0.3

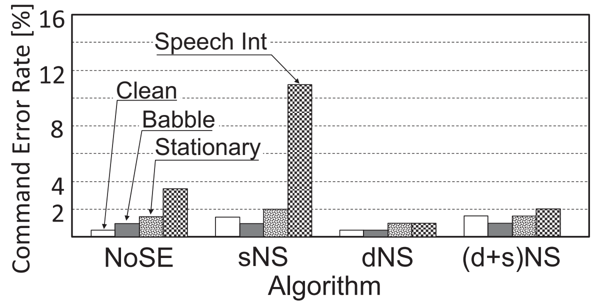

Fig. 6. Command error rate (CER) for 200 commands.

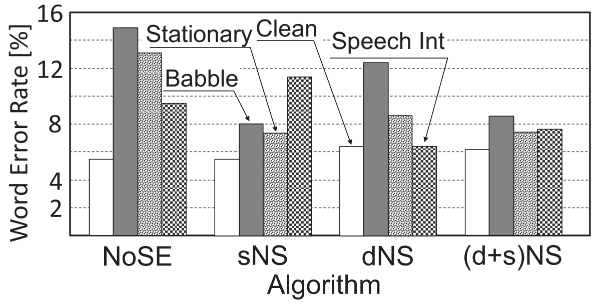

Fig. 7. Word error rate (WER) for dictation with 1185 words.

3.1. Signal enhancement for speech recognition

The recorded 2-channel signals were processed by the proposed multichannel directional NS with parameter settings shown in Tab. 1. A directional gain was designed with a constant beamwidth along frequency. The directional gain $G_d(k,l)$ and the spectral gain $G_f(k,l)$ were set to unity interchangeably to implement a conventional noise suppressor (spectral NS with $G_d(k,l) = 1$) and a pure directive selectivity (directional NS with $G_f(k,l) = 1$). Noise power was estimated by [5] and a spectral gain $G_f(l,k)$ was calculated by [2]. Evaluations were performed for four different conditions, namely, clean speech (Clean), babble noise (Babble), stationary noise (Stationary), and speech interference (Speech Int). They are to model an ideal environment, a party environment, a car environment, and an interfering-talker environment.

Figures 6 shows command error rate (CER) by no speech enhancement (NoSE), spectral NS (sNS), directional NS (dNS), and (d+s)NS. In any method, a short bar exhibits a low error rate and good performance. In case of CER, dNS achieves an error rate comparable to or lower than no processing. sNS provides almost comparable error rate to no processing except for speech interference. The error is even bigger than no processing, which means sNS introduces speech distortion whatever small it may be. This drawback is inherited in (d+s)NS with a degraded CER for clean speech. However, CER for speech interference is improved from NoSE, which is due to dNS.

A corresponding word error rate (WER) to CER is depicted in Fig. 7. dNS performs also well in the case of WER. However, for

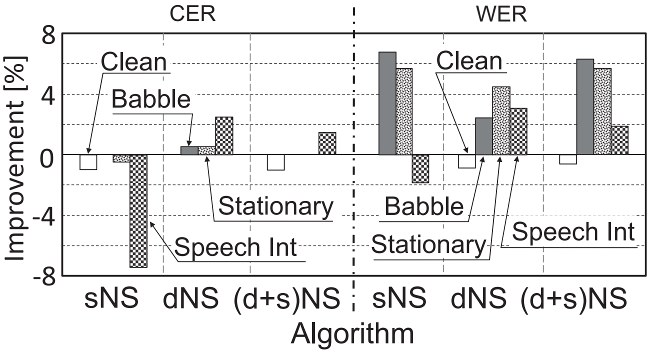

Fig. 8. CER and WER error-rate improvements.

clean speech, the error is increased by 1%, which is not significant for 1185 words in 200 dictations. Moreoer, the WER for babble noise by dNS is much higher than that by sNS. This is because babble noise interferes the target speech from non-look directions. It is again shown that sNS is weak against a directional interference with an increased error rate. Although the noise is suppressed, sNS introduces speech distortion like CER.

These characteristics are better demonstrated in Fig. 8 which shows CER and WER error-rate improvements for sNS, dNS, and (d+s)NS. Because this metric means a difference from the error by NoSE, there is no score for NoSE itself. It should be noted that a negative value represents degradation from NoSE. At a glance, it is not easy to see which is the best among sNS, dNS, or (d+s)NS. Slight degradation for clean speech is a drawback of sNS, while dNS is not very effective for diffuse noise such as babble noise. Overall, (d+s)NS is a good compromise over NoSE with negligible degradation for clean speech and sufficient improvement for babble noise.

3.2. Signal enhancement with sound localization

The evaluation scenario was slightly modified from the speech interference scenario by replacing the interfering speech by babble noise. A female speech sampled at 16 kHz was played back in front of the PC (0 degrees) with a measured SNR of 15.6 dB.

Figure 9 shows SNRI (signal-to-noise ratio improvement) and PESQ-I (PESQ improvement) for sNS (spectral NS with $G_d(l,k) = 1$), dNS (directional NS with $G_f(l,k) = 1$), and (d+s)NS (proposed). "Improvement" is defined as the score difference at the input and the output of processing. dNS achieves 13 dB SNRI and 0.4 PESQ-I, which are 9 dB and 0.1 better than sNS. SNRI and PESQ-I of (d+s)NS are 32 dB and 0.5 in both channels. It should be noted that the SNRI of sNS+dNS integrated structure is better than a simple addition of sNS and dNS SNRIs.

Effect on stereo presentation was evaluated by ILD (interaural level difference) difference between the input and the output signals with the target alone. Shown in Fig. 10 are ILD difference of a look direction at 0 degrees (a), at left 18 degrees (b), and left 36 degrees (c). In target-signal sections such as an area highlighted by a round-corner square in a dashed line, ILD difference is close to 0 for a 0-degree target direction. As the target direction is shifted toward left like 18 and 36 degrees, ILD difference exhibits larger fluctuations. When the target direction is at left 36 degrees, ILD difference significantly fluctuates over 2 dB. This result indicates that stereo image is well preserved when the target is in the look direction, *i.e.*, around 0 degrees. Otherwise, beam steering should be applied and a similar result of stereo presentation to the 0 degree look direction is expected.

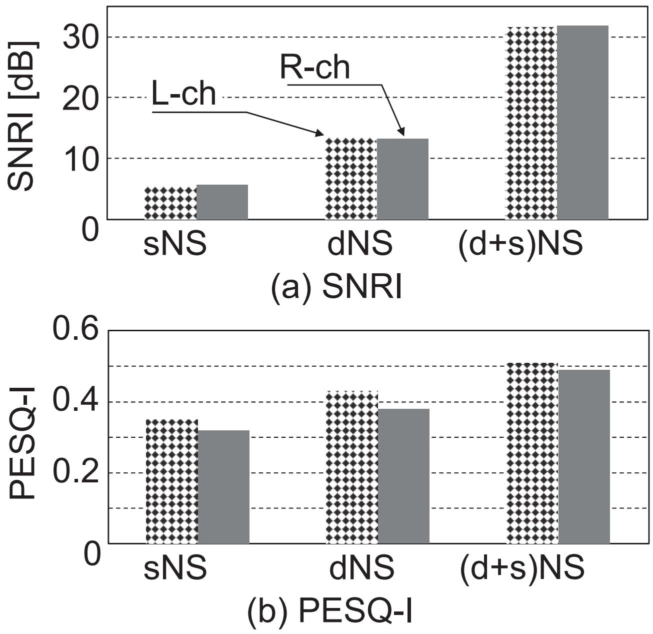

Fig. 9. Comparison of SNRI and PESQ-I. dNS: Directional NS, sNS: Spectral NS.

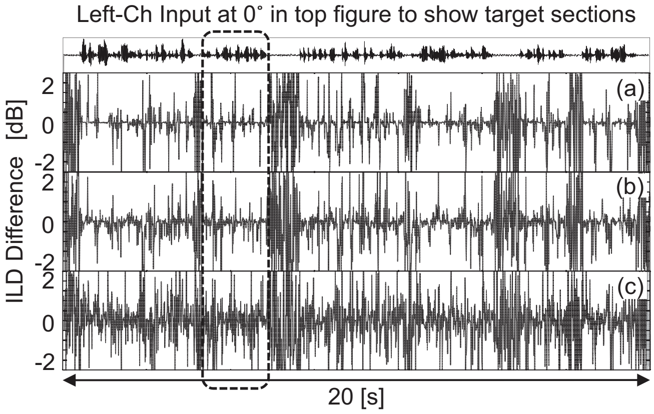

Fig. 10. ILD difference between input and output. Target female voice at 0 deg.(a), left 18 deg.(b), left 36 deg.(c).

4. CONCLUSION

A directional noise suppressor with an adjustable constant beamwidth for multichannel signal enhancement has been proposed. A directional gain based on inter-channel phase difference has been combined with a spectral gain commonly used in single-channel noise suppressors (NS). A design procedure of a directional gain to implement constant beamwidth has been established as specified passband edges and its constraint. A monaural structure has been extended to apply the directional gain to all channel signals so that multichannel perception is preserved. Evaluation with signals recorded by a commercial PC has demonstrated that the signal-to-noise ratio improvement (SNRI) and the PESQ score for the enhanced signal are equally improved in two channels by 26.1 dB and 0.2 over a conventional NS. ILD difference between the input and the output has shown to be small in target-signal sections which demonstrate good preservation of multichannel perception.

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] R. Martin, "Spectral subtraction based on minimum statistics," EUSIPCO'94, pp.1182–1185, Sep. 1994.

[4] I. Cohen, "Noise Spectrum Estimation in Adverse Environments: Improved Minima Controlled Recursive Averaging," IEEE Trans. Speech and Audio Processing, Vol. 11, No. 5, pp. 466–475, Sep. 2003.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

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

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

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

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

[22] 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.

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

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

[25] 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.

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

[27] 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.

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

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

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

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

[32] 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.

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

[34] 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.

[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] A. Sugiyama and R. Miyahara, "A directional noise suppressor with a specified beamwidth," Proc. ICASSP2015, to appear, Apr. 2015.

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