AUTOMATIC GAIN CONTROL WITH INTEGRATED SIGNAL ENHANCEMENT FOR SPECIFIED TARGET AND BACKGROUND-NOISE LEVELS

Akihiko Sugiyama and Ryoji Miyahara†

Data Science Research Laboratories, NEC Corporation
†Embedded System Division, NEC Engineering, Ltd.
1753 Shimonumabe, Nakahara-ku, Kawasaki 211-8666, Japan

ABSTRACT

This paper proposes automatic gain control with integrated signal enhancement for specified target and background-noise levels. The input signal is separated into a target signal and background noise by signal enhancement. The processing gains for the target and the noise are calculated from their estimated magnitudes and desirable output levels. After independent amplification/attenuation of the target signal and the noise, they are summed up to construct the output signal with their specified levels. Because of independent gain control, a more suitable signal for user needs can be obtained. Evaluation results demonstrate that the proposed gain control provides as much as 6 dB better gain than a commercial voice recorder with comparable signal quality.

Index Terms— AGC, Gain control, Distant signal, Noise level, Signal enhancement

1. INTRODUCTION

Gain control is a versatile function which is applied to a wide variety of products [1, 2]. An input signal is amplified or attenuated with a factor of $G_s(l)$ with a time index l so that any signal will have a power that is appropriate for a specific application. Typical applications include recording of audio signals arriving at a microphone from different distances, teleconferencing with several participants seated around the table [3], speakerphones [4], voice control systems [5], word spotting [6], and hearing aids [7].

When gain control is applied to a speakerphone, it is necessary to amplify a distant speech to make it comparable in power to a near-field speech. One of the most popular approaches is a nonlinear input-output gain function with voice activity detection [4, 5]. In order to avoid overflow in fixed-point arithmetic implementation, a large input signal is assigned a smaller gain than unity. However, design of the input-output gain function should reflect the input-signal dynamic range and is not an easy task. The overall performance is dependent on insufficient voice activity detection.

In noisy environment, if a gain is simply applied to the distant signal, both the target signal and the background noise are equally amplified. As a result, the noise components are more audible and annoying although the target signal is amplified and easier to listen to. This degradation of the amplified signal is more noticeable in non-speech sections where only the amplified noise exists. It is therefore important to keep the background-noise level equal to or smaller than that before amplification. On the other hand, the target-signal level should be adjusted independent of the background-noise level. However, there is no literature about gain control that enables independent control of the target signal and the background noise levels. Taking into account recent evolution of signal enhancement

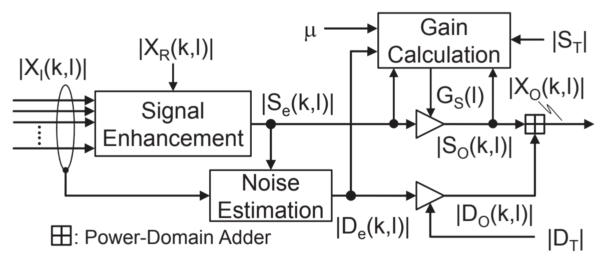

Fig. 1. Proposed automatic gain control for specified signal levels: A general structure.

with a single [8]-[12], dual [13]-[27], and multiple microphone scenarios [28]–[33], it is possible to separate the target signal and the noise and apply independent gains for optimum gain control.

This paper proposes automatic gain control with integrated signal enhancement for specified target and background-noise levels. The following section discusses a general structure of the proposed gain control with an emphasis on gain calculation for the target signal, which can be integrated with various types of signal enhancement, followed by a specific gain control structure with an integrated noise suppressor as a simple example. In Section 3, evaluation results are demonstrated to show superior control capability and comparable signal quality to a commercial voice recorder.

2. PROPOSED AUTOMATIC GAIN CONTROL

2.1. General Structure

A general structure of the proposed gain control is depicted in Fig. 1. It consists of Signal Enhancement, Noise Estimation, Gain Calculation, and gain multiplications for the target signal and the background noise. Assuming frequency-domain processing including gain control, Fig. 1 represents processing in one frequency bin between a forward and an inverse transform pair such as Fourier transform. An input signal $X_I(k,l)$ is defined as a sum of a target signal $S(k,l)$ and a noise signal (*i.e.* non-target) $D(k,l)$ as

$$X_I(k,l) = S(k,l) + D(k,l), \tag{1}$$

where k and l represent a frequency and a frame index. The input signal magnitude $|X_I(k,l)|$ may be provided in multiple channels for multichannel signal enhancement. One or more "reference" noise signals like $|X_R(k,l)|$ may be available for adaptive noise cancellation to estimate the true noise accurately.

At the output of Signal Enhancement, an enhanced target signal or a target signal estimate $|S_e(k,l)|$ is obtained. Assuming that $S(k,l)$ and $D(k,l)$ are not correlated, it is straightforward to calculate a noise or interference estimate $|D_e(k,l)|$ based on (2) once

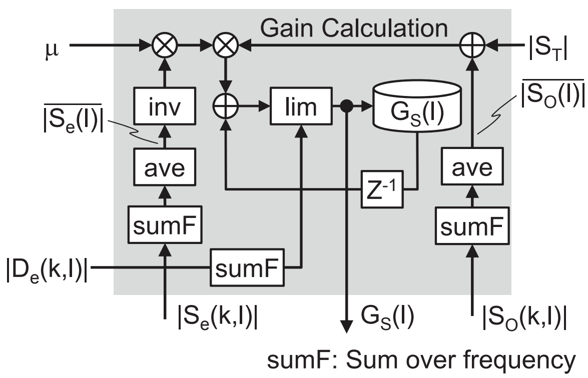

Fig. 2. Gain calculation for target signal.

$|S_e(k,l)|$ becomes available.

$$\mathrm{E}[|X_I(k,l)|^2] = \mathrm{E}[|S(k,l)|^2] + \mathrm{E}[|D(k,l)|^2]. \tag{2}$$

Calculation of a processing gain $G_D(l)$ for the noise estimate $|D_e(k,l)|^2$ is easier than $G_s(l)$ with a given desirable level $|D_T|$. This is because an estimated noise is relatively stationary by some averaging in estimation. $|D_e(k,l)|$ is first averaged over frequency to make it a frequency independent value $|D_e(l)|$ as

$$|D_e(l)| = \frac{1}{k2-k1+1}\sum_{k=k1}^{k2}|D_e(k,l)|, \tag{3}$$

$$\overline{|D_e(l)|} = \lambda\overline{|D_e(l-1)|} + (1-\lambda)|D_e(l)|, \tag{4}$$

with a constant λ. k_1 and k_2 may span the fullband or a subband which has a narrower bandwidth. The spectral shape of the background noise should be maintained by a frequency independent gain $G_D(l)$ and its level should change slowly with time.

$$G_D(l) = |D_T|/\overline{|D_e(l)|}, \tag{5}$$

so that a gain-controlled noise $|D_O(k,l)|$ is given by

$$|D_O(k,l)| = G_D(l)\cdot|D_e(k,l)| \approx |D_T|. \tag{6}$$

Practically, the desirable noise level is set as

$$|D_T| = \alpha_D\cdot\overline{|D_e(l)|}, \tag{7}$$

with $\alpha_D \leq 1$ so that the noise is not amplified.

A processing gain $G_s(l)$ for the target signal is more complex as is explained in the following subsection. $G_s(l)$ is commonly used for all frequencies to maintain the original spectral shape of the target signal, thus, it has no frequency index. With $G_s(l)$, a gain-controlled target $|S_O(k,l)|$ is given by

$$|S_O(k,l)| = G_s(l)\cdot|S_e(k,l)|. \tag{8}$$

The gain-controlled output signal $|X_O(k,l)|$ is given by

$$|X_O(k,l)| = \sqrt{|S_O(k,l)|^2 + |D_O(k,l)|^2}. \tag{9}$$

Because of this addition, imperfection of the target-noise separation is offset to some extent. $|X_O(k,l)|$ is combined with a phase, typically, a noisy phase of the input signal, and applied the inverse transform for construction of a time-domain gain-controlled signal.

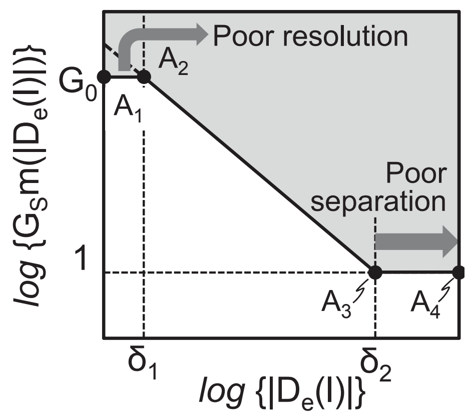

Fig. 3. Upper limit G_sm of $G_s(l)$ to avoid distortion by over-amplification.

2.2. Gain Calculation for Target Signal

A processing gain $G_s(l)$ for the target is calculated based on signals $|S_e(k,l)|$ and $|S_O(k,l)|$ before and after scaling as depicted in Fig. 2. $|S_e(k,l)|$ is averaged over frequency as

$$|S_e(l)| = \frac{1}{k2-k1+1}\sum_{k=k1}^{k2}|S_e(k,l)|, \tag{10}$$

It is then averaged over time to obtain its average $\overline{|S_e(l)|}$.

$$\overline{|S_e(l)|} = \lambda\overline{|S_e(l-1)|} + (1-\lambda)|S_e(l)|. \tag{11}$$

The actual value of λ is controlled with two settings to implement "fast attack and slow decay" as

$$\lambda = \begin{cases} \lambda_{up} & |S_e(l)| \geq \overline{|S_e(l-1)|} \\ \lambda_{down} & otherwise \end{cases}. \tag{12}$$

$\lambda_{up} \leq \lambda_{down}$ is maintained to make the tracking speed faster and to avoid overflow when the processing gain should decrease at an attack. $\overline{|S_O(l)|}$ is caluculated similarly from $|S_O(k,l)|$.

The processing gain $G_s(l)$ is updated by a single-tap NLMS algorithm with an averaged input signal $\overline{|S_e(l)|}$ and an averaged output signal $\overline{|S_O(l)|}$ as

$$G_s(l+1) = G_s(l) + \mu\cdot\frac{\overline{|S_e(l)|}\cdot\{\overline{|S_O(l)|} - |S_T|\}}{\overline{|S_e(l)|}^2}, \tag{13}$$

where $|S_T|$ is a desirable target signal level.

$G_s(l)$ is further applied an upper limit $G_Sm(|D_e(l)|)$ to avoid over-amplification as illustrated in Fig. 3. Over-amplification results in distortions with degraded signal separation by a large noise magnitude ($\overline{A_3A_4}$) or with insufficient precision by a combination of small noise and target-signal magnitudes ($\overline{A_1A_2}$). Assuming poor signal separation for $|D_e(l)| \geq \delta_2$, $G_Sm(|D_e(l)|) = 1$ for such values of estimated noise $|D_e(l)|$ ($\overline{A_3A_4}$). For example, with $\delta_2 = 0.5\cdot|S_T|$, good signal separation is assumed for signal-to-noise ratios (SNRs) greater than 6dB. From A_3 toward left, $G_Sm(|D_e(l)|)$ can take a larger value as the noise decreases with a negative slope of $\overline{A_2A_3}$. For $|D_e(l)| \leq \delta_1$ (*i.e.* $\overline{A_1A_2}$), noise is small and big amplification is needed. It means that the target signal has a small magnitude and poor resolution for unlimited amplification. $G_Sm(|D_e(l)|)$ should have an upper limit G_0 which depends on the arithmetic precision N_B, a head room N_H, and a minimum target-signal level S_{min} given by N_B.

$$G_0\cdot S_{min} \leq 2^{N_B-1-N_H}. \tag{14}$$

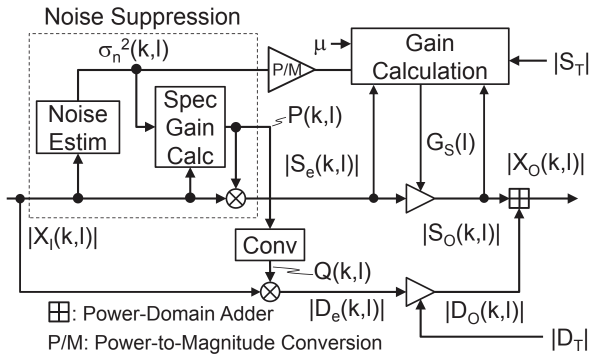

Fig. 4. Proposed automatic gain control for specified signal levels: A structure with an integrated noise suppressor.

N_H is for the variance of the target signal from its average $\overline{|S_e(l)|}$. Assuming that S_{min} is comparable to half the full-scale magnitude without N_H, *i.e.* $S_{min} = 2^{INT\{(N_B-1-N_H)/2\}}$ with an integer operator $INT\{\cdot\}$,

$$G_0 = 2^{N_B-1-N_H-INT\{(N_B-1-N_H)/2\}}. \tag{15}$$

From Fig. 3, $G_sm(|D_e(l)|)$ is given by

$$\log\{G_sm(|D_e(l)|)\} = -\frac{\log(G_0/1)}{\log(\delta_2/\delta_1)} \log \frac{|D_e(l)|}{\delta_1} + \log G_0, \tag{16}$$

and limited by

$$G_sm(|D_e(l)|) = \begin{cases} G_0 & G_sm(|D_e(l)|) > G_0 \\ 1 & G_sm(|D_e(l)|) < 1 \end{cases}. \tag{17}$$

$G_s(l)$ is finally limited by the following equation

$$G_s(l) = \min\{G_s(l), G_sm(|D_e(l)|)\}, \tag{18}$$

where $\min\{\cdot\}$ is a minimum-value operator.

2.3. Structure with an Integrated Noise Suppressor

After any signal enhancement, the enhanced signal should have a much higher SNR than the input noisy signal. For such a signal, a single channel enhancement is often applied as post processing. Therefore, design of automatic gain control integrated with a noise suppressor is essential.

Figure 4 shows a blockdiagram of the proposed automatic gain control integrated with a noise suppressor. It consists of Noise Estimation, Spectral Gain Calculation, and Gain Multiplications for the target and the background noise. Compared to the general structure in Fig. 1, Signal Enhancement is replaced with Noise Estim, Spec. Gain Calc, and a multiplier to generate $|S_e(k,l)|$. Noise Estimation is also replaced with Conv and a multiplier to generate $|D_e(k,l)|$.

An estimated noise magnitude, $\sqrt{\sigma_n^2(k,l)}$ is supplied to Gain Calculation in place of a reconstructed noise estimate $|D_e(k,l)|$. $\sigma_n^2(k,l)$ is generally better than $|D_e(k,l)|^2$ because (2) is valid based on an assumption that $S(k,l)$ and $D(k,l)$ are uncorrelated, which is often violated. Conv calculates a noise-separation gain $Q(k,l)$ from the target-separation gain $P(k,l)$ by

$$Q(k,l) = \sqrt{1 - P^2(k,l)}. \tag{19}$$

Other operations of Fig. 4, including Gain Calculation, are same as in Fig. 1.

Table 1. Parameters.

λ_{up}	0.87096	λ_{down}	0.93326
N_H	2	μ	0.03
$\lvert S_T \rvert$	-27 dBov	$\lvert D_T \rvert$	$0.4 \times \lvert D_e(l) \rvert$ ($\alpha_D = 0.4$)
δ_1	δ_2/G_0	δ_2	$0.5 \times \lvert S_T \rvert$ (-33 dBov)

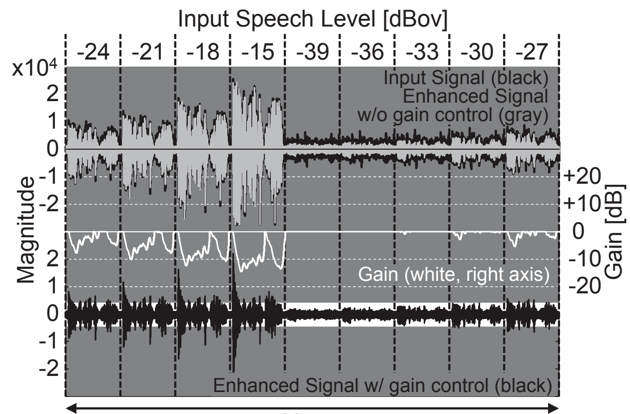

Fig. 5. Gain and signals with a noise level of -27 dBov.

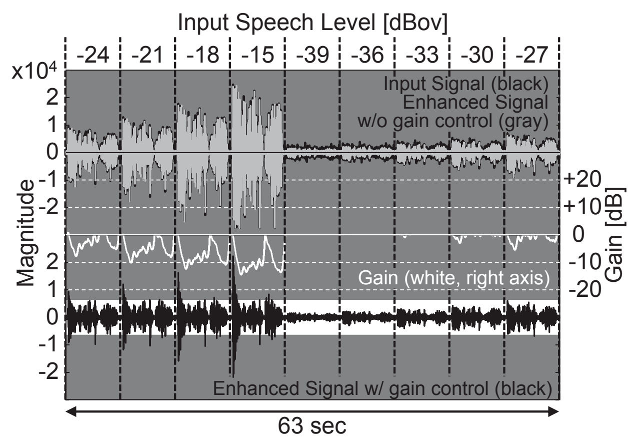

Fig. 6. Gain and signals with a noise level of -33 dBov.

3. EVALUATION

Evaluations were performed using a recorded speech and a recorded noise signal sampled at 48 kHz. The speech level was changed with time to implement different signal levels before being mixed with noise with different levels. A noise suppressor [12], with its demonstrated good performance, was used for signal enhancement to enable independent adjustment of the target signal level and the noise level. The frame size was set to 480 with a 50% overlap and 64 sample zero padding to apply a 1024 point DFT (discrete Fourier transform). N_B was set to 16. Parameters are summarized in Tab. 1.

3.1. Evaluation for Different Noise Levels with Changing Target-Signal Levels

Figures 5 through 8 demonstrate an input signal with no scaling (black, top), an enhanced (noise-suppressed) signal with no gain (gray, top), a gain $G_s(l)$ for the target signal (white, center), and an enhanced signal with gain control (black, bottom) for noise levels of -27, -33, -39, and -45 dBov. The speech level was changed from -24 to -15 dBov and from -39 to -27 dBov (left to right). Because the speech level was increased from the center toward the right, continuing from the left toward the center, the enhanced signal

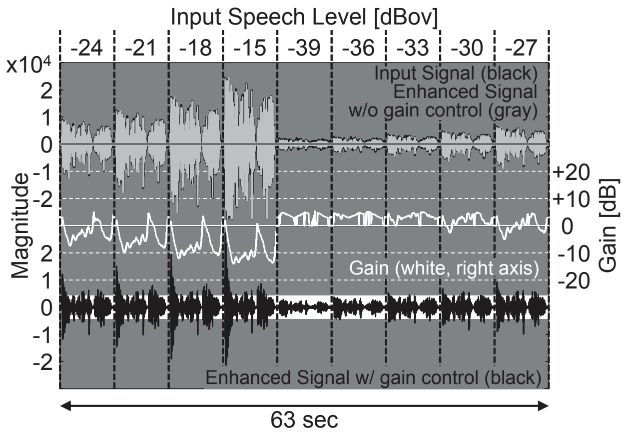

Fig. 7. Gain and signals with a noise level of -39 dBov.

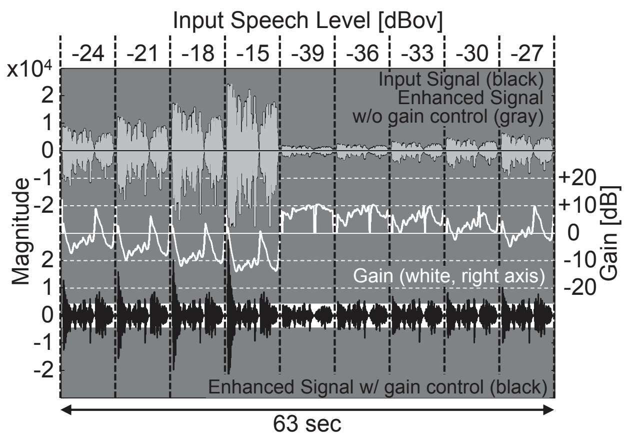

Fig. 8. Gain and signals with a noise level of -45 dBov.

in gray without gain control has a higher power level as it goes in that order. To compensate for this increase, the target signal gain $G_s(l)$ successfully takes a smaller value toward the right with a sharp jump at the center. Thus, the enhanced signal with gain control has almost constant power level on average. Figures 5 and 6, which represent a high noise level, exhibit the unit gain (0 dB) from the center to the right. This is because the target signal gain $G_s(l)$ is limited to avoid over-amplification of the speech as shown in $\overline{A_3A_4}$ of Fig. 3. A larger gain is observed for -39 to -45 dBov noise levels in Figs. 7 and 8. The enhanced signal in black at the bottom of the figures has almost constant power except big spikes in the left half figure. These spikes were generated by insufficient tracking speed of the target signal gain $G_s(l)$. Introduction of delay in the processing will remove these spikes. However, informal listening showed that too short spike-existing durations make them inaudible with no delay.

3.2. Comparison with a Commercial Product

The proposed gain control was compared to Voice Balancer on Olympus `VP-10` which was released in April 2015 and seems to have a nonlinear input-output gain. Other conventional methods do not disclose parameter values and are not available for comparison.

Figure 9 demonstrates (a) an input signal, (b) an output by Voice Balancer, and (c) an output by the proposed gain control for the noise-free case that is most preferable for Voice Balancer. To model distant talkers, multiple input speech levels were included. It is clear that the proposed gain control results in constant output power for a wide range of input speech levels. Shown in Tab. 2 are RMS (root-mean-square) output level of Voice Balancer in Fig. 9 normalized to that of the proposed gain control. The Voice Balancer output is 4.4 dB under-amplification and 1.6 dB over-amplification for input speech levels of -39 and -15 dB, respectively. This result indicates that Voice Balancer may result in 6 dB under-amplification for a distant talker when the speech level of the near-end talker is appropriately maintained. The proposed AGC does not commit over-amplification in such a case as demonstrated in Fig. 9 for a speech level of -15 dB.

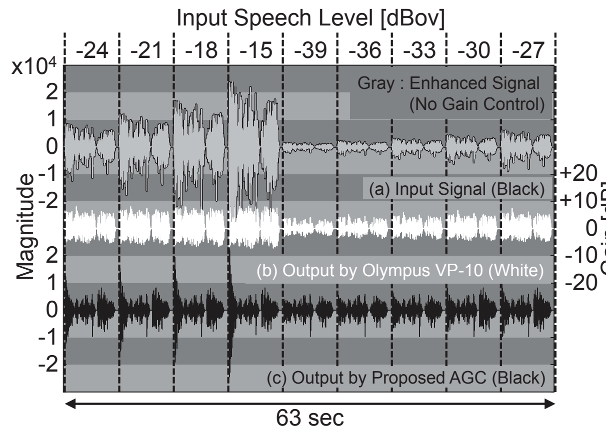

Fig. 9. Comparison with a commercial product by output signals. (a) Input signal with no gain control, (b) Output signal by a commercial voice recorder, (c) Output signal by proposed gain control.

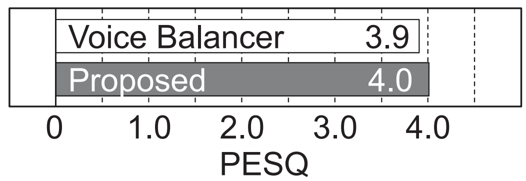

Fig. 10. PESQ comparison of the gain controlled signal in the noise-free condition.

Table 2. RMS output ratio of Voice Balancer to proposed gain control in dB. with values in Fig. 9.

Input Speech [dBov]	-39	-33	-27	-21	-15
RMS ratio to proposed	-4.4	-1.1	0	1.2	1.6

From a viewpoint of the amplified target-signal quality, PESQ scores for Voice Balancer and the proposed gain control are compared in Fig. 10 for the noise-free case to eliminate possible distortion by signal enhancement. The comparable scores demonstrate that the proposed gain control provides as much as 6 dB better distant talker amplification without sacrificing the local speech quality.

4. CONCLUSION

Automatic gain control with integrated signal enhancement for specified target signal and background-noise levels has been proposed. Thanks to target-noise separation and their independent gain control, an appropriate target-signal level has been achieved without affecting the background noise. Evaluation results have demonstrated superior gain control for a wide range of target-signal and noise levels. Comparison of output signals has shown that the proposed gain control provides as much as 6 dB better gain control with comparable signal quality to a commercial voice recorder.

5. REFERENCES

[1] G. R. Steber, "Digital signal processing in automatic gain control," IECON'88, pp. 381–384, Oct. 1988.

[2] D. Ellis, "Time-frequency automatic gain control," web resource, available: http://labrosa.ee.columbia.edu/matlab/tf-agc/ .

[3] P. L. Chu, "Voice-activated AGC for teleconferencing,h Proc of ICASSP96, pp. 929–932, Apr. 1996.

[4] A. Terasawa and H. Takeyama, "Automatic gain control sound system with automatically selected multiple gain functions," Tech. Rep. of Matsushita Elec. Works, pp.70–74, Feb. 2003 (in Japanese).

[5] I. I. Papp, Z. M. *Šarić*, N. Dj. Teslić, "Hands-free voice communication with TV," IEEE Trans. Consumer Elec., Vol. 57, No. 2, pp.606–614, May 2011.

[6] R. Prabhavalkar, R. Alvarez, C. Parada, P. Nakkiran, and T. N. Sainath, "Automatic gain control and multi-style training for robust small-footprint keyword spotting with deep neural networks,h Proc. ICASSP2015, Apr. 2015.

[7] P. P. Khing and B. A. Swanson, "The effect of automatic gain control structure and release time on cochlear implant speech intelligibility," PLoS ONE vol. 8, No. 11, Nov. 2013.

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

[9] 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.

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

[11] 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.

[12] M. Kato, A. Sugiyama, S. Serizawa, "A low-complexity noise suppressor with nonuniform subbands and a frequency-domain highpass filter," Proc. ICASSP2006, pp.473–476, May 2006.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

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

[20] 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.

[21] 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.

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

[23] 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. G. Shi, P. Aarabi and H. Jiang, "Phase-based dual-microphone speech enhancement using a prior speech model," IEEE Trans. ASLP, Vol. 15, No. 1, pp.109–118, Jan. 2007.

[24] S. Jeong, H. Yang, and M. Hahn, "Two-channel noise reduction for robust speech recognition in car environment," Electronics Letters, Vo. 44, No. 17, Aug. 2008.

[25] 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.

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

[27] A. Sugiyama and R. Miyahara, "A directional noise suppressor with a constant beamwidth for multichannel signal enhancement," Proc. EUSIPCO2015, Sep. 2015.

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

[29] O. Hoshuyama and A. Sugiyama, "Robust adaptive beamforming," Chap. 5, "Microphone arrays," M. Brandstein and D. Ward (Eds), Springer, 2001.

[30] 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.

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

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

[33] 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.