

A Noise Suppression Method for Speech Signal by Jointly Using Bayesian Estimation and Fuzzy Theory

Akira Ikuta¹, Hisako Orimoto¹, Kouji Hasegawa²

¹Department of Management Information Systems, Prefectural University of Hiroshima, Hiroshima, Japan

²Western Region Industrial Research Center, Hiroshima Prefectural Technology Research Institute, Kure, Japan

Email: ikuta@pu-hiroshima.ac.jp

How to cite this paper: Ikuta, A., Orimoto, H. and Hasegawa, K. (2021) A Noise Suppression Method for Speech Signal by Jointly Using Bayesian Estimation and Fuzzy Theory. *Journal of Software Engineering and Applications*, **14**, 631-645.
<https://doi.org/10.4236/jsea.2021.1412037>

Received: November 25, 2021

Accepted: December 28, 2021

Published: December 31, 2021

Copyright © 2021 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Speech recognition systems have been applied to inspection and maintenance operations in industrial factories to recording and reporting routines at construction sites, etc. where hand-writing is difficult. In these actual circumstances, some countermeasure methods for surrounding noise are indispensable. In this study, a new method to remove the noise for actual speech signal was proposed by using Bayesian estimation with the aid of bone-conducted speech and fuzzy theory. More specifically, by introducing Bayes' theorem based on the observation of air-conducted speech contaminated by surrounding background noise, a new type of algorithm for noise removal was theoretically derived. In the proposed noise suppression method, bone-conducted speech signal with the reduced high-frequency components was regarded as fuzzy observation data, and a stochastic model for the bone-conducted speech was derived by applying the probability measure of fuzzy events. The proposed method was applied to speech signals measured in real environment with low SNR, and better results were obtained than an algorithm based on observation of only air-conducted speech.

Keywords

Air- and Bone-Conducted Speeches, Noise Suppression, Bayesian Estimation, Fuzzy Data

1. Introduction

Speech recognition systems have been applied to various fields, for example, to inspection and maintenance operations in industrial factories and at construction sites, etc. where hand-writing is difficult. For speech recognition in such

actual circumstances, some suppression methods for surrounding noises are indispensable.

Previously reported methods for noise reduction in speech recognition can be classified into two categories. One is based on a single microphone [1] [2], and the other uses a microphone array [3]. Since the latter requires *a priori* information on the number of noise sources, and the number of microphones larger than that of the noise sources is needed in the case of multi-noise sources, this category demands large scale systems. Therefore, the former based on a single microphone is more advantageous than the latter [4] [5]. In such a noise suppression task for speech signals based on a single microphone, many algorithms applying the Kalman filter have been proposed up to now [6] [7] [8] [9]. However, the Kalman filter is originally based on the assumption of Gaussian white noise [10]. The actual noises show complex fluctuation forms with non-Gaussian and non-white properties.

From the above viewpoint, in our previously reported study, a noise suppression algorithm for the actual speech signals without requirement of the assumption of Gaussian white noise has been proposed [11]. The method can be applied to actual complex situation where both the noise statistics and the fluctuation forms of speech signal are unknown. By applying the algorithm to real speech signals with several kinds of noises, its effectiveness has been experimentally confirmed in comparison with the Kalman filter.

Furthermore, signal processing methods to remove the noise for actual speech signals have been proposed by jointly using the measured data of bone- and air-conducted speech signals [12] [13]. However, the algorithms of the previous methods were introduced a simple additive model of the original speech signal and surrounding noise for the air-conducted speech observation. Furthermore, the derived algorithms have applied to only the signals mixed with noises on computer, and not to signals in real environment under existence of noises.

In this study, a new noise suppression method for speech signals is proposed by using Bayes theorem after employing a posterior distribution based on the air-conducted speech observation contaminated by surrounding noise. In the proposed algorithm, in order to improve the accuracy of estimation of speech signal, an expansion expression of conditional probability density function reflecting all linear and non-linear correlation information between original speech signal and air-conducted speech observation is adopted as the model of the speech observation. Then, a probability distribution with parameters estimated from the bone-conducted speech is adopted as the prior distribution. Furthermore, the algorithm proposed in this study is applied to signals measured in real environment under existence of noises.

Though the bone-conducted speech signal is a kind of solid propagation sound with less effect by the surrounding noise, the high frequency components of the signal are reduced through the propagation process [14]. After considering the bone-conducted speech signal with the reduction of higher components as fuzzy data, applying the probability measure of fuzzy events [15], a new simplified

noise suppression method is derived by reflecting the air- and bone-conducted speech signals.

The effectiveness of the proposed method is confirmed by applying it to bone- and air-conducted speech measured in a real environment under the existence of surrounding noise.

2. Theoretical Consideration

2.1. Stochastic Model for Air- and Bone-Conducted Speech Signals by Introducing Fuzzy Theory

In the actual environment with a surrounding noise, let x_k , y_k and z_k be the original speech signal, the observations of air- and bone-conducted speech signals at a discrete time k . The observation y_k is contaminated by a surrounding noise v_k . In our previous studies, a simple additive model was considered for the air-conducted speech observation y_k [12] [13]. In this study, in order to improve the accuracy of estimation of speech signal x_k , an expansion expression of conditional probability density function $P(y_k | x_k)$ [11] reflecting all linear and non-linear correlation information between x_k and y_k is adopted as the model of air-conducted speech observation.

$$\begin{aligned} P(y_k | x_k) &= P(x_k, y_k) / P(x_k) \\ &= P(y_k) \sum_{r=0}^{\infty} \sum_{s=0}^{\infty} A_{rs} \theta_r^{(1)}(x_k) \theta_s^{(2)}(y_k) \end{aligned} \quad (1)$$

with

$$A_{rs} \equiv \left\langle \theta_r^{(1)}(x_k) \theta_s^{(2)}(y_k) \right\rangle, \quad (2)$$

where $\langle \rangle$ denotes the averaging operation on variables.

As the probability density functions x_k and y_k showing non-Gaussian distribution, the following statistical orthonormal expansion series expressions are adopted.

$$P(x_k) = N(x_k; \mu_x, \sigma_x^2) \sum_{i=0}^{\infty} B_i \frac{1}{\sqrt{i!}} H_i \left(\frac{x_k - \mu_x}{\sigma_x} \right), \quad (3)$$

$$P(y_k) = N(y_k; \mu_y, \sigma_y^2) \sum_{i=0}^{\infty} C_i \frac{1}{\sqrt{i!}} H_i \left(\frac{y_k - \mu_y}{\sigma_y} \right) \quad (4)$$

with

$$\begin{aligned} \mu_x &\equiv \langle x_k \rangle, \quad \sigma_x^2 \equiv \langle (x_k - \mu_x)^2 \rangle, \\ B_i &\equiv \left\langle \frac{1}{\sqrt{i!}} H_i \left(\frac{x_k - \mu_x}{\sigma_x} \right) \right\rangle, \quad B_0 = 1, \quad B_1 = B_2 = 0, \\ C_i &\equiv \left\langle \frac{1}{\sqrt{i!}} H_i \left(\frac{y_k - \mu_y}{\sigma_y} \right) \right\rangle, \quad C_0 = 1, \quad C_1 = C_2 = 0, \\ N(x; \mu, \sigma^2) &\equiv \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\}, \end{aligned} \quad (5)$$

where $H_i(\cdot)$ is a Hermite polynomial with i th order. Functions $\theta_r^{(1)}(x_k)$ and $\theta_s^{(2)}(y_k)$ are orthonormal polynomials having weighting functions $P(x_k)$ and $P(y_k)$, respectively. These orthonormal polynomials can be decomposed into linearly independent series as

$$\theta_r^{(1)}(x_k) = \sum_{i=0}^r \lambda_{ri}^{(1)} \frac{1}{\sqrt{i!}} H_i\left(\frac{x_k - \mu_x}{\sigma_x}\right), \tag{6}$$

$$\theta_s^{(2)}(y_k) = \sum_{i=0}^s \lambda_{si}^{(2)} \frac{1}{\sqrt{i!}} H_i\left(\frac{y_k - \mu_y}{\sigma_y}\right). \tag{7}$$

The coefficients $\lambda_{ri}^{(1)}$ and $\lambda_{si}^{(2)}$ are calculated beforehand by using Schmidt's orthogonalization algorithm [16]. The expansion coefficients A_{rs} with order $r \leq R, s \leq S$ can be obtained from the correlation relationship between original speech signal x_k and noisy observation of air-conducted speech y_k . Since the original speech signal is unknown in the presence of noise, these coefficients have to be estimated on the basis of the observation y_k . Let's regard the expansion coefficients A_{rs} as unknown parameter vector \mathbf{a} :

$$\mathbf{a} \equiv (a_{11}, \dots, a_{R1}, a_{12}, \dots, a_{R2}, \dots, a_{1S}, \dots, a_{RS}),$$

$$a_{rs} \equiv A_{rs}, \quad (r = 1, 2, \dots, R; s = 1, 2, \dots, S), \tag{8}$$

the following simple dynamical model is introduced for the simultaneous estimation of the parameters with the specific signal x_k :

$$\mathbf{a}_{k+1} = \mathbf{a}_k, \tag{9}$$

Next, in order to express the relationship between the original speech signal and bone-conducted speech, after regarding the bone-conducted speech as fuzzy data, the conditional probability distribution function $P(x_k | z_k)$ can be obtained by applying the probability measure of fuzzy events [15] to (1), as follows.

$$P(x_k | z_k) = P(x_k, z_k) / P(z_k)$$

$$= \int m_{\bar{y}_k}(y_k) P(x_k, y_k) dy_k / \int m_{\bar{y}_k}(y_k) P(y_k) dy_k \quad (\equiv N(x_k, z_k) / D(z_k)) \tag{10}$$

where $m_{\bar{y}_k}(y_k)$ is a membership function of the bone-conducted speech z_k , and a Gaussian type function:

$$m_{\bar{y}_k}(y_k) = \exp\{-\alpha(y_k - \bar{y}_k)^2\}, \quad (\bar{y}_k \equiv a + bz_k), \tag{11}$$

where a and b are constants and $\alpha (> 0)$ is a parameter, is adopted. Accordingly, by considering $P(x_k, y_k)$ in Equation (1) and $P(y_k)$ in Equation (4), and the membership function in Equation (11), the numerator of Equation (10) can be expressed as follows:

$$N(x_k, z_k) = P(x_k) \frac{e^{K_3}}{\sqrt{2K_1\sigma_y^2}} \int \left(\frac{1}{\sqrt{\pi/K_1}} \right) \exp\left\{-\frac{(y_k - K_2)^2}{1/K_1}\right\}$$

$$\cdot \sum_{i=0}^{\infty} C_i \frac{1}{\sqrt{i!}} H_i\left(\frac{y_k - \mu_y}{\sigma_y}\right) \sum_{r=0}^{\infty} \sum_{s=0}^{\infty} A_{rs} \theta_r^{(1)}(x_k) \theta_s^{(2)}(y_k) dy_k \tag{12}$$

with

$$K_1 \equiv (2\alpha\sigma_y^2 + 1) / (2\sigma_y^2), \quad K_2 \equiv (2\alpha\sigma_y^2 \bar{y}_k + \mu_y) / (2\alpha\sigma_y^2 + 1),$$

$$K_3 \equiv K_1 \left(K_2^2 - \frac{2\alpha\sigma_y^2 \bar{y}_k^2 + \mu_y^2}{2\alpha\sigma_y^2 + 1} \right). \quad (13)$$

After considering the equality on Hermite polynomial:

$$H_i \left(\frac{y_k - \mu_y}{\sigma_y} \right) = \sum_{j=0}^i d_{ij} H_j \left(\frac{y_k - K_2}{\sqrt{1/2K_1}} \right), \quad (14)$$

where d_{ij} are expansion coefficients reflecting bone-conducted speech signal, and using the orthonormal condition:

$$\int N(y_k; K_2, 1/2K_1) H_j \left(\frac{y_k - K_2}{\sqrt{1/2K_1}} \right) H_{j'} \left(\frac{y_k - K_2}{\sqrt{1/2K_1}} \right) dy_k = j! \cdot \delta_{jj'}, \quad (15)$$

the integral in Equation (12) can be calculated. Thus, the following expression is derived

$$N(x_k, z_k) = P(x_k) \frac{e^{K_3}}{\sqrt{2K_1\sigma_y^2}} \sum_{i=0}^{\infty} \frac{1}{\sqrt{i!}} C_i \sum_{r=0}^R \sum_{s=0}^S F_{st}(z_k) a_{rs,k} \theta_r^{(1)}(x_k), \quad (16)$$

$$F_{st}(z_k) \equiv \sum_{t=0}^a \sum_{j=0}^{\min\{i,t\}} \lambda_{st}^{(2)} \frac{1}{\sqrt{t!}} d_{ij} d_{ij} j!. \quad (17)$$

Furthermore, through the similar calculation process, the denominator of Equation (10) can be derived as follows:

$$D(z_k) = \frac{e^{K_3}}{\sqrt{2K_1\sigma_y^2}} G(z_k), \quad G(z_k) \equiv \sum_{i=0}^{\infty} \frac{1}{\sqrt{i!}} C_i d_{i0}. \quad (18)$$

Therefore, by substituting Equations (16) and (18) into Equation (10), the conditional probability distribution function $P(x_k | z_k)$ can be expressed explicitly.

2.2. Derivation of Noise Suppression Algorithm Based on Bayesian Estimation

To derive an estimation algorithm for the speech signal x_k , the Bayes' theorem for the conditional probability distribution [17] is first considered. Since the parameter \mathbf{a} is also unknown, the conditional joint probability distribution of x_k and \mathbf{a}_k is expressed as

$$P(x_k, \mathbf{a}_k | Y_k) = P(x_k, \mathbf{a}_k, y_k | Y_{k-1}) / P(y_k | Y_{k-1}), \quad (19)$$

where $Y_k (\equiv \{y_1, y_2, \dots, y_k\})$ is a set of air-conducted speech data up to time k . By expanding the conditional joint probability distribution $P(x_k, \mathbf{a}_k, y_k | Y_{k-1})$ in a statistical orthogonal expansion series on the basis of the well-known Gaussian distribution and calculating the conditional expectation, the estimates of x_k and $a_{rs,k}$ for mean can be derived as follows:

$$\begin{aligned} \hat{x}_k &\equiv \langle x_k | Y_k \rangle \\ &= \sum_{n=0}^{\infty} \{ B_{00n} E_{00}^{10} + B_{10n} E_{10}^{10} \} \frac{1}{\sqrt{n!}} H_n \left(\frac{y_k - y_k^*}{\sqrt{\Omega_k}} \right) \Big/ \sum_{n=0}^{\infty} B_{00n} \frac{1}{\sqrt{n!}} H_n \left(\frac{y_k - y_k^*}{\sqrt{\Omega_k}} \right) \end{aligned} \quad (20)$$

$$\begin{aligned} \hat{a}_{rs,k} &\equiv \langle a_{rs,k} | Y_k \rangle \\ &= \sum_{n=0}^{\infty} \{ B_{00n} E_{00}^{01} + B_{01n} E_{01}^{01} \} \frac{1}{\sqrt{n!}} H_n \left(\frac{y_k - y_k^*}{\sqrt{\Omega_k}} \right) \Big/ \sum_{n=0}^{\infty} B_{00n} \frac{1}{\sqrt{n!}} H_n \left(\frac{y_k - y_k^*}{\sqrt{\Omega_k}} \right) \end{aligned} \quad (21)$$

with

$$\begin{aligned} E_{00}^{10} &= x_k^* (\equiv \langle x_k | Y_{k-1} \rangle), \quad E_{10}^{10} = \sqrt{\Gamma_{x_k}}, \quad \Gamma_{x_k} \equiv \langle (x_k - x_k^*)^2 | Y_{k-1} \rangle, \\ E_{00}^{01} &= a_{rs,k}^* (\equiv \langle a_{rs,k} | Y_{k-1} \rangle), \quad E_{01}^{01} = \sqrt{\Gamma_{a_{rs,k}}}, \quad \Gamma_{a_{rs,k}} \equiv \langle (a_{rs,k} - a_{rs,k}^*)^2 | Y_{k-1} \rangle, \\ y_k^* &\equiv \langle y_k | Y_{k-1} \rangle, \quad \Omega_k \equiv \langle (y_k - y_k^*)^2 | Y_{k-1} \rangle, \end{aligned}$$

$$B_{lmn} \equiv \left\langle \frac{1}{\sqrt{l!}} H_l \left(\frac{x_k - x_k^*}{\sqrt{\Gamma_{x_k}}} \right) \prod_{r=0}^R \prod_{s=0}^S \frac{1}{\sqrt{m_{rs}!}} H_{m_{rs}} \left(\frac{a_{rs,k} - a_{rs,k}^*}{\sqrt{\Gamma_{a_{rs,k}}}} \right) \frac{1}{\sqrt{n!}} H_n \left(\frac{y_k - y_k^*}{\sqrt{\Omega_k}} \right) \Big| Y_{k-1} \right\rangle. \quad (22)$$

Furthermore, the estimate of $a_{rs,k}$ for variance is derived as follows:

$$\begin{aligned} P_{a_{rs,k}} &\equiv \langle (a_{rs,k} - \hat{a}_{rs,k})^2 | Y_k \rangle \\ &= \sum_{n=0}^{\infty} \{ B_{00n} E_{00}^{02} + B_{01n} E_{01}^{02} + B_{02n} E_{02}^{02} \} \frac{1}{\sqrt{n!}} H_n \left(\frac{y_k - y_k^*}{\sqrt{\Omega_k}} \right) \Big/ \sum_{n=0}^{\infty} B_{00n} \frac{1}{\sqrt{n!}} H_n \left(\frac{y_k - y_k^*}{\sqrt{\Omega_k}} \right) \end{aligned} \quad (23)$$

with

$$E_{00}^{02} = \Gamma_{a_{rs,k}} + (a_{rs,k}^* - \hat{a}_{rs,k})^2, \quad E_{01}^{02} = 2\sqrt{\Gamma_{a_{rs,k}}} (a_{rs,k}^* - \hat{a}_{rs,k}), \quad E_{02}^{02} = 2\Gamma_{a_{rs,k}}. \quad (24)$$

Using Equation (1) and the orthonormal property of $\theta_s^{(2)}(y_k)$, variables y_k^* and Ω_k in Equations (20) (21) and (23) can be calculated as follows:

$$\begin{aligned} y_k^* &= \left\langle \int y_k P(y_k | x_k) dy_k \Big| Y_{k-1} \right\rangle \\ &= \left\langle \sum_{r=0}^{\infty} \sum_{s=0}^1 e_{1s} A_{rs} \theta_r^{(1)}(x_k) \Big| Y_{k-1} \right\rangle \\ &= \sum_{r=0}^R \sum_{s=0}^1 e_{1s} a_{rs,k}^* \langle \theta_r^{(1)}(x_k) | Y_{k-1} \rangle \end{aligned} \quad (25)$$

$$\begin{aligned} \Omega_k &= \left\langle \int (y_k - y_k^*)^2 P(y_k | x_k) dy_k \Big| Y_{k-1} \right\rangle \\ &= \sum_{r=0}^R \sum_{s=0}^2 e_{2s} a_{rs,k}^* \langle \theta_r^{(1)}(x_k) | Y_{k-1} \rangle \end{aligned} \quad (26)$$

with

$$\begin{aligned} e_{10} &= \mu_y, \quad e_{11} = \sigma_y, \\ e_{20} &= f_{20} - \left(\frac{f_{21}}{\lambda_{11}^{(2)}} - \frac{f_{22}}{\lambda_{11}^{(2)} \lambda_{22}^{(2)}} \lambda_{21}^{(2)} \right) \lambda_{10}^{(2)} - \frac{f_{22}}{\lambda_{22}^{(2)}} \lambda_{20}^{(2)}, \\ e_{21} &= \frac{f_{21}}{\lambda_{11}^{(2)}} - \frac{f_{22}}{\lambda_{11}^{(2)} \lambda_{22}^{(2)}} \lambda_{21}^{(2)}, \quad e_{22} = \frac{f_{22}}{\lambda_{22}^{(2)}}, \end{aligned}$$

$$f_{20} = (\mu_y - y_k^*)^2 + \sigma_y^2, \quad f_{21} = 2\sigma_y(\mu_y - y_k^*), \quad f_{22} = \sqrt{2}\sigma_y^2. \quad (27)$$

Furthermore, by considering Equations (10) (16) (18) and orthonormal property of $\theta_r^{(1)}(x_k)$, variables x_k^* , Γ_{x_k} in Equation (22) and the conditional expectation in Equations (25) (26) can be calculated as follows:

$$\begin{aligned} x_k^* &= \left\langle \int x_k P(x_k | z_k) dx_k \middle| Y_{k-1} \right\rangle \\ &= \sum_{i=0}^{\infty} \frac{1}{\sqrt{i!}} C_i \sum_{r=0}^1 \sum_{s=0}^S h_{1r} F_{si}(z_k) a_{rs,k}^* / G(z_k) \end{aligned} \quad (28)$$

$$\begin{aligned} \Gamma_{x_k} &= \left\langle \int (x_k - x_k^*)^2 P(x_k | z_k) dx_k \middle| Y_{k-1} \right\rangle \\ &= \sum_{i=0}^{\infty} \frac{1}{\sqrt{i!}} C_i \sum_{r=0}^2 \sum_{s=0}^S h_{2r} F_{si}(z_k) a_{rs,k}^* / G(z_k) \end{aligned} \quad (29)$$

$$\begin{aligned} \langle \theta_r^{(1)}(x_k) \middle| Y_{k-1} \rangle &= \left\langle \int \theta_r^{(1)}(x_k) P(x_k | z_k) dx_k \middle| Y_{k-1} \right\rangle \\ &= \sum_{i=0}^{\infty} \frac{1}{\sqrt{i!}} C_i \sum_{s=0}^S F_{si}(z_k) a_{rs,k}^* / G(z_k) \end{aligned} \quad (30)$$

with

$$\begin{aligned} h_{10} &= \mu_x, \quad h_{11} = \sigma_x, \\ h_{20} &= p_{20} - \left(\frac{P_{21}}{\lambda_{11}^{(1)}} - \frac{P_{22}}{\lambda_{11}^{(1)} \lambda_{22}^{(1)}} \lambda_{21}^{(1)} \right) \lambda_{10}^{(1)} - \frac{P_{22}}{\lambda_{22}^{(1)}} \lambda_{20}^{(1)}, \\ h_{21} &= \frac{P_{21}}{\lambda_{11}^{(1)}} - \frac{P_{22}}{\lambda_{11}^{(1)} \lambda_{22}^{(1)}} \lambda_{21}^{(1)}, \quad h_{22} = \frac{P_{22}}{\lambda_{22}^{(1)}}, \\ p_{20} &= (\mu_x - x_k^*)^2 + \sigma_x^2, \quad p_{21} = 2\sigma_x(\mu_x - x_k^*), \quad p_{22} = \sqrt{2}\sigma_x^2. \end{aligned} \quad (31)$$

Since Equations (28) (29) and (30) can be evaluated by measuring bone-conducted speech z_k , no time transition models of x_k are necessary. Therefore, computation time of the proposed algorithm can be reduced than the previous one [12]. Furthermore, by considering Equation (9), two parameters $a_{rs,k}^*$ and $\Gamma_{a_{rs,k}}$ in Equation (22) are given by the estimates of $a_{rs,k}$ at the discrete time $k-1$, as follows:

$$a_{rs,k}^* = \hat{a}_{rs,k-1}, \quad \Gamma_{a_{rs,k}} = P_{a_{rs,k-1}}. \quad (32)$$

Finally, considering Equations (1) (9) and (10), the expansion coefficients B_{lmn} in the estimation algorithm in Equations (20) (21) and (23) are given by the measurement of bone-conducted speech z_k , estimates of parameter $a_{rs,k}$ at the discrete time $k-1$, through the similar calculation process to Equations (25)-(30). Therefore, recursive estimation of the speech signal x_k can be achieved.

3. Application to Speech Signal in Real Environment

In order to confirm the actual usefulness of the proposed noise suppression algorithm, it was applied to speech signals in real noise environment. Though, in

the previous studies [12] [13], the noisy air-conducted speeches were created on a computer by mixing the original air-conducted speech signal measured in a noise-free environment, the algorithm proposed in this study was applied to signals measured in real environment under existence of actual noises. For a female and a male speech signals digitized with sampling frequency of 10 kHz and quantization of 16 bits, we estimated the speech signal based on the observation corrupted by additive noise.

More specifically, air-conducted speeches were measured in real environment under existence of a white noise generated from a noise generator and an actual machine noise. The bone-conducted speech was simultaneously measured by use of an acceleration sensor with the air-conducted speech. By setting roughly the amplitude of the noises at two levels, the proposed algorithm was applied to extremely difficult situations with low SNR (noise-free air-conducted speech signal to noise ratio defined by $\text{SNR} = 10 \log_{10} \left(\frac{\sum x_k^2}{\sum v_k^2} \right)$) being approximately -3 dB and -5 dB.

Using the observed bone-conducted speech and noisy observation on air-conducted speech, constants a and b are first calculated by introducing the linear regression model in Equation (11) and applying the least squared method to this model. Secondly, the parameter α of the membership function is obtained by calculating the standard deviation σ of y_k around \bar{y}_k , as $\alpha = 2\sigma$ after assuming Gaussian distribution for the deviation.

The observed signals on air-conducted female speech contaminated by the white noise and machine noise are shown in **Figure 1** and **Figure 2**. Furthermore, for the male speech signal, noisy air-conducted speech observations are shown in **Figure 3** and **Figure 4** respectively.

The estimated results by using the algorithm based on Equations (20)-(24) are shown in **Figure 5** and **Figure 6** for the female speech signal and in **Figure 7** and **Figure 8** for the male speech signal. For comparison, the estimated results of the female and male speech signals by using the estimation algorithm based on only the observation of air-conducted speech are shown in **Figures 9-12**.

By comparing **Figures 5-8** with **Figures 9-12**, it is obvious that the proposed method can suppress the effects of white noise and real machine noise better than the method based on observation of only air-conducted speech.

The air-conducted female and male speech signals spoken by the same speakers in the different situation without any noises are shown in **Figure 13** and **Figure 14** as references. By comparing these speech signals measured in noise-free circumstance with the estimated results by the proposed method and the results by using the algorithm based on the observation of only air-conducted signal, the effectiveness of the proposed method is obvious. Furthermore, the computation time of the proposed method was reduced by 55.2% of the algorithm based on the only air-conducted observation, because it is unnecessary for the proposed method to calculate recursively the estimate of variance of x_k based on the air-conducted speech y_k .

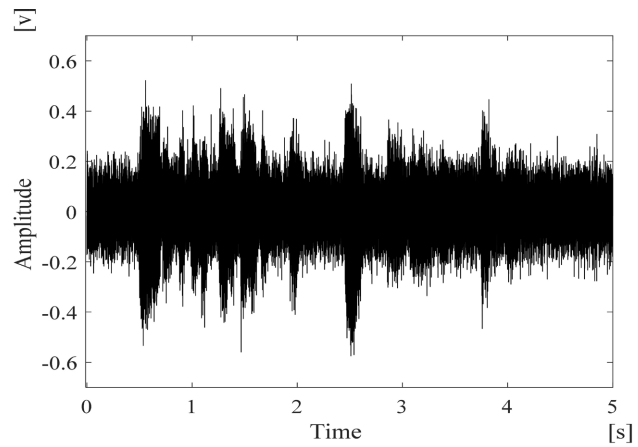


Figure 1. Observed female speech signal contaminated by white noise with $\text{SNR} \cong -3$ dB .

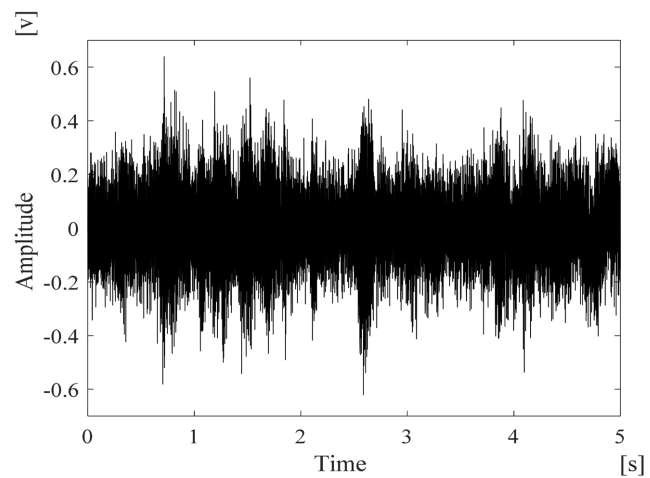


Figure 2. Observed female speech signal contaminated by machine noise with $\text{SNR} \cong -5$ dB .

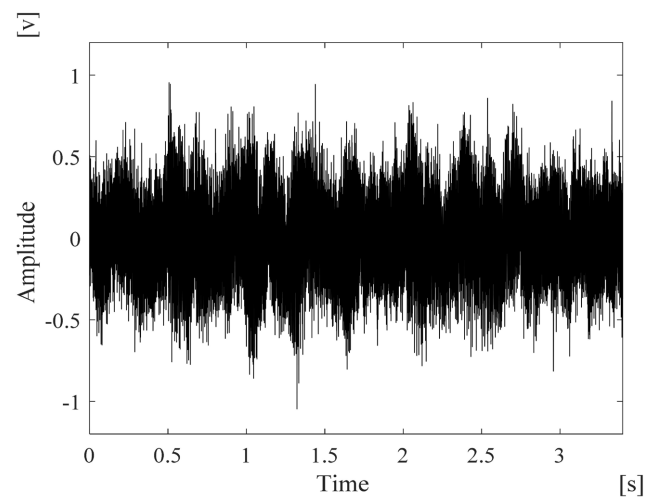


Figure 3. Observed male speech signal contaminated by white noise with $\text{SNR} \cong -3$ dB .

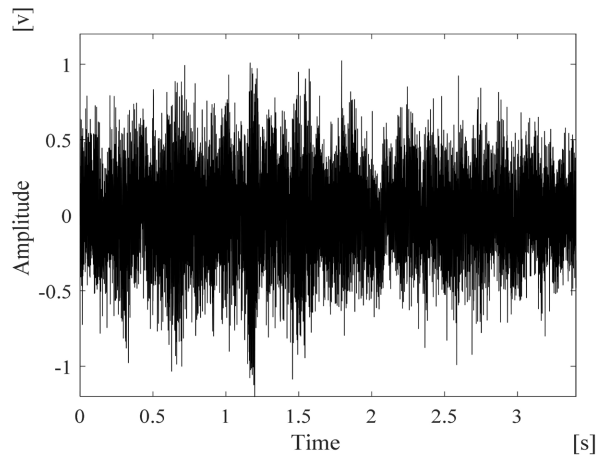


Figure 4. Observed male speech signal contaminated by machine noise with $SNR \cong -5$ dB .

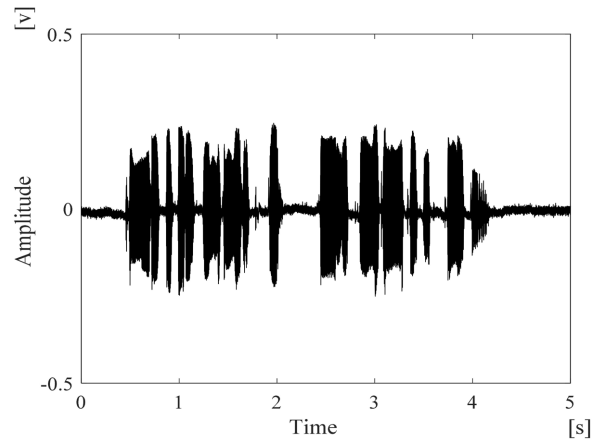


Figure 5. Estimated female speech signal by use of the proposed method based on observation contaminated by white noise with $SNR \cong -3$ dB .

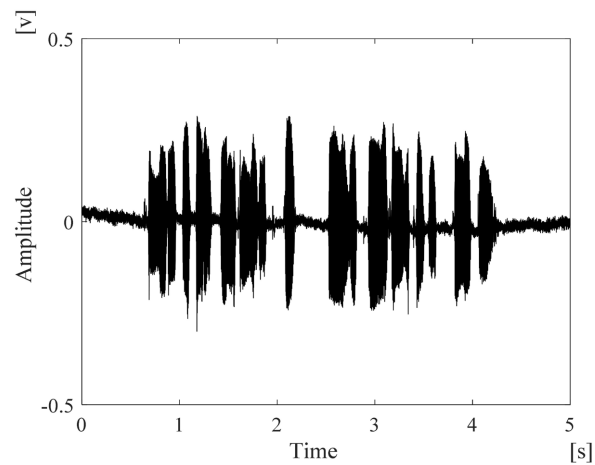


Figure 6. Estimated female speech signal by use of the proposed method based on observation contaminated by machine noise with $SNR \cong -5$ dB .

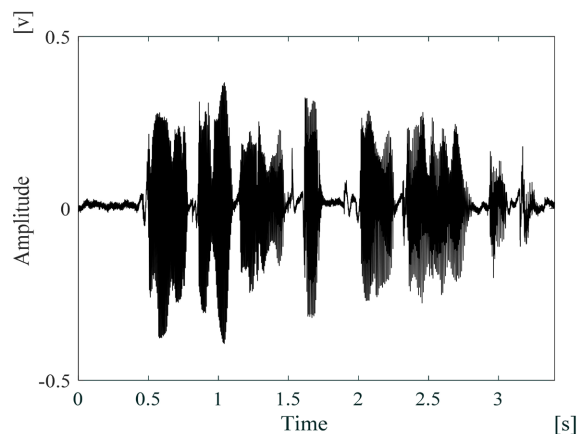


Figure 7. Estimated male speech signal by use of the proposed method based on observation contaminated by white noise with $\text{SNR} \cong -3 \text{ dB}$.

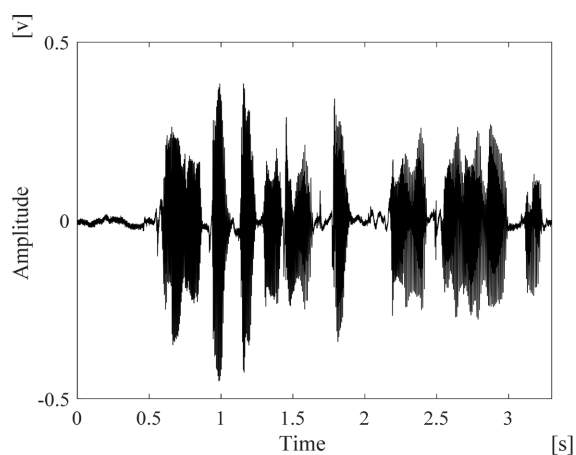


Figure 8. Estimated male speech signal by use of the proposed method based on observation contaminated by machine noise with $\text{SNR} \cong -5 \text{ dB}$.

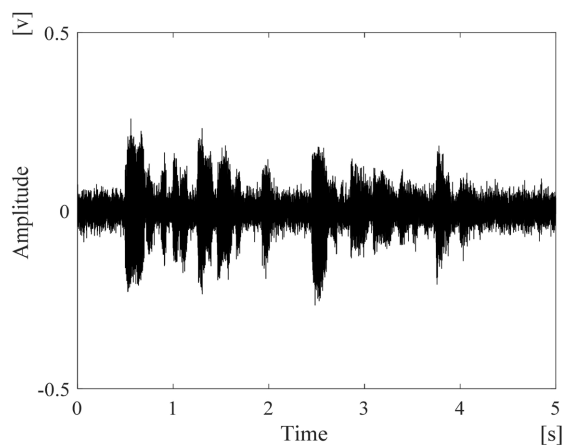


Figure 9. Estimated female speech signal by use of the method based on only air-conducted observation contaminated by white noise with $\text{SNR} \cong -3 \text{ dB}$.

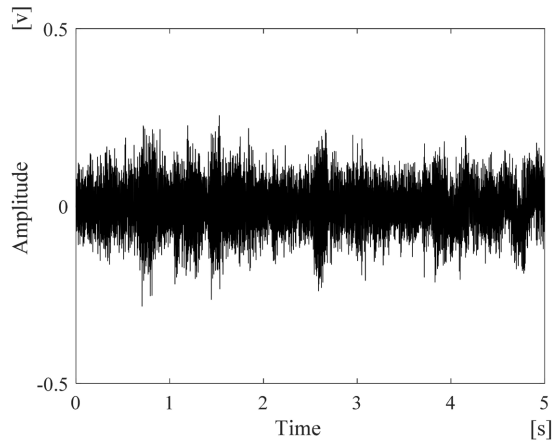


Figure 10. Estimated female speech signal by use of the method based on only air-conducted observation contaminated by machine noise with $SNR \cong -5$ dB .

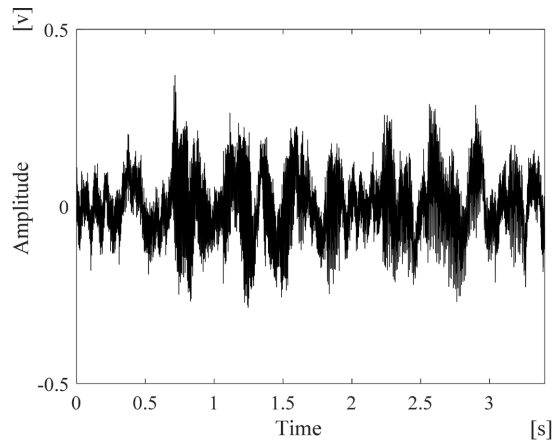


Figure 11. Estimated male speech signal by use of the method based on only air-conducted observation contaminated by white noise with $SNR \cong -3$ dB .

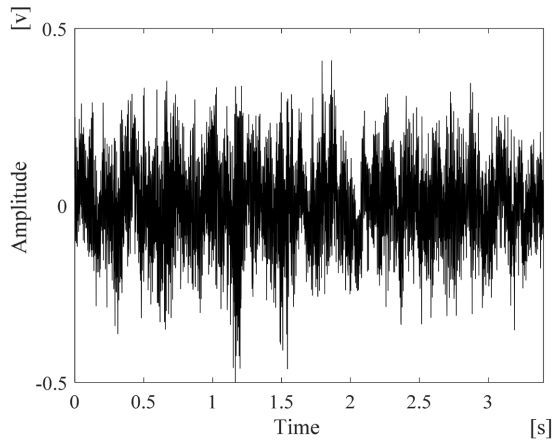


Figure 12. Estimated male speech signal by use of the method based on only air-conducted observation contaminated by machine noise with $SNR \cong -5$ dB .

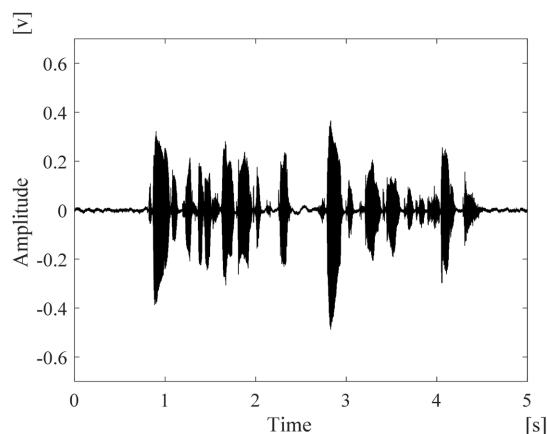


Figure 13. Air-conducted female speech signal in the different situation without any noises.

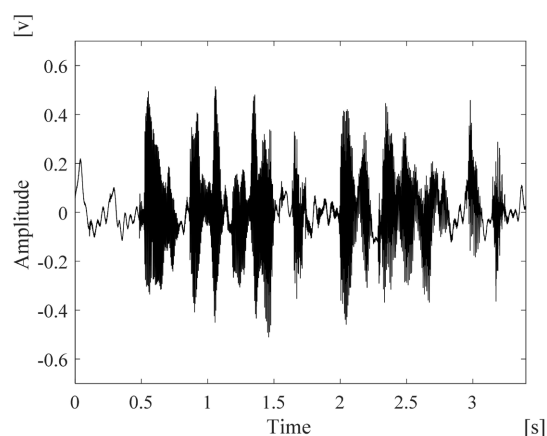


Figure 14. Air-conducted male speech signal in the different situation without any noises.

4. Conclusions

In this paper, after considering the bone-conducted speech signal with the reduction of higher components as fuzzy data, applying the probability measure of fuzzy events, a new noise suppression method is derived on the basis of Bayes' theorem as the fundamental principle of estimation. Furthermore, the proposed algorithm has been applied to real speech signals contaminated by noises measured in actual environment with low SNR. As a result, it has been revealed by experiments that better estimation results may be obtained by the proposed algorithm as compared with the method based on only air-conducted observations.

The proposed approach is quite different from the traditional standard techniques. However, we are still in an early stage of development, and a number of practical problems are yet to be investigated in the future. These include: 1) application to a diverse range of speech signals in actual noise environment, 2) extension to cases with multi-noise sources, and 3) finding an optimal number of expansion terms for the expansion-based probability expressions adopted.

Acknowledgements

The authors are grateful to Ms. Yui Maeda of the Prefectural University of Hiroshima for her help during this study. This work was supported in part by fund from the Grant-in-Aid for Scientific Research No. 19K04428 from the Ministry of Education, Culture, Sports, Science and Technology-Japan.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Yamashita, K. and Shimamura, T. (2005) Nonstationary Noise Estimation Using Low-Frequency Regions for Spectral Subtraction. *IEEE Signal Processing Letters*, **12**, 465-468. <https://doi.org/10.1109/LSP.2005.847864>
- [2] Plapous, C., Marro, C. and Scalart, P. (2006) Improved Signal-to-Noise Ratio Estimation for Speech Enhancement. *IEEE Transactions on Speech and Audio Processing*, **14**, 2098-2108. <https://doi.org/10.1109/TASL.2006.872621>
- [3] McCowan, I.A. and Bourlard, H. (2003) Microphone Array Post-Filter Based on Noise Field Coherence. *IEEE Transactions on Speech and Audio Processing*, **11**, 709-716. <https://doi.org/10.1109/TSA.2003.818212>
- [4] Kawamura, A., Fujii, K., Itoh, Y. and Fukui, Y. (2002) A Noise Reduction Method Based on Linear Prediction Analysis. *IEICE Transactions on Fundamentals*, **J85-A**, 415-423. <https://doi.org/10.1109/ICASSP.2002.1004860>
- [5] Kawamura, A., Fujii, K. and Itoh, Y. (2005) A Noise Reduction Method Based on Linear Prediction with Variable Step-Size. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, **E88-A**, 855-861. <https://doi.org/10.1093/ietfec/e88-a.4.855>
- [6] Kim, W. and Ko, H. (2001) Noise Variance Estimation for Kalman Filtering of Noisy Speech. *IEICE Transactions on Information and Systems*, **E84-D**, 155-160.
- [7] Li, H., Wang, X., Dai, B. and Lu, W. (2007) A Kalman Smoothing Algorithm for Speech Enhancement Based on the Properties of Vocal Tract Varying Slowly. *Proceedings of Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel Distributed Computing*, Qingdao, 30 July-1 Aug. 2007, 832-836.
- [8] Tanabe, N., Furukawa, T. and Tsuji, S. (2008) Robust Noise Suppression Algorithm with the Kalman Filter Theory for White and Colored Disturbance. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, **E91-A**, 818-829. <https://doi.org/10.1093/ietfec/e91-a.3.818>
- [9] Jia, H., Zhang, X. and Jin, C. (2009) A Modified Speech Enhancement Algorithm Based on the Subspace. *Proceedings of 2009 Second International Symposium on Knowledge Acquisition and Modeling*, Wuhan, 30 November-1 December 2009, 344-347. <https://doi.org/10.1109/KAM.2009.19>
- [10] Candy, J.V. (2009) *Bayesian Signal Processing: Classical, Modern, and Particle Filtering Methods*. John Wiley & Sons Ltd., Hoboken. <https://doi.org/10.1002/9780470430583>
- [11] Ikuta, A. and Orimoto, H. (2011) Adaptive Noise Suppression Algorithm for Speech Signal Based on Stochastic System Theory. *IEICE Transactions on Fundamentals of*

Electronics, Communications and Computer Sciences, **E94-A**, 1618-1627.

<https://doi.org/10.1587/transfun.E94.A.1618>

- [12] Ikuta, A., Orimoto, H. and Gallagher, G. (2018) Noise Suppression Method by Jointly Using Bone- and Air-Conducted Speech Signals. *Noise Control Engineering Journal*, **66**, 472-488. <https://doi.org/10.3397/1/376640>
- [13] Orimoto, H., Ikuta, A. and Hasegawa, K. (2021) Speech Signal Detection Based on Bayesian Estimation by Observing Air-Conducted Speech under Existence of Surrounding Noise with the Aid of Bone-Conducted Speech. *Intelligent Information Management*, **13**, 199-213. <https://doi.org/10.4236/iim.2021.134011>
- [14] Shin, H.S., Kang, H.G. and Fingscheidt, T. (2012) Survey of Speech Enhancement Supported by a Bone Conduction Microphone. *Proceedings of 10th ITG Conference on Speech Communication*, Braunschweig, 26-28 September 2012, 47-50.
- [15] Ikuta, A. and Orimoto, H. (2014) Fuzzy Signal Processing of Sound and Electromagnetic Environment by Introducing Probability Measure of Fuzzy Events. *Proceedings of International Conference on Fuzzy Computation Theory and Applications*, Rome, 22-24 October 2014, 5-13.
- [16] Orimoto, H. and Ikuta, A. (2012) Prediction of Response Probability Distribution by Considering Additive Property of Energy and Evaluation in Decibel Scale for Sound Environment System with Unknown Structure. *Transactions of the Society of Instrument and Control Engineers*, **48**, 830-836. <https://doi.org/10.9746/sicetr.48.830>
- [17] Orimoto, H. and Ikuta, A. (2019) State Estimation for Sound Environment System with Nonlinear Observation Characteristics by Introducing Wide-Sense Particle Filter. *Intelligent Information Management*, **11**, 87-101. <https://doi.org/10.4236/iim.2019.116008>