

Noise Cancellation Algorithm Based on Air- and Bone-Conducted Speech Signals by Considering an Unscented Transformation Method

Hisako Orimoto*, Akira Ikuta

Prefectural University of Hiroshima, Faculty of Management and Information System, 734-8338, Japan

ARTICLE INFO

Article history:

Received: 22 February, 2019

Accepted: 10 April, 2019

Online: 12 April, 2019

Keywords:

Noise Cancellation

Speech Signal

Air- and Bone-conducted

UT Method

ABSTRACT

Noise control is essential when applying speech recognition in noisy environments such as factories. In this study, a signal processing for noise cancellation is proposed by using a noise-insensitive bone-conducted speech signal together with an air-conducted speech signal. The speech signal is generally expressed by a nonlinear model. The extended Kalman filter is very famous as a state estimation method for nonlinear systems. However, this filter needs a linearized approximation model for the nonlinear systems. By using the sample point called Sigma point, the unscented Kalman filter (UKF) can be applied to the nonlinear system model without linear approximation. In this study, new type method is proposed based on the UKF. Although UKF considers Gaussian noise, an extended UKF considering non-Gaussian noise is proposed. A noise cancellation method is derived by use of air- and bone-conducted speech signals. The validity of this method is investigated by using both conducted speech signals measured in a noisy real environment.

1 Introduction

Recently, speech recognition systems are used in car navigation systems and smart speakers etc.. However, speech recognition can not be performed effectively in circumstances with heavy noises. Therefore, some countermeasures against surrounding noise are indispensable in such situations.

Kalman filter has been applied to many noisy circumstances as a noise cancellation method for speech signal [1],[2],[3]. This filter assumes a linear model subject to white and Gaussian noise as the system equation and the observation equation [4],[5]. Though, the extended Kalman filter (EKF) [6] can be applied to nonlinear systems, linear approximation models of nonlinear systems are required. Therefore, many improvements are necessary for the noise cancellation method to apply it to actual speech signal processing. From the above viewpoint, in our previously reported study, a noise cancellation method was proposed by using air- and bone-conducted speech signals in the situation contaminated by non-Gaussian and non-white noises [7].

However, since the calculation of the expansion coefficients in the previous algorithm was very complicated, a simplified method is required.

On the other hand, the unscented Kalman filter (UKF) by use of unscented transformation (UT) method can be applied to nonlinear system[9]. The UT method is a technique for

calculating the statistics of a random variable that has been nonlinearly transformed. The set of samples on so-called sigma points(σ -points) are chosen so that they capture the specific properties of the underlying distribution. Therefore, this method can be applied to arbitrary nonlinear systems. In our previous study, a noise cancellation method based on only air conducted speech signal has been proposed by applying the UT method [8].

In this study, a new noise cancellation algorithm based on air- and bone-conducted speech signals is proposed by considering the UT method. The relationship between air-conducted speech signal and background noise is expressed as an additive model based on the additive property of sound pressure. However, propagation mechanism of bone-conducted speech signal is complicated and has to be considered as an unknown system in general. Therefore, a system model including unknown parameters is introduced in this study. More specifically, the sample points obtained by using the UT method are introduced. The noise cancellation algorithm is derived by use of an expansion expression of Bayes' theorem. This method can be considered non-Gaussian properties of noises and nonlinear correlation information between the speech signal and observation. Furthermore, the validity of the proposed method is experimentally confirmed by applying it to real speech signal with noises.

*Corresponding Author: Hisako Orimoto, Address, +81 82 251 9804 & orimoto@pu-hiroshima.ac.jp

2 Theory

2.1 Modeling of Air- and Bone-Conducted Speech Signals

We consider the original speech signal x_k , observed air-conducted speech signal y_k and bone-conducted speech signal z_k at discrete time k . The observation y_k is contaminated by a surrounding noise v_k . According to the additive property of sound pressure, the following relationship can be established.

$$y_k = x_k + v_k, \quad (1)$$

where the mean and variance of v_k are known. In order to derive the propagation model of the bone-conducted speech signal, the correlation information between x_k and z_k is required. However, it is difficult to obtain prior information on the unknown speech signal x_k . In this study, a new adaptive algorithm for noise cancellation is proposed by introducing a propagation model with unknown parameters between x_k and z_k as the bone-conducted speech signal model for z_k :

$$z_k = a_k x_k + b_k w_k, \quad (2)$$

where w_k is a random noise (*mean* : 0, *variance* 1) and a_k and b_k are unknown parameters.

2.2 Estimation Method Combined Bayes' Theorem with UT Method

The conditional joint probability distribution of the specific signal x_k and the unknown parameters a_k and b_k is expressed by using expansion expression of Bayes' theorem[10]. To simplify the derivation process of the estimation algorithm, σ -points and the weighting coefficients are introduced in expansion coefficients.

The conditional probability distribution of x_k , a_k and b_k is expressed as

$$\begin{aligned} &P(x_k, a_k, b_k | Y_k, Z_k) \\ &= P_0(x_k | Y_{k-1}, Z_{k-1}) P_0(a_k | Y_{k-1}, Z_{k-1}) P_0(b_k | Y_{k-1}, Z_{k-1}) \\ &\quad \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \sum_{s=0}^{\infty} \sum_{t=0}^{\infty} A_{lmnst} \\ &\quad \varphi_l^{(1)}(x_k) \varphi_m^{(2)}(a_k) \varphi_n^{(3)}(b_k) \varphi_s^{(4)}(y_k) \varphi_t^{(5)}(z_k) \\ &\quad / \sum_{s=0}^{\infty} \sum_{t=0}^{\infty} A_{000st} \varphi_s^{(4)}(y_k) \varphi_t^{(5)}(z_k), \end{aligned} \quad (3)$$

with

$$A_{lmnst} \equiv \left\langle \varphi_l^{(1)}(x_k) \varphi_m^{(2)}(a_k) \varphi_n^{(3)}(b_k) \varphi_s^{(4)}(y_k) \varphi_t^{(5)}(z_k) | Y_{k-1}, Z_{k-1} \right\rangle, \quad (4)$$

where $Y_k = \{y_1, y_2, \dots, y_k\}$ and $Z_k = \{z_1, z_2, \dots, z_k\}$ are sets of air- and bone-conducted speech signal data up to time k . The above five functions $\varphi_l^{(1)}(x_k)$, $\varphi_m^{(2)}(a_k)$, $\varphi_n^{(3)}(b_k)$, $\varphi_s^{(4)}(y_k)$ and $\varphi_t^{(5)}(z_k)$ are orthonormal polynomials of degrees l, m, n, s and t with weighting functions $P_0(x_k | Y_{k-1}, Z_{k-1})$, $P_0(a_k | Y_{k-1}, Z_{k-1})$, $P_0(b_k | Y_{k-1}, Z_{k-1})$, $P_0(y_k | Y_{k-1}, Z_{k-1})$, $P_0(z_k | Y_{k-1}, Z_{k-1})$, which can be chosen as the probability functions describing the dominant part of the fluctuation. As

the examples of standard probability functions, Gaussian distribution is adopted:

$$\begin{cases} P_0(x_k | Y_{k-1}, Z_{k-1}) = N(x_k; x_k^*, \Gamma_{x_k}), \\ P_0(a_k | Y_{k-1}, Z_{k-1}) = N(a_k; a_k^*, \Gamma_{a_k}), \\ P_0(b_k | Y_{k-1}, Z_{k-1}) = N(b_k; b_k^*, \Gamma_{b_k}), \\ P_0(y_k | Y_{k-1}, Z_{k-1}) = N(y_k; y_k^*, \Omega_k), \\ P_0(z_k | Y_{k-1}, Z_{k-1}) = N(z_k; z_k^*, \Phi_k), \end{cases} \quad (5)$$

with

$$\begin{aligned} N(x; \mu, \Gamma) &= \frac{1}{\sqrt{2\pi\Gamma}} e^{-\frac{(x-\mu)^2}{2\Gamma}}, \\ x_k^* &\equiv \langle x_k | Y_{k-1}, Z_{k-1} \rangle, \\ \Gamma_{x_k} &\equiv \langle (x_k - x_k^*)^2 | Y_{k-1}, Z_{k-1} \rangle, \\ a_k^* &\equiv \langle a_k | Y_{k-1}, Z_{k-1} \rangle, \\ \Gamma_{a_k} &\equiv \langle (a_k - a_k^*)^2 | Y_{k-1}, Z_{k-1} \rangle, \\ b_k^* &\equiv \langle b_k | Y_{k-1}, Z_{k-1} \rangle, \\ \Gamma_{b_k} &\equiv \langle (b_k - b_k^*)^2 | Y_{k-1}, Z_{k-1} \rangle, \\ y_k^* &\equiv \langle y_k | Y_{k-1}, Z_{k-1} \rangle, \\ \Omega_k &\equiv \langle (y_k - y_k^*)^2 | Y_{k-1}, Z_{k-1} \rangle, \\ z_k^* &\equiv \langle z_k | Y_{k-1}, Z_{k-1} \rangle, \\ \Phi_k &\equiv \langle (z_k - z_k^*)^2 | Y_{k-1}, Z_{k-1} \rangle. \end{aligned} \quad (6)$$

The orthonormal polynomials[11] with five weighting probability distributions in Eq. (5) are then specified as

$$\varphi_l^{(1)}(x_k) = \frac{1}{\sqrt{l!}} H_l \left(\frac{x_k - x_k^*}{\sqrt{\Gamma_{x_k}}} \right), \quad (7)$$

$$\varphi_m^{(2)}(a_k) = \frac{1}{\sqrt{m!}} H_m \left(\frac{a_k - a_k^*}{\sqrt{\Gamma_{a_k}}} \right), \quad (8)$$

$$\varphi_n^{(3)}(b_k) = \frac{1}{\sqrt{n!}} H_n \left(\frac{b_k - b_k^*}{\sqrt{\Gamma_{b_k}}} \right), \quad (9)$$

$$\varphi_s^{(4)}(y_k) = \frac{1}{\sqrt{s!}} H_s \left(\frac{y_k - y_k^*}{\sqrt{\Omega_k}} \right), \quad (10)$$

$$\varphi_t^{(5)}(z_k) = \frac{1}{\sqrt{t!}} H_t \left(\frac{z_k - z_k^*}{\sqrt{\Phi_k}} \right). \quad (11)$$

The estimates for mean and variance (i.e., conditional mean and variance) of x_k , a_k , b_k , which are the first and second order statistics, can be expressed as follows:

$$\begin{aligned} \hat{x}_k &\equiv \langle x_k | Y_k, Z_k \rangle \\ &= \frac{\sum_{l=0}^1 \sum_{s=0}^{\infty} \sum_{t=0}^{\infty} A_{l00st} c_{1l} \varphi_s^{(4)}(y_k) \varphi_t^{(5)}(z_k)}{\sum_{s=0}^{\infty} \sum_{t=0}^{\infty} A_{000st} \varphi_s^{(4)}(y_k) \varphi_t^{(5)}(z_k)}, \end{aligned} \quad (12)$$

$$\begin{aligned} \hat{a}_k &\equiv \langle a_k | Y_k, Z_k \rangle \\ &= \frac{\sum_{m=0}^1 \sum_{s=0}^{\infty} \sum_{t=0}^{\infty} A_{0m0st} d_{1m} \varphi_s^{(4)}(y_k) \varphi_t^{(5)}(z_k)}{\sum_{s=0}^{\infty} \sum_{t=0}^{\infty} A_{000st} \varphi_s^{(4)}(y_k) \varphi_t^{(5)}(z_k)}, \end{aligned} \quad (13)$$

$$\hat{b}_k \equiv \langle b_k | Y_k, Z_k \rangle = \frac{\sum_{n=0}^1 \sum_{s=0}^{\infty} \sum_{t=0}^{\infty} A_{00nst} e_{1n} \varphi_s^{(4)}(y_k) \varphi_t^{(5)}(z_k)}{\sum_{s=0}^{\infty} \sum_{t=0}^{\infty} A_{000st} \varphi_s^{(4)}(y_k) \varphi_t^{(5)}(z_k)}, \quad (14)$$

$$P_{x_k} \equiv \langle (x_k - \hat{x}_k)^2 | Y_k, Z_k \rangle = \frac{\sum_{l=0}^2 \sum_{s=0}^{\infty} \sum_{t=0}^{\infty} A_{l00st} c_{2l} \varphi_s^{(4)}(y_k) \varphi_t^{(5)}(z_k)}{\sum_{s=0}^{\infty} \sum_{t=0}^{\infty} A_{000st} \varphi_s^{(4)}(y_k) \varphi_t^{(5)}(z_k)}, \quad (15)$$

$$P_{a_k} \equiv \langle (a_k - \hat{a}_k)^2 | Y_k, Z_k \rangle = \frac{\sum_{m=0}^2 \sum_{s=0}^{\infty} \sum_{t=0}^{\infty} A_{0m0st} d_{2m} \varphi_s^{(4)}(y_k) \varphi_t^{(5)}(z_k)}{\sum_{s=0}^{\infty} \sum_{t=0}^{\infty} A_{000st} \varphi_s^{(4)}(y_k) \varphi_t^{(5)}(z_k)}, \quad (16)$$

$$P_{b_k} \equiv \langle (b_k - \hat{b}_k)^2 | Y_k, Z_k \rangle = \frac{\sum_{n=0}^2 \sum_{s=0}^{\infty} \sum_{t=0}^{\infty} A_{00nst} e_{2n} \varphi_s^{(4)}(y_k) \varphi_t^{(5)}(z_k)}{\sum_{s=0}^{\infty} \sum_{t=0}^{\infty} A_{000st} \varphi_s^{(4)}(y_k) \varphi_t^{(5)}(z_k)}. \quad (17)$$

Where coefficients are appropriate constant satisfying the following equality:

$$\begin{aligned} x_k &= \sum_{l=0}^1 c_{1l} \varphi_l^{(1)}(x_k), (x_k - \hat{x}_k)^2 = \sum_{l=0}^2 c_{2l} \varphi_l^{(1)}(x_k), \\ a_k &= \sum_{m=0}^1 d_{1m} \varphi_m^{(2)}(a_k), (a_k - \hat{a}_k)^2 = \sum_{m=0}^2 d_{2m} \varphi_m^{(2)}(a_k), \\ b_k &= \sum_{n=0}^1 e_{1n} \varphi_n^{(2)}(b_k), (b_k - \hat{b}_k)^2 = \sum_{n=0}^2 e_{2n} \varphi_n^{(3)}(b_k). \end{aligned} \quad (18)$$

By using the UT method, the expansion coefficients defined by (4) can be realized for arbitrary nonlinear systems. When the UT method is applied to approximate the means and variances of x_k, a_k, b_k, y_k and z_k , the σ -points $x_k^{*(i)}, a_k^{*(i)}, b_k^{*(i)}, y_k^{*(i)}$ and $z_k^{*(i)}$ are obtained as sample points, as follows:

$$\begin{aligned} x_k^{*(0)} &= x_k^*, \\ x_k^{*(1)} &= x_k^* + \sqrt{(1 + \lambda)\Gamma_{x_k}}, x_k^{*(2)} = x_k^* - \sqrt{(1 + \lambda)\Gamma_{x_k}}, \\ a_k^{*(0)} &= a_k^*, \\ a_k^{*(1)} &= a_k^* + \sqrt{(1 + \lambda)\Gamma_{a_k}}, a_k^{*(2)} = a_k^* - \sqrt{(1 + \lambda)\Gamma_{a_k}}, \\ b_k^{*(0)} &= b_k^*, \\ b_k^{*(1)} &= b_k^* + \sqrt{(1 + \lambda)\Gamma_{b_k}}, b_k^{*(2)} = b_k^* - \sqrt{(1 + \lambda)\Gamma_{b_k}}, \\ y_k^{*(0)} &= y_k^*, \\ y_k^{*(1)} &= y_k^* + \sqrt{(1 + \lambda)\Omega_k}, y_k^{*(2)} = y_k^* - \sqrt{(1 + \lambda)\Omega_k}, \\ z_k^{*(0)} &= z_k^*, \\ z_k^{*(1)} &= z_k^* + \sqrt{(1 + \lambda)\Phi_k}, z_k^{*(2)} = z_k^* - \sqrt{(1 + \lambda)\Phi_k}, \end{aligned} \quad (19)$$

The σ -points are decided so as to obtain the approximately same mean and variance as original variables. Where λ is a regulation parameter. The weights to be used are obtained as follows:

$$\begin{aligned} W^{(0)} &= \frac{\lambda}{(1 + \lambda)}, \\ W^{(1)} &= \frac{1}{2(1 + \lambda)}, \\ W^{(2)} &= \frac{1}{2(1 + \lambda)}. \end{aligned} \quad (20)$$

Here, the weighing coefficients $W^{(i)}$ have to satisfy the normalization constraint.

$$\sum_{i=0}^2 W^{(i)} = 1. \quad (21)$$

Each expansion coefficient A_{lmnst} defined by (4) is obtained specifically by substituting σ -points $x_k^{*(i)}, a_k^{*(i)}, b_k^{*(i)}, y_k^{*(i)}$ and $z_k^{*(i)}$ into the conditional expectation of x_k, y_k and z_k .

$$\begin{aligned} A_{10010} &= \frac{1}{\sqrt{\Gamma_{x_k}} \sqrt{\Omega_k}} \sum_{i=0}^2 W^{(i)} (x_k^{*(i)} - x_k^*) (y_k^{*(i)} - y_k^*), \\ A_{10020} &= \frac{1}{\sqrt{2} \sqrt{\Gamma_{x_k}} \Omega_k} \sum_{i=0}^2 W^{(i)} (x_k^{*(i)} - x_k^*) \{ (y_k^{*(i)} - y_k^*)^2 - \Omega_k \}, \\ A_{20010} &= \frac{1}{\sqrt{2} \Gamma_{x_k} \sqrt{\Omega_k}} \sum_{i=0}^2 W^{(i)} \{ (x_k^{*(i)} - x_k^*)^2 - \Gamma_{x_k} \} (y_k^{*(i)} - y_k^*), \\ A_{20020} &= \frac{1}{2\Gamma_{x_k} \Omega_k} \sum_{i=0}^2 W^{(i)} \{ (x_k^{*(i)} - x_k^*)^2 - \Gamma_{x_k} \} \cdot \{ (y_k^{*(i)} - y_k^*)^2 - \Omega_k \}, \\ A_{10001} &= \frac{1}{\sqrt{\Gamma_{x_k}} \sqrt{\Phi_k}} \sum_{i=0}^2 W^{(i)} (x_k^{*(i)} - x_k^*) (z_k^{*(i)} - z_k^*), \\ A_{10002} &= \frac{1}{\sqrt{2} \sqrt{\Gamma_{x_k}} \Phi_k} \sum_{i=0}^2 W^{(i)} (x_k^{*(i)} - x_k^*) \{ (z_k^{*(i)} - z_k^*)^2 - \Phi_k \}, \\ A_{20001} &= \frac{1}{\sqrt{2} \Gamma_{x_k} \sqrt{\Phi_k}} \sum_{i=0}^2 W^{(i)} \{ (x_k^{*(i)} - x_k^*)^2 - \Gamma_{x_k} \} (z_k^{*(i)} - z_k^*), \\ A_{20002} &= \frac{1}{2\Gamma_{x_k} \Phi_k} \sum_{i=0}^2 W^{(i)} \{ (x_k^{*(i)} - x_k^*)^2 - \Gamma_{x_k} \} \{ (z_k^{*(i)} - z_k^*)^2 - \Phi_k \}, \\ A_{10011} &= \frac{1}{\sqrt{\Gamma_{x_k}} \sqrt{\Omega_k} \sqrt{\Phi_k}} \sum_{i=0}^2 W^{(i)} (x_k^{*(i)} - x_k^*) (y_k^{*(i)} - y_k^*) (z_k^{*(i)} - z_k^*), \\ A_{20011} &= \frac{1}{\sqrt{2} \Gamma_{x_k} \sqrt{\Omega_k} \sqrt{\Phi_k}} \sum_{i=0}^2 W^{(i)} \{ (x_k^{*(i)} - x_k^*)^2 - \Gamma_{x_k} \} \cdot (y_k^{*(i)} - y_k^*) (z_k^{*(i)} - z_k^*), \\ A_{10021} &= \frac{1}{\sqrt{2} \sqrt{\Gamma_{x_k}} \Omega_k \sqrt{\Phi_k}} \sum_{i=0}^2 W^{(i)} (x_k^{*(i)} - x_k^*) \cdot \{ (y_k^{*(i)} - y_k^*)^2 - \Omega_k \} (z_k^{*(i)} - z_k^*), \\ A_{10012} &= \frac{1}{\sqrt{2} \sqrt{\Gamma_{x_k}} \sqrt{\Omega_k} \Phi_k} \sum_{i=0}^2 W^{(i)} (x_k^{*(i)} - x_k^*) \cdot (y_k^{*(i)} - y_k^*) \{ (z_k^{*(i)} - z_k^*)^2 - \Phi_k \}, \\ A_{20012} &= \frac{1}{2\Gamma_{x_k} \sqrt{\Omega_k} \Phi_k} \sum_{i=0}^2 W^{(i)} \{ (x_k^{*(i)} - x_k^*)^2 - \Gamma_{x_k} \} \cdot (y_k^{*(i)} - y_k^*) \{ (z_k^{*(i)} - z_k^*)^2 - \Phi_k \}, \\ A_{10022} &= \frac{1}{2 \sqrt{\Gamma_{x_k}} \Omega_k \Phi_k} \sum_{i=0}^2 W^{(i)} (x_k^{*(i)} - x_k^*) \cdot \{ (y_k^{*(i)} - y_k^*)^2 - \Omega_k \} \{ (z_k^{*(i)} - z_k^*)^2 - \Phi_k \}, \\ A_{20021} &= \frac{1}{2\Gamma_{x_k} \Omega_k \sqrt{\Phi_k}} \sum_{i=0}^2 W^{(i)} \{ (x_k^{*(i)} - x_k^*)^2 - \Gamma_{x_k} \} \cdot \{ (y_k^{*(i)} - y_k^*)^2 - \Omega_k \} (z_k^{*(i)} - z_k^*), \\ A_{20022} &= \frac{1}{2 \sqrt{2} \Gamma_{x_k} \Omega_k \Phi_k} \sum_{i=0}^2 W^{(i)} \{ (x_k^{*(i)} - x_k^*)^2 - \Gamma_{x_k} \} \cdot \{ (y_k^{*(i)} - y_k^*)^2 - \Omega_k \} \{ (z_k^{*(i)} - z_k^*)^2 - \Phi_k \}, \end{aligned} \quad (22)$$

with

$$A_{00000} = 1, A_{10000} = A_{00010} = A_{00001} = A_{20000} = A_{00020} = A_{00002} = 0.$$

Furthermore, expansion coefficients of a_k, y_k and z_k are

expressed as follows:

$$\begin{aligned}
 A_{01010} &= \frac{1}{\sqrt{\Gamma_{a_k} \sqrt{\Omega_k}}} \sum_{i=0}^2 w^{(i)} (a_k^{*(i)} - a_k^*) (y_k^{*(i)} - y_k^*), \\
 A_{01020} &= \frac{1}{\sqrt{2} \sqrt{\Gamma_{a_k} \Omega_k}} \sum_{i=0}^2 w^{(i)} (a_k^{*(i)} - a_k^*) \left\{ (y_k^{*(i)} - y_k^*)^2 - \Omega_k \right\}, \\
 A_{02010} &= \frac{1}{\sqrt{2} \Gamma_{a_k} \sqrt{\Omega_k}} \sum_{i=0}^2 w^{(i)} \left\{ (a_k^{*(i)} - a_k^*)^2 - \Gamma_{a_k} \right\} (y_k^{*(i)} - y_k^*), \\
 A_{02020} &= \frac{1}{2 \Gamma_{a_k} \Omega_k} \sum_{i=0}^2 w^{(i)} \left\{ (a_k^{*(i)} - a_k^*)^2 - \Gamma_{a_k} \right\} \\
 &\quad \cdot \left\{ (y_k^{*(i)} - y_k^*)^2 - \Omega_k \right\}, \\
 A_{01001} &= \frac{1}{\sqrt{\Gamma_{a_k} \sqrt{\Phi_k}}} \sum_{i=0}^2 w^{(i)} (a_k^{*(i)} - a_k^*) (z_k^{*(i)} - z_k^*), \\
 A_{01002} &= \frac{1}{\sqrt{2} \sqrt{\Gamma_{a_k} \Phi_k}} \sum_{i=0}^2 w^{(i)} (a_k^{*(i)} - a_k^*) \left\{ (z_k^{*(i)} - z_k^*)^2 - \Phi_k \right\}, \\
 A_{02001} &= \frac{1}{\sqrt{2} \Gamma_{a_k} \sqrt{\Phi_k}} \sum_{i=0}^2 w^{(i)} \left\{ (a_k^{*(i)} - a_k^*)^2 - \Gamma_{a_k} \right\} (z_k^{*(i)} - z_k^*), \\
 A_{02002} &= \frac{1}{2 \Gamma_{a_k} \Phi_k} \sum_{i=0}^2 w^{(i)} \left\{ (a_k^{*(i)} - a_k^*)^2 - \Gamma_{a_k} \right\} \\
 &\quad \cdot \left\{ (z_k^{*(i)} - z_k^*)^2 - \Phi_k \right\}, \\
 A_{01011} &= \frac{1}{\sqrt{\Gamma_{a_k} \sqrt{\Omega_k} \sqrt{\Phi_k}}} \sum_{i=0}^2 w^{(i)} (a_k^{*(i)} - a_k^*) (y_k^{*(i)} - y_k^*) (z_k^{*(i)} - z_k^*), \\
 A_{02011} &= \frac{1}{\sqrt{2} \Gamma_{a_k} \sqrt{\Omega_k} \sqrt{\Phi_k}} \sum_{i=0}^2 w^{(i)} \left\{ (a_k^{*(i)} - a_k^*)^2 - \Gamma_{a_k} \right\} \\
 &\quad \cdot (y_k^{*(i)} - y_k^*) (z_k^{*(i)} - z_k^*), \\
 A_{01021} &= \frac{1}{\sqrt{2} \sqrt{\Gamma_{a_k} \Omega_k} \sqrt{\Phi_k}} \sum_{i=0}^2 w^{(i)} (a_k^{*(i)} - a_k^*) \\
 &\quad \cdot \left\{ (y_k^{*(i)} - y_k^*)^2 - \Omega_k \right\} (z_k^{*(i)} - z_k^*), \\
 A_{01012} &= \frac{1}{\sqrt{2} \sqrt{\Gamma_{a_k} \sqrt{\Omega_k} \Phi_k}} \sum_{i=0}^2 w^{(i)} (a_k^{*(i)} - a_k^*) (y_k^{*(i)} - y_k^*) \\
 &\quad \cdot \left\{ (z_k^{*(i)} - z_k^*)^2 - \Phi_k \right\}, \\
 A_{02012} &= \frac{1}{2 \Gamma_{a_k} \sqrt{\Omega_k} \Phi_k} \sum_{i=0}^2 w^{(i)} \left\{ (a_k^{*(i)} - a_k^*)^2 - \Gamma_{a_k} \right\} \\
 &\quad \cdot (y_k^{*(i)} - y_k^*) \left\{ (z_k^{*(i)} - z_k^*)^2 - \Phi_k \right\}, \\
 A_{01022} &= \frac{1}{2 \sqrt{\Gamma_{a_k} \Omega_k} \Phi_k} \sum_{i=0}^2 w^{(i)} (a_k^{*(i)} - a_k^*) \\
 &\quad \cdot \left\{ (y_k^{*(i)} - y_k^*)^2 - \Omega_k \right\} \left\{ (z_k^{*(i)} - z_k^*)^2 - \Phi_k \right\}, \\
 A_{02021} &= \frac{1}{2 \Gamma_{a_k} \Omega_k \sqrt{\Phi_k}} \sum_{i=0}^2 w^{(i)} \left\{ (a_k^{*(i)} - a_k^*)^2 - \Gamma_{a_k} \right\} \\
 &\quad \cdot \left\{ (y_k^{*(i)} - y_k^*)^2 - \Omega_k \right\} (z_k^{*(i)} - z_k^*), \\
 A_{02022} &= \frac{1}{2 \sqrt{2} \Gamma_{a_k} \Omega_k \Phi_k} \sum_{i=0}^2 w^{(i)} \left\{ (a_k^{*(i)} - a_k^*)^2 - \Gamma_{a_k} \right\} \\
 &\quad \cdot \left\{ (y_k^{*(i)} - y_k^*)^2 - \Omega_k \right\} \left\{ (z_k^{*(i)} - z_k^*)^2 - \Phi_k \right\}.
 \end{aligned} \tag{23}$$

The expansion coefficients of b_k , y_k and z_k ($A_{00110}, A_{00120}, A_{00210}, A_{00220}, A_{00101}, A_{00102}, A_{00201}, A_{00202}, A_{00111}, A_{00211}, A_{00121}, A_{00112}, A_{00212}, A_{00122}, A_{00221}, A_{00222}$) are calculated through the same manners.

After substituting (1) (2) into the definition of four parameters y_k^* , Ω_k , z_k^* and Φ_k in (6), the following expressions can be derived.

$$y_k^* = x_k^* + \langle v_k \rangle, \tag{24}$$

$$\Omega_k = \Gamma_{x_k} + \langle (v_k - \langle v_k \rangle)^2 \rangle, \tag{25}$$

$$z_k^* = a_k^* \lambda_k^*, \tag{26}$$

$$\Phi_k = (\Gamma_{a_k} + a_k^{*2}) (\Gamma_{x_k} + x_k^{*2}) + (\Gamma_{b_k} + b_k^{*2}) \langle w_k^2 \rangle - a_k^{*2} x_k^{*2}. \tag{27}$$

In order to derive the predicted values of the speech signal x_k and the unknown parameters a_k, b_k , the time transition of the speech signal x_k is expressed as follows.

$$x_{k+1} = F x_k + G u_k, \tag{28}$$

where, u_k is a random input with mean 0 and variance 1. Parameters F and G are calculated from time correlation information of x_k and x_{k+1} :

$$F = \frac{\langle x_{k+1} x_k \rangle}{\langle x_k^2 \rangle}, \quad G = \sqrt{(1 - F^2) \langle x_k^2 \rangle}. \tag{29}$$

Therefore, x_{k+1}^* and $\Gamma_{x_{k+1}}$ can be expressed as follows:

$$x_{k+1}^* = F \hat{x}_k + G \langle u_k \rangle, \tag{30}$$

$$\Gamma_{x_{k+1}} = F^2 P_{x_k} + G^2 \langle (u_k - \langle u_k \rangle)^2 \rangle, \tag{31}$$

Since the parameters a_k and b_k are constants, time transition models are introduced for the recursive estimation.

$$a_{k+1} = a_k, b_{k+1} = b_k. \tag{32}$$

By using these relationships, the predictions are given as follows,

$$a_{k+1}^* = \hat{a}_k, \quad \Gamma_{a_{k+1}} = P_{a_k}, \tag{33}$$

$$b_{k+1}^* = \hat{b}_k, \quad \Gamma_{b_{k+1}} = P_{b_k}. \tag{34}$$

The state estimation algorithm with expansion coefficient A_{lmnst} reflecting linear and nonlinear correlation information among variables and statistics of non-Gaussian noise is completed.

3 Experiment

In order to confirm the validity of the proposed noise cancellation algorithm, we compared it with the method using only the air-conducted speech signal. The compared method was derived by considering the following conditional probability distribution.

$$\begin{aligned}
 &P(x_k, a_k, b_k | Y_k) \\
 &= P_0(x_k | Y_{k-1}) P_0(a_k | Y_{k-1}) P_0(b_k | Y_{k-1}) \\
 &\quad \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \sum_{s=0}^{\infty} A_{lmns} \\
 &\quad \varphi_l^{(1)}(x_k) \varphi_m^{(2)}(a_k) \varphi_n^{(3)}(b_k) \varphi_s^{(4)}(y_k) \\
 &\quad / \sum_{s=0}^{\infty} \sum_{t=0}^{\infty} A_{000s} \varphi_s^{(4)}(y_k).
 \end{aligned} \tag{35}$$

Based on (35), the estimates of mean and variance of x_k are derived as follow:

$$\begin{aligned}
 \hat{x}_k &\equiv \langle x_k | Y_k \rangle \\
 &= \frac{\sum_{l=0}^1 \sum_{s=0}^{\infty} A_{l00s} c_{1l} \varphi_s^{(4)}(y_k)}{\sum_{s=0}^{\infty} A_{000s} \varphi_s^{(4)}(y_k)},
 \end{aligned} \tag{36}$$

$$\begin{aligned}
 P_{x_k} &\equiv \langle (x_k - \hat{x}_k)^2 | Y_k \rangle \\
 &= \frac{\sum_{l=0}^2 \sum_{s=0}^{\infty} A_{l00s} c_{2l} \varphi_s^{(4)}(y_k)}{\sum_{s=0}^{\infty} A_{000s} \varphi_s^{(4)}(y_k)}.
 \end{aligned} \tag{37}$$

Male and female speech signals were used in the experiment. The speech signal data were measured in the anechoic

chamber in the acoustic laboratory. The observed speech signal are contaminated with the white noise, the pink noise and the machine noise respectively. The spectra of these noises are shown in Figures 1-3. Furthermore, the observation data of air-conducted speech signal were created by mixing noises with speech signal on a computer.

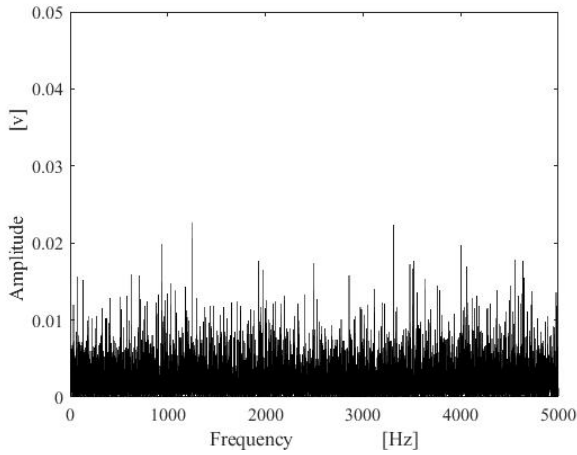


Figure 1: Spectrum of white noise.

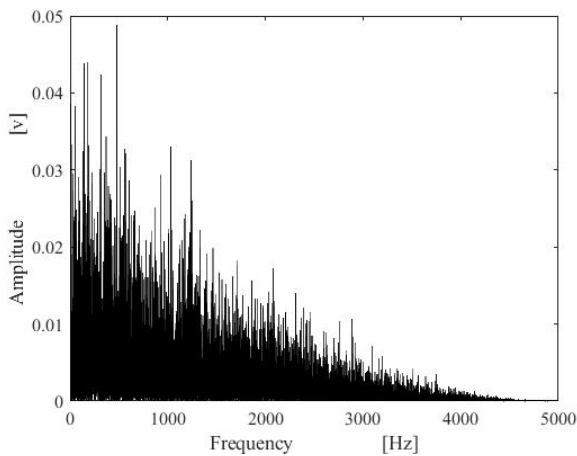


Figure 2: Spectrum of pink noise.

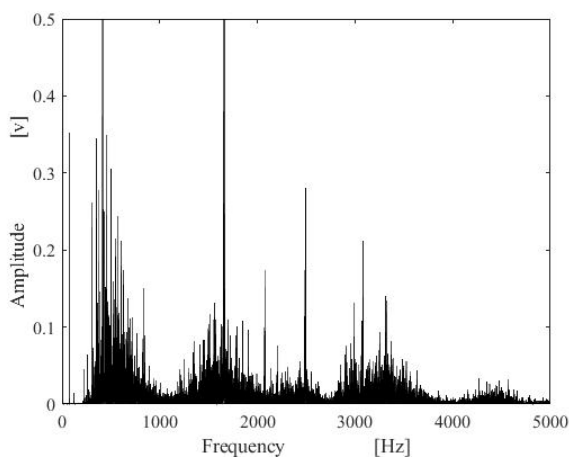


Figure 3: Spectrum of machine noise.

processing time for speech signal of about 3.5 seconds in length was from 0.5 to 0.8 seconds.

Table 1: The specifications of the personal computer.

	Specification
PC	Dell Vostro 3650
CPU	Intel Core i7-6700 @ 3.40GHz
MEMORY	8.00G
OS	Win 10 Pro 64bit

As an evaluation method of estimation result, the Root Mean Square Error (RMSE) and Performance Evaluation Index (PEI) are adopted.

$$RMSE = \sqrt{\frac{\sum(x_k - \hat{x}_k)^2}{N}}, \quad (38)$$

$$PEI = 10 \log_{10} \frac{\sum x_k^2}{\sum(x_k - \hat{x}_k)^2} [dB]. \quad (39)$$

As the RMS Error is smaller value, the better estimation result is obtained. On the other hand, the larger the PEI is, the better the estimation is. Table 2 and Table 3 show the results for the male speech signal and a female speech signal respectively. In the cases of lower noises, almost the same estimation results are obtained in the proposed method and the compared method. On the other hand, in the case of higher noises, the proposed method obtains better results than the compared method. Furthermore, in comparison with our previous method [7] with complicated algorithm, almost the same accurate estimation results are obtained as the proposed method. Therefore, the superiority of the proposed method adopting the simplified algorithm could be confirmed.

Table 2: Comparisons of RMSE and PEI for a male speech signal.

(a) white noise				
	Proposed method		Compared method	
S/N	RMSE	PEI	RMSE	PEI
1/1	0.0160	6.8438	0.0163	6.7057
1/2	0.0225	3.8693	0.0269	2.3286
1/3	0.0267	2.4111	0.0450	-2.1282
1/4	0.0330	0.5615	0.0838	-7.5301
1/5	0.0480	-2.6982	0.1338	-11.6004

(b) pink noise				
	Proposed method		Compared method	
S/N	RMSE	PEI	RMSE	PEI
1/1	0.0192	5.2693	0.0207	4.6124
1/2	0.0281	1.9414	0.0345	0.1846
1/3	0.0322	0.7773	0.0626	-4.9946
1/4	0.0419	-1.5062	0.0823	-7.3776
1/5	0.0669	-5.5803	0.1334	-11.5703

(c) machine noise				
	Proposed method		Compared method	
S/N	RMSE	PEI	RMSE	PEI
1/1	0.0238	3.4024	0.0272	2.2381
1/2	0.0313	1.0116	0.0489	-2.8538
1/3	0.0399	-1.0931	0.0767	-6.7613
1/4	0.0586	-4.4310	0.1080	-9.7397
1/5	0.0686	-5.8003	0.1409	-12.0497

Table 1 shows the specifications of the personal computer for signal processing in the experiment. The signal

Some of the waveform summarized in Tables 2 and 3 are shown in Figures 4-23. Figures 4 and 14 show the original speech signals of male and female, respectively. Figures 5, 8, 11, 15, 18 and 21 show the speech signals contaminated by noises with amplitude of 3 times larger than the original signals. The estimated results by using of proposed method are shown in Figures 6, 9, 12, 16, 19 and 22. On the other hand, the comparison results are shown in Figures 7, 10, 13, 17, 20 and 23. In the cases of contaminated by white noise and pink noise, the better results are obtained by the proposed method than the compared method.

Table 3: Comparisons of RMSE and PEI for a female speech signal.

(a) white noise

S/N	Proposed method		Compared method	
	RMSE	PEI	RMSE	PEI
1/1	0.0134	4.8382	0.0104	7.0371
1/2	0.0159	3.3213	0.0160	3.2779
1/3	0.0180	2.2390	0.0195	1.5406
1/4	0.0198	1.4447	0.0273	-1.3608
1/5	0.0260	-0.9548	0.0457	-5.8342

(b) pink noise

S/N	Proposed method		Compared method	
	RMSE	PEI	RMSE	PEI
1/1	0.0136	4.656	0.0128	5.2366
1/2	0.0176	2.4293	0.0211	0.8777
1/3	0.0231	0.0790	0.0271	-1.297
1/4	0.0241	-0.2875	0.0374	-4.1092
1/5	0.0264	-1.0874	0.0510	-6.8009

(c) machine noise

S/N	Proposed method		Compared method	
	RMSE	PEI	RMSE	PEI
1/1	0.0147	3.9846	0.0168	2.8276
1/2	0.0207	1.0503	0.0237	-0.1545
1/3	0.0293	-1.9655	0.0499	-6.6083
1/4	0.0456	-5.8243	0.0553	-7.5024
1/5	0.0425	-5.2075	0.0660	-9.0326

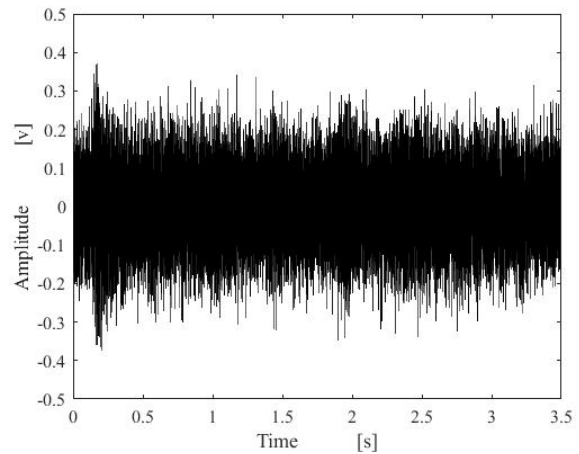


Figure 5: Male speech signal containing a white noise

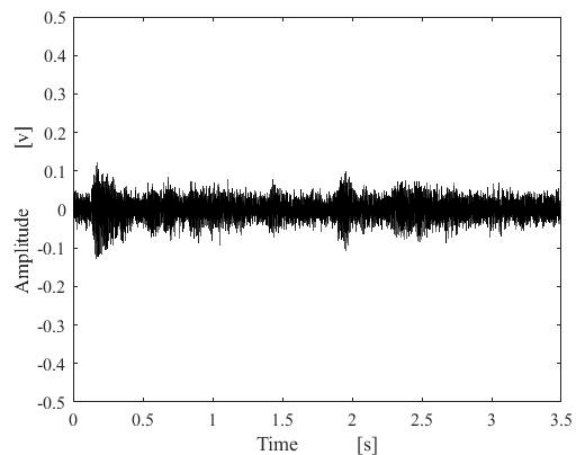


Figure 6: Estimated results by using proposed method.

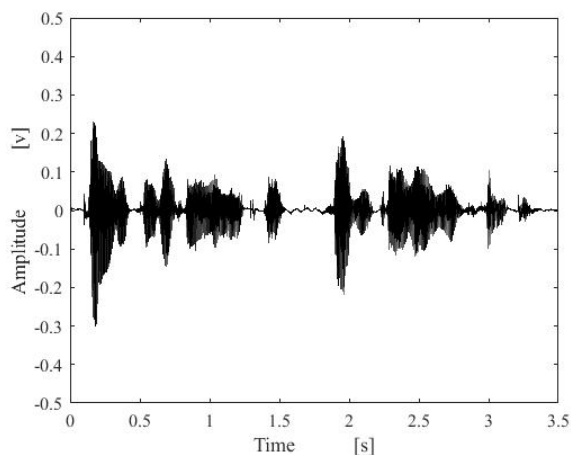


Figure 4: Original male speech signal

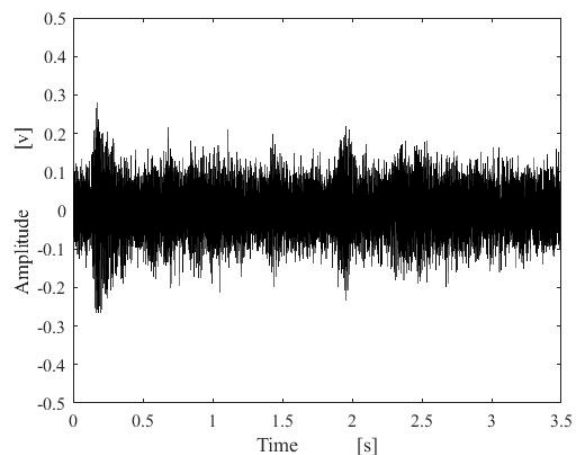


Figure 7: Estimated results by using compared method.

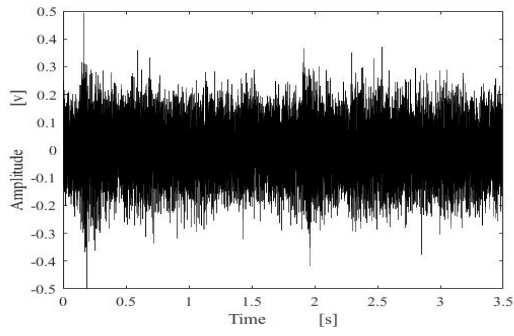


Figure 8: Male speech signal containing a pink noise

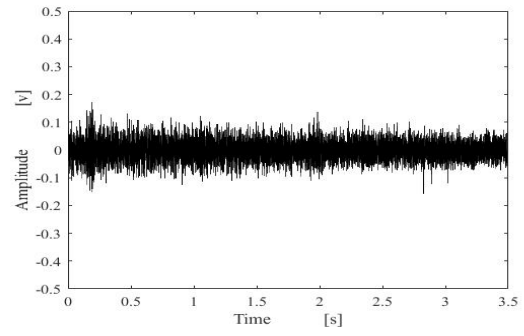


Figure 12: Estimated results by using proposed method.

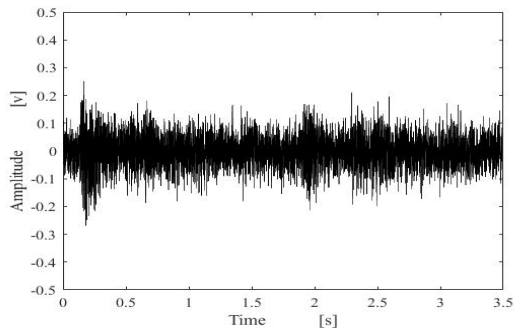


Figure 9: Estimated results by using proposed method.

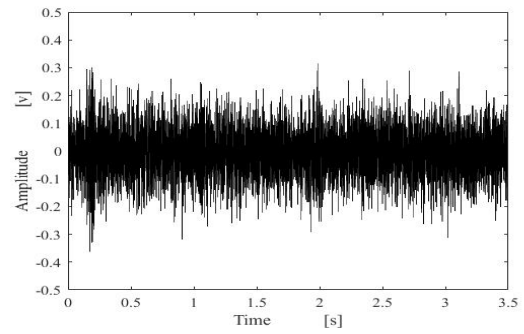


Figure 13: Estimated results by using compared method.

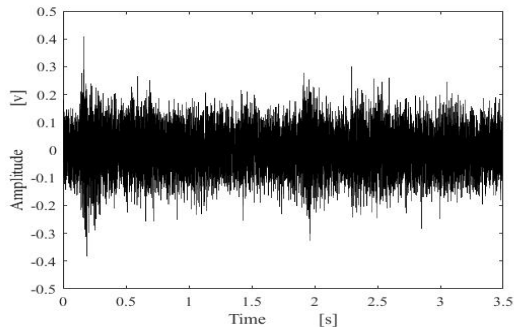


Figure 10: Estimated results by using compared method.

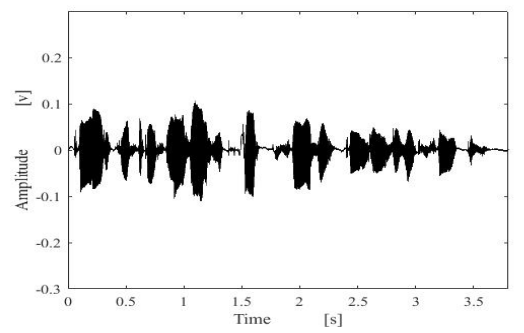


Figure 14: Original female speech signal

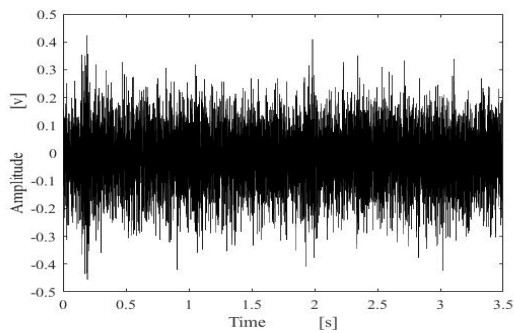


Figure 11: Male speech signal containing a machine noise

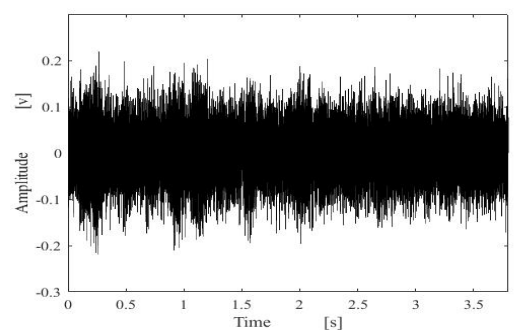


Figure 15: Female speech signal containing a white noise

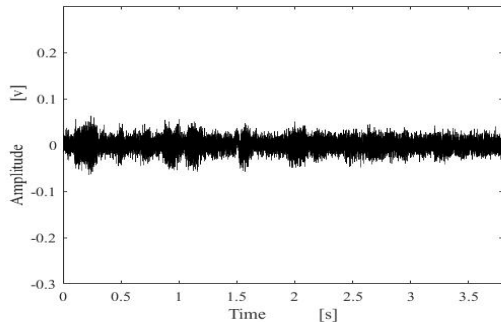


Figure 16: Estimated results by using proposed method.

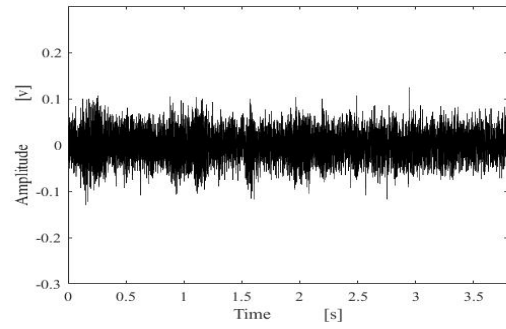


Figure 20: Estimated results by using compared method.

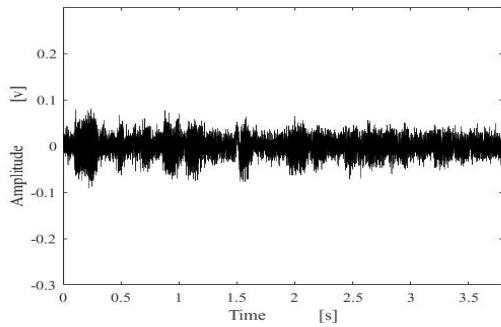


Figure 17: Estimated results by using compared method.

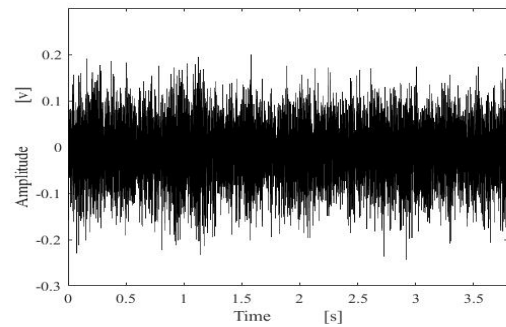


Figure 21: Female speech signal containing a machine noise

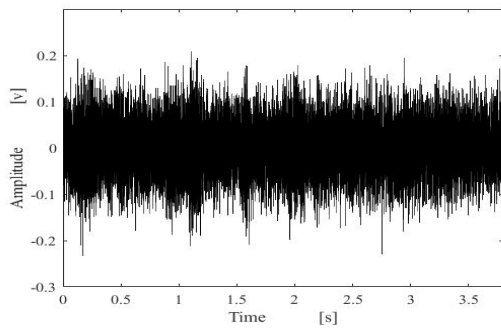


Figure 18: Female speech signal containing a pink noise

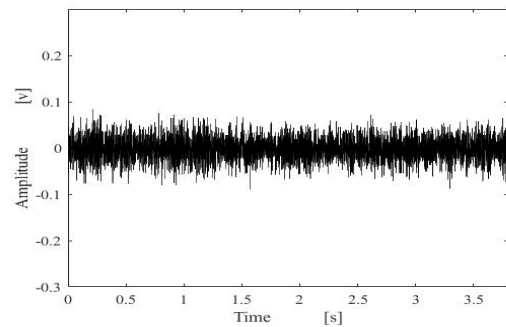


Figure 22: Estimated results by using proposed method.

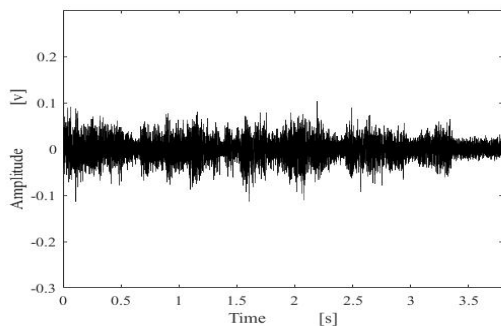


Figure 19: Estimated results by using proposed method.

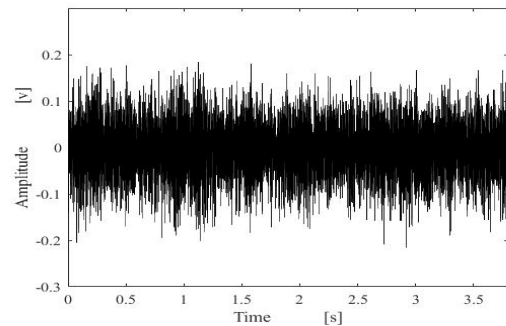


Figure 23: Estimated results by using compared method.

4 Conclusion

In this paper, a new method to suppress noise for speech signal has been proposed, which is applicable to actual en-

vironment with non-Gaussian and non-white noises. The aim of the proposed method is to improve the accuracy of estimation by using air- and bone-conducted speech signals.

The proposed method considered σ -points of not only x_k but also unknown parameter a_k , b_k and observation values y_k , z_k . Moreover, this study has proposed a method including the higher order correlation information between σ -points. Our algorithm has been realized by utilizing the Bayes' theorem as the fundamental principle of estimation and UT method using σ -points. Application of our algorithm has been made to real speech signal contaminated by noises. It has been revealed by experiments that better estimation results could be obtained by the proposed algorithm as compared with the method without using bone-conducted speech signal. However, we have not tried to apply the proposed algorithm to real speech recognition by use of a voice recognition software. Therefore, by applying the algorithm to speech recognition system, the effectiveness of the theory has to be confirmed experimentally.

The proposed approach is quite different from those 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: (i) Introduction of a realistic nonlinear model expressing the actual propagation characteristics of bone-conducted speech signal instead of the simple model in (2). (ii) Consideration of higher order expansion coefficients A_{lmnst} ($l, m, n, s, t \geq 3$), in the estimation algorithm. (iii) Selection of an optimal point to put the sensor to measure the bone-conducted speech signal.

References

- [1] M. Gabrea, E. Grivel, and M. Najim, "A single microphone Kalman filter-based noise canceller", *IEEE Signal Process. Lett.*, 6 (3), 55-57, 1999.
- [2] W. Kim, and H. Ko, "gNoise variance estimation for Kalman filtering of noisy speech", *IEICE Trans. Inf. and Syst.*, E84-D (1), 155-160, 2001.
- [3] N. Tanabe, T. Furukawa, and S. Tsuji, "Robust noise suppression algorithm with the Kalman filter theory for white and colored disturbance", *IEICE Trans. Fundamentals*, E91-A (3), 818-829, 2008.
- [4] R. E. Kalman, "A new approach to linear filtering and prediction problems", *Trans. ASME, Series, D, J. Basic Engineering*, 82 (1), 35-45, 1960.
- [5] R. E. Kalman and R. S. Buch, "New results in linear filtering and prediction theory", *Trans. ASME Series D, J. Basic Engineering*, 83 (1), 95-108, 1961.
- [6] H. J. Kushner, "Approximations to optimal nonlinear filter", *IEEE Trans. on Automatic Control*, 12 (5), 546-556, 1967.
- [7] A. Ikuta, H. Orimoto and G. Gallagher, "Noise suppression method by jointly using bone- and air-conducted speech signals", *Noise Control Engr. J.* 66 (6), 472-488, 2018.
- [8] H. Orimoto and A. Ikuta, "Signal processing for Noise cancellation method of speech signal by using an extension type UKF", *Proceedings of SIGNAL PROCESSING algorithms, architectures, arrangements, and applications (SPA)*, 304-309, 2018.
- [9] James V. Candy, *Bayesian Signal Processing Classical, Modern, and Particle Filtering Methods*, Wiley-IEEE Press, 2009.
- [10] M. Ohta and H. Yamada, "New Methodological Trials of Dynamical State Estimation for the Noise and Vibration Environmental System", *Acustica*, 55 (4), 199-212, 1984.
- [11] M. Ohta and T. Koizumi, "General statistical treatment of the response of a non-linear rectifying device to a stationary random input", *IEEE Trans. Inf. Theory*, 14 (4), 595-598, (1968).