Intelligent Information Management, 2021, 13, 199-213
https://www.scirp.org/journal/iim
ISSN Online: 2160-5920
ISSN Print: 2160-5912

Speech Signal Detection Based on Bayesian Estimation by Observing Air-Conducted Speech under Existence of Surrounding Noise with the Aid of Bone-Conducted Speech

Hisako Orimoto[1], Akira Ikuta[1], Kouji Hasegawa[2]

[1]Department of Management Information Systems, Prefectural University of Hiroshima, Hiroshima, Japan
[2]Western Region Industrial Research Center, Hiroshima Prefectural Technology Research Institute, Kure, Japan
Email: orimoto@pu-hiroshima.ac.jp

How to cite this paper: 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

Received: May 31, 2021
Accepted: July 4, 2021
Published: July 7, 2021

Abstract

In order to apply speech recognition systems to actual circumstances such as inspection and maintenance operations in industrial factories to recording and reporting routines at construction sites, etc. where hand-writing is difficult, some countermeasure methods for surrounding noise are indispensable. In this study, a signal detection method to remove the noise for actual speech signals is proposed by using Bayesian estimation with the aid of bone-conducted speech. 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 is theoretically derived. In the proposed speech detection method, bone-conducted speech is utilized in order to obtain precise estimation for speech signals. The effectiveness of the proposed method is experimentally confirmed by applying it to air- and bone-conducted speeches measured in real environment under the existence of surrounding background noise.

Keywords

Speech Signal Detection, Bayesian Estimation, Air- and Bone-Conducted Speeches, Surrounding Noise

1. Introduction

Many kinds of speech recognition systems have been developed according to the progress of digital information technique. For example, these systems are applied to inspection and maintenance operations in industrial factories and to re-

cording and reporting routines at construction sites, etc. For speech recognition in such actual circumstances, some countermeasure 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 prior information on the number of noise sources, and the number of microphones needed is larger than that of the noise sources 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 reduction task for speech signals based on a single microphone, many algorithms applying Kalman filter have been proposed up to now by assuming Gaussian white noise [6] [7] [8]. 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 [9].

Furthermore, in our previous study, a signal processing method to remove the noise for actual speech signals was proposed by jointly using the measured data of bone- and air-conducted speeches [10]. However, the algorithm of the previous method was highly complicated because it utilized lower and higher order correlations between the original speech signals, bone- and air-conducted speeches. Therefore, large computation time was required in the application to real speech signals data. Furthermore, a time transition model (*i.e.*, system equation) of the speech signals was needed for recursive estimation, and it had to be established for each speech signal in advance.

In this study, a method to detect the speech signals is proposed by applying the Bayesian estimation based on a posterior probability with observation data of air-conducted speech contaminated by surrounding background noise. In the proposed algorithm, by regarding the probability distribution with parameters based on the measurement of bone-conducted speech as a prior probability distribution, the precise estimation of the speech signals can be achieved. Though the bone-conducted speech is a kind of solid propagation sound with less effect by the surrounding noise, the high-frequency components of the signal are damped through the propagation process [11]. On the other hand, the air-conducted speech contains all frequency components though the signal is strongly affected by the surrounding noise. Therefore, by using jointly both air- and bone-conducted speeches, more accurate estimations of the speech signals can be expected whilst recovering the high-frequency components of the speech signals even in a very noisy circumstance.

The algorithm derived in this study does not require any time transition models for speech signals, and can be applied to speech signals with arbitrary fluctuation forms. Furthermore, since only the correlation information between the

speech signals and the observation of air-conducted speech is utilized in the proposed method, the estimation algorithm of the speech signals can be simplified, and the online processing can be expected due to the large reduction of the computation time. The effectiveness of the proposed method is confirmed by applying it to air- and bone-conducted speeches measured in an anechoic room at Hiroshima Prefectural Technology Research Institute cooperated with Prefectural University of Hiroshima, under the existence of surrounding background noise.

2. Detection Method for Air- and Bone-Conducted Speeches

2.1. Stochastic Model for Air- and Bone-Conducted Speeches

In the actual environment with a surrounding noise, let x_k, y_k and z_k be the original speech signal, the observation of air- and bone-conducted speech signals at a discrete time *k*. The observation y_k is contaminated by a surrounding background noise v_k. According to the additive property of sound pressure, the following relationship can be established.

$$y_k = x_k + v_k, \tag{1}$$

where the statistics of v_k are assumed to be known.

In order to express the relationship between the original speech signal and bone-conducted speech, the correlation information between x_k and z_k is necessary in general. However, it is difficult to find the information in advance because x_k is an unknown signal to be estimated. In this study, a conditional probability distribution function in orthogonal expansion series is adopted as the relationship between x_k and z_k:

$$P(x_k \mid z_k) = \frac{P(x_k, z_k)}{P(z_k)} = P(x_k) \sum_{r=0}^{\infty} \sum_{s=0}^{\infty} A_{rs} \theta_r^{(1)}(x_k) \theta_s^{(2)}(z_k) \tag{2}$$

with

$$A_{rs} \equiv \left\langle \theta_r^{(1)}(x_k) \theta_s^{(2)}(z_k) \right\rangle, \tag{3}$$

where $\langle\ \rangle$ denotes the averaging operation on variables. The linear and nonlinear correlations between x_k and z_k are reflected hierarchically in each expansion coefficient A_{rs}. From the definition of (3), the expansion coefficient satisfies the following conditions:

$$A_{00} = 1, \quad A_{r0} = A_{0s} = 0, \quad (r, s \geq 1). \tag{4}$$

Functions $\theta_r^{(1)}(x_k)$ and $\theta_s^{(2)}(z_k)$ are orthonormal polynomials having weighting functions $P(x_k)$ and $P(z_k)$ respectively, and can be composed as follows:

$$\theta_r^{(1)}(x_k) = \sum_{i=0}^{r} \lambda_{ri}^{(1)} x_k^i, \quad \theta_s^{(2)}(z_k) = \sum_{i=0}^{s} \lambda_{si}^{(2)} z_k^i, \tag{5}$$

where $\lambda_{ri}^{(1)}$ and $\lambda_{si}^{(2)}$ are coefficients calculated by using Schmidt's orthogonalization algorithm [12]. The expansion coefficients A_{rs} with order $r \leq R$,

$s \le S$ can be obtained from the correlation information between speech signal x_k and bone-conducted speech z_k. Since the speech signal is unknown in the presence noises, 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 $\mathbf{a}$.

$$\mathbf{a} \equiv \left(a_{11}, \cdots, a_{R1}, a_{12}, \cdots, a_{R2}, \cdots, a_{1S}, \cdots, a_{RS}\right)',$$
$$a_{rs} \equiv A_{rs}, \quad \left(r = 1, 2, \cdots, R; s = 1, 2, \cdots, S\right), \tag{6}$$

where $'$ denotes the transpose of a matrix, and $R \cdot S$ is the number of unknown parameters to be estimated. Then a simple dynamical model:

$$\mathbf{a}_{k+1} = \mathbf{a}_k, \tag{7}$$

is introduced for the simultaneous estimation of the parameter and the clean speech signal x_k.

2.2. Derivation of Speech Signal Detection Algorithm Based on Bayesian Estimation

To derive an estimation algorithm for the speech signal x_k, we place our basis on Bayes' theorem for the conditional probability distribution [13]. Since the parameter $\mathbf{a}_k$ is also unknown, the conditional probability distribution of x_k, $\mathbf{a}_k$ is expressed by

$$P\left(x_k, \mathbf{a}_k \mid Y_k\right) = \frac{P\left(x_k, \mathbf{a}_k, y_k \mid Y_{k-1}\right)}{P\left(y_k \mid Y_{k-1}\right)}, \tag{8}$$

where $Y_k \left(= \{y_1, y_2, \cdots, y_k\}\right)$ is a set of air-conducted speech data up to time *k*. By expanding the conditional joint probability distribution $P\left(x_k, \mathbf{a}_k, y_k \mid Y_{k-1}\right)$ in a statistical orthogonal expansion series on the basis of the well-known standard probability distributions, which describe the dominant part of the actual fluctuation, the following expression is derived.

$$\begin{aligned} P\left(x_k, \mathbf{a}_k \mid Y_k\right) = {} & P_0\left(x_k \mid Y_{k-1}\right) P_0\left(\mathbf{a}_k \mid Y_{k-1}\right) \sum_{l=0}^{\infty} \sum_{\mathbf{m}=\mathbf{0}}^{\infty} \sum_{n=0}^{\infty} B_{l\mathbf{m}n} \\ & \cdot \varphi_l^{(1)}\left(x_k\right) \varphi_{\mathbf{m}}^{(2)}\left(\mathbf{a}_k\right) \varphi_n^{(3)}\left(y_k\right) \Big/ \sum_{n=0}^{\infty} B_{00n} \varphi_n^{(3)}\left(y_k\right) \\ & \left(\sum_{\mathbf{m}=\mathbf{0}}^{\infty} \equiv \sum_{m_{11}=0}^{\infty} \cdots \sum_{m_{RS}=0}^{\infty}, \mathbf{m} \equiv \left(m_{11}, \cdots, m_{RS}\right) \right) \end{aligned} \tag{9}$$

with

$$B_{l\mathbf{m}n} \equiv \left\langle \varphi_l^{(1)}\left(x_k\right) \varphi_{\mathbf{m}}^{(2)}\left(\mathbf{a}_k\right) \varphi_n^{(3)}\left(y_k\right) \mid Y_{k-1} \right\rangle. \tag{10}$$

The above three functions $\varphi_l^{(1)}\left(x_k\right)$, $\varphi_{\mathbf{m}}^{(2)}\left(\mathbf{a}_k\right)$ and $\varphi_n^{(3)}\left(y_k\right)$ are orthonormal polynomials of degrees *l*, $\mathbf{m}$ and *n* with weighting functions $P_0\left(x_k \mid Y_{k-1}\right)$, $P_0\left(\mathbf{a}_k \mid Y_{k-1}\right)$ and $P_0\left(y_k \mid Y_{k-1}\right)$.

As examples of standard probability functions for the speech signal, the parameters and observations of the air-conducted speech, we adopt Gaussian distri-

butions, as

$$P_0\left(x_k \mid Y_{k-1}\right) = N\left(x_k; x_k^*, \Gamma_{x_k}\right),$$

$$P_0\left(\mathbf{a}_k \mid Y_{k-1}\right) = \prod_{r=1}^{R}\prod_{s=1}^{S} N\left(a_{rs,k}; a_{rs,k}^*, \Gamma_{a_{rs,k}}\right),$$

$$P_0\left(y_k \mid Y_{k-1}\right) = N\left(y_k; y_k^*, \Omega_k\right) \tag{11}$$

with

$$N\left(x;\mu,\sigma^2\right) \equiv \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{\left(x-\mu\right)^2}{2\sigma^2}\right\},$$

$$x_k^* \equiv \left\langle x_k \mid Y_{k-1}\right\rangle,\ \ \Gamma_{x_k} \equiv \left\langle \left(x_k - x_k^*\right)^2 \mid Y_{k-1}\right\rangle,$$

$$a_{rs,k}^* \equiv \left\langle a_{rs,k} \mid Y_{k-1}\right\rangle,\ \ \Gamma_{a_{rs,k}} \equiv \left\langle \left(a_{rs,k} - a_{rs,k}^*\right)^2 \mid Y_{k-1}\right\rangle,$$

$$y_k^* \equiv \left\langle y_k \mid Y_{k-1}\right\rangle,\ \ \Omega_k \equiv \left\langle \left(y_k - y_k^*\right)^2 \mid Y_{k-1}\right\rangle. \tag{12}$$

The orthonormal polynomials with three weighting probability distributions in (11) are then specified as

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

$$\varphi_{\mathbf{m}}^{(2)}\left(\mathbf{a}_k\right) = \prod_{r=}^{R}\prod_{s=1}^{S} \frac{1}{\sqrt{m_{rs}!}} H_{m_{rs}}\left(\frac{a_{rs,k} - a_{rs,k}^*}{\sqrt{\Gamma_{a_{rs,k}}}}\right),$$

$$\varphi_n^{(3)}\left(y_k\right) = \frac{1}{\sqrt{n!}} H_n\left(\frac{y_k - y_k^*}{\sqrt{\Omega_k}}\right), \tag{13}$$

where $H_l(\,)$ denotes the Hermite polynomial with lth order [14]. The non-Gaussian properties of the speech signal and observations of the air-conducted speech are reflected in each expansion coefficient $B_{\mathbf{lm}n}$.

Based on (9), the estimates of x_k and $a_{rs,k}$ for mean can be expressed as

$$\hat{x}_k \equiv \left\langle x_k \mid Y_k\right\rangle = \frac{\sum_{n=0}^{\infty}\left\{B_{\mathbf{00}n} C_{\mathbf{00}}^{1,\mathbf{0}} + B_{\mathbf{10}n} C_{\mathbf{10}}^{1,\mathbf{0}}\right\}\varphi_n^{(3)}\left(y_k\right)}{\sum_{n=0}^{\infty} B_{\mathbf{00}n}\varphi_n^{(3)}\left(y_k\right)}, \tag{14}$$

$$\hat{a}_{rs,k} \equiv \left\langle a_{rs,k} \mid Y_k\right\rangle = \frac{\sum_{n=0}^{\infty}\left\{B_{\mathbf{00}n} C_{\mathbf{0}0}^{0,1} + B_{\mathbf{01}n} C_{\mathbf{01}}^{0,1}\right\}\varphi_n^{(3)}\left(y_k\right)}{\sum_{n=0}^{\infty} B_{\mathbf{00}n}\varphi_n^{(3)}\left(y_k\right)} \tag{15}$$

with

$$C_{\mathbf{00}}^{1,\mathbf{0}} = x_k^*,\ \ C_{\mathbf{10}}^{1,\mathbf{0}} = \sqrt{\Gamma_{x_k}},\ \ C_{\mathbf{00}}^{0,1} = a_{rs,k}^*,\ \ C_{\mathbf{01}}^{0,1} = \sqrt{\Gamma_{a_{rs,k}}}. \tag{16}$$

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

$$P_{a_{rs,k}} \equiv \left\langle \left(a_{rs,k} - \hat{a}_{rs,k} \right)^2 \mid Y_k \right\rangle = \frac{\sum_{n=0}^{\infty} \left\{ B_{00n} C_{00}^{0,2} + B_{01n} C_{01}^{0,2} + B_{02n} C_{02}^{0,2} \right\} \varphi_n^{(3)}\left(y_k \right)}{\sum_{n=0}^{\infty} B_{00n} \varphi_n^{(3)}\left(y_k \right)} \tag{17}$$

with

$$C_{00}^{0,2} = \Gamma_{a_{rs,k}} + \left(a_{rs,k}^* - \hat{a}_{rs,k} \right)^2, \; C_{01}^{0,2} = 2\sqrt{\Gamma_{a_{rs,k}}} \left(a_{rs,k}^* - \hat{a}_{rs,k} \right), \; C_{02}^{0,2} = \sqrt{2}\Gamma_{a_{rs,k}}. \tag{18}$$

Using the property of conditional expectation, (1) (2) and (7), the variables in (14) can be calculated as follows:

$$y_k^* = \left\langle x_k + v_k \mid Y_{k-1} \right\rangle = x_k^* + \bar{v}_k, \;\; \left(\bar{v}_k \equiv \left\langle v_k \right\rangle \right), \tag{19}$$

$$\Omega_k = \left\langle \left(x_k + v_k - x_k^* - \bar{v}_k \right)^2 \mid Y_{k-1} \right\rangle = \Gamma_{x_k} + R_k, \;\; \left(R_k \equiv \left\langle \left(v_k - \bar{v}_k \right)^2 \right\rangle \right), \tag{20}$$

$$\begin{aligned} x_k^* &= \left\langle \int x_k P\left(x_k \mid z_k \right) \mathrm{d}x_k \mid Y_{k-1} \right\rangle \\ &= \left\langle \sum_{r=0}^{1} \sum_{s=0}^{\infty} d_{1r} A_{rs} \theta_s^{(2)}\left(z_k \right) \mid Y_{k-1} \right\rangle \\ &= \sum_{r=0}^{1} \sum_{s=0}^{\infty} d_{1r} a_{rs,k}^* \theta_s^{(2)}\left(z_k \right), \end{aligned} \tag{21}$$

$$\begin{aligned} \Gamma_{x_k} &= \left\langle \int \left(x_k - x_k^* \right)^2 P\left(x_k \mid z_k \right) \mathrm{d}x_k \mid Y_{k-1} \right\rangle \\ &= \left\langle \sum_{r=0}^{2} \sum_{s=0}^{\infty} d_{2r} A_{rs} \theta_s^{(2)}\left(z_k \right) \mid Y_{k-1} \right\rangle \\ &= \sum_{r=0}^{2} \sum_{s=0}^{\infty} d_{2r} a_{rs,k}^* \theta_s^{(2)}\left(z_k \right), \end{aligned} \tag{22}$$

$$a_{rs,k}^* = \left\langle a_{rs,k-1} \mid Y_{k-1} \right\rangle = \hat{a}_{rs,k-1}, \tag{23}$$

$$\Gamma_{a_{rs,k}} = \left\langle \left(a_{rs,k-1} - \hat{a}_{rs,k-1} \right)^2 \mid Y_{k-1} \right\rangle = P_{a_{rs,k-1}}. \tag{24}$$

The coefficients d_{1r} and d_{2r} in (21) and (22) are determined in advance by expanding x_k and $\left(x_k - x_k^* \right)^2$ in the orthogonal series of $\theta_r^{(1)}\left(x_k \right)$, as follows:

$$\begin{aligned} &d_{10} = -\lambda_{10}^{(1)} / \lambda_{11}^{(1)}, \;\; d_{11} = 1/\lambda_{11}^{(1)}, \\ &d_{20} = x_k^{*2} + \left(\lambda_{10}^{(1)} / \lambda_{11}^{(1)} \right) \left(2x_k^* + \lambda_{21}^{(1)} / \lambda_{22}^{(1)} \right) - \lambda_{20}^{(1)} / \lambda_{22}^{(1)}, \\ &d_{21} = -\left(1/\lambda_{11}^{(1)} \right) \left(2x_k^* + \lambda_{21}^{(1)} / \lambda_{22}^{(1)} \right), \;\; d_{22} = 1/\lambda_{22}^{(1)}. \end{aligned} \tag{25}$$

Furthermore, substituting (1) into (13) and using an additive theorem of Hermite polynomial:

$$\begin{aligned} &\frac{\left(\xi_1^2 + \xi_2^2 + \cdots + \xi_\varsigma^2 \right)^{n/2}}{n!} H_n \left(\frac{\xi_1 X_1 + \xi_2 X_2 + \cdots + \xi_\varsigma X_\varsigma}{\sqrt{\xi_1^2 + \xi_2^2 + \cdots + \xi_\varsigma^2}} \right) \\ &= \sum_{\eta_1 + \eta_2 + \cdots + \eta_\varsigma = n} \prod_{i=1}^{\varsigma} \frac{\xi_i^{\eta_i}}{\eta_i!} H_{\eta_i}\left(X_i \right), \end{aligned} \tag{26}$$

the orthonormal polynomial $\varphi_n^{(3)}(y_k)$ can be expressed as follows:

$$\varphi_n^{(3)}(y_k) = \frac{1}{\sqrt{n!}} H_n\left(\frac{\sqrt{\Gamma_{x_k}}\left(\frac{x_k - x_k^*}{\sqrt{\Gamma_{x_k}}}\right) + \sqrt{R_k}\left(\frac{v_k - \overline{v}_k}{\sqrt{R_k}}\right)}{\sqrt{\Gamma_{x_k} + R_k}}\right), \quad (27)$$

$$= \frac{1}{\sqrt{n!}} \sum_{i=0}^{n} \binom{n}{i} \left(\frac{\Gamma_{x_k}}{\Omega_k}\right)^{\frac{n-i}{2}} \left(\frac{R_k}{\Omega_k}\right)^{\frac{i}{2}} H_{n-i}\left(\frac{x_k - x_k^*}{\sqrt{\Gamma_{x_k}}}\right) H_i\left(\frac{v_k - \overline{v}_k}{\sqrt{R_k}}\right)$$

Therefore, using (2) and (27), the expansion coefficient $B_{\mathbf{l}\mathbf{m}n}$ defined by (10) can be calculated as follows:

$$\begin{aligned} B_{\mathbf{l}\mathbf{m}n} &= \frac{1}{\sqrt{n!}} \sum_{i=0}^{n} \binom{n}{i} \left(\frac{\Gamma_{x_k}}{\Omega_k}\right)^{\frac{n-i}{2}} \left(\frac{R_k}{\Omega_k}\right)^{\frac{i}{2}} \\ &\cdot \left\langle \varphi_{\mathbf{m}}^{(2)}(\mathbf{a}_k) \int \varphi_l^{(1)}(x_k) H_{n-i}\left(\frac{x_k - x_k^*}{\sqrt{\Gamma_{x_k}}}\right) P(x_k \mid z_k) \mathrm{d}x_k \mid Y_{k-1} \right\rangle \left\langle H_i\left(\frac{v_k - \overline{v}_k}{\sqrt{R_k}}\right) \right\rangle \\ &= \frac{1}{\sqrt{n!}} \sum_{i=0}^{n} \binom{n}{i} \left(\frac{\Gamma_{x_k}}{\Omega_k}\right)^{\frac{n-i}{2}} \left(\frac{R_k}{\Omega_k}\right)^{\frac{i}{2}} \\ &\cdot \left\langle \varphi_{\mathbf{m}}^{(2)}(\mathbf{a}_k) \sum_{r=0}^{l+n-i} \sum_{s=0}^{\infty} d_{l+n-i,r} A_{rs} \theta_s^{(2)}(z_k) \mid Y_{k-1} \right\rangle \left\langle H_i\left(\frac{v_k - \overline{v}_k}{\sqrt{R_k}}\right) \right\rangle \\ &= \sum_{i=0}^{n} \frac{\sqrt{n!}}{i!(n-i)!} \left(\frac{\Gamma_{x_k}}{\Omega_k}\right)^{\frac{n-i}{2}} \left(\frac{R_k}{\Omega_k}\right)^{\frac{i}{2}} \sum_{r=0}^{l+n-i} \sum_{s=0}^{\infty} d_{l+n-i,r} \\ &\cdot \left\langle \varphi_{\mathbf{m}}^{(2)}(\mathbf{a}_k) a_{rs,k} \mid Y_{k-1} \right\rangle \left\langle H_i\left(\frac{v_k - \overline{v}_k}{\sqrt{R_k}}\right) \right\rangle \theta_s^{(2)}(z_k) \end{aligned} \quad (28)$$

where $d_{l+n-i,r}$ is appropriate coefficient that satisfies the following equality:

$$\varphi_l^{(1)}(x_k) H_{n-i}\left(\frac{x_k - x_k^*}{\sqrt{\Gamma_{x_k}}}\right) = \sum_{j=0}^{l+n-i} d_{l+n-i,j} \theta_j^{(1)}(x_k). \quad (29)$$

From (19)-(22) and (28), the variables y_k^*, Ω_k and the expansion coefficient $B_{\mathbf{l}\mathbf{m}n}$ in the estimation algorithms (14)-(18) are given by the measurement data of bone-conducted speech z_k, estimates of parameter a_{rs} at the discrete time $k-1$ and statistics of the surrounding noise v_k. Therefore, the estimation of the speech signal can be performed by observing air-conducted speech y_k in a recursive way.

The flow chart of the proposed speech signal detection algorithm is illustrated in **Figure 1**. As compared with the previously reported algorithm [10], time transition model for the speech signal is not required in the proposed algorithm and the calculation process of the algorithm can be fairly simplified.

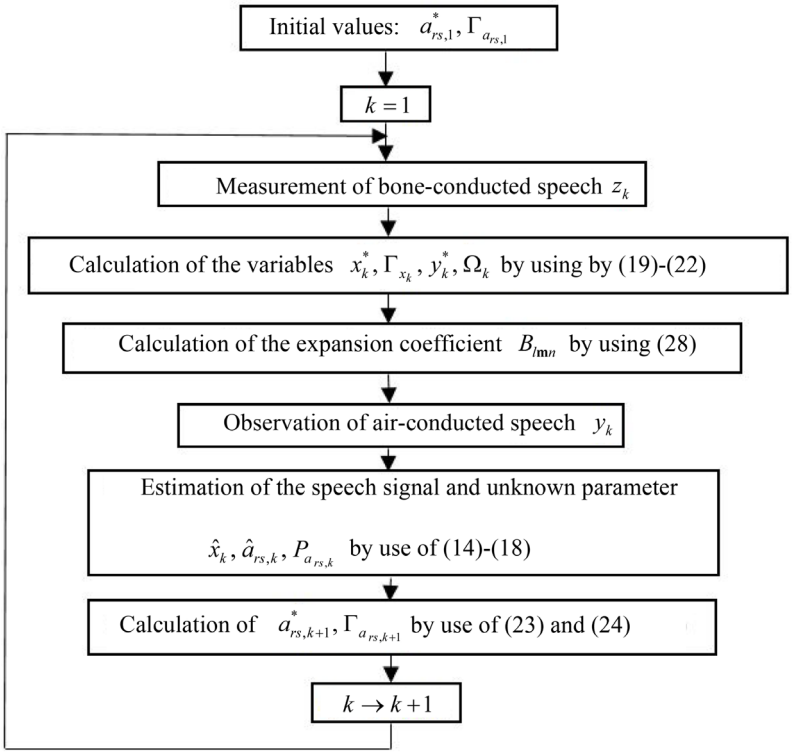

Figure 1. Flow chart of the proposed signal detection algorithm.

3. Application to Real Speech Signal

In order to confirm the effectiveness of the proposed signal detection algorithm, it was applied to real speech signals. The speech signal data were measured in the anechoic chamber in the acoustic laboratory building of the West Region Industrial Research Centre, Hiroshima Prefectural Technology Research Institute. For a male and a female 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, we created noisy air-conducted speeches on a computer by mixing the original air-conducted speech signal measured in a noise-free environment with machine noise recorded in advance, as an example of actual surrounding noise. By setting the amplitude (*i.e.*, mean squared value of instantaneous signal) of the noise to 1, 2, 3, 4, 5 and 10 times of that of the noise-free speech signals, we have applied the proposed algorithm to extremely difficult situations with low SNR. Furthermore, the bone-conducted speech was simultaneously measured by use of an acceleration sensor with the air-conducted speech. The noise-free air-conducted male speech signal and the created noisy air-conducted speech observation by using machine noise with the same amplitude as the noise-free speech signal are shown in **Figure 2** and **Figure 3**, and the observed wave of the bone-conducted speech is shown in **Figure 4**. Furthermore, for the female speech signal, the noise-free air-conducted speech signal, noisy air-conducted speech observation and bone-conducted speech are respectively shown in **Figures 5-7**.

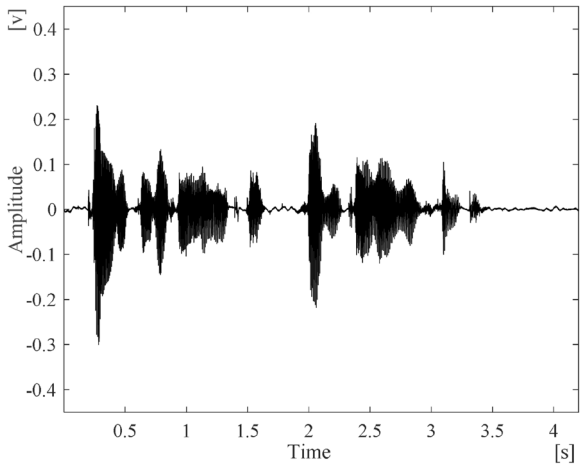

Figure 2. Noise free male speech signal.

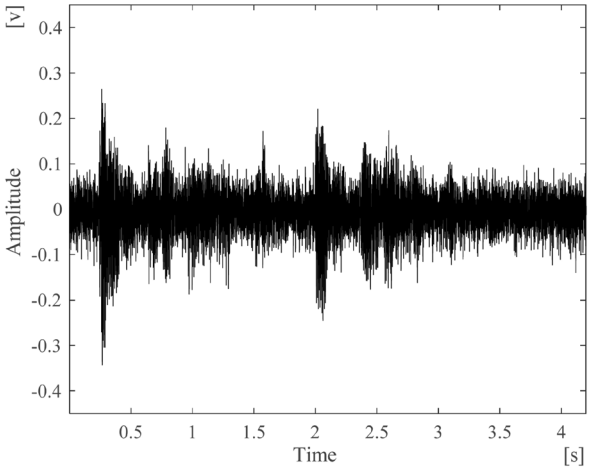

Figure 3. Noisy air-conducted speech observation by using machine noise with the same amplitude as the noise-free male speech signal.

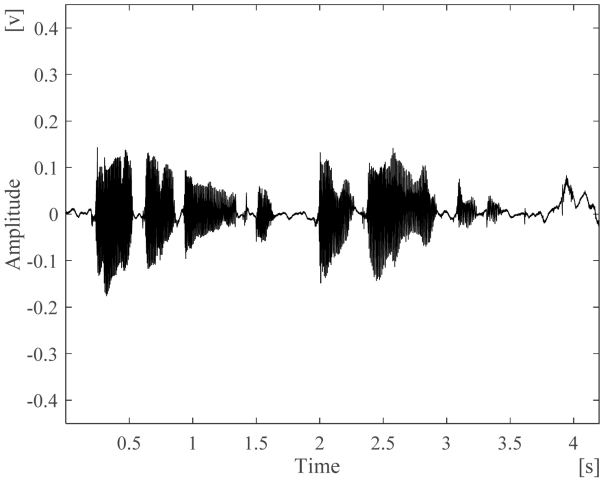

Figure 4. The observed wave of the bone-conducted male speech.

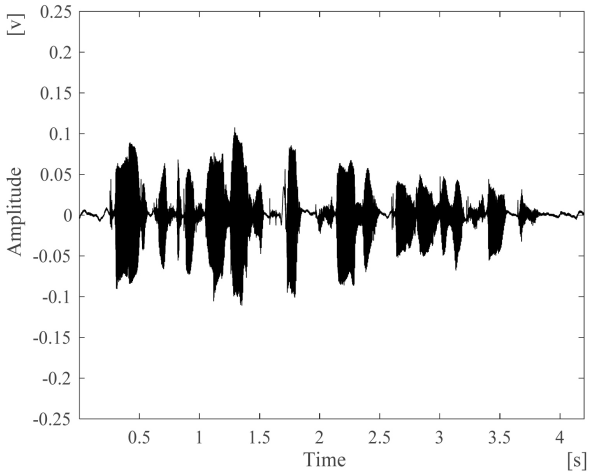

Figure 5. Noise-free female speech signal.

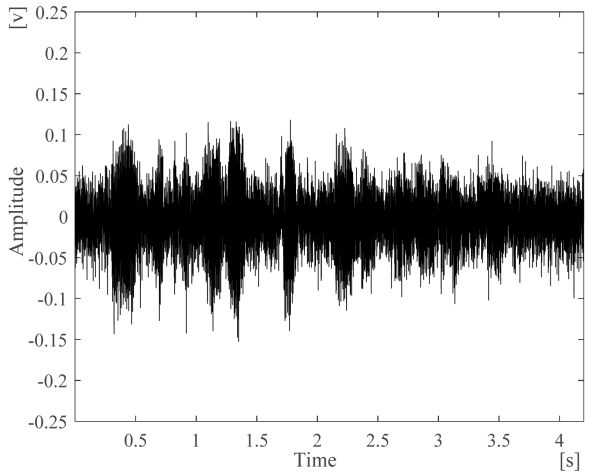

Figure 6. Noisy air-conducted speech observation by using machine noise with the same amplitude as the noise-free female speech signal.

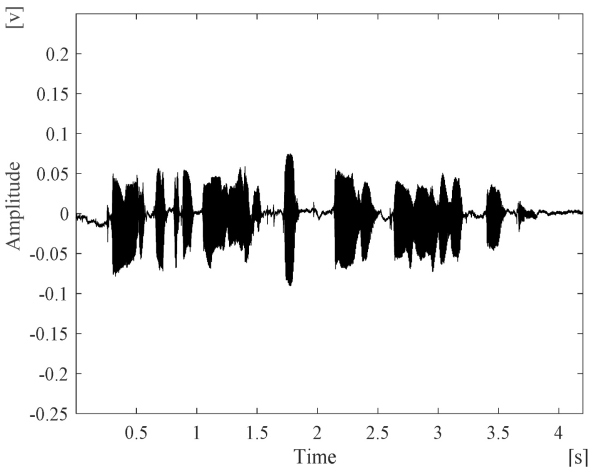

Figure 7. The observed wave of the bone-conducted female speech.

The estimated results by using the algorithm based on (14)-(18) are shown in Figure 8 for the male speech signal and in Figure 9 for the female speech signal. For comparison, the estimated results of the male and female speech signals by using the estimation algorithm based on only the observation of air-conducted speech are shown in Figure 10 and Figure 11.

Furthermore, the estimated results by the previously reported method [10] are shown in Figure 12 for the male speech signal and in Figure 13 for the female speech signal.

By comparing Figure 8, Figure 10 and Figure 12 with the original male speech signal shown in Figure 2, and comparing Figure 9, Figure 11, Figure 13 with Figure 5, it is obvious that the proposed method can suppress the effects of real machine noise better than the method based on observation of only air-conducted speech and the previously reported method.

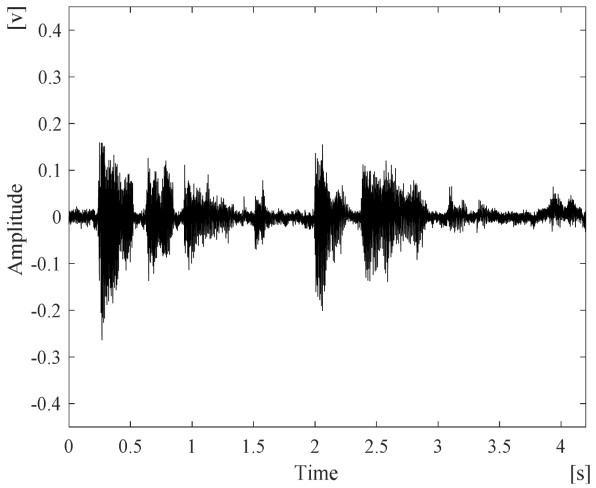

Figure 8. Estimated male speech signal by use of the proposed method.

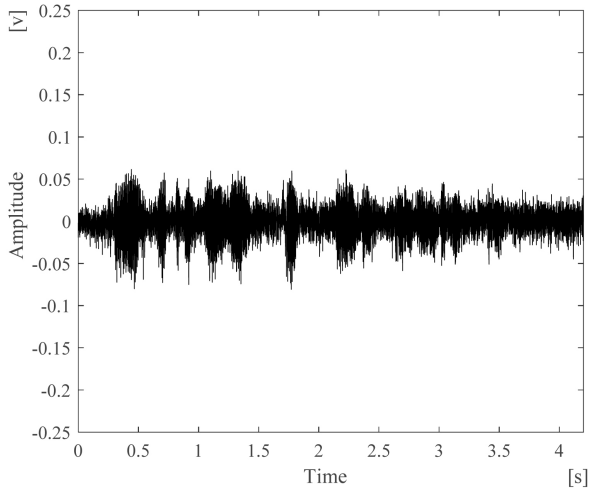

Figure 9. Estimated female speech signal by use of the proposed method.

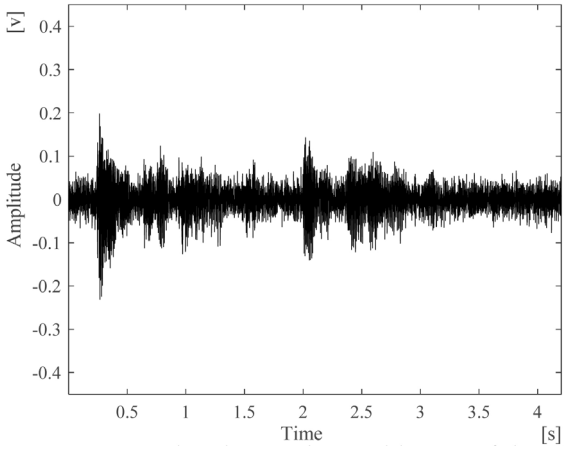

Figure 10. Estimated male speech signal by use of the method based on only the observation of air-conducted speech.

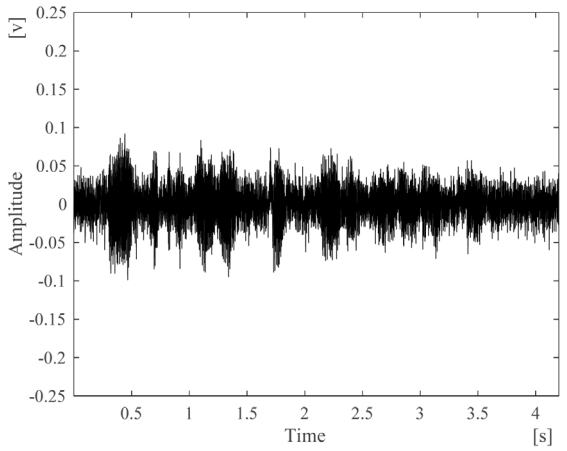

Figure 11. Estimated female speech signal by use of the method based on only the observation of air-conducted speech.

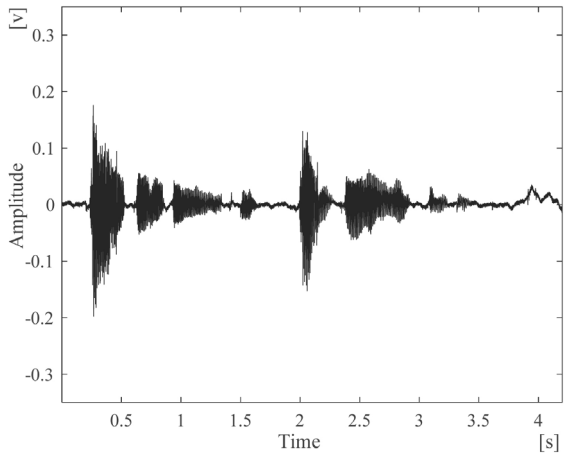

Figure 12. Estimated male speech signal by use of the previous method.

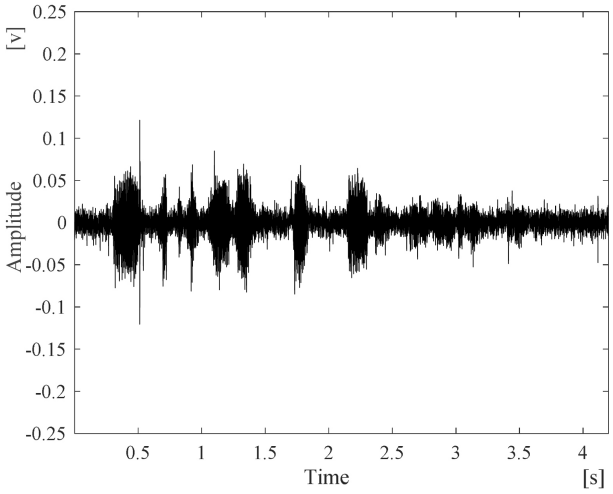

Figure 13. Estimated female speech signal by use of the previous method.

Table 1. Performance comparisons for a male speech signal contaminated by machine noise.

S/N Ratio	RMS Error			PEI		
	Proposed Method	Compared Method	Previous Method	Proposed Method	Compared Method	Previous Method
1/1	0.017125	0.026065	0.019172	5.4578	1.8092	4.4769
1/2	0.019116	0.031325	0.019757	4.5022	0.21255	4.2161
1/3	0.020016	0.034177	0.020262	4.1030	−0.54429	3.9967
1/4	0.020533	0.036012	0.020798	3.8813	−0.99868	3.7657
1/5	0.020870	0.037291	0.022099	3.7400	−1.3018	3.2428
1/10	0.021611	0.040187	0.022310	3.4369	−1.9514	3.1602

Table 2. Performance comparisons for a female speech signal contaminated by machine noise.

S/N Ratio	RMS Error			PEI		
	Proposed Method	Compared Method	Previous Method	Proposed Method	Compared Method	Previous Method
1/1	0.015122	0.018922	0.014828	3.3381	1.3909	3.5083
1/2	0.017412	0.022791	0.017592	2.1131	-0.22529	2.0424
1/3	0.018452	0.024837	0.019537	1.6092	-0.97169	1.1132
1/4	0.019044	0.026115	0.020103	1.3350	-1.4075	0.86505
1/5	0.019422	0.026977	0.020456	1.1645	-1.6899	0.71357
1/10	0.020310	0.028752	0.021072	0.8014	-2.2431	0.47891

The estimation RMS (root mean square) error and the PEI (performance evaluation index) defined by

$$\text{RMS Error} = \frac{1}{N}\sum_{k=1}^{N}\left(x_k - \hat{x}_k\right)^2, \tag{30}$$

$$\mathrm{PEI} = 10\log_{10}\left(\frac{\sum x_k^2}{\sum\left(x_k - \hat{x}_k\right)^2}\right)[\mathrm{dB}]. \tag{31}$$

are shown in Table 1 (the male speech signal) and Table 2 (the female speech signal).

Furthermore, the computation time of the proposed method was reduced by 39.3% of the previous method. From these results, the improved effectiveness of the proposed method in the simplified algorithm with the aid of bone-conducted speech can be clearly noticed in comparison with the estimation by the compared method based on the observation of only air-conducted speech and the previous method in the complicated algorithm.

4. Conclusions

4.1. Novel Contribution

In this study, a new method to detect speech signals under the existence of surrounding noise has been proposed from the viewpoint of Bayesian estimation by observing air-conducted speech with the aid of measurement of bone-conducted speech. Furthermore, it has been revealed by experiments that the proposed method is more effective than the method based on the observation of only air-conducted speech and the previous method in the complicated algorithm, to remove the surrounding noise in real noise environment.

4.2. Future Researches

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; 3) finding an optimal number of expansion terms for the expansion-based probability expression adopted; and 4) improvement of estimation precision by considering higher order statistics of surrounding noise.

Acknowledgements

The authors are grateful to Mr. Daishi Takagi for his help during this study. This work was supported in part by the fund from the Grant-in-Aid for Scientific Research No.19 K04428 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] Boll, S.F. (1979) Suppression of Acoustic Noise in Speech Using Spectral Subtrac-

tion. *IEEE Transactions on Acoustics*, *Speech*, *and Signal Processing*, **27**, 113-120. https://doi.org/10.1109/TASSP.1979.1163209

[2] Virag, N. (1999) Single Channel Speech Enhancement Based on Masking Properties of the Human Auditory System. *IEEE Transactions on Speech and Audio Processing*, **7**, 126-137. https://doi.org/10.1109/89.748118

[3] Kaneda, Y. and Ohga, J. (1986) Adaptive Microphone-Array System for Noise Reduction. *IEEE Transactions on Acoustics*, *Speech*, *and Signal Processing*, **34**, 1391-1400. https://doi.org/10.1109/TASSP.1986.1164975

[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.

[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] Gabrea, M. Griel, E. and Najim, M. (1999) A Single Microphone Kalman Filter-Based Noise Canceller. *IEEE Signal Processing Letters*, **6**, 55-57. https://doi.org/10.1109/97.744623

[7] 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.

[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] 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

[10] 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

[11] Shin, H.S. Kang, H.G. and Fingscheidt, T. (2012) Survey of Speech Enhancement Supported by a Bone Conduction Microphone. *Proceedings of* 10*th ITG Conference on Speech Communication*, Braunschweig, January 2012, 47-50.

[12] Ohta, M. and Yamada. H. (1984) New Methodological Trials of Dynamical State Estimation for the Noise and Vibration Environmental System—Establishment of General Theory and Its Application to Urban Noise Problems. *Acta Acustica United with Acustica*, **55**, 199-212. https://www.ingentaconnect.com/content/dav/aaua/1984/00000055/00000004/art00003

[13] Ikuta, A. Tokhi, M.O. and Ohta, M. (2011) A Cancellation Method of Background Noise for a Sound Environment System with Unknown Structure. *IEICE Transactions on Fundamentals of Electronics*, *Communications and Computer Sciences*, **E84-A**, 457-466.

[14] Ohta, M. and Koizumi, T. (1968) General Statistical Treatment of the Response of a Nonlinear Rectifying Device to a Stationary Random Input (Corresp.). *IEEE Transactions on Information Theory*, **14**, 595-598. https://doi.org/10.1109/TIT.1968.1054178