A BAYESIAN APPROACH FOR NOISE SUPPRESSION OF SPEECH SIGNAL IN REAL ENVIRONMENT

Akira Ikuta, Hisako Orimoto, and Yegui Xiao

Department of Management Information Systems, Prefectural University of Hiroshima
Ujina-higashi 1-1-71, Minami-ku, 734-8558, Hiroshima, Japan
phone: + (81) 82 251 9763, fax: + (81) 82 251 9405, email: {ikuta, orimoto, xiao}@pu-hiroshima.ac.jp
http://www.pu-hiroshima.ac.jp/

ABSTRACT

Numerous noise suppression methods for speech signals have been developed up to now. In this paper, a new method to suppress noise in speech signals is proposed, which requires a single microphone only and doesn't need any priori information on both noise spectrum and pitch. It works in the presence of noise with high amplitude. More specifically, an adaptive noise suppression algorithm applicable to real-life speech recognition is proposed without assuming the Gaussian white noise, which performs effectively even though the noise statistics and the fluctuation form of speech signal are unknown. The effectiveness of the proposed method is confirmed by applying it to real speech signals contaminated by noises.

1. INTRODUCTION

Speech recognition systems have been applied in various fields due to recent development of digital signal processing technique. For example, these systems are applied to inspection and maintenance operations in industrial factories and to recording and reporting routines at construction sites, etc. where hand-writing is difficult. For speech recognition in such actual circumstances, some countermeasure methods for surrounding noises are indispensable.

Previously reported methods for noise reduction in the speech recognition can be classified into two categories. One is based on single microphone [1, 2] and the other uses of microphone array [3]. Since the latter requires priori 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 this paper, in order to suppress the noises that inevitably exist in the observed speech signal of actual environment and probably have large amplitude, a practical method based on single microphone is considered, which does not require any priori information on noise spectrum and pitch.

In such a noise suppression task for speech signal, many algorithms applying Kalman filter have been proposed up to now [6-8]. However, the Kalman filter is originally based on the assumption of Gaussian white noise, and the information of noise variance must be given in advance [9]. However, the actual noises show complex fluctuation forms with non-Gaussian and non-white properties, and the statistics of the noises are unknown in general. Furthermore, in many previous algorithms, time series models such as AR (auto-regressive) model [10] are introduced for speech signal, and the parameters of the models have to be estimated in advance. Speech enhancement methods based on Bayesian estimation have been proposed by introducing simplified approximate models such as Gaussian and AR models [11, 12]. Therefore, the previously reported noise suppression methods have many problems if applied to real speech signals.

In this paper, a new noise suppression algorithm for actual speech signals contaminated by non-Gaussian and non-white noises is theoretically proposed. More specifically, first, expansion expressions of the conditional probability distribution reflecting the information on linear and nonlinear correlation among the time series of the speech signal and noisy observation are adopted to express the system and observation characteristics. Next, a method to estimate adaptively the time series of the speech signal is derived on the basis of the Bayes' theorem of probability. The proposed method can be applied adaptively to actual complex situation where both the noise statistics and the fluctuation forms of speech signal are unknown. Furthermore, by applying the proposed algorithm to real speech signals with several kinds of noises, its effectiveness is experimentally confirmed.

Throughout the paper, expectation operation on variable x and expectation on x conditioned by y are expressed as $<x>$ and $<x \mid y>$ respectively.

2. NOISE SUPPRESSION ALGORITHM FOR SPEECH SIGNAL

2.1 Stochastic Model for Speech Signal

In the actual environment with a surrounding noise, let x_k and y_k be the speech signal and the observation at a discrete time k. The observation y_k is contaminated by a noise with unknown statistics. In this paper, an adaptive signal processing method to estimate x_k based on the successive observation of y_k is theoretically derived.

In order to estimate the parameters of linear and/or nonlinear time series models for speech signal, the correlation information on time series of speech signal x_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. Furthermore, since the speech signal usually shows complicated fluctuation forms, the probability distribution reflecting the whole information on the fluctuation must be introduced and utilized effectively and wisely in order to achieve precise estimates for x_k. In this paper, a new adaptive algorithm for noise suppression is proposed by introducing a time transition probability of x_k.

In order to estimate recursively the speech signal based on the noisy observation, the correlation information among x_k, x_{k+1} and y_k has to be utilized. Therefore, attention is focused on the joint probability distribution function $P(x_k, x_{k+1}, y_k)$ reflecting all linear and non-linear correlation information among x_k, x_{k+1} and y_k. Expanding the joint probability distribution function $P(x_k, x_{k+1}, y_k)$ in an orthogonal form based on the product of $P(x_k)$, $P(x_{k+1})$ and $P(y_k)$, the following expression can be derived.

$$P(x_k, x_{k+1}, y_k) = P(x_k)P(x_{k+1})P(y_k) \cdot \sum_{r=0}^{\infty}\sum_{s=0}^{\infty}\sum_{t=0}^{\infty} A_{rst}\theta_r^{(1)}(x_k)\theta_s^{(1)}(x_{k+1})\theta_t^{(2)}(y_k) \quad (1)$$

with

$$A_{rst} \equiv <\theta_r^{(1)}(x_k)\theta_s^{(1)}(x_{k+1})\theta_t^{(2)}(y_k)> \quad (2)$$

The linear and non-linear correlation information among x_k, x_{k+1} and y_k is reflected hierarchically in each expansion coefficient A_{rst}. Functions $\theta_r^{(1)}(x_k)$ and $\theta_t^{(2)}(y_k)$ are orthonormal polynomials having weighting functions $P(x_k)$ and $P(y_k)$, respectively. These orthonormal polynomials can be decomposed by using Schmidt's orthogonalization algorithm [13]. From (1), the time transition probability of x_k is given as

$$P(x_{k+1} \mid x_k) = P(x_k, x_{k+1}) / P(x_k) = P(x_{k+1})\sum_{r=0}^{\infty}\sum_{s=0}^{\infty} A_{rs0}\theta_r^{(1)}(x_k)\theta_s^{(1)}(x_{k+1}). \quad (3)$$

Furthermore, the conditional probability distribution function $P(y_k \mid x_k)$ of the observation y_k is given by

$$P(y_k \mid x_k) = P(x_k, y_k) / P(x_k) = P(y_k)\sum_{r=0}^{\infty}\sum_{t=0}^{\infty} A_{r0t}\theta_r^{(1)}(x_k)\theta_t^{(2)}(y_k). \quad (4)$$

The expansion coefficients A_{rs0} and A_{r0t} with orders $r \le R$, $s \le S$, $t \le T$ can be obtained from the correlation information of time series x_k, and the correlation relationship between speech signal x_k and noisy observation y_k. Since the speech signal is unknown in the presence of noises, these coefficients have to be estimated on the basis of the observation y_k. Let's regard the expansion coefficients A_{rs0} and A_{r0t} as unknown parameters $\mathbf{a}$ and $\mathbf{b}$ (column vectors):

$$\mathbf{a} \equiv (a_1,\ a_2,\ ...,\ a_I)' \equiv (\mathbf{a}_{(1)}',\ \mathbf{a}_{(2)}',\ ...,\ \mathbf{a}_{(S)}')',$$

$$\mathbf{a}_{(s)} \equiv (A_{1s0},\ A_{2s0},\ ...,\ A_{Rs0})',\ (s = 1,\ 2,\ ...,\ S), \quad (5)$$

$$\mathbf{b} \equiv (b_1,\ b_2,\ ...,\ b_J)' \equiv (\mathbf{b}_{(1)}',\ \mathbf{b}_{(2)}',\ ...,\ \mathbf{b}_{(T)}')',$$

$$\mathbf{b}_{(t)} \equiv (A_{10t},\ A_{20t},\ ...,\ A_{R0t})',\ (t = 1,\ 2,\ ...,\ T), \quad (6)$$

where ' denotes the transpose of a matrix, and $I\ (= R \cdot S)$, $J\ (= R \cdot T)$ are the number of unknown expansion coefficients to be estimated. Then simple dynamical models:

$$\mathbf{a}_{k+1} = \mathbf{a}_k, \quad \mathbf{b}_{k+1} = \mathbf{b}_k, \quad (7)$$

may be introduced for the simultaneous estimation of both the parameters and the clean speech signal x_k. Since the parameters $\mathbf{a}$ and $\mathbf{b}$ reflect the correlation information between x_k, x_{k+1}, and x_k, y_k respectively, time constant models of (7) are introduced for simplification of the algorithm.

2.2 Derivation of Noise Suppression Algorithm Based on Bayes' Theorem

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

$$P(x_k, \mathbf{a}_k, \mathbf{b}_k \mid Y_k) = \frac{P(x_k, \mathbf{a}_k, \mathbf{b}_k, y_k \mid Y_k)}{P(y_k \mid Y_k)}, \quad (8)$$

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

$$P(x_k, \mathbf{a}_k, \mathbf{b}_k \mid Y_k) = \sum_{l=0}^{\infty}\sum_{\mathbf{m}=\mathbf{0}}^{\infty}\sum_{\mathbf{n}=\mathbf{0}}^{\infty}\sum_{q=0}^{\infty} B_{l\mathbf{mn}q} P_0(x_k \mid Y_{k-1}) \cdot P_0(\mathbf{a}_k \mid Y_{k-1}) P_0(\mathbf{b}_k \mid Y_{k-1}) \varphi_l^{(1)}(x_k)\varphi_{\mathbf{m}}^{(2)}(\mathbf{a}_k) \cdot \varphi_{\mathbf{n}}^{(3)}(\mathbf{b}_k)\varphi_q^{(4)}(y_k) \Big/ \sum_{q=0}^{\infty} B_{\mathbf{000}q}\varphi_q^{(4)}(y_k) \quad (9)$$

with

$$B_{l\mathbf{mn}q} \equiv <\varphi_l^{(1)}(x_k)\varphi_{\mathbf{m}}^{(2)}(\mathbf{a}_k)\varphi_{\mathbf{n}}^{(3)}(\mathbf{b}_k)\varphi_q^{(4)}(y_k) \mid Y_{k-1}> \quad (10)$$

The above four functions $\varphi_l^{(1)}(x_k)$, $\varphi_{\mathbf{m}}^{(2)}(\mathbf{a}_k)$, $\varphi_{\mathbf{n}}^{(3)}(\mathbf{b}_k)$ and $\varphi_q^{(4)}(y_k)$ are orthonormal polynomials of degrees l, $\mathbf{m}$, $\mathbf{n}$ and q with weighting functions $P_0(x_k \mid Y_{k-1})$, $P_0(\mathbf{a}_k \mid Y_{k-1})$, $P_0(\mathbf{b}_k \mid Y_{k-1})$ and $P_0(y_k \mid Y_{k-1})$. As examples of standard probability functions for the speech signal, the parameters and observation, here we adopt Gaussian distributions with means and variances, as

$$x_k^* \equiv <x_k \mid Y_{k-1}>,\ \Gamma_{x_k} \equiv <(x_k - x_k^*)^2 \mid Y_{k-1}>,$$

$$a_{i,k}^* \equiv <a_{i,k} \mid Y_{k-1}>,\ \Gamma_{a_{i,k}} \equiv <(a_{i,k} - a_{i,k}^*)^2 \mid Y_{k-1}>,$$

$$b_{i,k}^* \equiv <b_{i,k} \mid Y_{k-1}>,\ \Gamma_{b_{i,k}} \equiv <(b_{i,k} - b_{i,k}^*)^2 \mid Y_{k-1}>,$$

$$y_k^* \equiv <y_k \mid Y_{k-1}>,\ \Omega_k \equiv <(y_k - y_k^*)^2 \mid Y_{k-1}>. \quad (11)$$

The non-Gaussian properties of speech signal and observation are reflected in each expansion coefficient $B_{l\mathbf{mn}q}$. The orthonormal polynomials with the weighting probability distributions are then specified as Hermite polynomial [13].

Based on (9), the estimate of the polynomial function $f_{L,\mathbf{M},\mathbf{N}}(x_k,\mathbf{a}_k,\mathbf{b}_k)$ of x_k, $\mathbf{a}_k$ and $\mathbf{b}_k$ with $(L,\mathbf{M},\mathbf{N})$ th order can be derived as follows.

$$\hat{f}_{L,\mathbf{M},\mathbf{N}}(x_k,\mathbf{a}_k,\mathbf{b}_k) \equiv < f_{L,\mathbf{M},\mathbf{N}}(x_k,\mathbf{a}_k,\mathbf{b}_k) \mid Y_k > = \frac{\sum_{l=0}^{L}\sum_{\mathbf{m}=\mathbf{0}}^{\mathbf{M}}\sum_{\mathbf{n}=\mathbf{0}}^{\mathbf{N}}\sum_{q=0}^{\infty} C_{l\mathbf{mn}}^{L\mathbf{MN}} B_{l\mathbf{mn}q}\varphi_q^{(4)}(y_k)}{\sum_{q=0}^{\infty} B_{\mathbf{000}q}\varphi_q^{(4)}(y_k)}, \quad (12)$$

where $C_{l\mathbf{mn}}^{L\mathbf{MN}}$ is an appropriate constant satisfying the following equality:

$$f_{L,\mathbf{M},\mathbf{N}}(x_k,\mathbf{a}_k,\mathbf{b}_k) = \sum_{l=0}^{L}\sum_{\mathbf{m}=\mathbf{0}}^{\mathbf{M}}\sum_{\mathbf{n}=\mathbf{0}}^{\mathbf{N}} C_{l\mathbf{mn}}^{L\mathbf{MN}}\varphi_l^{(1)}(x_k) \cdot \varphi_{\mathbf{m}}^{(2)}(\mathbf{a}_k)\varphi_{\mathbf{n}}^{(3)}(\mathbf{b}_k). \quad (13)$$

Using the property of conditional expectation, (3), (4) and (7), the variables in (11) can be calculated as follows:

$$x_k^* = < \int x_k P(x_k \mid x_{k-1})dx_k \mid Y_{k-1} > = < \sum_{r=0}^{\infty}\sum_{s=0}^{1} e_{1s}A_{rs0}\theta_r^{(1)}(x_{k-1}) \mid Y_{k-1} > = \sum_{s=0}^{1} e_{1s} < \mathbf{A}'_{(s),k}\Theta(x_{k-1}) \mid Y_{k-1} >, \quad (14)$$

$$\Gamma_{x_k} = \sum_{s=0}^{2} e_{2s} < \mathbf{A}'_{(s),k}\Theta(x_{k-1}) \mid Y_{k-1} >, \quad (15)$$

$$a_{i,k}^* = < a_{i,k-1} \mid Y_{k-1} > = \hat{a}_{i,k-1}, \quad (16)$$

$$\Gamma_{a_{i,k}} = < (a_{i,k-1}-\hat{a}_{i,k-1})^2 \mid Y_{k-1} > = P_{a_{i,k-1}}, \quad (17)$$

$$b_{i,k}^* = < b_{i,k-1} \mid Y_{k-1} > = \hat{b}_{i,k-1}, \quad (18)$$

$$\Gamma_{b_{i,k}} = < (b_{i,k-1}-\hat{b}_{i,k-1})^2 \mid Y_{k-1} > = P_{b_{i,k-1}}, \quad (19)$$

$$y_k^* = < \int y_k P(y_k \mid x_k)dy_k \mid Y_{k-1} > = < \sum_{r=0}^{\infty}\sum_{t=0}^{1} d_{1t}A_{r0t}\int \theta_r^{(1)}(x_k)P(x_k \mid x_{k-1})dx_k \mid Y_{k-1} > = \sum_{r=0}^{\infty}\sum_{t=0}^{1} d_{1t} < A_{r0t}\mathbf{A}'_{(r),k}\Theta(x_{k-1}) \mid Y_{k-1} >, \quad (20)$$

$$\Omega_k = \sum_{r=0}^{\infty}\sum_{t=0}^{2} d_{2t} < A_{r0t}\mathbf{A}'_{(r),k}\Theta(x_{k-1}) \mid Y_{k-1} > \quad (21)$$

with

$$\mathbf{A}_{(r),k} \equiv (0,\ \mathbf{a}_{(r),k}')', \quad (r = 1,\ 2,\ \cdots),$$
$$\mathbf{A}_{(0),k} \equiv (1,\ 0,\ 0,\ \cdots,\ 0)',$$
$$\Theta(x_k) \equiv (\theta_0^{(1)}(x_k),\ \theta_1^{(1)}(x_k),\ \cdots,\ \theta_R^{(1)}(x_k))'. \quad (22)$$

The coefficients e_{1s}, e_{2s}, d_{1t}, d_{2t} in (14), (15), (20) and (21) are determined in advance by expanding x_k, $(x_k - x_k^*)^2$, y_k and $(y_k - y_k^*)^2$ in the following orthogonal series forms:

$$x_k = \sum_{i=0}^{1} e_{1i}\theta_i^{(1)}(x_k), \quad (x_k - x_k^*)^2 = \sum_{i=0}^{2} e_{2i}\theta_i^{(1)}(x_k), \quad (23)$$

$$y_k = \sum_{i=0}^{1} d_{1i}\theta_i^{(2)}(y_k), \quad (y_k - y_k^*)^2 = \sum_{i=0}^{2} d_{2i}\theta_i^{(2)}(y_k). \quad (24)$$

Furthermore, using (3), (4) and the orthonormal condition of $\theta_r^{(1)}(x_k)$ and $\theta_s^{(2)}(y_k)$, each expansion coefficient $B_{l\mathbf{mn}q}$ defined by (10) can be obtained as follows:

$$B_{l\mathbf{mn}q} = < \varphi_l^{(1)}(x_k)\varphi_{\mathbf{m}}^{(2)}(\mathbf{a}_k)\varphi_{\mathbf{n}}^{(3)}(\mathbf{b}_k) \cdot \int \varphi_q^{(4)}(y_k)P(y_k \mid x_k)dy_k \mid Y_{k-1} >$$
$$= < \varphi_{\mathbf{m}}^{(2)}(\mathbf{a}_k)\ \varphi_{\mathbf{n}}^{(3)}(\mathbf{b}_k) \cdot \sum_{r=0}^{\infty}\sum_{t=0}^{q} d_{qt}A_{r0t}\int \varphi_l^{(1)}(x_k)\theta_r^{(1)}(x_k)P(x_k \mid x_{k-1})dx_k \mid Y_{k-1} >$$
$$= \sum_{r=0}^{\infty}\sum_{t=0}^{q}\sum_{j=0}^{l+r} d_{qt}w_{l+r,j} < \varphi_{\mathbf{m}}^{(2)}(\mathbf{a}_k)\varphi_{\mathbf{n}}^{(3)}(\mathbf{b}_k) \cdot A_{r0t}\mathbf{A}'_{(j),k}\Theta(x_{k-1}) \mid Y_{k-1} >, \quad (25)$$

where d_{qi} and $w_{l+r,j}$ are appropriate coefficients that satisfy the following equalities:

$$\varphi_q^{(4)}(y_k) = \sum_{i=0}^{q} d_{qi}\theta_i^{(2)}(y_k),$$
$$\varphi_l^{(1)}(x_k)\theta_r^{(1)}(x_k) = \sum_{j=0}^{l+r} w_{l+r,j}\theta_j^{(1)}(x_k). \quad (26)$$

Substituting the dynamical models of $\mathbf{a}_k$ and $\mathbf{b}_k$ in (7) into (14)-(21) and (25), the parameters x_k^*, Γ_{x_k}, $a_{i,k}^*$, $\Gamma_{a_{i,k}}$, $b_{i,k}^*$, $\Gamma_{b_{i,k}}$, y_k^*, Ω_{y_k} and the expansion coefficient $B_{l\mathbf{mn}q}$ may be then given in functional forms on estimations of x_{k-1}, $\mathbf{a}_{k-1}$ and $\mathbf{b}_{k-1}$. Therefore, the estimation of the speech signal can be performed in a recursive way. The flow chart of the proposed adaptive noise suppression algorithm is illustrated in Fig. 1.

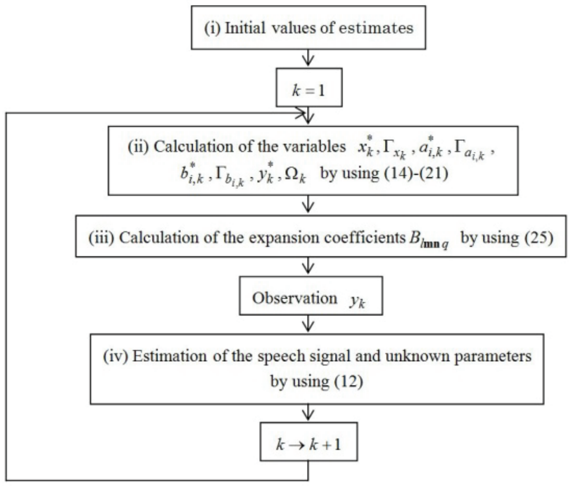

Fig. 1 Flow chart of the proposed noise suppression algorithm.

3. APPLICATION TO SPEECH SIGNAL IN REAL ENVIRONMENT

In order to confirm the effectiveness of the proposed adaptive noise suppression algorithm, it is applied to actual speech signals. More specifically, for a female and a male speech signals digitized with sampling frequency of 44.1 kHz, time duration of 2.72 sec and quantization of 16 bits, we estimated the speech signal based on the observation corrupted with additive noises. The following three kinds of noises were adopted: (a) white noise, (b) colored noise generated by the AR(1) model: $v_{k+1} = -0.5v_k + e_k$, where e_k is a white noise, and (c) machine noise. By setting the amplitude (i.e., mean squared value of instantaneous signal) of noises to 1, 2, 3, 4, 5 and 10 times of that of the speech signals, we have applied the proposed algorithm to extremely difficult situations with low S/N (signal to noise) Ratio. For comparison, noise suppression by introducing the linear systems:

$$y_k = \alpha_k x_k + \beta_k v_k\,,\quad x_{k+1} = F_k x_k + G_k u_k\,, \qquad (27)$$

are considered. In (27), v_k and u_k are random noises with zero-mean and variance 1, and α_k, β_k, F_k and G_k are unknown parameters. Therefore, the extended Kalman filter [14] is applied to estimate simultaneously the speech signal x_k, the parameters α_k, β_k, F_k and G_k by introducing the dynamic models of the parameters:

$$\alpha_{k+1} = \alpha_k\,,\ \beta_{k+1} = \beta_k\,,\ F_{k+1} = F_k\,,\ G_{k+1} = G_k\,. \qquad (28)$$

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

$$10\log_{10}(\sum x_k^2 / \sum (x_k - \hat{x}_k)^2)\ \ [\text{dB}]\,, \qquad (29)$$

are shown in Table 1 (the female speech signal) and Table 2 (the male speech signal) . From Tables 1 and 2, it is obvious that the proposed method performs more accurately than the extended Kalman filter does. Though the extended Kalman filter assumes an ideal situation of Gaussian white noise, real environmental noises are not such ideal one. The spectrum of the observed female speech signal contaminated by a machine noise is shown in Fig. 2. Furthermore, the spectra of the estimated signals by the proposed method and the extended Kalman filter are shown in Figs. 3 and 4 respectively. By comparing the estimated results with the spectrum of original female speech signal shown in Fig. 5, it is obvious that the proposed method can be suppressed the effects by the real machine noise than the extended Kalam filter.

Table 1 Performance comparisons for a female speech signal contaminated by (a) white noise, (b) colored noise, and (c) machine noise.

(a) White Noise

S/N Ratio	RMS Error		Performance Evaluation Index	
	Proposed Method	Extended Kalman Filter	Proposed Method	Extended Kalman Filter
1/1	0.02504	0.03028	2.263	0.6107
1/2	0.02566	0.03181	2.049	0.1826
1/3	0.02603	0.03259	1.926	-0.02864
1/4	0.02648	0.03238	1.775	0.03003
1/5	0.02705	0.03234	1.593	0.03865
1/10	0.02840	0.03241	1.168	0.02004

(b) Colored Noise

S/N Ratio	RMS Error		Performance Evaluation Index	
	Proposed Method	Extended Kalman Filter	Proposed Method	Extended Kalman Filter
1/1	0.02325	0.03263	2.905	-0.03844
1/2	0.02446	0.03289	2.466	-0.10728
1/3	0.02689	0.03263	1.643	-0.03826
1/4	0.02730	0.03297	1.511	-0.1281
1/5	0.02773	0.03249	1.374	-0.000007972
1/10	0.02905	0.03249	0.9701	-0.0004417

(c) Machine Noise

S/N Ratio	RMS Error		Performance Evaluation Index	
	Proposed Method	Extended Kalman Filter	Proposed Method	Extended Kalman Filter
1/1	0.02584	0.03249	1.990	-0.0003864
1/2	0.02777	0.03249	1.362	-0.0002714
1/3	0.02825	0.03249	1.214	-0.0002203
1/4	0.02901	0.03249	0.9827	-0.0001898
1/5	0.02976	0.03249	0.7623	-0.0001690
1/10	0.03161	0.03249	0.2386	-0.0001172

Table 2 Performance comparisons for a male speech signal contaminated by (a) white noise, (b) colored noise, and (c) machine noise.

(a) White Noise

S/N Ratio	RMS Error		Performance Evaluation Index	
	Proposed Method	Extended Kalman Filter	Proposed Method	Extended Kalman Filter
1/1	0.04802	0.06094	2.084	0.01407
1/2	0.05105	0.06095	1.553	0.01188
1/3	0.05298	0.06098	1.229	0.008043
1/4	0.05414	0.06100	1.042	0.005607
1/5	0.05491	0.06100	0.9183	0.005189
1/10	0.06195	0.06681	0.6580	0.002479

(b) Colored Noise

S/N Ratio	RMS Error		Performance Evaluation Index	
	Proposed Method	Extended Kalman Filter	Proposed Method	Extended Kalman Filter
1/1	0.04464	0.06104	2.718	-0.0006239
1/2	0.04684	0.06118	2.300	-0.01953
1/3	0.04839	0.06167	2.016	-0.08950
1/4	0.04953	0.06104	1.815	-0.00001808
1/5	0.05044	0.6104	1.655	-0.0001241
1/10	0.05326	0.06104	1.184	-0.00003442

(c) Machine Noise

S/N Ratio	RMS Error		Performance Evaluation Index	
	Proposed Method	Extended Kalman Filter	Proposed Method	Extended Kalman Filter
1/1	0.04671	0.06104	2.324	-0.0002187
1/2	0.05054	0.06104	1.639	-0.0001547
1/3	0.05282	0.06104	1.256	-0.0001263
1/4	0.05409	0.06104	1.050	-0.0001093
1/5	0.05494	0.06104	0.9138	-0.00009776
1/10	0.05730	0.06104	0.5487	-0.00006900

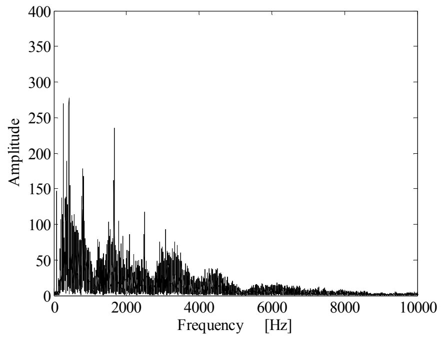

Fig. 2 Spectrum of observed signal contaminated by a machine noise.

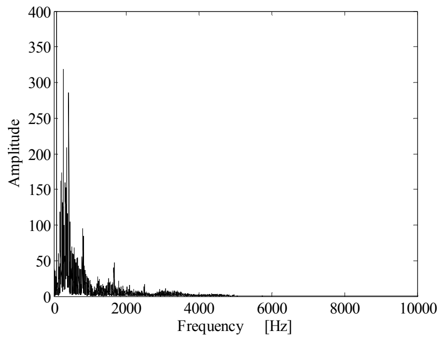

Fig. 3 Spectrum of the estimated speech signal by use of the proposed method for the observation contaminated by a machine noise.

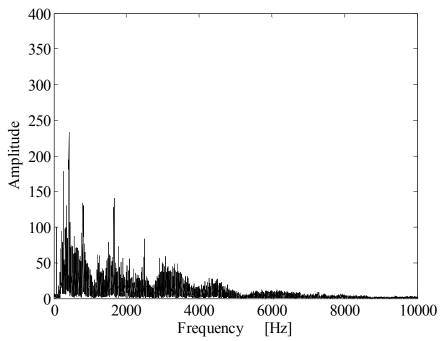

Fig. 4 Spectrum of the estimated speech signal by use of the extended Kalman filter for the observation contaminated by a machine noise.

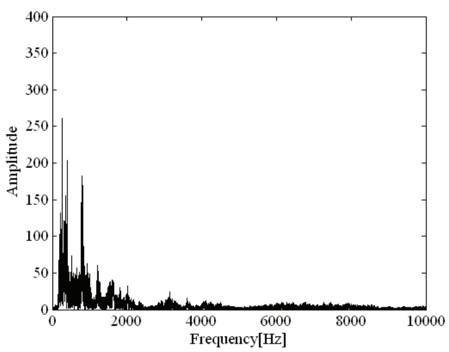

Fig. 5 Spectrum of original female speech signal.

4. CONCLUSIONS

In this paper, a new noise suppression algorithm for speech signal has been proposed, which is applicable to actual environment with non-Gaussian and non-white noises. The proposed method can be applied to real-life situations that both the noise statistics and the time transition relationship of speech signal are unknown. Our algorithm has been realized by introducing a conditional probability expression as the system model and utilizing the Bayes' theorem as the fundamental principle of estimation. Application of our algorithm has been made to real speech signals contaminated by noises. As a result, it has been revealed by experiments that better estimation results may be obtained by our algorithm as compared with the extended Kalman filter.

REFERENCES

[1]S.F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoustics, Speech, and Signal Process.*, vol.27, pp.113-120, Feb. 1979.

[2] N. Virag, "Single channel speech enhancement based on masking properties of the human auditory system," *IEEE Trans. Speech Audio Process.*, vol.7, pp.126-137, Feb. 1999.

[3] Y. Kaneda, and J. Ohga, "Adaptive microphone-array system for noise reduction," *IEEE Trans. Acoustics, Speech, and Signal Process.*, vol.34, pp.1391-1400, Jun. 1986.

[4] A. Kawamura, K. Fujii, Y. Itoh, and Y. Fukui, "A noise reduction method based on linear prediction analysis," *IEICE Trans. Fundamentals*, vol.J85-A, pp.415-423, 2002.

[5] A. Kawamura, Y. Iiguni, and Itoh, "A noise reduction method based on linear prediction with variable step-size," *IEICE Trans. Fundamentals*, vol.E88-A, pp.855-861, Apl. 2005.

[6] M. Gabrea, E. Grivel, and M. Najim, "A single microphone Kalman filter-based noise canceller," *IEEE Signal Process. Lett.*, vol.6, pp.55-57, Mar. 1999.

[7] W. Kim, and H. Ko, "Noise variance estimation for Kalman filtering of noisy speech," *IEICE Trans. Inf. & Syst.*, vol.E84-D, pp.155-160, Jan. 2001.

[8] N. Tanabe, T. Furukawa, and S. Tsuji, "Robust noise suppression algorithm with the Kalman filter theory for white and colored disturbance," *IEICE Trans. Fundamentals,* vol.E91-A, pp.818-829, Mar. 2008.

[9] R.E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME, Series D, J. Basic Engineering,* vol.82, pp.35-45, Jan. 1960.

[10] P.H. Franses, *Time Series Models for Business and Economic Forecasting*, Cambridge University Press, Cambridge, 1998.

[11] J. Hao, H. Attias, S. Nagarajan, and T.W. Lee, "Speech enhancement, gain, and noise spectrum adaptation using approximate Bayesian estimation," *IEEE Trans. Audio Speech Lang. Process.*, vol.17, pp.24-37, Jan. 2009

[12] S. Srinivasan, J. Samuelsson, and W.B. Kleijin, "Codebook-based Bayesian speech enhancement for nonstationary environment," *IEEE Trans. Audio Speech Lang. Process.*, vol.15, pp. 441-452, Feb. 2007

[13] A. Ikuta, M.O. Tokhi, and M. Ohta, "A cancellation method of background noise for a sound environment system with unknown structure," *IEICE Trans. Fundamentals,* vol.E84-A, pp.457-466, Feb. 2001.

[14] H.J. Kushner, "Approximations to optimal nonlinear filter," *IEEE Trans. Autom. Control*, vol.12, pp.546-556, May 1967