International Journal of Innovative Computing, Information and Control
Volume **8**, Number **7(B)**, July **2012**

 ISSN 1349-4198
pp. **4979–4994**

RELIABILITY-WEIGHTED HMM CONSIDERING INEXACT OBSERVATION AND ITS VALIDATION IN SPEAKER IDENTIFICATION

Md Tariquzzaman[1], Jin Young Kim[1], Seung You Na[1] and Hyoung-Gook Kim[2,*]

[1]School of Electronics and Computer Engineering
Chonnam National University
No. 300, Yongbong Dong, buk-Gu, gwangju, 500-757, South Korea
tareq@moiza.chonnam.ac.kr; { beyondi; syna }@jnu.ac.kr

[2]Department of Wireless Communication Engineering
Kwangwoon University
Wolgye-Dong, Nown-Gu, Seoul 139-701, South Korea
*Corresponding author: hkim@kw.ac.kr

Received May 2011; revised September 2011

Abstract. *The Hidden Markov Model (HMM) is widely used in pattern recognition areas such as speech and speaker recognition, handwritten recognition, gesture recognition. In this paper, we present a reliability-weighted HMM (RW-HMM) approach considering inexact observation. We introduce a weighting factor – confidence measures of observations – in HMM target function and we derive a training algorithm based on the traditional EM algorithm for optimizing a modified HMM target function. To verify the effectiveness of our proposed method, we performed speaker identification (SI) experiments using the ETRI and YOHO databases. The experimental results show that our proposed approach significantly outperforms the conventional approaches particularly to deal with the noisy environments.*
Keywords: Observation reliability, Hidden Markov model, EM algorithm, Speaker identification

1. **Introduction.** Signal observation with high precision is very important in signal processing and pattern recognition. In real applications, however, it is inevitable to experience corrupted signals. In general, signals are distorted by additive noises, channel characteristics or audio and video codecs. Thus, distorted signal observations degrade the recognizer's performance. There have been a number of approaches to cope with the problem of corrupted observations. Nevertheless, devising a solution to such an issue still remains an open research area.

The conventional approaches, specially applied in the pattern recognition field, can be grouped into three main categories: robust feature extraction (RFE) [1-9], model adaptation (MA) [10-16] and missing feature approach (MFA) [17-21]. In speech signal processing, cepstral mean subtraction (CMS) [1,3] and RASTA filtering [4] are the representative methods of RFE. Further, MA methods, known as speaker adaptation, are successfully applied in the speech recognition area [11,15]. MFAs are basically based on partial spectrographic information, where the two basic approaches termed as feature-vector imputation and marginalization are adopted in literature. The later approaches have achieved remarkable performance enhancements; nonetheless, the state-of-the-art approaches still have some limitations for application in mobile environment.

First, the previous approaches focus or deal with the problem of corrupted signals in the recognition phase under the assumption that the training signals are free from any signal distortions. Nevertheless, due to the flood of mobile devices in our society, the recognition developers currently face the abundance of distorted signals. Thus, it is necessary to develop training tools to treat the distorted signals in the training stage.

Second, even though RFE, MA and MFA are applied, there are still uncertainties about the features and recognition process. That is, speech features obtained via the CMS process are still affected by environmental noises.

Third, according to [21], MFA achieved the best performance enhancement. In spite of the successful results, MFA is hard to be adopted, as the implementation process is comparatively complex.

Thus, we need a novel and simple algorithm, which could cope with the corrupted signal problem in training and service phases. Recently, a promising modified algorithm based on observation confidence has been developed in GMM training and applied to speaker identification [22], which leads to a new dimension in research to conquer the mismatch condition problem in pattern recognition.

On the other hand, for modeling time series data, the Hidden Markov Model (HMM) [23-26] is an ever-present tool, which is broadly used in speech signal processing, computational molecular biology and other areas of artificial intelligence and pattern recognition. There are different research approaches that have focused on the structure of HMM for enhancing classification performance [23,24]. In the traditional HMM training, the observation vectors are treated evenly considering that the observations are free from any kind of uncertainty or inexactness due to signal distortion. In state-of-the-art research, some approaches have been focused on the uncertainty of observation in HMM aspect for robust pattern recognition [28-31]. Nevertheless, dealing with the uncertain observation is still a key area of research, particularly in HMM based pattern recognition domain. Specifically, during the likelihood calculation in HMM training, the uncertain observation should be considered based on the degree of uncertainty.

In this paper, we propose a reliability-weighted HMM as an extension of the commonly used classical HMM by introducing a weighting factor in the HMM target function. The weighting factors could be interpreted as the reliability values of the features obtained from distorted observations. We define a modified HMM object function and derive a training algorithm based on EM algorithm, and consequently the decoding algorithm is modified based on the reliability. Our proposed method does not change the basic structure of HMM, but it enhances performance by suppressing the contributions of observations with a comparatively high degree of corruption in probability calculation. To verify the performance of our proposal, the modified approach, RW-HMM, is applied to the subject of both text-dependent and text-independent speaker identification with the observation confidence, defined as a function of signal-to-noise ratio (SNR). We evaluate the proposed method using the ETRI and YOHO databases constructed for testing SI systems.

2. The Reliability-Weighted Hidden Markov Model (RW-HMM).

2.1. Motivation and definition of a modified object function.

Generally, signals are corrupted by noises in signal capture environments and distorted while being transmitted through channels. So, a captured signal $X(t)$ at time t can be expressed as $X(t) = h(t) * u(t) + \eta(t)$, where $u(t)$ is a pure input signal, $h(t)$ is the channel or convolutional noise and $\eta(t)$ is the additive noise signal at time t, respectively. The additive noises could be of different types and different magnitudes at the respective environment

such as the environment with babble noises (BN) or factory noises (FN). In a pattern recognition domain, it is typically assumed that the training database is a set of clean signals. Undoubtedly, we can get a clean database in many applications. However, with the advent of mobile and internet applications, the need for processing corrupted database as a training DB increases. Thus, it is required to train pattern recognition systems with corrupted data.

Let us assume that we have a function for measuring the uncertainty or inexactness of observed signals. The output of measure's function is termed as the reliability of the observed signals. Let us define ρ_t as the reliability of the tth observation. The reliability ρ_t can be regarded as a fuzzy membership function of the observation, which satisfies the condition that $0 \leq \rho_t \leq 1$. Then each observation should be controlled to have contribution depending on the corresponding reliability value in training (or modeling) and recognition processes. Now let us consider the simple problem of weighted sample mean and variance with a data set $[x_1, x_2, \cdots, x_n]$ and non-negative weights $[w_1, w_2, \cdots, w_n]$. The weighted mean $\bar{x}$ and variance $\bar{\sigma}$ are calculated as follows [35]:

$$\bar{x} = \frac{w_1 x_1 + w_2 x_2 + \cdots + w_n x_n}{w_1 + w_2 + \cdots + w_n}$$
$$\bar{\sigma} = \frac{w_1 (x_1 - \bar{x})^2 + w_2 (x_2 - \bar{x})^2 + \cdots + w_n (x_n - \bar{x})^2}{w_1 + w_2 + \cdots + w_n}$$

Therefore, the data elements with a high weight contribute more to the weighted mean comparing the data elements with a low weight imposing a constraint that the weights cannot be negative. In pattern recognition domain, the weight can be considered as the degree of data importance or each item's population. The weighted mean is used, for example, to aggregate a set of scores (e.g., examination scores on different subjects) to a single resultant score. In our approach the observation reliability is considered as the degree of data importance. Typically, the observations with high reliability values have high weight values. Considering our reliability condition we can interpret the reliability values as the weights without any modification.

By the way we can obtain the equations of the weighted sample mean and variance for the random variable with Gaussian probability density function by a simple modification of maximum likelihood estimation (MLE). For Gaussian case the following theorem is satisfied.

Theorem 2.1. *Let $\{X_t\} = [X_1, X_2, \cdots, X_T]$ be T observations, coming from Gaussian distribution, and $\{\rho_t\} = [\rho_1, \rho_2, \cdots, \rho_T]$, be corresponding weights (or reliability values). Then the weighted observation mean $\bar{\mu}_X$ and variance $\bar{\sigma}_X$ are obtained by maximizing the modified conditional probability, i.e.,*

$$\max_\theta \tilde{P}(X|\bar{\mu}_X, \bar{\sigma}_X) = \max_\theta \Pi_t P_t^{\rho_t}(X_t|\bar{\mu}_X, \bar{\sigma}_X),$$

where $P(X_t) = \dfrac{1}{\sigma_X \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_t - \bar{\mu}_X}{\sigma}\right)^2}$. The process is ML estimation for mean and variance.

Proof: The likelihood function of the parameters are given by

$$L(x_1, x_2, \cdots, x_T|\bar{\mu}_X, \bar{\sigma}_X) = \Pi_t P_t^{\rho_t}(X_t|\bar{\mu}_X, \bar{\sigma}_X) = \Pi_t \left[\left(\frac{1}{\sigma_X \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{X_t - \bar{\mu}_X}{\bar{\sigma}_X}\right)^2} \right)^{\rho_t} \right].$$

By applying log to the likelihood we obtain the log-likelihood function of

$$\lambda = \ln(L) = -\frac{T}{2} \sum\nolimits_{t=1}^{T} \rho_t \ln(2\pi) - T\rho_t \ln \bar{\sigma}_X - \frac{1}{2} \sum\nolimits_{t=1}^{T} \left(\frac{X_t - \bar{\mu}_X}{\bar{\sigma}_X} \right)^2 \rho_t.$$

The log-likelihood function is maximized with the parameters of $\bar{\mu}_X$ and $\bar{\sigma}_X$ as follows.

$$\frac{d(\lambda)}{d\bar{\mu}_X} = \frac{1}{\bar{\sigma}_X^2} \sum\nolimits_{t=1}^{T} (X_t - \bar{\mu}_X)\rho_t = 0,$$

and

$$\frac{d(\lambda)}{d\bar{\sigma}_X} = -\frac{T}{\bar{\sigma}_X} \sum\nolimits_{t=1}^{T} \rho_t + \frac{1}{\bar{\sigma}_X^3} \sum\nolimits_{t=1}^{T} (X_t - \bar{\mu}_X)^2 \rho_t = 0.$$

Solving the above equations, we get the following expression for mean $\bar{\mu}_X$ and variance $\bar{\sigma}_X$

$$\bar{\mu}_X = \frac{\sum_{t=1}^{T} \rho_t X_t}{\sum_{t=1}^{T} \rho_t},$$

and

$$\bar{\sigma}_X = \sqrt{\frac{\sum_{t=1}^{T} (X_t - \bar{\mu}_X)^2 \rho_t}{T \sum_{t=1}^{T} \rho_t}}.$$

In Theorem 2.1, the maximization is done for the cost function of $\Pi_t P_t^{\rho_t}(X_t|\mu, \sigma)$ instead of $\Pi_t P_t(X_t|\mu, \sigma)$. That is, the observation probability of P_t is modified into $P_t^{\rho_t}$. Thus we can conclude that the contribution of the tth observation with the reliability ρ_t could be controlled by introducing weighted probability of $P_t^{\rho_t}$ and thus statistically for modified MLE it is seen that the observations with low reliability values have low influences in the maximization process.

Now, let us consider the problem of applying the reliability function to HMM training. Before deriving the proposed algorithm, we define some notations for the HMM and RW-HMM as shown in Table 1.

HMM is defined as $\lambda = (A, B, \pi)$, where $A = [a_{ij}]$ is the state transition matrix, $\pi = [\pi_j]$ is the vector of initial probabilities, and $B = [b_{jm}] = [\mu_{jm}, \sum_{jm}, c_{jm}]$ is the emission matrix or the set of N_S output distribution mixtures [26]. The probability density function of continuous observation is a mixture of the form $b_j(o_t)$ as described in Equation (1).

$$b_j(o_t) = \sum\nolimits_{m=1}^{M} g_{jm} \phi\left(o_t; \mu_{jm}, \sum\nolimits_{jm}\right), \quad 1 \le j \le N \tag{1}$$

In Equation (1), the standard stochastic restraints are applied, such that the conditions in the set of Equation (2) are fulfilled.

$$g_{jm} \ge 0, \quad 1 \le j \le S, \quad 1 \le m \le M; \quad \sum\nolimits_{m=1}^{M} g_{jm} = 1, \quad 1 \le j \le N \tag{2}$$

In general, the kernel function ϕ is a Gaussian density. For an observation sequence with a length of T, the object function for calculating the probability of the observations is represented as in Equation (3).

$$L_\lambda(O) = \sum\nolimits_{s} a_{q_0} \Pi_{t=1}^{T} a_{q_{t-1} q_t} a\{b_{q_t}(o_t)\} \tag{3}$$

The object function in Equation (3) tells us that each observation is treated equally without any consideration of its reliability. It is because, in the classical HMM training, the observation vectors are considered to be clean or free of distortion. Now, let us assume that each observation has a different reliability value. Then, we can introduce the reliability for calculating Equation (3) as explained above. The object function in Equation (3) is thus modified as in Equation (4).

$$\tilde{L}_\lambda(O) = \sum\nolimits_{s} a_{q_0} \Pi_{t=1}^{T} a_{q_{t-1} q_t} a\{b_{q_t}(o_t)\}^{\rho_t} \tag{4}$$

TABLE 1. Notations for HMM and RW-HMM

$S = \{s_1, s_2, \cdots .s_N\}$. Set of states of a hidden Markov model.
N. Number of states.
M. Number of clusters within a state.
$O = (o_1, o_2, \cdots, o_T)$. A sequence of external observations.
o_t. A variable representing the external observations at time t.
T. The number of time steps in the sequence.
$q = (q_1, q_2, \cdots, q_T)$. A sequence of internal states.
q_t. A variable representing the internal state at time t.
$\lambda = (A, B, \pi)$. A hidden Markov model as defined by its A, B and π matrices.
a_{ij}. State transition probability from i to j state for a model λ.
$A = \{a_{ij}\}$. The $N \times N$ matrix of transition probabilities.
$b_j(o_t) = P_\lambda(o_t \mid q_t = j)$.
$\pi_i = P_\lambda(q_1 = i)$. Initial state probability for model λ.
$\pi = \{\pi_i\}$. The vector of initial state probabilities.
m_t. Variable identifying the cluster index of the cluster that occurred at time t.
$m = (m_1, m_2, \cdots, m_t)$. A sequence of cluster indexes, one per time step.
$g_{jk} = P_\lambda(m_t = k \mid q_t = j)$. Emission probability (gain) of kth cluster by state s_j.
$\phi(.)$. A kernel function (e.g., Gaussian) to model the probability density of a cluster.
μ_{ik}. The centroids or prototype of the kth cluster of ith state.
$\sum_{ik}$. The covariance matrix of the kth cluster of ith state.
$\alpha_t(i) = P_\lambda(o_1, o_2, \cdots, o_T \mid q_t = i)$. The forward variable for the sequence o at time t for state i.
$\beta_t(i) = P_\lambda(o_{t+1}, \cdots, o_T \mid q_t = i)$. The backward variable for the sequence o at time t for state i.
$\sim$. Indicates that a factor with the tilde ($\sim$) is modified so that the observation reliability is considered.
$\rho = (\rho_1, \rho_2, \cdots, \rho_T)$. A sequence of observation reliability values.
ρ_t. The observation reliability of tth observation.

In Equation (4), the probabilities are weighted by the corresponding reliability values. So, we can term the HMM model of Equation (4) as the reliability-weighted HMM. Equation (4) could be interpreted as a soft or partial marginalization of observation probability in the aspect of missing feature theory. Also, we can consider Equation (4) as a hybrid likelihood formulation with fuzzy-membership-based weighting. Therefore, the RW-HMM model presented has the same structure as the conventional HMM, $\lambda = (A, B, \pi)$. However, at the training stage of HMM, we need two input features, i.e., (O, ρ). With the definition of the RW-HMM object function, we need to modify the HMM training procedure to optimize Equation (4). We will describe the proposed overall training procedure in the next Subsection 2.2.

2.2. **The training algorithm for RW-HMM.** First, let us calculate the modified posterior likelihood $\tilde{L}_\lambda(O)$ of Equation (4) with the consideration of the observation confidence ρ_t. The next goal is to maximize $\tilde{L}_\lambda(O)$ over all the parameters of HMM, λ. In the following, we briefly describe the procedure to inductively re-estimate the parameters so that a monotonic increase in the likelihood is achieved.

To implement the re-estimation equations, we need to evaluate $\tilde{L}_\lambda(O)$ as well as the posterior probabilities, based on the current model parameters of λ, each of the states $j = 1, \ldots, N$ at each time $t = 1, \ldots, T$. The re-estimation method was developed by means of simple forward and backward induction. This inductive calculation was discussed in the

contexts of [25] and [26]. The same inductive calculation is used for driving the modified training method.

2.2.1. *Inductive calculation of posterior probabilities.* Let's begin by decomposing $\tilde{L}_\lambda(O)$ by states at an arbitrary time t, i.e.,

$$\tilde{L}_\lambda(O) = \sum\nolimits_{j=1}^{N} \tilde{L}_\lambda(O, q_t = j),$$

identically in t. Then, the summand can be written as

$$\tilde{L}_\lambda(O, q_t = j) = \tilde{L}_\lambda(o_1, \cdots, o_t, q_t = j).\tilde{L}_\lambda(o_{t+1}, \cdots, o_T | o_1, \cdots, o_t, q_t = j).$$

The conditional likelihood of $o_{t+1}, \cdots, o_T$ given that $q_t = j$ is independent of $o_1, \cdots, o_t$. So, we have

$$\tilde{L}_\lambda(O, q_t = j) = \tilde{\alpha}_t(j)\tilde{\beta}_t(j),$$

where $\tilde{\alpha}_t(j) = \tilde{L}_\lambda(o_1, \cdots, o_t, q_t = j)$ and $\tilde{\beta}_t(j) = \tilde{L}_\lambda(o_{t+1}, \cdots, o_T | q_t = j)$. Moreover, $\tilde{\alpha}_t(j) = \sum_{i=1}^{N} \tilde{L}_\lambda(o_1, \cdots, o_t, q_t = j, q_{t-1} = i)$, and

$$\tilde{L}_\lambda(o_1, \cdots, o_t, q_t = j, q_{t-1} = i) = \tilde{\alpha}_{t-1}(i)\{b_j(o_t)\}^{\rho^t} a_{ij}.$$

Therefore,

$$\tilde{\alpha}_t(j) = \sum\nolimits_{i=1}^{N} \tilde{\alpha}_{t-1}(i) a_{ij} \{b_j(o_t)\}^{\rho^t} \tag{5}$$

and

$$\tilde{\beta}_t(j) = \sum\nolimits_{k=1}^{N} \tilde{\beta}_{t+1}(k) a_{jk} \{b_k(o_{t+1})\}^{\rho_t}, \text{ with } \tilde{\beta}_T(j) \triangleq 1, \quad j = 1, \cdots, N \tag{6}$$

Based on these functions in Equations (5) and (6), the re-estimation will be performed.

2.2.2. *The training algorithm based on EM algorithm.* In an EM algorithm, the auxiliary function $\tilde{Q}(\lambda, \bar{\lambda})$ of current parameters of λ and new parameters of $\bar{\lambda}$ is defined by

$$\tilde{Q}(\lambda, \bar{\lambda}) = \sum\nolimits_q \sum\nolimits_m \tilde{P}_\lambda(qm|O) \log \tilde{P}_{\bar{\lambda}}(qmO),$$

where the summand is the overall possible q-m sequences of high level and low level hidden states. Also, q and m are the state and mixture indices, respectively. Since we have

$$\begin{aligned} \tilde{P}_{\bar{\lambda}}(qmO) \;=\; & \bar{\pi}_{q_1} \left(\bar{g}_{q_1 m_1} \phi \left(o_1, \bar{\mu}_{q_1 m_1}, \overline{\textstyle\sum}_{q_1 m_1} \right) \right)^{\rho_1}, \cdots, \\ & a_{q_{T-1} q_T} \left(\bar{g}_{q_T m_T} \phi \left(o_T, \bar{\mu}_{q_T m_T}, \overline{\textstyle\sum}_{q_T m_T} \right) \right)^{\rho_T}, \end{aligned}$$

The following Equation (7) is obtained.

$$\begin{aligned} \tilde{Q}(\lambda, \bar{\lambda}) \;=\; & \textstyle\sum_q \sum_m \tilde{P}_\lambda(qm|O) \log \bar{\pi}_{q_1} \\ & + \textstyle\sum_{t=1}^{T-1} \sum_q \sum_m \tilde{P}_\lambda(qm|O) \log \bar{a}_{q_t q_{t+1}} \\ & + \textstyle\sum_{t=1}^{T} \sum_q \sum_m \rho_t \tilde{P}_\lambda(qm|O) \log \bar{g}_{q_t m_t} \\ & + \textstyle\sum_{t=1}^{T} \sum_q \sum_m \tilde{P}_\lambda(qm|O) \rho_t \log \phi(o_t, \bar{\mu}_{q_t m_t}, \overline{\textstyle\sum}_{q_t m_t}) \end{aligned} \tag{7}$$

In the above expression, the first two terms are independent of m and are simplified as follows:

$$\sum\nolimits_q \sum\nolimits_m \tilde{P}_\lambda(qm|O) \log \bar{\pi}_{q_1} = \sum\nolimits_q \tilde{P}_\lambda(q|O) \log \bar{\pi}_{q_1} \tag{8}$$

$$\sum\nolimits_{t=1}^{T-1} \sum\nolimits_q \sum\nolimits_m \tilde{P}_\lambda(qm|O) \log \bar{a}_{q_t q_{t+1}} = \sum\nolimits_{t=1}^{T-1} \sum\nolimits_q \tilde{P}_\lambda(q|O) \log \bar{a}_{q_t q_{t+1}} \tag{9}$$

The part of $\tilde{Q}(.)$, depending on cluster gains, can be transformed as

$$\begin{aligned} & \sum_{t=1}^{T}\sum_{q}\sum_{m}\rho_t\tilde{P}_\lambda(qm|O)\log\bar{g}_{q_tm_t} \\ = & \sum_{t=1}^{T}\sum_{i=1}^{S}\sum_{k=1}^{M}\sum_{q}\sum_{m}\rho_t\tilde{P}_\lambda(qm|O)\log\bar{g}_{q_tm_t}\delta(i,q_t)\delta(j,m_t) \\ = & \sum_{t=1}^{T}\sum_{i=1}^{S}\sum_{k=1}^{M}\rho_t\tilde{P}_\lambda(q_t=i,m_t=k|O)\log\bar{g}_{q_tm_t} \end{aligned} \tag{10}$$

Thus, the part of $\tilde{Q}(.)$ that depends on $\bar{g}_{q_tm_t}$ is of the form of $w_j\log x_j$ with $x_j=\bar{g}_{ik}$ and with $\sum x_j=1$ and $x_j\geq 0$, with a maximum achieved for

$$x_j=\frac{w_j}{\sum w_j} \tag{11}$$

Thus,

$$\bar{g}_{ik}=\frac{\sum_{t=1}^{T}\rho_t\tilde{P}_\lambda(q_t=i,m_t=k|O)}{\sum_{t=1}^{T}\sum_{k=1}^{M}\rho_t\tilde{P}_\lambda(q_t=i,m_t=k|O)}=\frac{\sum_{t=1}^{T}\rho_t\tilde{P}_\lambda(q_t=i,m_t=k|O)}{\sum_{t=1}^{T}\rho_t\tilde{P}_\lambda(q_t=i|O)} \tag{12}$$

Also, the term of $\tilde{Q}(.)$, depending on the cluster centroids and variances, can be re-written as

$$\begin{aligned} & \sum_{t=1}^{T}\sum_{q}\sum_{m}\tilde{P}_\lambda(qm|O)\rho_t\log\phi\left(o_t,\bar{\mu}_{q_tm_t},\overline{\sum}_{q_tm_t}\right) \\ = & \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{k=1}^{M}\sum_{q}\sum_{m}\tilde{P}_\lambda(qm|O)\rho_t\log\phi\left(o_t,\bar{\mu}_{q_tm_t},\overline{\sum}_{q_tm_t}\right)\delta(i,q_t)\delta(j,m_t) \\ = & \sum_{t=1}^{T}\sum_{i=1}^{N}\sum_{k=1}^{M}\sum_{m}\tilde{P}_\lambda(q_t=i,m_t=k|O)\rho_t\log\phi\left(o_t,\bar{\mu}_{q_tm_t},\overline{\sum}_{q_tm_t}\right) \end{aligned} \tag{13}$$

In the M-step, the parameters of $\bar{\pi}_i$, $\bar{a}_{ij}$, $\bar{\mu}_{ikn}$ and $\bar{\sigma}^2_{ikn}$ are obtained by setting the derivatives of the auxiliary function $\tilde{Q}(\lambda,\bar{\lambda})$ about $\bar{\pi}_i$, $\bar{a}_{ij}$, $\bar{\mu}_{ikn}$ and $\bar{\sigma}^2_{ikn}$, to zero.

$$\frac{\partial\tilde{Q}(\lambda,\bar{\lambda})}{\partial\bar{\pi}_i}=0,\ \frac{\partial\tilde{Q}(\lambda,\bar{\lambda})}{\partial\bar{a}_{ij}}=0,\ \frac{\partial\tilde{Q}(\lambda,\bar{\lambda})}{\partial\bar{\mu}_{ikn}}=0,\ \text{and}\ \frac{\partial\tilde{Q}(\lambda,\bar{\lambda})}{\partial\overline{\sum}^2_{ikn}}=0$$

A derivative process similar to the one done in the original HMM is needed, so we can skip the process [25]. Summarizing the E-M optimization rules for the case of RW-HMM with Gaussian densities having diagonal covariance matrices, we get the following equations:

$$\bar{\pi}_i=\tilde{P}_\lambda(q_1=j|O) \tag{14}$$

$$\bar{a}_{ij}=\frac{\sum_{t=1}^{T-1}\tilde{P}_\lambda(q_t=i,q_{t+1}=j|O)}{\sum_{t=1}^{T-1}\tilde{P}_\lambda(q_t=i|O)} \tag{15}$$

$$\bar{g}_{ik}=\frac{\sum_{t=1}^{T}\tilde{P}_\lambda(q_t=i,m_t=k|O)}{\sum_{t=1}^{T}\tilde{P}_\lambda(q_t=i|O)} \tag{16}$$

$$\bar{\mu}_{ikn}=\frac{\sum_{t=1}^{T}\tilde{P}_\lambda(q_t=i,m_t=k|O)\rho_to_t}{\sum_{t=1}^{T}\tilde{P}_\lambda(q_t=i,m_t=k|O)\rho_t} \tag{17}$$

$$\bar{\sigma}^2_{ikn}=\frac{\sum_{t=1}^{T}\tilde{P}_\lambda(q_t=i,m_t=k|O)\rho_t(o_{tn}-\bar{\mu}_{ikn})^2}{\sum_{t=1}^{T}\tilde{P}_\lambda(q_t=i,m_t=k|O)\rho_t} \tag{18}$$

where $t=1,\ldots,T$ indices the time within a training sequence, $i=1,\ldots,N$ and $j=1,\ldots,N$ indices the hidden state, $k=1,\ldots,M$ indices the cluster embedded within a state, and $n=1,\ldots,P$ indices the dimension of the continuous vector of observations. The three terms of $\tilde{P}_\lambda(q_t=i,m_t=k|O)$, $\tilde{P}_\lambda(q_t=j|O)$ and $\tilde{P}_\lambda(q_t=i,q_{t+1}=j|O)$ are

obtained from the forward-backward algorithm in accordance with the following equations:

$$\tilde{P}_\lambda(q_t = i, m_t = k|O) = \tilde{P}_\lambda(q_t = j|O)\frac{\bar{g}_{ik\phi}(o_t, \mu_{ik}, \sum_{ik})}{b_j(o_t)} \tag{19}$$

$$\tilde{P}_\lambda(q_t = j|O) = \frac{\tilde{\alpha}_t(j)\tilde{\beta}_t(j)}{\sum_i \tilde{\alpha}_t(i)\tilde{\beta}_t(i)} \tag{20}$$

$$\tilde{P}_\lambda(q_t = i, q_{t+1} = j|O) = \tilde{\alpha}_t(i)a_{ij}\tilde{\beta}_{t+1}(j)[b_j(o_t)]^{\rho_t} \tag{21}$$

Thus, during the HMM model creation through EM optimization, the contribution of highly uncertain observation gets suppressed depending on its reliability value.

2.3. **The RW-HMM decoder.** Now, we will calculate the probability $P(O|\lambda)$ with the inclusion of the proposed reliability weighing to observation. We know already that $q = [q_1, q_2, \cdots, q_t, \cdots, q_T]$ represents any general state sequence. The forward variable $\tilde{\alpha}_t(j)$ is defined as:

$$\tilde{\alpha}_t(j) = P(o_1, o_2, \cdots, o_t, q_t = j|\lambda), \tag{22}$$

which is the probability of the partial observation sequence from time 1 to time t and state q_j at time t, given the model λ, respectively. The following recursive routine is used to give a solution for $\alpha_t(i)$:

a) Initialization of $\alpha_1(j)$ for each of the N states, which is the joint probability of the observation o_1 and the model being in the state j at time $t = 1$:

$$\alpha_1(j) = \pi_j b_j(o_1)^{\rho_1}, \quad 1 \le j \le N \tag{23}$$

b) The recursive step:

$$\alpha_{t+1}(j) = \left(\sum\nolimits_{i=1}^{N} \alpha_t(i)a_{ij}\right)\{b_j(o_{t+1})\}^{\rho_{t+1}}, \quad 1 \le t \le T-1, \quad 1 \le j \le N \tag{24}$$

c) The termination step finds the probability $P(O|\lambda)$, which can be achieved by calculating the termination forward variables:

$$\alpha_T(j) = P(o_1, o_2, \cdots, o_t, \cdots, o_T, q_t = j|\lambda) \tag{25}$$

d) and summing up these over the S possible termination states, gives:

$$P(O|\lambda) = \sum\nolimits_{j=1}^{N} \alpha_T(j) \tag{26}$$

3. **Validation of RW-HMM in Speaker Identification Domain.** To validate the effectiveness of RW-HMM, we apply this proposed method to the text-dependent and text-independent SI. In this section, we describe the implemented SI system and the experimental results.

3.1. **Overview of the speaker identification system.** The schematic diagram of the entire speaker identification process is shown in Figure 1. In the diagram, the solid line indicates the baseline speaker identification of a classical HMM algorithm, while including the dotted line represents the modified speaker identification process adopting RW-HMM. The training and testing procedures are as follows:

I. *Pre-processing*: We have applied a simple energy-based VAD [31] in our system. As a result, only those speech parts are taken for feature extraction, the energy of which is higher than the average energy of five segments at the initial silent region.

II. *Feature extraction*: Mel-cepstrum analysis is performed and cepstral mean subtraction (CMS) is applied to mimic the channel noise to obtain robust features. For all of the experiments, we have taken seventeen mel-cepstrum features and also included the zero's order cepstral co-efficient and log-energy per frame, and thus obtained the audio features vector of dimension nineteen.
III. *Reliability calculation*: First, the segmental SNR is calculated, and it is then transformed into the observation reliability value using the designated sigmoid function.
IV. *HMM model training*: Using the RW-HMM training algorithm, the HMM models are trained for each speaker.
V. *SI testing*: Using the trained RW-HMM based speaker models and RW-HMM decoder, the probability calculation is performed. The speaker identity (ID) having the maximum probability is, hence, selected as the identified speaker.

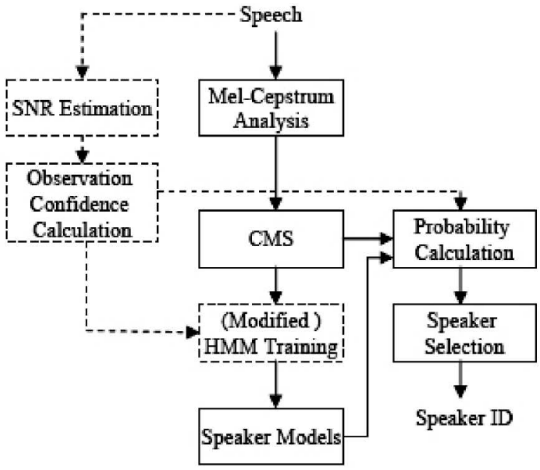


FIGURE 1. Schematic diagram of the RW-HMM based speaker identification system

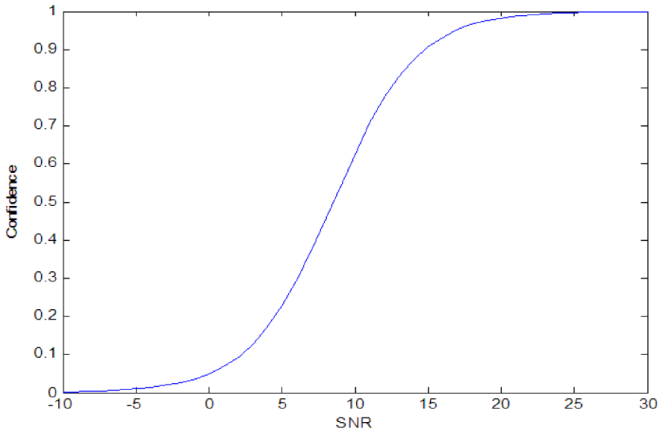


FIGURE 2. Observation reliability based on sigmoid function ($a = 0.35, b = 8.5$)

In our experiments, as shown in Figure 1, the reliability ρ_t is calculated from the signal-to-noise ratio (SNR). That means, $\rho_t = f(SNR)$. The function $f(.)$ must satisfy the fuzzy membership conditions. We apply a simple sigmoid function as the transformation from SNR to reliability as follows:

$$\rho_t = \frac{1}{1 + e^{-a(SNR_t - b)}} \tag{27}$$

where a is the scaling parameter, b is the shift parameter in dB, and SNR_t is the estimated SNR of tth observation. Figure 2 shows an example of $f(.)$ with the parameters $a = 0.35$ and $b = 8.5$.

First, we need to calculate the segmental SNR from the noisy speech and thereafter, using Equation (27), the value of observation reliability is calculated with the predefined values of parameters a and b. However, the predefined values of a and b might not be picked to maximize the performance of SI.

To optimize the parameters of a and b, we set an object function using the identification rate as in Equation (28).

$$IR(a,b) = \frac{\sum_{k=1}^{K}\sum_{l=1}^{L_k}\delta\left(\arg Max_m(P_m(O_{kl})),k\right)}{\sum_{k=1}^{K}L_k} \tag{28}$$

where IR means identification rate, $\delta(i,j)$ is the delta function, O_{kl} is the lth speech feature vector of the kth speaker, and P_m is the observation probability of the given feature sequence for the mth speaker. The parameter $\arg Max_m P_m$ is the index of the speaker with the maximum probability and P_m is defined by Equation (26). Optimizing Equation (28) is not a linear problem. Thus, the particle swarm optimization (PSO) approach is adopted [32]. It is expected that the optimized parameter obtained by PSO will enhance the performance of speaker identification.

3.2. **The experiments and results.** To evaluate the proposed method based-SI-system shown in Figure 1, we have used ETRI and YOHO database. The ETRI speech database was constructed by ETRI (Electronics and Telecommunication Research Institute, South Korea) for SI system evaluation. The ETRI DB was developed for text dependent speaker recognition in mobile environments. The DB contains the speeches of 49 speakers and 20 utterances per speech. The speech DB has the quality of 8 bit μ-law PCM and 8 kHz sampling. Further, the utterances were recorded in noise free environments. Each speaker's utterances are partitioned into three data sets (DS): G1 (1-10) for training, G2 (11-15) for membership function training and G3 (16-20) for testing the SI performances. For text independent speaker identification, we have used the publicly available YOHO database. The utterances (1-10) denoted by S1 from session 1 are used for training stage, whereas utterances (1-5) termed as S4 from session 4 are employed in the test stage.

There are different kinds of noises such as white, car, factory noise (FN) and babble noise (BN). As we have previously mentioned, the purpose of the SI experiments is to show the effectiveness of the proposed RW-HMM. For convenience sake, we test the RW-HMM approach with white noises at first with ETRI DB. To do so, white Gaussian noises are added to the training and testing utterances as follows:

$$X(t) = u(t) + \alpha\eta(t), \tag{29}$$

where $X(t)$ is the digitized noisy speech signal sample at time t, $u(t)$ is the clean speech signal sample and $\eta(t)$ is the noise signal at the respective time t. α is a gain controlling parameter of the depending on the degree of signal-to-noise ratio.

In addition, to validate our RW-HMM in real environment conditions, we have merged the real noises, i.e., NOISEX-92 [36] with the YOHO clean speech database. In our NOISEX-92 based experiments, we projected different values of α to give different degree of signal-to-noise ratios in dB. We have randomly chosen the initial frame of the real noises as of limited duration of each of the real noises in comparison with the YOHO speech database.

The estimated segmental SNR is simply calculated by Equation (30).

$$SNR(t) = 10 \log \frac{\hat{\sigma}_X^2(t) - \hat{\sigma}_\eta^2}{\hat{\sigma}_\eta^2}, \tag{30}$$

where $\hat{\sigma}_X^2(t)$ is the noisy speech signal power over time frame t and $\hat{\sigma}_\eta^2$ is the noise power obtained from a silence region, i.e., the average noise power of initial five segments [29].

In applying RW-HMM to the SI domain, we can consider two different noise environments as follows:

Case I: Training and testing environments have the same degrees of signal-to-noise ratios. That is, the noise level at the training stage is equal to that at the testing stage. In this case, all the added noises are white Gaussian noises while the ETRI DB is involved.

Case II: Training and testing environments are different from each other. Also, in training DB, the utterances have variable SNRs. For this case, we have chosen both the white Gaussian noises and real noises (NOISEX-92) with the ETRI DB and YOHO DB respectively. In the experiments with the YOHO DB, we compare the proposed approach with the conventional HMM and frame-selection approach [33,34]. The frame selection approach is one of the noise-robust recognition methods, which are based on the binary decision of input observations. The FS approach can be considered as a marginalization method applied to whole feature vectors.

3.2.1. *Case I.* For Case I, we have performed the experiments with respective SNRs of 30, 20 and 10 dB. Table 2 and Table 3 show the performances of the classical HMM termed as baseline HMM (BHMM) and the proposed RW-HMM, respectively.

As observed in Table 2, the identification performances are not severely degraded when the noise environments of training and testing stages are the same. According to Table 3, RW-HMM has enhanced the identification performance by $2 \sim 4\%$.

To make a rule for the values of the parameters 'a' and 'b', we performed more experiments for various SNR cases. From the experiments, we get the curves of a and b with the average SNRs. Figure 3 and Figure 4 show the values of the parameters 'a' and 'b'. In the following identification experiments, we estimate the values of 'a' and 'b' using Figure 3 and Figure 4.

Table 2. ETRI DB based SI performances with BHMM

Average SNR of Training DS (dB)	DS for BHMM training	Average SNR of Test DS	Test DS	SI rate (%)
30	G1	30	G3	93.88
20	G1	20	G3	93.84
10	G1	10	G3	91.84

Table 3. ETRI DB based SI performances with RW-HMM

Average SNR for training and testing DS (dB)	DS for RW-HMM training	Training DS for membership function optimization	Optimized parameters of membership function		Testing DS	SI rate (%)
			a_2	b_2		
30	G1	G2	0.83	19.4	G3	97.96
20	G1	G2	0.81	13.5	G3	95.92
10	G1	G2	0.55	1.3	G3	93.88

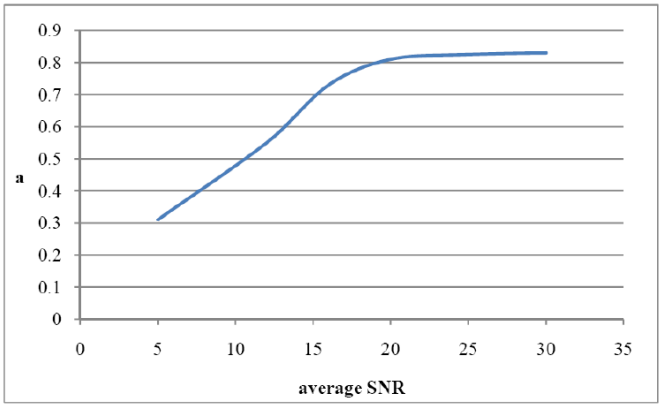


FIGURE 3. Graphical views of the values of 'a' of the sigmoid function w.r.t. average SNR

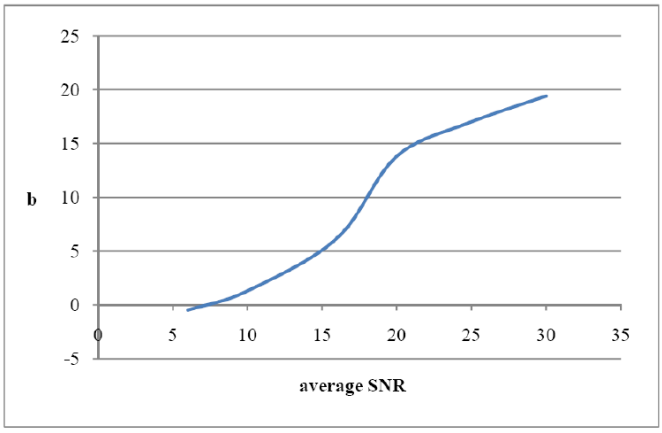


FIGURE 4. Graphical views of the values of 'b' of the sigmoid function w.r.t. average SNR

3.2.2. *Case II.* In Case II, the training utterances have various SNRs from 12dB to 30dB and the testing utterances have 4dB of average SNR. Table 4 shows the SI performances of the baseline HMM (BHMM) for Case II.

TABLE 4. ETRI DB based experimental results with BHMM training

Average SNR for Training DS (dB)	Dataset for the added SNR	Train DS for BHMM training	SNR for Test DS (dB)	Test DS	SI rate (%)
30	1-3	G1	04	G3	46.94
20	4-6				
16	7-8				
12	9-10				

As shown in Table 4, the testing environments are quite different from the training environments, and the training environments are very dynamic due to the different noise levels. Adopting the BHMM, the identification rate is significantly reduced to 46.94% with because of the mismatch of training and testing environments. Table 5 shows the experimental results based on the proposed method, RW-HMM.

Comparing the SI rate of Table 5 with that of Table 4, we observe that the RW-HMM performance was greatly enhanced from 46.94% to 81.63% by introducing an SNR estimation method, despite of the observation uncertainty.

TABLE 5. ETRI DB based experimental results of RW-HMM for the mismatch condition

Average SNR for training DS (dB)	Dataset for the added SNR	Train DS for BHMM training	average SNR to test DS (dB)	Test DS	SI rate (%)
30	1-3	G1	04	G3	81.63
20	4-6				
16	7-8				
12	9-10				

The following Tables 6 and 7, show the performance comparisons of Baseline HMM (BHMM), frame selection based BHMM (BHMM + FS) [33,34] and RW-HMM with YOHO database. In the experiments, the same type of noise environment and different types of noise environments are merged with the YOHO clean speech, respectively. In the experiments, the parameters values of the membership function are taken with approximation fit from Figure 4 and Figure 5 as of the respective degree of average SNR. We have simply selected the reliable frame based on the degree of segmental reliability, i.e., the segmental SNR value in dB. In frame selection approach, we have adopted the overall average SNR as the threshold value to accept or reject the segment of speech features. In the reliable frame selection method, only those segments are accepted as a reliable frame, which has higher SNR than the threshold SNR in dB. In this process, on average, about 25% to 35% frames are treated as unreliable segments in respect to different degrees of threshold SNR.

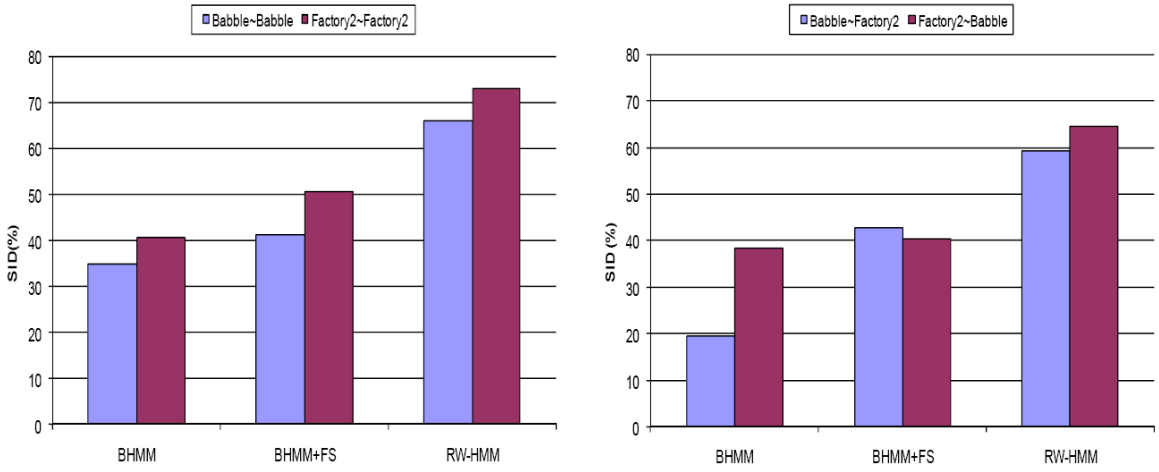
TABLE 6. YOHO DB based experimental results of BHMM, BHMM + FS and RW-HMM with the same real noise type in the train and test stages

Train DS	Noise Type to Train DS	Avg. SNR of Train DS (in dB)	Threshold SNR for frame selection	Test DS	Noise Type to Test DS	Avg. SNR of Test DS (in dB)	Threshold SNR for Frame Selection	HMM Type	IR (%)
S1	Factory2	11.72	–	S4	Factory2	4.72	–	BHMM	40.57
S1	Factory2	11.72	11.72	S4	Factory2	4.72	4.72	BHMM + FS	50.72
S1	Factory2	11.72	–	S4	Factory2	4.72	–	RW-HMM	73.18
S1	Babble	13.26	–	S4	Babble	6.14	–	BHMM	34.78
S1	Babble	13.26	13.26	S4	Babble	6.14	6.14	BHMM + FS	41.30
S1	Babble	13.26	–	S4	Babble	6.14	–	RW-HMM	65.94

From Table 6 to Table 7, it is seen that the speaker identification performances in percentage of the classical HMM (BHMM) and classical HMM with reliable frame selection approach (BHMM + FS) are respectively, 40.57, 34.78, 38.41, and 19.59; 50.72, 41.30, 40.58, and 42.75, whereas in respect to those environmental conditions, the RW-HMM performances are 73.18, 65.94, 64.49, and 59.42, respectively. Eventually, Figure 5 shows the performance comparisons of the different approaches graphically based on YOHO DB. From the experimental result, it can be concluded that the proposed approach outperforms the conventional HMM and FS-based HMM under real environment and thus validate the RW-HMM.

TABLE 7. YOHO DB based experimental results of BHMM, BHMM + FS and RW-HMM experimental results of RW-HMM with different real noise type in the train and test stages

Train DS	Noise Type to Train DS	Avg. SNR of Train DS (in dB)	Threshold SNR for frame selection	Test DS	Noise Type to Test DS	Avg. SNR of Test DS (in dB)	Threshold SNR for frame selection	HMM Type	IR(%)
S1	Factory2	11.72	–	S4	Babble	6.14	–	BHMM	38.41
S1	Factory2	11.72	11.72	S4	Babble	6.14	6.14	BHMM + FS	40.57
S1	Factory2	11.72	–	S4	Babble	6.14	–	RW-HMM	64.49
S1	Babble	13.26	–	S4	Factory2	4.72	–	BHMM	19.59
S1	Babble	13.26	13.26	S4	Factory2	4.72	4.72	BHMM + FS	42.75
S1	Babble	13.26	–	S4	Factory2	4.72	–	RW-HMM	59.42



(a) (b)

FIGURE 5. Speaker identification performances using YOHO DB: (a) with same real noises in the train and test stages, (b) with different real noises in the train and test stages

3.3. **Discussion.** From Table 2 through 5 and Table 6 through 7, it is evident that RW-HMM enhances the performance of the SI system regardless of the consistency of training and testing environments. As a result, we can conclude that RW-HMM is a very useful modification or an extension of the classical HMM, even though the validation domain is very specific. However, to adopt the RW-HMM in diverse pattern recognition problems, a proper reliability function should be defined so that the reliability can be calculated from observations. In this paper, we have used the SNR-derived reliability function and adopted a sigmoid function as the transformation because SNR is one of the appropriate information for measuring the degree of noise distortion. Fitting or making universality of the OMF parameters in the sigmoid function in signal-to-noise ratio and defining reliability functions, however, remain to be an open research problem. Especially, in the pattern recognition problems defined on image or video information, the main source of quality

degradation is illumination changes. Therefore, to apply RW-HMM to image processing, a reliability measure adequate to an application domain should be defined. All in all, through the SI experimental results, we can confirm that the proposed RW-HMM has been successfully working in the pattern recognition field.

4. **Conclusion.** In this study, we proposed the reliability-weighted HMM considering inexact observation. We defined a modified object function by introducing the factor of observation confidence. For the proposed RW-HMM modeling, we suggested the training algorithm based on an EM approach. The effectiveness of the proposed scheme was confirmed in the speaker identification task domain. The experimental results showed that the designed RW-HMM could be one of the most promising solution approaches to the diverse field of pattern recognition problem with inexact observation, where the classical HMM currently plays an important role. The current developed approach has two limitations: defining a proper reliability function to deal with the diverse pattern recognition problems and making universality of the OMF parameters in the sigmoid function. In future work, we intend to implement the RW-HMM in speech processing domains including speech recognition, multi-modal speaker recognition and lip reading. In particular, we need to devise adequate reliability functions to specific application domains and to develop universal OMF parameters in the sigmoid function.

Acknowledgment. This paper was partially supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD, Basic Research Promotion Fund) in 2007 and 2008, and the Research Grant of Kwangwoon University in 2011.

REFERENCES

[1] B. S. Atal, Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification, *Journal of the Acoustical Society of America*, vol.55, pp.1304-1312, 1974.

[2] J. P. Campbell, Speaker recognition: A tutorial, *Proc. of the IEEE*, vol.85, pp.1437-1462, 1997.

[3] S. Furui, Cepstral analysis technique for automatic speaker verification, *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol.29, pp.254-272, 1981.

[4] H. Hermansky and N. Morgan, RASTA processing of speech, *IEEE Trans. on Speech and Audio Processing*, vol.2, no.4, pp.578-589, 1994.

[5] R. J. Mammone, X. Zhang and R. P. Ramachandran, Robust speaker recognition: A feature-based approach, *IEEE Signal Processing Magazine*, vol.13, pp.58-71, 1996.

[6] D. A. Reynolds, An overview of automatic speaker recognition technology, *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol.4, pp.4072-4075, 2000.

[7] D. Stephane and R. Christophe, Robust feature extraction and acoustic modeling at multitel: Experiments on the Aurora databases, *Proce. of Eurospeech*, pp.1789-1792, 2003.

[8] T. Guan and Q. Gong, A study on the effects of spectral information encoding in mandarin speech recognition in white noise, *ICIC Express Letters*, vol.3, no.3(A), pp.415-420, 2009.

[9] X. Wang, J. Lin, Y. Sun, H. Gan and L. Yao, Applying feature extraction of speech recognition on VoIP auditing, *International Journal of Innovative Computing, Information and Control*, vol.5, no.7, pp.1851-1856, 2009.

[10] W. Campbell, D. Sturim and D. Reynolds, SVM based speaker verification using a GMM supervector kernel and NAP variability compensation, *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp.637-640, 2005.

[11] R. Chengalvarayan and L. Deng, A maximum a posteriori approach to speaker adaptation using the trended hidden Markov model, *IEEE Trans. on Speech and Audio Processing*, vol.9, pp.549-557, 2001.

[12] P. Kenny, Joint factor analysis of speaker and session variability: Theory and algorithms, *Technical Report CRIM-06/08-13*, 2006.

[13] C. H. Lee, C. H. Lin and B. H. Juang, A study on speaker adaptation on the parameters of continuous density hidden Markov models, *IEEE Trans. on Signal Processing*, vol.39, pp.806-814, 1991.

[14] E. Mengusoglu, Confidence measure based model adaptation for speaker verification, *Proc. of the 2nd IASTED International Conference on Communications, Internet and Information Technology*, pp.408-411, 2000.
[15] C. H. Sit, M. W. Mak and S. Y. Kung, Maximum likelihood and maximum a posteriori adaptation for distributed speaker recognition systems, *Proc. of the 1st International Conference on Biometric Authentication*, pp.640-647, 2004.
[16] K. Yiu, M. Mak and S. Kung, Environment adaptation for robust speaker verification, *Proc. of Eurospeech*, pp.2973-2976, 2003.
[17] M. Cooke, A. Morris and P. Green, Missing data techniques for robust speech recognition, *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp.863-866, 1997.
[18] B. Raj, M. L. Seltzer and R. M. Stern, Reconstruction of missing features for robust speech recognition, *Speech Communication*, vol.43, pp.275-296, 2004.
[19] A. C. Morris, J. Barker and H. Bourlard, From missing data to may be useful data: Soft data modelling for noise robust ASR, *Proc. of WISP Workshop on Innovative Methods in Speech Recognition*, pp.153-164, 2001.
[20] P. Renevey and A. Drygajlo, Introduction of a reliability measure in missing data approach for robust speech recognition, *Proc. of the 10th EUSIPCO*, pp.473-476, 2000.
[21] B. Raj and R. M. Stern, Missing-feature approaches in speech recognition, *IEEE Signal Processing Magazine*, pp.101-116, 2005.
[22] J. Y. Kim et al., Modified GMM training for inexact observation and its application to speaker identification, *Speech Sciences*, vol.14, pp.163-175, 2007.
[23] M. Brand, N. Oliver and A. Pentland, Coupled hidden Markov models for complex action recognition, *Proc. of IEEE International Conference on Computer Vision and Pattern Recognition*, pp.994-999, 1997.
[24] Z. Ghahramani and M. I. Jordan, Factorial hidden Markov models, *Machine Learning*, vol.29, pp.245-273, 1997.
[25] L. A. Liporace, Maximum likelihood estimation for multivariate observations of Markov sources, *IEEE Trans. on Information Theory*, vol.28, pp.729-734, 1982.
[26] R. Rabiner and B. Juang, An introduction to hidden Markov models, *IEEE Acoustics, Speech, and Signal Processing Magazine*, vol.3, pp.4-16, 1986.
[27] J. A. Arrowood and M. A. Clements, Using observation uncertainty in HMM decoding, *Proc. of International Conference on Spoken Language Processing*, Denver, CO, pp.1561-1564, 2002.
[28] J. Droppo, A. Acero and L. Deng, Uncertainty decoding with splice for noise robust speech recognition, *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Orlando, FL, vol.1, pp.57-60, 2002.
[29] T. Kristjansson and B. J. Frey, Accounting for uncertainty in observations: A new paradigm for robust speech recognition, *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Orlando, FL, pp.61-64, 2002.
[30] H. Liao and M. J. F. Gales, Adaptive training with joint uncertainty decoding for robust recognition of noisy data, *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, Honolulu, HI, pp.389-392, 2007.
[31] M. Vondrasek and P. Pollak, Methods for speech SNR estimation: Evaluation tool and analysis of VAD dependency, *Journal of Radio Engineering*, vol.14, pp.6-11, 2005.
[32] R. Eberhart and J. Kennedy, A new optimizer using particle swarm theory, *Proc. of the 6th International Symposium on Micro Machine and Human Science*, pp.39-43, 1997.
[33] R. P. Lippmann and B. A. Carlson, Robust speech recognition with time-varying filtering, interruptions, and noise, *Proc. of the IEEE Speech Recognition and Understanding Workshop*, pp.365-372, 1997.
[34] M. Nosratighods et al., P-value segment selection technique for speaker verification, *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol.4, Honolulu, HI, pp.269-272, 2007.
[35] J. E. But and G. M. Narber, *Elementary Statistics for Grographers*, The Guilford Press, 1995
[36] A. Varga and H. J. M. Steeneken, Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems, *Speech Communication*, vol.12, no.3, pp.247-251, 1993.