A FUZZY APPROACH TO SPEAKER VERIFICATION*

DAT TRAN† and MICHAEL WAGNER‡

University of Canberra, School of Computing, ACT 2601, Australia
†Dat.Tran@canberra.edu.au
‡Michael.Wagner@canberra.edu.au

This paper proposes a fuzzy approach to speaker verification. For an input utterance and a claimed identity, most of the current methods compute a claimed speaker’s score, which is the ratio of the claimed speaker’s and the impostors’ likelihood functions, and compare this score with a given threshold to accept or reject this speaker. Considering the speaker verification problem based on fuzzy set theory, the claimed speaker’s score is viewed as the fuzzy membership function of the input utterance in the claimed speaker’s fuzzy set of utterances. Fuzzy entropy and fuzzy c-means membership functions are proposed as fuzzy membership scores, which are the ratios of functions of the claimed speaker’s and impostors’ likelihood functions. A likelihood transformation is also considered to relate current likelihood and fuzzy membership scores. We also proposed fuzzy scores using membership functions similar to those produced by noise-clustering-based method. This noise clustering concept provides very effective modifications to several methods, which can overcome some of the problems of ratio-type scores and greatly reduce the false acceptance rate. Experiments were performed to evaluate proposed normalization methods for speaker verification using the YOHO corpus. Experiments demonstrate that fuzzy methods and their noise clustering versions outperform conventional methods.

Keywords: Speaker verification; normalization method; fuzzy c-means clustering; fuzzy entropy clustering; noise clustering.

1. Introduction

Speaker verification techniques are used to verify the identity claimed by people accessing certain protected systems. It enables access control of various services by voice. Speaker verification is the process of accepting or rejecting the identity claim of a speaker. An identity claim is made by an unknown speaker, and an utterance of this unknown speaker is compared with the model for the speaker whose identity is claimed. If the match is good enough, i.e. above a given threshold, the identity claim is accepted. Most of the current methods compute the claimed speaker’s score as the ratio of the claimed speaker’s and the impostors’ likelihood functions. Speaker verification performance is strongly affected by variations in signal characteristics, therefore normalization methods have been applied to compensate for

*This work was supported by the University of Canberra, Australia, under the University Research Grants Scheme.
these variations. For practical implementations, the use of “cohort speakers” as a background speaker set that is representative of the impostors’ population close to the claimed speaker has been proposed. In all verification paradigms, there are two classes of errors: false rejections and false acceptances. A false rejection occurs when the system incorrectly rejects a true speaker and a false acceptance occurs when the system incorrectly accepts an impostor. An equal error rate condition is often used to adjust system parameters so that the two types of errors are equal.

Consider the false rejections of the claimed speaker and the false acceptances of impostors caused in the current likelihood ratio-based scores. A false rejection of the claimed speaker can arise because of the use of the background speaker set. The likelihood values of the cohort speakers are assumed to be equally weighted. However, this assumption is often not true as the similarity measures between each cohort speaker and the claimed speaker might be different. To overcome this drawback, different weights based on fuzzy integration were proposed. An alternative way is to consider the speaker verification problem in fuzzy set theory framework. Since input utterances belong to either the claimed speaker or impostors, the set of input utterances can be divided into two subsets for the claimed speaker and impostors. Computing the claimed speaker’s score for an input utterance can be viewed as a fuzzification process, where two above subsets are considered as fuzzy subsets and the score means the fuzzy membership function, which denotes the degree of belonging of the input utterance to the claimed speaker. Accepting (or rejecting) the claimed speaker is thus viewed as a defuzzification process, where the input utterance is (or is not) in the claimed speaker’s fuzzy subset if the fuzzy membership value is (or is not) greater than the given threshold. According to this fuzzy set theory-based viewpoint, currently used scores are viewed as fuzzy membership scores and inversely, other fuzzy memberships can be used as the claimed speaker’s scores. So the fuzzy c-means (FCM) and the fuzzy entropy (FE) memberships coming from fuzzy clustering are proposed and experimentally evaluated as efficient scores. As an extension, a transformation is established to relate these proposed scores to currently used likelihood ratio scores and subsequently, more efficient fuzzy membership scores are proposed for speaker verification.

On the other hand, false acceptances can arise because of the relativity of the ratio-based values. This is a problem similar to the one addressed by Chen et al. For example, the two ratios of 0.06/0.03 and 0.0000006/0.0000003 have the same value of 2. The first ratio can lead to a correct decision whereas the second one is unlikely since both likelihood values are very low. This problem can be overcome by applying the idea of the well-known noise clustering method in fuzzy clustering, where impostor’s utterances are considered as noisy data and thus should have arbitrarily small fuzzy membership scores in the claimed speaker’s fuzzy subset. This is implemented by simply adding to the normalization term a constant membership value, which denotes belonging of all input utterances to impostors’ fuzzy subset.

The general form of the proposed scores after considering the false acceptances
and the false rejections is proposed in this paper. Some transformation functions are also proposed and applied to speaker verification experiments. Comparing with the current normalization methods, better results for experiments performed on the YOHO corpus including 138 speakers are reported in this paper. All noise clustering-based methods show better results than the other methods.

The remainder of this paper is organized as follows. Section 2 briefly introduces a typical speaker verification system. Current normalization methods are reviewed in Sec. 3. Proposed normalization methods are presented in Sec. 4. Section 5 reports some experimental results to evaluate the proposed methods. Finally, we conclude the paper in Sec. 6.

2. A Speaker Verification System

Let $\lambda_0$ be the claimed speaker model and $\lambda$ be a model representing all other possible speakers, i.e. impostors. For a given input utterance $X$ and a claimed identity, the choice is between the hypothesis $H_0$: $X$ is from the claimed speaker $\lambda_0$, and the alternative hypothesis $H_1$: $X$ is from the impostors $\lambda$. A claimed speaker’s score $S(X)$ is computed to reject or accept the speaker claim. Depending on the meaning of the score, we can distinguish between similarity scores $L(X)$ and dissimilarity scores $D(X)$ between $X$ and $\lambda_0$. Likelihood scores are included in $L(X)$ and vector quantization (VQ) distortion scores are included in $D(X)$. These scores satisfy the following rules

$$L(X) \begin{cases} > \theta_L & \text{accept} \\ \leq \theta_L & \text{reject} \end{cases}$$

and

$$D(X) \begin{cases} < \theta_D & \text{accept} \\ \geq \theta_D & \text{reject} \end{cases}$$

where $\theta_L$ and $\theta_D$ are the decision thresholds. We define equivalent scores as scores giving the same EER even though they possibly use different thresholds. For example, $S_a(X) = P(X|\lambda_0)$ and $S_b(X) = \log P(X|\lambda_0)$ are equivalent scores, but use thresholds $\theta$ and $\log \theta$, respectively. Figure 1 presents a typical speaker verification system.

![Fig. 1. A typical speaker verification system.](image)

*aIn this context, we use the term impostor for any speaker other than the claimed speaker and without any implication of fraudulent intent or active voice manipulation.*
3. Current Normalization Methods

The simplest method of scoring is to use the absolute likelihood score (unnormalized score) of an utterance. In the log domain, that is

\[ L_0(X) = \log P(X|\lambda_0). \]  

(3)

This score is strongly influenced by variations in the test utterance such as the speaker’s vocal characteristics, the linguistic content and the speech quality. It is very difficult to set a common decision threshold to be used over different tests. This drawback is overcome to some extent by using normalization. According to the Bayes decision rule for minimum risk, a likelihood ratio

\[ L_1(X) = \frac{P(X|\lambda_0)}{P(X|\lambda)} \]  

(4)

is used. This ratio produces a relative score which is less volatile to nonspeaker utterance variations.\(^{11}\) In the log domain, (4) is equivalent to the following normalization technique, proposed by Higgins et al. in 1991

\[ L_1(X) = \log P(X|\lambda_0) - \log P(X|\lambda). \]  

(5)

The term \( \log P(X|\lambda) \) in (5) is called the normalization term and requires calculation of all impostors’ likelihood functions. An approximation of this method is to use only the closest impostor model for calculating the normalization term\(^6\)

\[ L_2(X) = \log P(X|\lambda_0) - \max_{\lambda \neq \lambda_0} \log P(X|\lambda). \]  

(6)

However when the size of the population increases, both of these normalization methods \( L_1(X) \) and \( L_2(X) \) are unrealistic since all impostors’ likelihood functions must be calculated for determining the value of the normalization term. Therefore a subset of the impostor models is used. This subset consists of \( B \) “background” speaker models \( \lambda_i, i = 1, \ldots, B \) and is representative of the population close to the claimed speaker, i.e. the “cohort speaker” set.\(^{12}\) Depending on the approximation of \( P(X|\lambda) \) in (4) by the likelihood functions of the background model set \( P(X|\lambda_i), i = 1, \ldots, B \), we obtain different normalization methods. An approximation\(^{11}\) has been applied that is the arithmetic mean (average) of the likelihood functions of \( B \) background speaker models. The corresponding score for this approximation is

\[ L_3(X) = \log P(X|\lambda_0) - \log \left\{ \frac{1}{B} \sum_{i=1}^{B} P(X|\lambda_i) \right\}. \]  

(7)

If the claimed speaker’s likelihood function is also included in the above arithmetic mean, we obtain the normalization method based on \( a \) posteriori probability\(^8\)

\[ L_4(X) = \log P(X|\lambda_0) - \log \sum_{i=0}^{B} P(X|\lambda_i). \]  

(8)
Note that, $i = 0$ in (8) denotes the claimed speaker model and the constant term $1/B$ is accounted for in the decision threshold. If the geometric mean is used instead of the arithmetic mean to approximate $P(X|\lambda)$, we obtain the normalization method\(^6\) as follows

$$L_5(X) = \log P(X|\lambda_0) - \frac{1}{B} \sum_{i=1}^{B} \log P(X|\lambda_i).$$ \hspace{1cm} (9)

Normalization methods can also be applied to the likelihood function of each vector $x_t$, $t = 1, \ldots, T$ in $X$, and such methods are called frame level normalization methods. Such a method has been proposed as follows\(^7\)

$$L_6(X) = \sum_{t=1}^{T} \left[ \log P(x_t|\lambda_0) - \log \sum_{i=1}^{B} P(x_t|\lambda_i) \right].$$ \hspace{1cm} (10)

For VQ-based speaker verification systems, the following score is widely used

$$D_1(X) = D(X, \lambda_0) - \frac{1}{B} \sum_{i=1}^{B} D(X, \lambda_i)$$ \hspace{1cm} (11)

where

$$D(X, \lambda_i) = \sum_{t=1}^{T} d_t^{(i)}$$ \hspace{1cm} (12)

where $d_t^{(i)}$ is the VQ distance between vector $x_t \in X$ and the nearest codevector in the codebook $\lambda_i$. It can be seen that this score is equivalent to $L_5(X)$ if we replace $D(X, \lambda_i)$ in (11) by $[-\log P(X|\lambda_i)]$, $i = 0, 1, \ldots, B$.

4. Proposed Normalization Methods

4.1. Fuzzy membership scores

Consider the speaker verification problem in the framework of fuzzy set theory. To accept or reject the claimed speaker, the task is to make a decision whether the input utterance $X$ is either from the claimed speaker $\lambda_0$ or from the set of impostors $\lambda$, based on comparing the score for $X$ and a decision threshold $\theta$. Thus the space of input utterances can be considered as consisting of two fuzzy sets: $C$ for the claimed speaker and $I$ for impostors. Degree of belonging of $X$ to the fuzzy set $C$ is denoted by the fuzzy membership function and can be regarded as a claimed speaker’s score satisfying the rule in (1). Making a (hard) decision is thus a defuzzification process where $X$ is completely in $C$ if the fuzzy membership of $X$ in $C$ is sufficiently high, i.e. greater than the threshold $\theta$.

In theory, there are many ways to define the fuzzy membership function; therefore, it can be said that this fuzzy approach can accommodate more general scores than the current likelihood ratio scores for speaker verification. These are termed fuzzy membership scores, which can denote the belonging of $X$ to the claimed
speaker. Based on this discussion, all of the above-mentioned likelihood-based scores $L_i(X)$, $i = 1, \ldots, 5$, can also be viewed as fuzzy membership scores if their values are scaled into the interval $[0, 1]$. Although the dissimilarity scores $D_i(X)$ cannot be viewed as fuzzy membership scores but as shown below, a similar approach can be obtained for those scores. The next task is to find more effective fuzzy membership scores which can reduce both false rejection and false acceptance errors.

4.2. A solution for the false rejection problem

Fuzzy membership scores can be taken from fuzzy clustering methods. The FE score $L_7(X)$ and the FCM score $L_8(X)$ are proposed as follows:

$$L_7(X) = \left[ P(X|\lambda_0) \right]^{1/n} / \sum_{i=0}^{B} \left[ P(X|\lambda_i) \right]^{1/n}$$  \hspace{1cm} (13)

$$L_8(X) = \frac{-\log P(X|\lambda_0)^{1/(1-m)}}{\sum_{i=0}^{B} [-\log P(X|\lambda_i)]^{1/(1-m)}}$$  \hspace{1cm} (14)

where $n > 0$ and $m > 1$ are degrees of fuzzy entropy and fuzziness, respectively. In the log domain, as $n = 1$, the score $L_7(X)$ reduces to $L_4(X)$ in (8). To find out more efficient fuzzy membership scores, we note that two above fuzzy scores are the ratios of functions of likelihood functions whereas scores in Sec. 3 are the ratios of likelihood functions. We thus consider a transformation applied to current scores to obtain new scores in the next section.

Consider a transformation $T: P \rightarrow T(P)$, where $P$ is the likelihood function and $T(P)$ is some continuous function of $P$. For example, $T[P(X|\lambda_0)] = \log P(X|\lambda_0)$. Applying this transformation to the likelihood ratio score $L_1(X)$ gives an alternative score $S(X)$

$$S(X) = \frac{T[P(X|\lambda_0)]}{T[P(X|\lambda)]}.$$  \hspace{1cm} (15)

With $T[P] = (P)^{1/n}$ and $T[P] = (-\log P)^{1/(1-m)}$, we obtain the scores $L_7(X)$ and $L_8(X)$, respectively. More efficient fuzzy membership scores can be found by considering an example of the score $L_3(X)$ including three cohort speakers $\lambda_1, \lambda_2, \lambda_3$. Assume that $X$ is an utterance coming from the claimed speaker, the threshold $\theta = 1$ and for simplicity we denote $P(X|\lambda_i) = P_i$, where $i = 0, \ldots, 3$ and $P_{123} = (P_1 + P_2 + P_3)/3$ is the normalization term. If $P_0 < P_{123}$, a false rejection case occurs since $P_0/P_{123} < 1$. Assuming there exists a transformation using a continuous increasing function $T: P_i \rightarrow T[P_i]$, $i = 0, \ldots, 3$ such that small values of $P_i$ are changed much more than larger values of $P_i$. In this case, we might have $T[P_0] > T[P_{123}]$ and thus a true acceptance occurs since $T[P_0]/T[P_{123}] > 1$, where we assume that the threshold $\theta_T = 1$ and $T[P_{123}] = (T[P_1] + T[P_2] + T[P_3])/3$. So the function $T(P)$ should be nonlinear. Moreover, for convenience in calculating products of probabilities, the function $T(P)$ should be related to the logarithm.
function. This is also convenient for applying these methods to VQ-based speaker verification since the distance in VQ can be defined as the negative logarithm of the corresponding likelihood function. Based on this discussion, we propose the following dissimilarity scores:

\[
D_2(X) = \frac{\log P(X|\lambda_0)}{\max_{\lambda \neq \lambda_0} \log P(X|\lambda)}
\]

\[
D_3(X) = \frac{\log P(X|\lambda_0)}{\log \left( \frac{1}{B} \sum_{i=1}^{B} P(X|\lambda_i) \right)}
\]

\[
D_4(X) = \frac{\log P(X|\lambda_0)}{\log \left( \frac{1}{B} \sum_{i=1}^{B} P(X|\lambda_i) \right)}
\]

\[
D_5(X) = \frac{\log P(X|\lambda_0)}{\frac{1}{B} \sum_{i=1}^{B} \log P(X|\lambda_i)}
\]

\[
D_6(X) = \frac{\arctan[\log P(X|\lambda_0)]}{\frac{1}{B} \sum_{i=1}^{B} \arctan[\log P(X|\lambda_i)]}
\]

where \(T[P(X|\lambda)]\) of the impostors is approximated by the transformed likelihood functions of background speakers \(D_2(X), \ldots, D_5(X)\) in the same manner as was applied to \(L_2(X), \ldots, L_5(X)\). Note that the factor \(1/B\) in \(D_4(X)\) is not accounted for in the decision threshold as for \(L_4(X)\) in (8). To apply these scores to VQ-based speaker verification systems, likelihood functions should be changed to VQ distortions as shown for the score \(D_1(X)\) in (11) and (12).

4.3. A solution for the false acceptance problem

As discussed in Sec. 1, the use of the background speaker set can cause false acceptances of impostors because of the relativity of the ratio-based values in current and proposed scores. Suppose, \(L_3(X)\) in (7) is used. Now consider the two equal likelihood ratios in the following example

\[
L_3(X_1) = \frac{0.06}{0.03} = L_3(X_2) = \frac{0.0000006}{0.0000003}
\]

Here both \(X_1\) and \(X_2\) are accepted if the given threshold is 1.0. The first ratio can lead to a correct decision that the input utterance \(X_1\) is from the claimed speaker (true acceptance). However, it is improbable that \(X_2\) is from the claimed speaker or from any of background speakers since both likelihood values in the second ratio are very low. \(X_2\) is probably from an impostor and thus a false acceptance can occur on the basis of the likelihood ratio. This is a problem similar to the one addressed by Chen et al.\(^2\) In this work, an absolute threshold has been set to reject any test utterance with a raw distortion score \(D(X)\) greater than the threshold.

A more robust method is proposed here using the fuzzy membership scores based on the noise clustering (NC) method.\(^3\) This method can be applied not only
to proposed fuzzy membership scores but also to likelihood ratio scores $L(X)$ as well as distortion scores $D(X)$. Therefore, this method is regarded as an extended version of the above-mentioned method proposed by Chen et al. Both these methods give the same results if only distortion scores are used. The proposed NC-based score can reduce the false acceptance error by forcing the membership value of the input utterance $X$ to become as small as possible if $X$ is really from impostors, not from the claimed speaker or background speakers. This fuzzy approach is simple but very effective. It just adds a suitable constant value $\epsilon > 0$ (similar to the constant distance $\delta$ in the NC method) to denominators of ratios, i.e. to the normalization terms as follows

$$\text{Change} \sum \cdots \text{ in the normalization term to } \sum \cdots + \epsilon . \quad (22)$$

Note that, the NC method can be applied not only to fuzzy membership scores but also to likelihood ratio scores $L(X)$ as well as to distortion scores $D(X)$. For example, NC-based versions of $L_i(X)$, $i = 3, \ldots, 8$ are as follows

$$L_{3nc}(X) = \log P(X|\lambda_0) - \log \left\{ \frac{1}{B + 1} \left[ \sum_{i=1}^{B} P(X|\lambda_i) + \epsilon_3 \right] \right\}$$

$$L_{4nc}(X) = \log P(X|\lambda_0) - \log \left[ \sum_{i=0}^{B} P(X|\lambda_i) + \epsilon_4 \right]$$

$$L_{5nc}(X) = \log P(X|\lambda_0) - \frac{1}{B + 1} \left[ \sum_{i=1}^{B} \log P(X|\lambda_i) + \log \epsilon_5 \right]$$

$$L_{6nc}(X) = \sum_{t=1}^{T} \left\{ \log P(x_t|\lambda_0) - \log \left[ \sum_{i=1}^{B} P(x_t|\lambda_i) + \epsilon_6 \right] \right\}$$

$$L_{7nc}(X) = \frac{[P(X|\lambda_0)]^{1/n}}{\sum_{i=0}^{B} [P(X|\lambda_i)]^{1/n} + \epsilon_7^{1/n}}$$

$$L_{8nc}(X) = \frac{[-\log P(X|\lambda_0)]^{1/(1-m)}}{\sum_{i=0}^{B} [-\log P(X|\lambda_i)]^{1/(1-m)} + (-\log \epsilon_8)^{1/(1-m)}} \quad (23)$$

where the index “nc” means “noise clustering”. Similarly, NC-based versions for the dissimilarity scores $D_i(X)$, $i = 2, \ldots, 6$ are also obtained with the same manner. For illustration, applying NC-based scores for the first example in (21) with $\epsilon_3 = 0.00001$ gives

$$L_{3nc}(X_1^c) = \frac{0.06}{0.03 + 0.00001} \approx 2.0 > L_{3nc}(X_2^c) = \frac{0.0000006}{0.0000003 + 0.00001} \approx 0.06 . \quad (24)$$
For both the false rejection and false acceptance problems, we propose a generalized form of fuzzy membership scores as follows:

$$S(X) = \frac{T[P(X|\lambda_0)]}{T[P(X|\lambda)] + \epsilon}.$$ (25)

The next section presents experiments performed to evaluate the proposed fuzzy membership scores. The speech data used in evaluation experiments is the YOHO corpus.

5. Evaluation Experiments and Results

5.1. Database description

The YOHO corpus was designed for speaker verification systems in office environments with limited vocabulary. There are 138 speakers, 108 males and 30 females. The vocabulary consists of 56 two-digit numbers ranging from 21 to 97 pronounced as “twenty-one”, “ninety-seven”, and spoken continuously in sets of three, for example “36-45-89”, in each utterance. There are four enrolment sessions per speaker, numbered 1 through 4, and each session contains 24 utterances. There are also ten verification sessions, numbered 1 through 10, and each session contains 4 utterances. All waveforms are low-pass filtered at 3.8 kHz and sampled at 8 kHz.

For the YOHO corpus, speech processing was performed using HTK V2.0, a toolkit for building hidden Markov models (HMMs). The data were processed in 32 ms frames at a frame rate of 10 ms. Frames were Hamming windowed and preemphasized. The basic feature set consisted of 12th-order mel-frequency cepstral coefficients (MFCCs) and the normalized short-time energy, augmented by the corresponding delta MFCCs to form a final set of feature vector with a dimension of 26 for individual frames.

5.2. Algorithmic issues

The Gaussian mixture model (GMM) regarded as the 1-state HMM and the VQ model were used in speaker verification experiments performed on the YOHO corpus in text-independent mode.

GMMs and VQ models are respectively initialized as follows:

- GMMs: mixture weights, mean vectors and covariance matrices were initialized with essentially random choices. Covariance matrices are diagonal, i.e. $[\Sigma_k]_{ii} = \sigma_k^2$ and $[\Sigma_k]_{ij} = 0$ if $i \neq j$, where $\sigma_k^2$, $1 \leq k \leq K$ are variances.
- VQ models: these models were trained by the LBG algorithm — a widely used version of the K-means algorithm, where starting from the codebook size of 1, a binary split procedure is performed to double the codebook size in end of several iterations.
Constraints on GMM and VQ parameters during training are as follows:

- GMMs: a variance limiting constraint was applied to all GMMs using diagonal covariance matrices. This constraint places a minimum variance value \( \sigma^2_{\text{min}} \) on elements of all variance vectors in the GMM, that is, \( \sigma_i^2 = \sigma^2_{\text{min}} \) if \( \sigma_i^2 \leq \sigma^2_{\text{min}} \). In our experiments, \( \sigma^2_{\text{min}} = 10^{-2} \). The chosen number of mixtures were 16, 32, and 64 to compare VQ models using the LBG algorithm.

- VQ models: Codebook sizes of 16, 32 and 64 were chosen for all experiments. There is no consistent method to find good values of fuzzy parameters for a given data set; therefore, we have tested speaker models using the training data set to choose good values. The EER was sensitive to values of fuzzy parameters \( m \) and \( n \). The parameter \( \epsilon \) was initially chosen as the minimum probability \( P(X|\lambda_i) \) and the maximum distortion \( D(X, \lambda_i) \) in experiments using GMMs and VQ codebooks, respectively, where \( X \in \) the training data set and \( i = 0, \ldots, B \). Fuzzy parameters were set to \( n = 5.0 \) for \( L_7(X) \) and \( L_{7nc}(X) \), \( m = 1.05 \) for \( L_8(X) \) and \( L_{8nc}(X) \), \( \epsilon = 10^{-9} \) for all NC-based methods using GMMs and \( \epsilon = 10^{-22} \) for all NC-based methods using VQ codebooks.

<table>
<thead>
<tr>
<th>Normalization Methods</th>
<th>GMM 16 Mixtures</th>
<th>GMM 32 Mixtures</th>
<th>GMM 64 Mixtures</th>
<th>VQ 16 Vectors</th>
<th>VQ 32 Vectors</th>
<th>VQ 64 Vectors</th>
</tr>
</thead>
<tbody>
<tr>
<td>( L_3(X) )</td>
<td>4.4</td>
<td>3.1</td>
<td>2.4</td>
<td>5.9</td>
<td>4.6</td>
<td>3.6</td>
</tr>
<tr>
<td>( L_{3nc}(X) )</td>
<td>4.2</td>
<td>2.9</td>
<td>2.0</td>
<td>5.0</td>
<td>4.3</td>
<td>3.5</td>
</tr>
<tr>
<td>( L_4(X) )</td>
<td>4.4</td>
<td>3.1</td>
<td>2.4</td>
<td>5.9</td>
<td>4.6</td>
<td>3.6</td>
</tr>
<tr>
<td>( L_{4nc}(X) )</td>
<td>4.2</td>
<td>2.9</td>
<td>2.0</td>
<td>5.0</td>
<td>4.2</td>
<td>3.5</td>
</tr>
<tr>
<td>( L_5(X) )</td>
<td>4.5</td>
<td>3.3</td>
<td>2.4</td>
<td>6.8</td>
<td>5.1</td>
<td>4.2</td>
</tr>
<tr>
<td>( L_{5nc}(X) )</td>
<td>3.9</td>
<td>2.7</td>
<td>1.9</td>
<td>4.5</td>
<td>3.4</td>
<td>2.9</td>
</tr>
<tr>
<td>( L_6(X) )</td>
<td>5.0</td>
<td>3.4</td>
<td>2.4</td>
<td>6.9</td>
<td>5.1</td>
<td>4.2</td>
</tr>
<tr>
<td>( L_{6nc}(X) )</td>
<td>4.5</td>
<td>2.9</td>
<td>2.1</td>
<td>2.1</td>
<td>4.9</td>
<td>4.0</td>
</tr>
<tr>
<td>( L_7(X) )</td>
<td>4.2</td>
<td>3.0</td>
<td>2.1</td>
<td>5.3</td>
<td>4.0</td>
<td>3.1</td>
</tr>
<tr>
<td>( L_{7nc}(X) )</td>
<td>4.2</td>
<td>2.9</td>
<td>2.0</td>
<td>4.5</td>
<td>3.4</td>
<td>3.0</td>
</tr>
<tr>
<td>( L_8(X) )</td>
<td>3.9</td>
<td>2.8</td>
<td>1.9</td>
<td>4.5</td>
<td>3.3</td>
<td>2.8</td>
</tr>
<tr>
<td>( L_{8nc}(X) )</td>
<td>3.9</td>
<td>2.8</td>
<td>1.8</td>
<td>4.3</td>
<td>3.3</td>
<td>2.7</td>
</tr>
<tr>
<td>( D_3(X) )</td>
<td>4.1</td>
<td>2.9</td>
<td>2.0</td>
<td>5.4</td>
<td>4.1</td>
<td>3.1</td>
</tr>
<tr>
<td>( D_{3nc}(X) )</td>
<td>4.3</td>
<td>2.9</td>
<td>2.0</td>
<td>4.8</td>
<td>3.8</td>
<td>3.1</td>
</tr>
<tr>
<td>( D_4(X) )</td>
<td>4.2</td>
<td>3.0</td>
<td>2.0</td>
<td>5.2</td>
<td>3.9</td>
<td>3.1</td>
</tr>
<tr>
<td>( D_{4nc}(X) )</td>
<td>4.2</td>
<td>2.9</td>
<td>2.0</td>
<td>4.7</td>
<td>3.8</td>
<td>3.0</td>
</tr>
<tr>
<td>( D_5(X) )</td>
<td>4.1</td>
<td>3.0</td>
<td>2.1</td>
<td>5.3</td>
<td>3.8</td>
<td>3.1</td>
</tr>
<tr>
<td>( D_{5nc}(X) )</td>
<td>3.8</td>
<td>2.8</td>
<td>1.9</td>
<td>4.5</td>
<td>3.3</td>
<td>2.7</td>
</tr>
<tr>
<td>( D_6(X) )</td>
<td>4.4</td>
<td>3.2</td>
<td>2.4</td>
<td>6.1</td>
<td>4.8</td>
<td>3.6</td>
</tr>
<tr>
<td>( D_{6nc}(X) )</td>
<td>4.3</td>
<td>3.1</td>
<td>2.3</td>
<td>5.1</td>
<td>4.4</td>
<td>3.5</td>
</tr>
</tbody>
</table>
5.3. Verification results

Experiments were performed on 138 speakers using each speaker as a claimed speaker with five closest background speakers and 132 mixed-gender impostors (excluding five background speakers) and rotating through all speakers. The total number of claimed test utterances and impostor test utterances are 5520 (138 claimed speakers × 40 test utterances) and 728640 ((138 × 132) impostors × 40 test utterances), respectively.

Table 1 shows results for GMMs and VQ codebooks to compare EERs obtained by using current and proposed normalization methods. For 16, 32 and 64 mixture GMMs, proposed normalization methods $D_3(X), D_4(X), \ldots, D_6(X)$, especially $D_6(X)$ (using the arctan function) produced lower EERs than current normalization methods $L_3(X), \ldots, L_6(X)$. Noise clustering-based normalization methods $L_{3nc}(X), \ldots, L_{8nc}(X)$ and $D_{3nc}(X), \ldots, D_{6nc}(X)$ also produced better results compared to current and proposed methods. The current normalization method $L_6(X)$ produced the highest EER of 5.0% for 16-mixture GMMs and the proposed method $L_{8nc}(X)$ produced the lowest EER of 1.8% for 64-mixture GMMs. The better performances were also obtained for proposed methods using VQ codebooks. The highest EER of 6.9% was obtained using the current method $L_6(X)$ for VQ codebook size of 16 and the lowest EER of 2.7% was obtained using proposed methods $L_{8nc}(X)$ and $D_{5nc}(X)$ for VQ codebook size of 64.

6. Summary and Conclusion

A fuzzy approach to speaker verification has been proposed in this paper. Based on fuzzy set theory, fuzzy entropy and fuzzy c-means membership scores are proposed. The likelihood transformation has been proposed to find new scores more general than current likelihood scores. This fuzzy approach also leads to a noise clustering-based version for all scores, which improves speaker verification performance markedly. Using the arctan function in computing the score illustrates a theoretical extension for normalization methods, where not only the logarithm function but also other functions can be used. Speaker verification experiments have been performed on the YOHO corpus to evaluate proposed normalization methods. The lowest EER for 138 speakers of the YOHO corpus was 1.8% for the noise clustering-based normalization method using the fuzzy c-means membership function as a similarity score for 64-mixture Gaussian mixture speaker models.

Acknowledgment

The authors would like to thank the reviewers for their comments and suggestions.
References

Dat Tran received the B.S. and M.S. degrees in theoretical physics from the University of Natural Sciences, Vietnam, in 1984 and 1994, respectively. He received the Grad. Dip. degree in information sciences and the Ph.D. in information sciences and engineering from the University of Canberra, Australia, in 1997 and 2001, respectively.

He was a lecturer in physics at the University of Natural Sciences, Vietnam from 1984 to 1995. Currently, he is a lecturer in computing at the University of Canberra, Australia.

He was a summer intern at the IBM T. J. Watson Research Center, NY in 2000. Dr. Tran is a member of the IEEE and the Australian Speech Science and Technology Association (ASSTA).

His main research interests are in the areas of speech recognition, speaker recognition, fuzzy clustering, fuzzy pattern recognition, and ensemble methods.

Michael Wagner received a Diplomphysiker degree from the University of Munich in 1973 and a Ph.D. in computer science from the Australian National University in 1979. He has held research and teaching positions at the Technical University of Munich, National University of Singapore, Nixdorf AG, the University of Wollongong, Australian Defence Force Academy and Australian National University. He has been a visiting fellow, among others, at the Universities of Hong Kong and Amsterdam and at Siemens Research & Development Laboratories in Munich. He has published in the areas of speech recognition, speaker verification, computer user authentication and computer security, and was the Foundation President of the Australian Speech Science and Technology Association from 1988 to 1992.

Currently, he is Professor of Computing at the University of Canberra and Director of UC’s Human–Computer Communication Laboratory. Prof. Wagner is a member of the Board of the International Speech Communication Association, fellow of the Institution of Engineers Australia, member of ASSTA and senior member of the IEEE.