Robust Speaker Identification System Based on Two-Stage Vector Quantization

Wan-Chen Chen\textsuperscript{1,2}, Ching-Tang Hsieh\textsuperscript{2,*} and Chih-Hsu Hsu\textsuperscript{3}

\textsuperscript{1}Department of Electronic Engineering, St. John’s University, Taipei, Taiwan 251, R.O.C.
\textsuperscript{2}Department of Electrical Engineering, Tamkang University, Tamsui, Taiwan 251, R.O.C.
\textsuperscript{3}Department of Information Technology, Ching Kuo Institute of Management and Health, Keelung, Taiwan 203, R.O.C.

Abstract

This paper presents an effective method for speaker identification system. Based on the wavelet transform, the input speech signal is decomposed into several frequency bands, and then the linear predictive cepstral coefficients (LPCC) of each band are calculated. Furthermore, the cepstral mean normalization technique is applied to all computed features in order to provide similar parameter statistics in all acoustic environments. In order to effectively utilize these multi-band speech features, we propose a multi-band 2-stage vector quantization (VQ) as the recognition model in which different 2-stage VQ classifiers are applied independently to each band and the errors of all 2-stage VQ classifiers are combined to yield total error and a global recognition decision. Finally, the KING speech database is used to evaluate the proposed method for text-independent speaker identification. The experimental results show that the proposed method gives better performance than other recognition models proposed previously in both clean and noisy environments.

Key Words: Speaker Identification, Wavelet Transform, Linear Predictive Cepstral Coefficient (LPCC), 2-Stage Vector Quantization

1. Introduction

In general, speaker recognition can be divided into two parts: speaker verification and speaker identification. Speaker verification refers to the process of determining whether or not the speech samples belong to some specific speaker. However, in speaker identification, the goal is to determine which one of a group of known voices best matches the input voice sample. Furthermore, in both tasks, the speech can be either text-dependent or text-independent. Text-dependent means that the text used in the test system must be the same as that used in the training system, while text-independent means that there is no limitation on the text used in the test system. Certainly, the method used to extract and model the speaker-dependent characteristics of the speech signal seriously affects the performance of a speaker recognition system.

Many researches have been done on the feature extraction of speech. The linear predictive cepstral coefficients (LPCC) were used because of their simplicity and effectiveness in speaker/speech recognition [1,2]. Other widely used feature parameters, namely, the mel-scale frequency cepstral coefficients (MFCC) [3], were calculated by using a filter-bank approach, in which the set of filters had equal bandwidths with respect to the mel-scale frequencies. This method was based on the fact that human perception of the frequency contents of sounds did not follow a linear scale. The above two most commonly used feature extraction techniques do not provide in-
variant parameterization of speech; the representation of the speech signal tends to change in various noise conditions. The performance of these speaker identification systems may be severely degraded when a mismatch between the training and testing environments occurs. Various types of speech enhancement and noise elimination techniques have been applied to feature extraction. Typically, Furui [4] used the cepstral mean normalization (CMN) technique to eliminate the channel bias by subtracting off the global average cepstral vector from each cepstral vector.

In past studies on recognition models, the hidden Markov model (HMM) [5,6], Gaussian mixture model (GMM) [7–10], and vector quantization (VQ) [11–13] were used to perform speaker recognition. HMM [5,6] is widely used in speech recognition, and it is also commonly used in text-dependent speaker verification. GMM [7–10] provides a probabilistic model of the underlying sounds of a person’s voice. It is computationally more efficient than HMM and has been widely used in text-independent speaker recognition. It has been shown that VQ [11–13] is very effective for speaker recognition. Although the performance of VQ is not as good as that of GMM [7], VQ is computationally more efficient than GMM. VQ is also widely used in coding of speech signals and the multi-stage VQ (MSVQ) [14] can achieve very low bit rates and storage complexity in comparison with conventional VQ.

Conventionally, feature extraction is carried out by computing acoustic feature vectors over the full band of the spectral representation of speech. The major drawback of this approach is that even a partial band-limited noise corruption affects all feature vector components. The multi-band approach dealt with this problem by performing acoustic feature analysis independently on a set of frequency subbands [15]. Since the resulting coefficients were computed independently, a band-limited noise signal did not spread over the entire feature space. In our previous works [16,17], we proposed a multi-band feature extraction method in which the LPCC features extracted from the full band and different subbands are combined to form a single feature vector. This feature extraction method was evaluated in a speaker identification system using VQ and GMM as identifiers. The experimental results showed that this multi-band feature was more effective and robust than the full-band LPCC and MFCC features, particularly in noisy environment. In other previous works, we proposed the multi-layer eigen-codebook VQ (MLECVQ) [18] and likelihood combination GMM (LCGMM) [19] recognition models. The general idea was to split the whole frequency band into full-band and a few subbands on which different eigen-codebook vector quantization (ECVQ) and GMM classifiers were independently applied and then recombined to yield global likelihood scores and a global recognition decision. The experimental results showed that the MLECVQ and LCGMM recognition models were more effective and robust than the conventional GMM using full-band LPCC and MFCC features, particularly in noisy environment.

In this study, the multi-band linear predictive cepstral coefficients (MBLPCC) proposed previously [16–19] were used as the front end of the speaker identification system. Then, the cepstral mean normalization was applied to these multi-band speech features to provide similar parameter statistics in all acoustic environments. In order to effectively utilize these multi-band speech features, we proposed a multi-band 2-stage VQ as the recognition model. Different 2-stage VQ classifiers were applied independently to each band, and then the errors of all 2-stage VQ classifiers were combined to yield total error and a global recognition decision. By evaluating the proposed method, we can see that the proposed method outperforms the MLECVQ and LCGMM recognition models proposed previously.

This paper is organized as follows. The proposed algorithm for extracting speech features is described in section 2. The multi-band speaker recognition model is presented in section 3. Experimental results and comparisons with the baseline GMM, MLECVQ and LCGMM models are presented in section 4. Concluding remarks are made in section 5.

2. Multi-Band Features Based on Wavelet Transform

The recent interest in the multi-band feature extraction approach has mainly been attributed to Allen’s paper [20], where it is argued that the human auditory system processes features from different subbands independently, and that the merging is done at some higher point of processing to produce a final decision. The advantages of
using multi-band processing are multifold and have been described in earlier publications [21–23]. The major drawback of a pure subband-based approach may be that information about the correlation between different subbands is lost. Therefore, we suggest that the full-band features are not to be ignored, but should be combined with subband features to maximizing the recognition accuracy. A similar approach that combined information from the full band and subbands at the recognition stage was found to improve recognition performance [23]. It is not a trivial matter to decide at which temporal level the subband features should be combined. In the multi-band approaches [21,22], different recognizers for each band were applied, and likelihood recombination was done at the HMM state, phone or word level to yield global score and a global recognition decision. In our approach, the full band and subband features were used in the recognition model.

Based on time-frequency multi-resolution analysis, the effective and robust MBLPCC features are used as the front end of the speaker identification system. First, the LPCC is extracted from the full-band input signal. Then the wavelet transform is applied to decompose the input signal into two frequency subbands: a lower frequency subband and a higher frequency subband. To capture the characteristics of an individual speaker, the LPCC of the lower frequency subband is calculated. There are two main reasons for using the LPCC parameters: their good representation of the envelope of the speech spectrum of vowels, and their simplicity. Based on this mechanism, we can easily extract the multi-resolution features from all lower frequency subband signals simply by iteratively applying the wavelet transform to decompose the lower frequency subband signals, as depicted in Figure 1. As shown in Figure 1, the wavelet transform can be realized by using a pair of finite impulse response (FIR) filters, $h$ and $g$, which are low-pass and high-pass filters, respectively, and by performing the down-sampling operation $\downarrow 2$. The down-sampling operation is used to discard the odd-numbered samples in a sample sequence after filtering is performed.

The schematic flow of the proposed feature extraction method is shown in Figure 2. After the full-band LPCC is extracted from the input speech signal, the discrete wavelet transform (DWT) is applied to decompose the input signal into a lower frequency subband, and the subband LPCC is extracted from this lower frequency subband. The recursive decomposition process enables us to easily acquire the multi-band features of the speech signal. Based on the concept of the proposed method, the number of subbands depends on the level of the decomposition process. If speech signals band-limited from 0 to 4000 Hz are recursively decomposed two times, then three bands signals, 0–4000 Hz, 0–2000 Hz, and 0–1000 Hz, will be generated. Since the spectra of the three bands will overlap in the lower frequency region, the proposed multi-band feature extraction method focuses on the spectrum of the speech signal in the low frequency region similar to MFCC features.

Finally, cepstral mean normalization is applied to normalize the feature vectors so that their short-time means are normalized to zero as follows:

$$\hat{X}_k(t) = X_k(t) - \mu_k$$

(1)

where $X_k(t)$ is the $k$th component of feature vector at time (frame) $t$, and $\mu_k$ is the mean of the $k$th component of the feature vectors of a specific speaker’s utterance.

![Figure 1. Two-band analysis tree for a discrete wavelet transform.](image)
3. Multi-Band Speaker Recognition Models

3.1 2-Stage Vector Quantization

In MSVQ, each stage consists of a stage codebook. A reproduction of an input vector \( x \) is formed by selecting from each stage codebook a stage code vector, \( q_i \), for stage \( i \), and forming their sum, i.e., \( \hat{x} = \sum_{i=1}^{P} q_i \), where \( P \) denotes the number of stages. For simplicity, MSVQ implemented in this paper is restricted to only two stages. The basic structure of 2-stage VQ is depicted in Figure 3. The quantizer \( Q_1 \) in first stage uses a codebook \( Y = \{y_1, y_2, \ldots, y_N\} \) and the second stage quantizer \( Q_2 \) uses a codebook \( Z = \{z_1, z_2, \ldots, z_M\} \). In this paper, a sequential search procedure is used in 2-stage VQ. The input vector \( x \) is quantized with the first stage codebook producing the selected first-stage code vector \( y = Q_1(x) \) where \( Q_i(\cdot) \) denotes the \( i \)th stage quantization operation. The code vector \( y \) in codebook \( Y \) nearest in Euclidean distance to the input \( x \) is found such that

\[
\| x - y \|^2 \leq \| x - y_i \|^2, \ i = 1, 2, \cdots, N
\]  

(2)

A residual vector \( e_1 \) is formed by subtracting \( y \) from \( x \).

\[
e_1 = x - y
\]  

Then residual \( e_1 \) is quantized with the code vector \( z = Q_2(e_1) \) which minimizes the Euclidean distance over all other code vectors in codebook \( Z \).

\[
\| e_1 - z_j \|^2 \leq \| e_1 - z_i \|^2, \ j = 1, 2, \cdots, M
\]  

(4)

The input vector \( x \) is quantized to \( \hat{x} \) where

\[
\hat{x} = y + z
\]  

(5)

The error \( d(x, \lambda) \) between the original vector \( x \) and reproduction vector \( \hat{x} \) by the 2-stage VQ \( \lambda \) is equal to the quantization error in the last (second) stage.

\[
d(x, \lambda) = \| x - \hat{x} \| = \| x - y - z \| = \| e_1 - z \|
\]  

(6)

In this paper, the conventional way of designing the stage codebooks for sequential search MSVQ with a mean-squared error (MSE) overall distortion criterion is to apply the Linde-Buzo-Gray (LBG) algorithm [24] stage-by-stage for a given training set of input vectors. The LBG algorithm is first applied to the training set to design the first stage codebook. A training set for the second stage is then constructed by pooling together the residual vectors resulting from the quantization of the
original training set with the first stage codebook. The second stage codebook is then designed with the LBG algorithm applied to the derived residual training set.

### 3.2 Multi-Band 2-Stage VQ Recognition Model

As described in Section 1, MSVQ is widely used in coding of speech signals and can achieve very low bit rates and storage complexity in comparison with conventional VQ. Here, we use 2-stage VQ as the classifier. Our multi-band scheme combines the error of the independent 2-stage VQ for each band, as illustrated in Figure 4. We name this recognition model the multi-band 2-stage VQ. First, the input signal is decomposed into \( L \) sub-bands. Then the LPCC features extracted from each band are further normalized to zero mean by using the cepstral mean normalization technique. Finally, different 2-stage VQ classifiers are applied independently to each band, and then the errors of all 2-stage VQ classifiers are combined to yield total error and a global recognition decision.

For speaker identification, a group of \( S \) speakers is represented by the multi-band 2-stage VQs, \( \lambda_1, \lambda_2, \ldots, \lambda_S \). Assume a given testing utterance \( X \) is divided into \( T \) frames, \( X = \{X^{(1)}, X^{(2)}, \ldots, X^{(T)}\} \). For the \( j \)th frame \( X^{(j)} \), \( X^{(0)} \) is decomposed into \( L \) subbands. Let \( X_{ij}^{(j)} \) and \( \lambda_{ki} \) be the feature vector of the \( j \)th frame and the associated 2-stage VQ of a specific speaker \( k \) in band \( i \), respectively. The \( j \)th frame error \( d(X^{(j)}, \lambda_k) \) for the multi-band 2-stage VQ of a specific speaker \( k \) is determined as the sum of the errors of all bands as follows:

\[
d(X^{(j)}, \lambda_k) = \sum_{i=0}^{L} d(X_{ij}^{(j)}, \lambda_{ki})
\]

When \( i = 0 \), \( d(X_{ij}^{(0)}, \lambda_{k0}) \) represents the \( j \)th frame error between the full-band feature and the associated 2-stage VQ for a specific speaker \( k \). Therefore, the total error for the \( T \)-frame testing speech utterance \( X \) is given by

\[
d(X, \lambda_k) = \sum_{j=1}^{T} d(X^{(j)}, \lambda_k) = \sum_{j=1}^{T} \sum_{i=0}^{L} d(X_{ij}^{(j)}, \lambda_{ki})
\]

For a given testing speech utterance \( X, \hat{S} \) is classified to belong to the speaker \( \hat{S} \) who has the minimum error \( d(X, \lambda_{\hat{S}}) \).

\[
\hat{S} = \arg \min_{1 \leq s \leq S} d(X, \lambda_s)
\]

### 4. Experimental Results

This section presents experiments conducted to evaluate application of the proposed multi-band 2-stage VQ to closed-set text-independent speaker identification. The first experiment studied the effect of the number of bands used in the proposed recognition model. The next experiment evaluated the effect of the number of code vectors under both clean and noisy environments. The last experiment compared the performance of the multi-band 2-stage VQ with that of the MLECVQ [18] and LCGMM [19] models proposed previously.

#### 4.1 Database Description and Parameters Setting

The proposed methods were evaluated using the KING speech database [25] for closed-set text-independent...
speaker identification. The KING database is a collection of conversational speech from 51 male speakers. For each speaker there are 10 sections of conversational speech that were recorded at different times. Each section consists of about 30 seconds of actual speech. The speech from a section was recorded locally using a microphone and was transmitted over a long distance telephone link, thus providing a high-quality (clean) version and a telephone quality version of the speech. The speech signals were recorded at 8 kHz and 16 bits per sample. In our experiments, the noisy speech was generated by adding Gaussian noise to the clean version speech at the desired signal to noise ratio (SNR). In order to eliminate the silence segments from an utterance, simple segmentation based on the signal energy of each speech frame was applied. All experiments were performed using five sections of speech from 20 speakers. For each speaker, 90 seconds of speech cut from four clean version sections provided the training utterances. Typically, we need utterances longer than 2 seconds to achieve adequate accuracy in speaker identification [7]. Hence, the other one section was divided into no overlapping segments 2 seconds in length and provided the testing utterances.

The evaluation of the average speaker identification performance was conducted in the following manner. The identified speaker of each test segment was compared with the actual speaker of the test utterance and the number of segments correctly identified was tabulated. The above steps were repeated for all test segments from each speaker in the population consisting of 20 speakers. The average identification rate was computed as follows:

\[
\text{identification rate} = \frac{N_{\text{correct}}}{N_{\text{total}}} \tag{10}
\]

where \( N_{\text{correct}} \) is the number of correctly identified segments, and \( N_{\text{total}} \) is the total number of test segments.

In all experiments conducted in this study, each frame of an analyzed utterance had 256 samples with 192 overlapping samples. Furthermore, 20 orders of LPCC for each frequency band were calculated and the first order coefficient was discarded, which leads to a 19 dimensional feature vector. For our multi-band approach, we used 2, 3 and 4 bands as follows:

- 2 bands: \((0-4000), (0-2000)\) Hz;
- 3 bands: \((0-4000), (0-2000), (0-1000)\) Hz;
- 4 bands: \((0-4000), (0-2000), (0-1000), (0-500)\) Hz.

### 4.2 Contribution of the Multi-Band

As explained in section 2, the number of subbands depends on the decomposition level of the wavelet transform. The first experiment evaluated the effect of the number of bands used in the multi-band 2-stage VQ model. In this experiment, the 2-stage VQ in each band had 128 code vectors in first stage and 32 code vectors in second stage. The experimental results are shown in Figure 5. In Figure 5, the single-band using only the full-band signal had the poorest performance. We could see that the identification rate initially increased as the number of bands increased, and the best performance could be achieved when the number of bands was set to be three. Because the signals were decomposed into several frequency subbands and the spectra of the subbands overlapped in the lower frequency region, the success of the MBLPCC features could be attributed to the emphasis on the spectrum of the signal in the low-frequency region. It was found that increasing the number of bands to more than three not only increased the computation time but also decreased the identification rate. In this case, the signals of the lowest frequency subband were located in the very low frequency region, which put too much emphasis on the lower frequency spectrum of speech. In addition, the number of samples within the lowest frequency subband was so small that the spectral characteristics of speech could not be estimated accurately. Consequently, the poor result in the lowest frequency subband degraded the system performance. The experimental results gave us a distinct guide for choosing the good decomposition level for the wavelet transform.

![Figure 5. Effect of number of bands on the identification performance of the multi-band 2-stage VQ model with 128 code vectors in first stage and 32 code vectors in second stage in clean environment.](image-url)
4.3 Effects of the Number of Code Vectors in First and Second Stage Codebooks

This experiment evaluated the effect of the number of code vectors in first and second stage codebooks of the 3-band 2-stage VQ in both clean and noisy environments. The experimental results are shown in Table 1. One could see that the 3-band 2-stage VQ with 128 code vectors in first stage and 32 code vectors in second stage achieved best identification rate except in 10 dB SNR condition. Consequently, the experimental results gave us a distinct guide for choosing the good number of code vectors in first and second stage codebooks.

4.4 Comparison with Other Existing Models

First, we would compare the performance of the 2-stage VQ model with that of the GMM and ECVQ [18] models using full-band LPCC features under Gaussian noise corruption. The parameters of the 2-stage VQ model were defined as follows: 128 code vectors in first stage and 32 code vectors in second stage, 19 dimensional LPCC feature vectors. The parameters of the GMM model were defined as follows: 50 mixtures, 19 dimensional LPCC feature vectors. The parameters of the ECVQ model were defined as follows: 64 code vectors, one projection basis vector, 15 dimensional LPCC feature vectors, the adjustment factor $R_j = 0.03$. The experimental results were shown in Figure 6. We could see that the performance of the GMM model was seriously degraded by Gaussian noise corruption. On the other hand, the 2-stage VQ and ECVQ models were more robust under Gaussian noise corruption. The 2-stage VQ had slightly better performance compared with ECVQ model except in 5 dB SNR condition.

In next experiment, we would compare the performance of the proposed model with other multi-band models. The performance of the 3-band 2-stage VQ recognition model with 128 code vectors in first stage and 32 code vectors in second stage was compared with that of the baseline GMM, MLECVQ [18] and LCGMM [19] models proposed previously under Gaussian noise cor-

<table>
<thead>
<tr>
<th>Number of code vectors in first stage codebook</th>
<th>Number of code vectors in second stage codebook</th>
<th>SNR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Clean</td>
</tr>
<tr>
<td>80</td>
<td>80</td>
<td>94.35%</td>
</tr>
<tr>
<td>100</td>
<td>60</td>
<td>95.35%</td>
</tr>
<tr>
<td>128</td>
<td>32</td>
<td>95.35%</td>
</tr>
<tr>
<td>144</td>
<td>16</td>
<td>95.02%</td>
</tr>
<tr>
<td>152</td>
<td>8</td>
<td>94.35%</td>
</tr>
</tbody>
</table>

Figure 6. Identification rates for GMM, ECVQ, and 2-stage VQ models using full-band LPCC features with white noise corruption.
ruption. The parameters of the baseline GMM model were defined as follows: 50 mixtures, the features of full band, 19 dimensional LPCC feature vectors. The parameters of the MLECVQ model were defined as follows: 64 code vectors, one projection basis vector, the features of four bands, 15 dimensional LPCC feature vectors, the adjustment factor $R_j = 0.03$. The parameters of the LCGMM model were defined as follows: 50 mixtures, the features of three bands, 19 dimensional LPCC feature vectors. The experimental results were shown in Table 2. We could see that the performance of the baseline GMM was seriously degraded as the SNR decreased and achieved the poorest performance among all the models since our previous work [19] indicated that the MBLPCC features was more robust than the full-band LPCC features used by the baseline GMM. The 3-band 2-stage VQ could yield at least comparable performance in clean and 5 dB SNR conditions while providing best robustness among all models in the case of noisy speech. The 3-band LCGMM achieved better performance in clean, 20dB, and 15 dB SNR conditions but poorer performance in lower SNR conditions compared with the 4-band MLECVQ. Since above experiment indicated that the ECVQ was more robust than GMM under lower SNR conditions in single band scheme, the MLECVQ would be more robust than LCGMM under lower SNR conditions in multi-band scheme. Based on these results, it can be concluded that the 3-band 2-stage VQ is effective and robust to represent the characteristics of individual speakers in additive Gaussian noise conditions.

Wu et al. [26] proposed an auditory model according to the mechanism of human auditory periphery and cochlear nucleus, based on which four channels of auditory features were extracted. For each channel of auditory feature, a GMM classifier was independently applied and then combined the scores of these channels to yield global likelihood scores. They used speech utterances of 49 speakers in clean and telephonic version of the KING speech database to evaluate the performance of the auditory model for text-independent speaker identification. For each speaker, 3 sections of speech (about 90 seconds) provided the training utterances. The segment length of testing utterances was 6.4 seconds. The experimental results were shown in Table 3. The auditory model achieved more robust performance below 20 dB SNR but poorer performance in clean and 30 dB SNR conditions compared with the GMM using MFCC features. Since the performance of the 3-band 2-stage VQ outperform that of the baseline GMM from the clean speech data down to 5 dB SNR, the proposed 3-band 2-stage VQ performed better than the auditory model [26].

5. Conclusions

In this study, the effective and robust MBLPCC features were used as the front end of a speaker identification system. In order to effectively utilize these multi-band speech features, we proposed a multi-band 2-stage VQ as the recognition model. Different 2-stage VQ classifiers were applied independently to each band, and then

<p>| Table 2. Identification rates for 4-band MLECVQ, 3-band LCGMM, and 3-band 2-stage VQ with white noise corruption |
|----|----------------|---------|---------|---------|---------|</p>
<table>
<thead>
<tr>
<th>SNR</th>
<th>Method</th>
<th>Clean</th>
<th>20 dB</th>
<th>15 dB</th>
<th>10 dB</th>
<th>5 dB</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseline GMM</td>
<td>95.02%</td>
<td>77.23%</td>
<td>54.87%</td>
<td>27.95%</td>
<td>8.49%</td>
<td></td>
</tr>
<tr>
<td>4-band MLECVQ</td>
<td>95.02%</td>
<td>88.04%</td>
<td>82.39%</td>
<td>70.43%</td>
<td>54.82%</td>
<td></td>
</tr>
<tr>
<td>3-band LCGMM</td>
<td>96.68%</td>
<td>92.36%</td>
<td>83.72%</td>
<td>65.12%</td>
<td>37.87%</td>
<td></td>
</tr>
<tr>
<td>3-band 2-stage VQ</td>
<td>95.35%</td>
<td>93.02%</td>
<td>86.38%</td>
<td>75.08%</td>
<td>53.49%</td>
<td></td>
</tr>
</tbody>
</table>

<p>| Table 3. Identification rates for auditory computational model [26] and GMM with white noise corruption |
|----|------|---------|---------|---------|-------|</p>
<table>
<thead>
<tr>
<th>SNR</th>
<th>Method</th>
<th>Clean</th>
<th>30 dB</th>
<th>20 dB</th>
<th>10 dB</th>
</tr>
</thead>
<tbody>
<tr>
<td>GMM + MFCC</td>
<td>88.09%</td>
<td>88.16%</td>
<td>43.76%</td>
<td>14.67%</td>
<td></td>
</tr>
<tr>
<td>Auditory model [26]</td>
<td>76.83%</td>
<td>73.10%</td>
<td>65.03%</td>
<td>50.59%</td>
<td></td>
</tr>
</tbody>
</table>
errors of all 2-stage VQ classifiers were combined to yield total error. Finally, the KING speech database was used to evaluate the proposed method for text-independent speaker identification. The experimental results show that the proposed method is more effective and robust than the baseline GMM, MLECVQ and LCGMM models proposed previously.

References


Manuscript Received: Jul. 8, 2007
Accepted: Mar. 10, 2008