EURASIP Journal on Advances in Signal Processing
a SpringerOpen Journal

RESEARCH Open Access

# Discriminative likelihood score weighting based on acoustic-phonetic classification for speaker identification

Youngjoo Suh* and Hoirin Kim

**Abstract**

In this paper, a new discriminative likelihood score weighting technique is proposed for speaker identification. The proposed method employs a discriminative weighting of frame-level log-likelihood scores with acoustic-phonetic classification in the Gaussian mixture model (GMM)-based speaker identification. Experiments performed on the Aurora noise-corrupted TIMIT database showed that the proposed approach provides meaningful performance improvement with an overall relative error reduction of 15.8% over the maximum likelihood-based baseline GMM approach.

**Keywords:** Discriminative training; Acoustic-phonetic classification; Score weighting; Speaker identification

## 1 Introduction

Speaker recognition mainly consists of two different tasks, speaker identification and speaker verification. The goal of speaker identification is to determine which one of a group of registered speakers best matches the input speech utterance [1], whereas that of speaker verification is to determine if the claim is true or false for the input speech utterance, a claim of identity, and the corresponding speaker model [2]. Speaker identification has potentially more applications such as access control, forensics, speech data management, personalization, and intelligent robot control than speaker verification [3]. While speaker verification has been much more studied and, as a result, produced successful outcomes by proposing a series of statistical techniques such as Gaussian mixture model (GMM) [1,4], hidden Markov model (HMM) [5], and support vector machine (SVM) [2,6-8], speaker identification has not reached such level of technical advancement. It is said that some causes of slower progress in speaker identification might be due to the increase in the expected error with growing population size and very high computational cost [3]. Most of the current speaker identification approaches are also based on the same statistical frameworks such as GMM [1,9] or SVM [2,6,10]. Although SVM has shown to be very effective in two-class classification problems such as speaker verification, it may need further algorithmic development in the multi-class tasks including speaker identification [11]. Under this circumstance, it can be said that the GMM method still plays an important role in speaker identification. In the GMM-based speaker identification framework, speaker information is extracted in the form of the probabilistic score for the corresponding speaker model, which is then compared with each other in the decision. In the scoring process, all speech frames in the given speech utterance are used to extract speaker information. However, it may be natural to assume that some speech frames can provide more speaker-specific information than others due to both acoustic-phonetic characteristics of speech signals and speaker's own voice characteristics. Nevertheless, conventional GMM-based speaker identification approaches extract registered speakers' information just by the equally weighted sum of their corresponding frame-level scores. Thus, we expect that the performance of speaker identification can be improved by applying discriminative weights on speech frames after taking into full account their acoustic-phonetic classes as well as speaker's voice characteristics in deriving speaker scores. Of course, a couple of discriminative weighting approaches to speaker identification based on the minimum classification error (MCE)

* Correspondence: yjsuh@kaist.ac.kr
Department of Electrical Engineering, Korea Advanced Institute of Science and Technology, 291 Daehak-Ro, Yuseong-Gu, Daejeon 305-701, Korea

criterion have already been introduced in the literature [5,12]. Most of these discriminative weighting approaches focus on the mixture weights in GMM. However, mixture weight-based discriminative weighting approaches have some limitations. First, training mixture weights using the MCE algorithm can cause the poor generalization problem due to the insufficient training data which can be common in speaker identification frameworks employing large-sized GMMs. Second, some discriminative mixture information can be lost by the mixture summation procedure during the likelihood calculation due to too many mixture components in large-sized GMMs. For these reasons, we propose a new discriminative weighting-based speaker identification technique. The proposed technique employs an acoustic-phonetic classification-driven discriminative weighting scheme for frame-level log-likelihood scores according to their acoustic-phonetic classes as well as speakers' voice characteristics. In the speaker identification experiments based on the Aurora noise-corrupted TIMIT database, the proposed technique yielded enhanced performance over the conventional GMM approach.

The organization of this paper is as follows. In Section 2, we introduce the basic algorithm of the conventional speaker identification system. In Section 3, we describe the overall principle of the proposed MCE-based discriminative score weighting technique which utilizes the acoustic-phonetic classification. Then, experimental results are discussed with our findings in Section 4. Finally, conclusions are given in Section 5.

## 2 Conventional speaker identification system

In a speaker identification system, a group of registered speakers $S = \{1, 2, 3, \ldots, S\}$ are modeled by a set of statistical models denoted $\mathbf{\Lambda} = \{\Lambda_1, \Lambda_2, \Lambda_3, \ldots, \Lambda_S\}$. The speaker identity is determined by finding a speaker with the maximum posterior probability for the given input feature vector sequence $X = \{x_{1\mathrm{cul}}, x_2, x_3, \ldots, x_T\}$ [1] as

$$\hat{S} = \arg\max_{1 \le j \le S} P(\Lambda_j|X). \tag{1}$$

Applying Bayes' rule, assuming equally likely speakers, and noting independence of $P(X)$ on the speakers, Equation 1 can be reduced to

$$\hat{S} = \arg\max_{1 \le j \le S} P(X|\Lambda_j). \tag{2}$$

Assuming statistical independence between feature vectors and taking logarithms, Equation 2 becomes

$$\hat{S} = \arg\max_{1 \le j \le S} \sum_{t=1}^{T} \log P(x_t|\Lambda_j). \tag{3}$$

## 3 MCE-based discriminative score weighting with acoustic-phonetic classification

It can be assumed that speech frames have different amount of speaker information according to their acoustic-phonetic classes [13] as well as speaker's voice characteristics. Under this assumption, a speech frame $x_t$ can be classified into its corresponding acoustic-phonetic class if some classification scheme is provided in advance. Of a variety of classification methods, we employed a hard classification approach based on the vector quantization technique with GMM for algorithmic simplicity. The unsupervised clustering capability of GMM can automatically provide a number of acoustic-phonetic classes for the whole acoustic space which spans the entire training data. The vector quantization-based hard classification approach can be defined by

$$Q(x_t) = \arg\max_{m} P(\psi_m|x_t), \tag{4}$$

where $\Psi_m$ denotes a set of Gaussian model parameters of the $m$th acoustic-phonetic class of a total of $M$ acoustic-phonetic classes which are given by a GMM estimated from the training data which are assumed to cover the whole acoustic-phonetic space of speech signals.

Then, the speaker identification rule in (3) can be represented with the concept of acoustic-phonetic classes by classifying each frame into its acoustic-phonetic class and computing the probabilistic scores on the basis of class as

$$\hat{S} = \arg\max_{1 \le j \le S} \sum_{m=1}^{M} \sum_{\forall x_t,\, x_t \in m} \log P(x_t|\Lambda_j). \tag{5}$$

Under this framework, each frame-level log-likelihood score can be discriminatively weighted on the basis of the acoustic-phonetic class as well as speaker to consider its acoustic-phonetic classes as well as speaker's voice characteristics in speaker identification. The speaker identification rule based on this discriminative score weighting (DSW) scheme is given by

$$\hat{S}_{\mathrm{DSW}} = \arg\max_{1 \le j \le S} \sum_{m=1}^{M} w_{jm} \sum_{\forall x_t,\, x_t \in m} \log P(x_t|\Lambda_j), \tag{6}$$

where $w_{jm}$ stands for the discriminative weight for the $m$th acoustic-phonetic class and the $j$th speaker model. The optimal weights for this speaker identification scheme can be obtained by using the MCE-based discriminative training algorithm [5,12,14-16], which aims at deriving a set of speaker models which minimizes classification errors, that is, speaker identification errors for training data. To train these weights discriminatively with the MCE criterion for speaker identification, we

define a discriminative function for each speaker which represents the log-likelihood of the feature vector sequence $X$ given model $\Lambda_j$ of speaker $j$ as

$$g_j(X,\Phi_W)=\sum_{m=1}^{M} w_{jm}\sum_{\forall x_t,x_t\in m}\log P\left(x_t|\Lambda_j\right), \tag{7}$$

where $\Phi_W$ stand for the weight parameters. In (7), the weights represent the amount of score contribution from their corresponding classes. In the equation, their integral sum needs to be normalized to avoid ill-training of the weights. According to these requirements, the weights need to satisfy such constraints [5,16] as

$$\sum_{m=1}^{M} w_{jm}=1,\quad w_{jm}\geq 0. \tag{8}$$

Then, the misclassification measure is defined for the true speaker that is the label information for the input feature vector sequence $k$ to measure how much the input feature sequence spoken by the true speaker is misclassified as

$$d_k(X,\Phi_W)=-g_k(X,\Phi_W)+G_k(X,\Phi_W), \tag{9}$$

with

$$G_k(X,\Phi_W)=\log\left(\frac{1}{S-1}\sum_{j\neq k}\exp\left[\eta g_j(X,\Phi_W)\right]\right)^{\frac{1}{\eta}}, \tag{10}$$

where $\eta$ is a positive constant for weight controlling of the competing speaker classes.

A loss function for approximating the empirical loss related to the soft count of classification errors is defined as

$$l_k(X,\Phi_W)=\frac{1}{1+\exp(-\gamma d_k(X,\Phi_W))}, \tag{11}$$

where $\gamma$ is a positive constant used to control the slope of the sigmoid function.

To satisfy the constraints in (8), we take logarithms as

$$\tilde{w}_{jm}=\log w_{jm}. \tag{12}$$

This new parameter set $\left\{\tilde{w}_{jm}\right\}$ is trained by using the generalized probabilistic descent (GPD) algorithm [16] as

$$\tilde{w}_{jm}^{(n+1)}=\tilde{w}_{jm}^{(n)}-\varepsilon\nabla l_k\left(\tilde{w}_{jm}^{(n)}\right), \tag{13}$$

where $\varepsilon$ is a step size of the GPD algorithm and $\Delta l_k\left(\tilde{w}_{jm}\right)$ is derived as

$$\nabla l_k\left(\tilde{w}_{jm}\right)=\frac{\partial l_k}{\partial d_k}\frac{\partial d_k}{\partial g_j}\frac{\partial g_j}{\partial \tilde{w}_{jm}}, \tag{14}$$

where

$$\frac{\partial l_k}{\partial d_k}=\gamma l_k(1-l_k), \tag{15}$$

$$\frac{\partial d_k}{\partial g_j}=\begin{cases}-1, & \text{if } j=k\\ \dfrac{\exp\left[\eta g_j(X,\Phi_W)\right]}{\sum_{i\neq k}\exp\left[\eta g_i(X,\Phi_W)\right]}, & \text{otherwise}\end{cases}, \tag{16}$$

$$\frac{\partial g_j}{\partial \tilde{w}_{jm}}=w_{jm}\sum_{\forall x_t,x_t\in m}\log P\left(x_t|\Lambda_j\right). \tag{17}$$

After $\tilde{w}_{jm}$ is updated, $w_{jm}$ is obtained by using the following transformation to satisfy the constraints in (8) as

$$w_{jm}=\frac{\exp\left(\tilde{w}_{jm}\right)}{\sum_{m=1}^{M}\exp\left(\tilde{w}_{jm}\right)}. \tag{18}$$

The pseudocode of this training algorithm for the discriminative score weights is given in Algorithm 1.

**Algorithm 1 Pseudocode of the MCE-based discriminative training algorithm**

```
1. Initialization step:
   For speaker j = 1 to S
      For mixture m = 1 to M
         Initialize w_jm to 1/M, logw_jm to log(w_jm)
   Initialize n to zero
2. Iteration step:
   While
      Initialize dSum to zero
      For all training utterances X's
         Set k to speaker ID of X
         For speaker j = 1 to S
            Compute g_j(X,Φ_W) in (7)
         Compute G_k(X,Φ_W) in (10)
         Add d_k(X,Φ_W) to dSum
         For speaker j = 1 to S
            For mixture component m = 1 to M
               Compute ∇l_k (w̃_jm) in (14)
               Compute w̃_jm^(n+1) in (13)
            For mixture component m = 1 to M
               Compute w_jm in (18)
         Add one to n
      Divide dSum by # X's;
      If dSum converged
         Go to termination step
      Else
         Continue
3. Termination step:
   Store w_jm
   Finish discriminative training
```

## 4 Experimental results

In the performance evaluation, we used a subset of the TIMIT database [17] consisting of 2,000 utterances spoken by 200 speakers evenly. To test the proposed

technique in the telephone-based various noisy environments, the original 16-kHz sampled clean speech data have been downsampled to 8-kHz sampling rate. These clean speech data are then artificially added with four kinds of the Aurora noise [18] composed of car, restaurant, subway, and street, with three signal-to-noise ratio (SNR) levels of 20, 10, and 0 dB, which results in a set of noisy speech data consisting of 12 noisy conditions, each with 2,000 utterances. The whole speech data used in the experiments consist of 26,000 utterances including the clean and noisy speech data. Of these data, half of them were used in the training and the remaining data are used for test, which eventually means that 65 utterances are assigned to each registered speaker in either side of the data. The baseline speaker identification system was built from the Gaussian mixture model-universal background model (GMM-UBM) by using the maximum *a posteriori* (MAP)-based adapted GMM algorithm [4]. According to this algorithm, UBM was estimated from the whole training data. Then, each registered speaker GMM is estimated by adapting UBM with its corresponding training data consisting of 65 clean and noisy speech utterances. In the feature extraction, a sequence of feature vectors, each of which consists of 12-dimensional mel-frequency cepstral coefficients (MFCCs) and a log energy, was extracted from each utterance with a frame length of 20 ms and a shifting interval of 10 ms. The parameters used in defining the MCE algorithm are chosen empirically such as weight control parameter $\eta = 750$, sigmoid slope parameter $\gamma = 0.1$, and step size parameter $\varepsilon = \left(1 - \frac{n}{100{,}000}\right)$.

For performance comparison, we also developed an SVM-based speaker identification system by using the LIBSVM version 3.11 toolkit [19]. The SVM for 200 registered speakers was trained with their GMM supervectors using 13,000 training utterances. Each GMM supervector was obtained by stacking the 13-dimensional adapted GMM mean vectors of the whole Gaussian mixture components [2]. As an SVM kernel, we used the linear kernel [2,8]. The multi-class classification algorithm adopted in SVM for our speaker identification was the one-versus-one method supported in the LIBSVM toolkit [19].

Figure 1 shows an example of how the proposed DSW scheme applies the weights on the log-likelihood scores and improves identification performance. The number of acoustic-phonetic classes was selected as 32 and the number of mixture components was 512 for all speaker models. For the given input speech signals depicted in Figure 1a, Figure 1b represents their corresponding acoustic-phonetic class ID sequence (i.e., class indices). Because the acoustic-phonetic classification is conducted on a single-frame basis, the class information is changing somewhat randomly even within the presumed phone boundaries. Based on this class information, the two weight contours drawn in Figure 1c show subtle but clear differences between the true speaker and the most competing speaker. In this figure, weight values seem smaller in silence regions and become larger in speech regions. Furthermore, weight levels in speech regions are different from each other according to their acoustic-phonetic classes and speakers. These results clearly indicate that the amount of speaker information varies with not only acoustic-phonetic classes but also speakers. Figure 1d shows the log-likelihood contours of the true speaker and the most competing speaker, their differences, and their difference accumulation that resulted from the speaker identification technique utilizing the conventional maximum likelihood (ML)-based GMM method. Without the frame-level score weighting scheme, the accumulated log-likelihood difference at the final frame is below zero, which means that the most competing speaker would be decided as the speaker identity in the decision process. On the contrary, in Figure 1e, the accumulated difference from the proposed DSW technique becomes higher than that of the ML approach as it approaches the final frame. At the final frame, it is above zero. This implies that the proposed DSW technique can reduce speaker identification errors effectively.

Figure 2 shows speaker identification results by the ML approach and the proposed DSW technique. We examined their performances with respect to the number of acoustic-phonetic classes for DSW and the number of mixture components in GMM for speaker identification. Compared with the conventional ML approach, the proposed DSW technique provides meaningful error reduction for all of the three GMMs consistently. In the three GMMs, the performance of the proposed technique reaches a peak value as the number of acoustic-phonetic classes approaches 32. With these classes, the proposed technique provides a maximum error reduction of 15.8% at the GMM of 128 mixture components. However, as the number of acoustic-phonetic classes gets smaller or larger than 32, speaker identification performance deteriorates gradually. The error reduction is negligible at the condition with 128 classes and 128 mixture components, in which the number of classes is equal to the number of mixture components. From these results, it can be concluded that the optimal number of acoustic-phonetic classes does not always correspond to the numbers of mixture components in GMM. This strongly indicates that we need to employ the proposed score weighting scheme, in which we can optimally select the number of classes given the speaker identification framework.

Table 1 represents speaker identification results from the four speaker identification methods, ML-based GMM (ML-GMM), MCE-based GMM (MCE-GMM), SVM, and

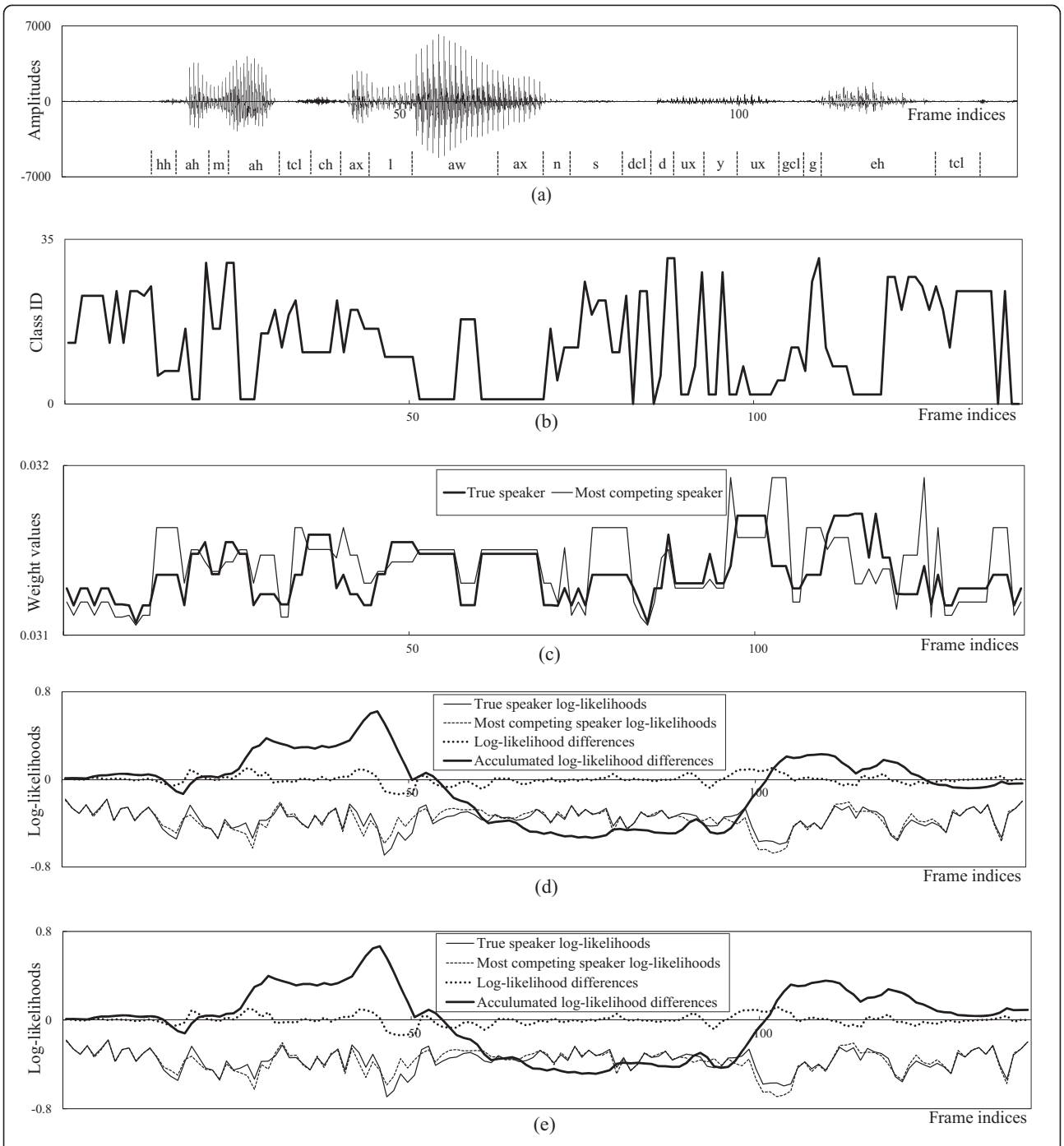


**Figure 1 Comparison of speaker identification results for noisy speech signals degraded by Aurora car noise with 20-dB SNR.** All horizontal axes represent time in 10-ms-long frame. **(a)** Speech signal waveform (a TIMIT sentence utterance spoken as 'How much allowance do you get?'). **(b)** Acoustic-phonetic class ID sequence. **(c)** Weight contours. **(d)** Unweighted scores. **(e)** Weighted scores.

the proposed DSW-based GMM (DSW-GMM), under various SNR conditions. Test data sets of the three SNR conditions except the clean condition include noisy speech data corrupted by four kinds of AURORA noise composed of car, restaurant, subway, and street. The number of mixture components in GMMs was all set to 512. We evaluated the performance of the MCE-based GMM to compare with that of the proposed DSW-based GMM. However, only mixture weights are trained discriminatively due to the excessive requirement of the training time in MCE-GMM. In SVM, the dimension of supervectors was 6,656 (i.e., $13 \times 512$) and the number of support

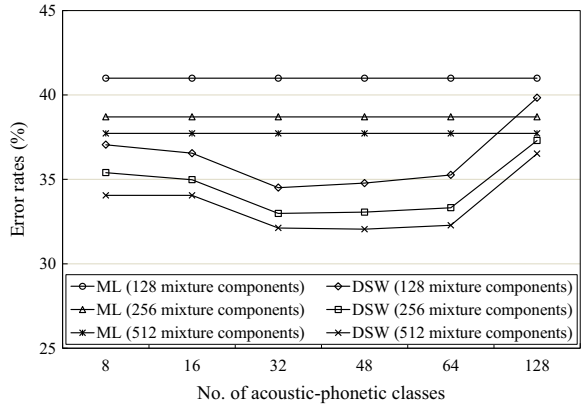


**Figure 2 Speaker identification results by the conventional ML approach and the proposed DSW technique.** The number in the parenthesis represents the number of mixture components in GMMs for speaker identification.

vectors was 12,644. In the table, MCE-GMM yields marginally improved performance compared with ML-GMM. This marginal gain by MCE-GMM supports our early argument mentioned in Section 1 that discriminative training of the mixture weights is not a promising approach when the number of mixture components is excessively large which is common in speaker recognition tasks. Experiments on the discriminative training of the mean vectors and covariance matrices of GMM were not conducted in our study due to their excessive computational load. When these experiments are performed as further works, the experimental results will confirm the effectiveness of MCE-GMM in large-sized GMMs more clearly. SVM shows meaningful performance improvement in the clean condition but provides somewhat inferior performance in the noisy conditions compared with ML-GMM. These inferior results are unexpected since SVM is generally reported to be superior to ML-GMM in pattern classification tasks. However, it should be noted that the number of utterances used to train the SVM parameters in each noisy condition, which means each noise and each SNR, amounts to five per speaker because 65 utterances from all 13 noisy conditions including the clean condition are used to train a speaker. This amount of speech data may be insufficient to estimate the SVM parameters reliably especially in the noisy condition, in which the boundary of each speaker model in the acoustic space can be more complex than in the clean condition. Unlike speaker verification, speaker identification is a multi-class classification task. Because SVM is basically a two-class classification technique, the SVM-based approach needs further algorithmic modification such as the one-versus-one method as employed in this study to cope with the multi-class task. However, it is known that one-versus-one method is prone to overfit unless the individual classifiers are carefully regularized [11]. Therefore, in spite of our experimental results from the SVM method, we still acknowledge that the SVM approach to speaker identification needs further experiments with a larger amount of training data to confirm its effectiveness compared to ML-GMM.

The proposed DSW-GMM approach produces significantly better performance over all SNR conditions compared with the other three methods. The error reductions gained by the DSW-GMM technique over the baseline ML-GMM method are above 10%. These results confirm the effectiveness of the proposed approach in improving speaker identification performance over various noisy conditions. The largest error reduction is achieved in the clean condition. The acoustic-phonetic classification becomes more accurate in the higher SNR conditions. Therefore, we believe that the largest error reduction in the clean condition largely resulted from the most accurate acoustic-phonetic classification.

## 5 Conclusions

In automatic speaker identification, a major technical goal can be to extract the speaker information from input speech signals as effective as possible and thus to maximize the identification performance. Speech signals consist of phones, and it can be said that each phone has different extent of speaker information. For this reason, it is expected that higher speaker identification performance can be achieved if scores involved in the decision process are treated discriminatively according to their acoustic or phonetic class. Discriminative training of mixture weights in GMM can be based on this approach. However, the number of mixture weights in GMMs for speaker identification is usually very large because each GMM should span not only speaker space but also phonetic space. The discriminative training of mixture weights in this large-sized GMM tends to produce low performance improvement. To overcome this drawback of the

**Table 1 Speaker identification error rates (%) by ML-GMM, MCE-GMM, SVM, and the proposed DSW-GMM**

| SNR | ML-GMM | MCE-GMM (mixture weights only) | SVM | DSW-GMM | Error reduction over ML-GMM |
|---|---|---|---|---|---|
| Clean | 10.90 | 10.90 | 9.00 | 8.30 | 23.85 |
| 20 dB | 17.70 | 17.05 | 25.43 | 15.80 | 10.73 |
| 10 dB | 30.02 | 29.35 | 42.45 | 26.45 | 11.89 |
| 0 dB | 72.13 | 68.53 | 69.53 | 59.22 | 17.90 |

conventional mixture weight-based discriminative training approach in the discriminative scoring process which is based on the acoustic or phonetic classes of input speech signals, we propose a new speaker identification technique which is based on the discriminative likelihood score weighting scheme using acoustic-phonetic classification. The proposed technique utilizes optimally weighted frame-level log-likelihood scores for speaker identification. In performance evaluation conducted on the Aurora noise-corrupted TIMIT database, the proposed method showed significantly higher performance compared with other well-known speaker identification approaches such as ML-GMM. MCE-GMM, and SVM.

#### Abbreviations

DSW: discriminative score weighting; DSW-GMM: discriminative score weighting-Gaussian mixture model; GMM: Gaussian mixture model; GMM-UBM: Gaussian mixture model-universal background model; GPD: generalized probabilistic descent; HMM: hidden Markov model; MAP: maximum *a posteriori*; MCE: minimum classification error; MCE-GMM: minimum classification error-Gaussian mixture model; MFCC: mel-frequency cepstral coefficient; ML: maximum likelihood; ML-GMM: maximum likelihood-Gaussian mixture model; SNR: signal-to-noise ratio; SVM: support vector machine; UBM: universal background model.

#### Competing interests

The authors declare that they have no competing interests.

#### Acknowledgements

This work was supported by the IT R&D program of MSIP/KEIT [10041807, Development of Original Software Technology for Automatic Speech Translation with Performance 90% for Tour/International Event focused on Multilingual Expansibility and based on Knowledge Learning].

Received: 20 January 2014 Accepted: 31 July 2014
Published: 10 August 2014

#### References

1. DA Reynolds, RC Rose, Robust text-independent speaker identification using Gaussian mixture speaker models. IEEE Trans. Speech Audio Process. **3**(1), 72–83 (1995)
2. WM Campbell, DE Sturim, DA Reynolds, Support vector machines using GMM supervectors for speaker verification. IEEE Signal. Proc. Let. **13**(5), 308–311 (2006)
3. T Kinnunen, E Karpov, P Fränti, Real-time speaker identification and verification. IEEE Trans. Audio Speech Lang. Process. **14**(1), 277–281 (2006)
4. DA Reynolds, TF Quatieri, RB Dunn, Speaker verification using adapted Gaussian mixture models. Digital Signal. Process. **10**, 19–41 (2000). doi:10.1006/dspr.1999.0361
5. C-S Liu, C-H Lee, W Chou, B-H Juang, AE Rosenberg, A study on minimum error discriminative training for speaker recognition. J. Acoust. Soc. Am. **97**(1), 637–648 (1995)
6. S Fine, J Navrátil, RA Gopinath, A hybrid GMM/SVM approach to speaker identification, in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing, Salt Lake City, 7–11 May 2001*, pp. 417–420
7. N Dehak, PJ Kenny, R Dehak, P Dumouchel, P Ouellet, Front-end factor analysis for speaker verification. IEEE Trans. Audio Speech Lang. Process. **19**(4), 788–798 (2011)
8. Y Suh, H Kim, Minimum classification error-based weighted support vector machine kernels for speaker verification. J. Acoust. Soc. Am. **133**(4), EL307–EL313 (2013)
9. S Kim, M Ji, H Kim, Robust speaker recognition based on filtering in autocorrelation domain and sub-band feature recombination. Pattern Recognition Lett. **31**, 593–599 (2010)
10. V Wan, WM Campbell, Support vector machines for speaker verification and identification, in *Proc IEEE Signal Processing Society Workshop, Sydney, 11–13 Dec 2000*, pp. 775–784
11. JC Platt, N Cristianini, J Shawe-Taylor, Large margin DAGs for multiclass classification, in *Advances in Neural Information Processing Systems, 12* (MIT Press, Cambridge, 2000), pp. 547–553
12. O Siohan, AE Rosenberg, S Parthasarathy, Speaker identification using minimum classification error training, in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing, Seattle, 12–15 May 1998*, pp. 109–112
13. Y Suh, M Ji, H Kim, Probabilistic class histogram equalization for robust speech recognition. IEEE Signal Proc. Let. **14**(4), 287–290 (2007)
14. B-H Juang, S Katagiri, Discriminative learning for minimum error classification. IEEE Trans. Signal Process. **40**(12), 3043–3054 (1992)
15. Y Suh, H Kim, Multiple acoustic model-based discriminative likelihood ratio weighting for voice activity detection. IEEE Signal Proc. Lett. **19**(8), 507–510 (2012)
16. B-H Juang, W Chou, C-H Lee, Minimum classification error rate methods for speech recognition. IEEE Trans. Speech Audio Process. **5**(3), 257–265 (1997)
17. WM Fisher, GR Doddington, KM Goudie-Marchall, The DARPA speech recognition research database: specifications and status, in *Proc. DARPA Workshop on Speech Recognition*, 1986, pp. 93–99
18. H-G Hirsch, D Pearce, The Aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions, in *Proc. Int. Conf. Spoken Language Processing, Beijing, 16–20 Oct 2000*, pp. 16–20
19. C-C Chang, C-J Lin, LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. **2**(27), 1–27:27 (2011). Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm

doi:10.1186/1687-6180-2014-126
**Cite this article as:** Suh and Kim: **Discriminative likelihood score weighting based on acoustic-phonetic classification for speaker identification.** *EURASIP Journal on Advances in Signal Processing* 2014 **2014**:126.