ORIENTAL JOURNAL OF
COMPUTER SCIENCE & TECHNOLOGY
An International Open Free Access, Peer Reviewed Research Journal
Published By: **Oriental Scientific Publishing Co., India.**
www.computerscijournal.org

ISSN: 0974-6471
April 2016,
Vol. 9, No. (1):
Pgs. 36-40

A Review on Text-Independent Speaker Verification Techniques in Realistic World

RENU SINGH[1]*, ARVIND KUMAR SINGH[1] and UTPAL BHATTACHARJEE[2]

[1]Department of Electrical Engineering, North Eastern Regional
Institute of Science and Technology, Itanagar, Arunachal Pradesh 791109, India.
[2]Department of Computer Engineering, Rajiv Gandhi University, Rono Hills, Doimukh,
Arunachal Pradesh, 791112, India.
*Corresponding author E-mail; renumona08@gmail.com

http://dx.doi.org/10.13005/ojcst/901.07

(Received: January 20, 2016; Accepted: February 25, 2016)

ABSTRACT

This paper presents a review of various speaker verification approaches in realistic world, and explore a combinational approach between Gaussian Mixture Model (GMM) and Support Vector Machine (SVM) as well as Gaussian Mixture Model (GMM) and Universal Background Model (UBM).

Keywords: Speaker Verification, Gaussian Mixture Model, Support Vector Machine, Universal Background Model, Likelihood-ratio.

INTRODUCTION

Speaker verification (SV) is the task of validating the claimed identity of a person from his/her voice. It is binary classifications problem in which we are distinguish between a true speaker and an imposter[1].

The Speaker Verification system works in two phases. The first phase - the Training Phase or the Enrollment Phase approximates the speaker to a Gaussian Mixture Model's parameters. The second phase – the verification phase represents the utilization of the system[2].

Applications of Speaker Verification

There are many applications to speaker verification. The applications cover almost all the areas where it is desirable to secure actions, transactions, or any type of interactions by identifying or authenticating the person making the transaction. We briefly review those various applications.

On-site Applications

On-site applications regroup all the applications where the user needs to be in front of the system to be authenticated. Typical examples are access control to some facilities (car, home,

warehouse), to some objects (locksmith), or to a computer terminal. Currently, ID verification in such context is done by mean of a key, a badge or a password, or personal identification number (PIN).

Remote Applications

Remote applications regroup all the applications where the access to the system is made through a remote terminal, typically a telephone or a computer. The aim is to secure the access to reserved services (telecom network, databases, web sites, etc.) or to authenticate the user making a particular transaction (e-trade, banking transaction, etc.).

Information Structuring

Organizing the information in audio documents is a third type of applications where speaker recognition technology is involved. Typical examples of the applications are the automatic annotation of audio archives, speaker indexing of sound tracks, and speaker change detection for automatic subtitling. The need for such applications comes from the movie industry and from the media related industry recognition is a key technology for audio indexing.

Games

Finally, another application area, rarely explored so far, is games: child toys, video games, and so forth. Indeed, games evolve toward a better interactivity and the use of player profiles to make the game more personal[3].

Techniques for Text –Independent Speaker Verification

GMM

As generative models, Gaussian Mixture Models (GMMs) have become the dominant modelling approach[4]. For text-independent speaker-verification, the Gaussian mixture model (GMM) based on statistical theory is the most widely used method. This is a powerful method, in which a likelihood-ratio detector is constructed according to the framework shown in Fig.2[5]

The principle of GMM is to abstract a random process from the speech, then to establish a probability model for each speaker. It is relatively independent between the various probability models. Assuming the variable M in the M-order GMM probability density function is the number of Gaussian probability density functions. And set X as the feature vector from feature extraction block of the speech[3].

$$P\left(X/_{\lambda}\right) = \sum_{i=1}^{M} \omega_{i} D_{i(X)}$$

SVM

Support Vector Machine is a powerful machine learning method, invented by Vapnik Essentially, SVM is a binary and classifier to search for the optimal decision boundary in two classes of data, which is based on the principle of structural risk minimization. Experimental results indicate that SVM can achieve a generalization performance that is greater than or equal to other classifiers, while requiring significantly less training data to achieve such an outcome. The principle of SVM relies on a linear separation in a high dimension feature space where the data have been previously mapped, in order to take into account the eventual non- linearities of the problem. An SVM classifier has the general form:

$$f(x)=\sum_{i=1}^{l} y_i \ \alpha_i \ \kappa(x_i, x) + b$$

Where $x_i \in R^n, i = 1,2,...,l$ is the training data .Each point of x_i belongs to one of the two c lasses identified by the label $y_i \in \{-1,1\}$. The coefficients á$_i$ and b are the solution of a quadratic programming problem. αi is non –zero for support vectors (SV) and is zero otherwise[4].

UBM

A Universal Background Model (UBM) is a model used in a biometric verification system to represent general, person-independent feature characteristics to be compared against a model of person-specific feature characteristics when making an accept or reject decision. For example, in a speaker verification system, the UBM is a speaker-independent Gaussian Mixture Model (GMM) trained with speech samples from a large set of speakers to represent general speech characteristics. Using a speaker-specific GMM trained with speech samples

from a particular enrolled speaker, a likelihood-ratio test for an unknown speech sample can be formed between the match score of the speaker-specific model and the UBM. The UBM may also be used while training the speaker[7]

The universal background model (UBM) is an effective frame-work that has found great success in speaker recognition. Conceptually, it is a large mixture of Gaussians that covers all speech, and in the context of speaker recognition, it is adapted to each speaker using a *maximum a* posteriori (MAP) scheme.

The UBM so far has received little attention from the automatic speech recognition (ASR) field. In this paper, we make a first attempt to apply UBM to acoustic modelling in ASR, and demonstrate substantial improvements at the maximum likelihood (ML) level. The basic idea is to adapt the UBM to each context-dependent phone rather than to each speaker. The context-dependent phones are not an unstructured collection of phones but are related via a tree structure; hence we devise a set of smoothing methods that can utilize this structure. Training is done through multiple iterations of EM rather than just one as in . Furthermore, in our best-performing system, a separate semi-tied co-variance (STC) transform is applied for each Gaussian in the UBM. One challenge in UBM-based speech recognition is the very large size of the resulting models. Entropy based pruning method similar to is used to address the problem. The results in this paper should be considered preliminary, as we have not had time to explore the many design choices involved.

The UBM is a Gaussian Mixture Model (GMM) whose parameters consist of *K* weights

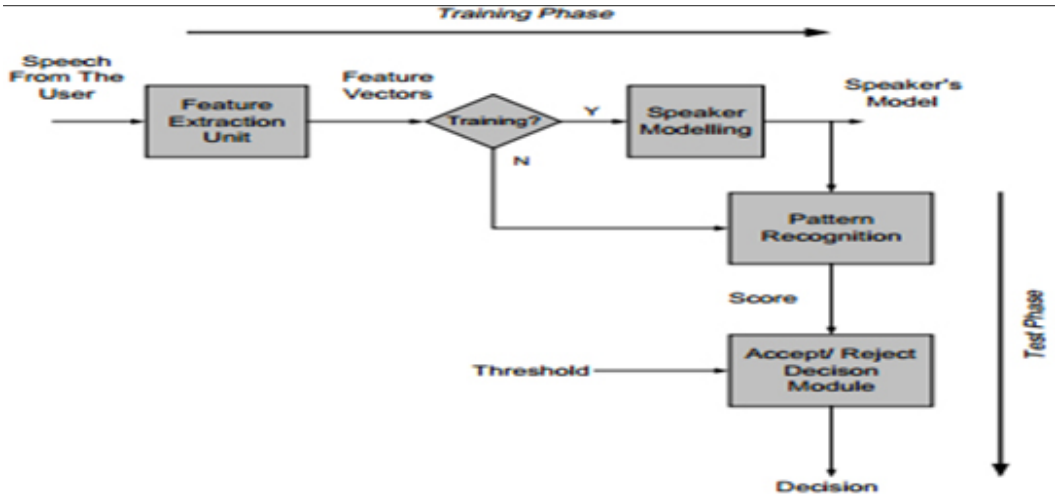

Fig.1: Bird's eye view of the Speaker Verification process

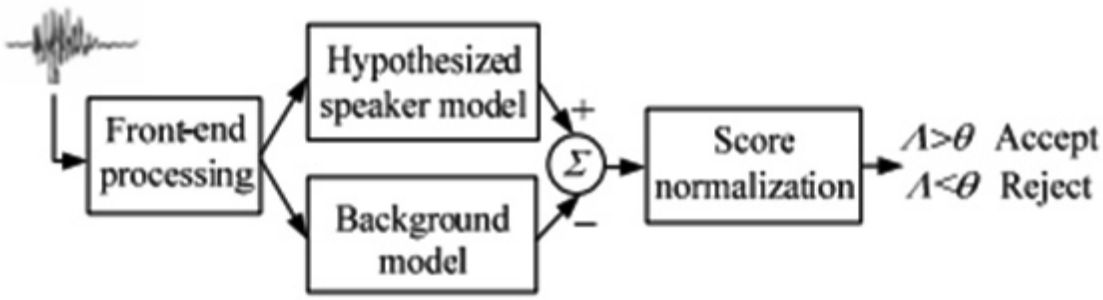

Fig. 2: Structure of a GMM likelihood-ratio text-independent speaker verification system

w_k, means μ_k and (diagonal) variances Σ_k, which in a speaker identification context are MAP adapted to each speaker's data to create a GMM for that speaker. In the speech recognition context, let us consider that the speech is already split up into many speech classes $j=1\cdots J$ based on the tree-clustered context dependent phones, and our reference transcriptions have been Viterbi-aligned given some previously existing models, so that we have (zero-one) phone posteriors, $\gamma_j(t)$. so we can treat the set of frames for which $\gamma_j(t)=1$. for some *j* as we would the data from a particular speaker.

It is helpful to consider a single iteration of standard EM up-date starting from the UBM[8]

GMM/SVM

A GMM-SVM system is a combination of a GMM-UBM and SVM systems. The GMM-UBM system serves as a means of feature extraction for the attached system .The SVM classifier is used to model the target speaker characteristics and to score the test utterances. The framework of the GMM-SVM system is shown in Fig.3[5]

GMM/UBM

Gaussian Mixture Model-Universal Background Model (GMM-UBM) is a standard reference classifier in speaker verification,[Comparing Maximum] The GMM-UBM system is the current state-of-the-art approach for text-independent speaker verification. The advantage of the approach is that both target speaker model and impostor model (UBM) have generalization ability to handle "unseen" acoustic patterns. However, since GMM-UBM uses a common anti-model namely UBM, for all target speakers, it tends to be weak in rejecting impostors' voices that are similar to the target speaker' voice. To overcome this limitation, we propose a discriminative feedback adaptation (DFA) framework that reinforces the discriminability between the target speaker model and the anti model, while preserves the generalization ability of the GMM-UBM approach. This is done by adapting

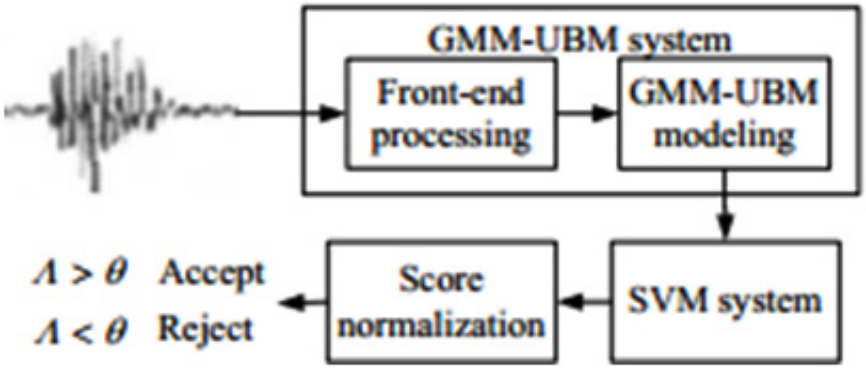

Fig. 3: Structure of GMM-SVM system

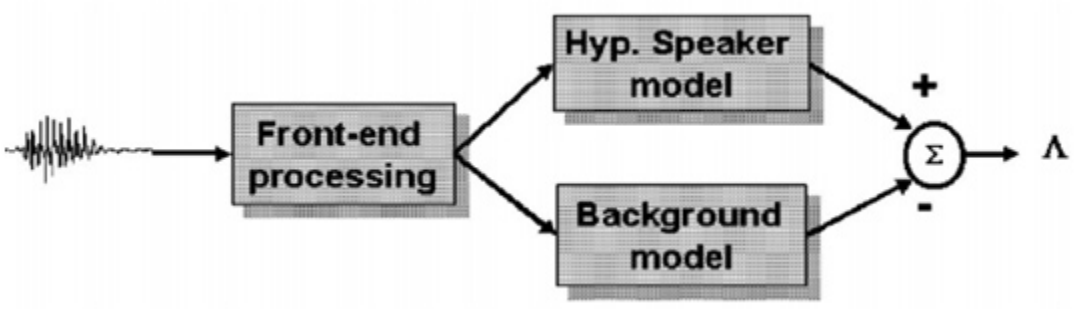

Fig. 4: Likelihood ratio-based speaker verification system

the UBM to a target-speaker dependent anti-model based on a minimum verification squared error criterion, rather than estimating from scratch by applying the conventional discriminative training schemes[8] Since NIST 1996 Speaker Recognition Evaluations (SRE), the Gaussian Mixture Model-Universal Background Model (UBM) speaker verification system has become dominating system because of its excellent performance in text-independent speaker recognition tasks .A GMM which used in speaker recognition applications represents multivariate probabilistic model makes it suitable for unconstrained text-independent applications.

The UBM is generally a GMM trained from a quite large pool of speech database to represent the speaker independent distribution of features including various speakers, category of language, handset types, ambient environment, channel variability, and so on. In the GMM-UBM system, we adapt the parameters of the UBM using the speaker's enrolment speech and Maximum A Posteriori (MAP)) estimation to derive the corresponding speaker model. During testing an unknown utterance, the system calculates the likelihood-ratio of producing the unknown utterance between the enrolment model and UBM Fig. 4 shows the framework of likelihood ratio-based speaker verification system[9]

CONCLUSION

In this paper, techniques for speaker verification like GMM ,SVM and UBM were discussed. Various hybrid speaker verification techniques like GMM/SVM and GMM/UBM were also discussed. Speaker can be identifying efficiently using the technique of feature extraction discussed. These techniques are able to authenticate the particular speaker, based on the individual information that is included in the voice signal.

REFERENCES

1. Sourjya Sarkar and K.Sreenivasa Rao,"Speaker Veriucatino in Noisy environment Using GMM Supervectors. IEEE, 2013.
2. S. Bhattacharyya , T. Srikanthan , and Pramod Krishnamurthy ," ideal gmm parametes & posterior log likelihood for speaker verification", IEEE,2001.
3. Tejal Chauhan, Hemant Soni, Sameena Zafar," A Review of Automatic Speaker Recognition System", *International Journal of Soft Computing and Engineering (IJSCE)* ISSN: 2231-2307, **3**(4):(2013)
4. Minghui Liu, Yanlu Xie, Zhiqiang Yao, Beiqian Dai," A New Hybrid GMM/SVM for Speaker Verification",IEEE,2006.
5. Zhao Jian , Dong Yuan , Zhao Xianyu,Yang Hao , LU Liang , WANG Haila ," Advances in SVM-Based System Using GMM Super Vectors for Text-Independent Speaker Verification", *tsinghua science and echnology* **13**(4); (2008).
6. A.K.Sarkar,S.P.RathandS.Umesh."VocalTractLengthNormalizationFactorBasedSpeaker-ClusterUBMforSpeakerVeriûcation", IEEE. .2010
7. Encyclopedia of Biometrics By Springer.
8. Daniel Povey Stephen M. Chu, and Balakrishnan Varadarajan," Universal background model based speech recognition 1",
9. Yi-Hsiang Chao, and Hsin-MinWang." discriminative feedback adaptation for gmmubm speaker verification", IEEE 2008.