IJREAT International Journal of Research in Engineering & Advanced Technology, Volume 1, Issue 6, Dec-Jan, 2014
ISSN: 2320 - 8791
www.ijreat.org

Efficient Speaker Verification System with Spoofing Attack

Mrs.Malathi Sharavanan[1], Mrs.M.Priya[2], C.P.Sangeetha[3]

[1]**Associate Professor, Prathyusha Institute Of Technology and Management, Thiruvallur, TamilNadu, India**

[2]**Associate Professor, Prathyusha Institute Of Technology and Management, Thiruvallur, TamilNadu, India**

[3]**M.E, Prathyusha Institute Of Technology and Management, Thiruvallur, TamilNadu, India**

ABSTRACT

In the existing system the state-of-the-art speaker verification system cannot distinguish the natural speech and converted speech. But in this project an attempt will be made to build an efficient speaker verification system to detect natural speech and synthesized speech. In this paper, we present new results evaluating the current state-of-the-art speaker verification system, Gaussian mixture model supervector with joint factor analysis (GMM-ZJFA) system, against spoofing attacks. In this the spoofing attacks are simulated by Gaussian mixture model based voice conversion technique. The results show that GMM-based conversion method which increases the false acceptance rate (FAR) from 3.24% to 17.33%. This suggests that GMM-JFA system is less vulnerable towards GMM-based conversion. The software used over here is matlab.

Keywords: *GMM, GMM-JFA, MFCC, EM Algorithm*

1. INTRODUCTION:

The objective of speaker verification is to make a binary decision to accept or reject a claim of identity based on the user's speech samples [1, 2]. In practice, speaker verification system can be used to verify a speaker's identity for control access to services such as telephone banking [1], voice mail [1, 2], and so on. On the other hand, the task of voice conversion is to modify one speaker's voice (source speaker) so that is sounds as if it has been uttered by another speaker (target speaker) [3, 4]. This paper studies voice conversion and speaker verification in an attack and defence experiment. We assume that the converted speech samples are obtained in telephony conversations where voice conversion is performed in one of the speakers. There have been multiple studies in voice conversion vs speaker verification. For example, speaker verification system against imposter's speech which are generated from HMMbased speech synthesis system [5] or adapted speech synthesis system with small size adaptation data [6], and voice conversion techniques [7, 8, 9]. These studies are all carried out on high quality speech. In telephone applications, such as telephone banking, the speaker verification system has to deal with telephone speech which is low quality and affected by channel variability. In our previous research, we conducted spoofing attack study using telephone speech [10] with five speaker verification systems: GMM-UBM (Gaussian mixture model with universal background model) system [11], VQ-UBM (vector quantized codebook with universal background model) system [12],GLDS-SVM (generalized linear discriminant sequence kernel support vector machine) system [13], GMM-SVM system [14], and GMM-JFA (Gaussian mixture model supervector with joint factor analysis) system [10]. Our previous results suggested that the GMM-JFA system, which is the current state-of-theart speaker verification system, obtained the best performance against spoofing attack simulated by a simple voice conversion technique. In this study, we continue the study of vulnerability of the current state-of-the-art speaker verification system by examining the performance of GMM-JFA system against spoofing attack. We will use two different voice conversion methods, GMM-based conversion method and unit selection based method, to simulate the spoofing attack. In the previous study, we used GMM-based voice conversion method to simulate spoofing attack. In the GMM-based voice conversion method, the transformation parameters is derived from Gaussian mixture models (GMM), and then the linear transformation is applied to the spectrum parameters of the source speech frames. Although GMM-based voice conversion techniques can generate speech with acceptable quality, the transformation is not perfect, and

hence may not transform the source feature vector to the target feature vector space. That is the reason why informal listening tests show that the converted speech may not resemble the target speaker, and the converted speech may sound like another

speaker who is neither source speaker nor target speaker. On the other hand, for telephone speech conversion, GMM-based conversion method can be viewed as a joint shift of channel factor and speaker characteristic. While in the unit-selection based conversion method, target speaker's feature vectors are directly used to synthesize the converted speech, without changing the original spectral envelop. If we consider the resulting speech a collection of speech frame regardless of the continuity and prosody of speech flow, unit selection should produce speech that sounds closer to the target speaker. Although informal listening tests show that converted speech from GMM-based conversion method is much smoother than that from unit-selection based method, current speaker verification systems, such as GMM-JFA system, are not considering the naturalness of the speech. One can expect that GMM-JFA speaker verification system is more vulnerable to unit selection based voice conversion. This paper is organized as follows. In sections 2, we will describe voice conversion techniques based on Gaussian mixture model and unit selection; the speaker verification system used in this study will be presented in section 3. In section 4, the experimental setups and results are presented and discussed. We conclude this article in section 5.

2. VOICE CONVERSION METHOD

We study a voice conversion technique in simulating the spoofing attacks. The voice conversion technique used is GMM-based conversion, which trains a mapping function between source and target, and requires a parallel corpus for training.

2.1. GMM-BASED VOICE CONVERSION

The most popular voice conversion method is based on joint density Gaussian mixture model (GMM), which is originally proposed in [4]. We apply this method to simulate spoofing attack in this study and describe as follows**.**

The training data of source speech contains N frames spectral vectors $X = [x_1^T, x_2^T, \ldots xN_T]^T$, and the training data of target speech contains M frames spectral vectors $Y = [y_1^T, y_2^T, \ldots, y_m^T, \ldots, y_M^T]^T$. For parallel data, we can use dynamic time warping algorithm to align source feature vectors to their counterparts in the target; for non-parallel data, nonparallel frame alignment method used in [15, 10] can be adopted to obtain feature vector pairs $Z = [z_1^T, z_2^T, \ldots, z_t^T, \ldots, z_T^T]^T$, where $z_t^T = [x_n^T, y_m^T]^T$.

The joint probability density of X and Y is modeled by GMM as in (1):

$$P(X,Y)=P(Z) = \sum_{i=1}^{L} w_l^{(z)} N(z \mid \mu_l^{(z)}, \sum_l^{(z)}) \quad (1)$$

where $\mu_1^{(z)} = \begin{pmatrix} \mu_l^{(x)} \\ \mu_l^{(y)} \end{pmatrix}$ and $\sum_l^{(x)} = \begin{bmatrix} \sum_l^{(xx)} & \sum_l^{(xy)} \\ \sum_l^{(yx)} & \sum_l^{(yy)} \end{bmatrix}$

these are the mean vector and covariance matrix of the multivariate Gaussian density $N(z \mid \mu_l^{(z)}, \sum_l^{(z)})$, respectively. Given the component l, $w_l^{(z)}$ is the prior probabilities of z, and $\sum_{l=1}^{L} w_l^{(z)} = 1$

In the training phase, the GMM parameters $\lambda^{(z)} = \{w_l^{(z)}, \mu_l^{(z)}, \sum_l^{(z)} \mid l = 1,2,\ldots,L\}$ were estimated using the expectation maximization (EM) algorithm in maximum likelihood sense.

In the conversion phase, a source speech feature vector x, the joint density model is adopted to formulate transformation function and hence to choose the target speakers feature vector $\hat{y} = F(x)$, the transformation function F(.) is given as follows:

$$F(x) = E(y \mid x)$$

$$= \sum_{l=1}^{L} P_I(x) \left(\mu_l^{(y)} + \sum_l^{(yx)} \left(\sum_l^{(xx)}\right)^{-1} \left(x - \mu_l^{(x)}\right)\right),$$

$$P_1(X) = \frac{w_l N(x \mid \mu_l^x, \sum_l^{xx})}{\sum_{k=1}^{l} w_k N(x \mid \mu_k^x, \sum_k^{xx})}$$

3. Speaker verification system

In this study, we use the GMM-JFA system. As in our previous study [10], we consider five speaker verification systems, GMM-UBM [11, 17], VQ-UBM [12], GLDS-SVM [13], GMM-SVM [14] and GMM-JFA [18], but the GMM-JFA system is the best system against spoofing attacks [10]. The GMM-JFA system, which

adopts joint factor analysis technique for modeling intersession and speaker variability in the GMM supervector space, is a widely recognized high performance system [18]. In this study, the GMM-JFA system uses 512 Gaussian mixtures. The Gaussian mixture model is trained using the HTK toolkit [19]. For score normalization, we use T-norm followed by Z-norm (TZ-norm).. 12 dimensions MFCCs with delta and delta-delta coefficients are computed via 27-channel mel-frequency filterbank. Then RASTA filtering, voice activity detection and utterance level cepstrum mean variance normalization techniques are applied to the extracted MFCCs. So the final feature vectors are 36-dimension MFCCs.

3.PROPOSED SCHEME

In this the voice conversion using GMM and the speaker verification system with an ability to discern between the natural and synthetic speech. But in the existing system the state-of-the-art speaker verification system cannot distinguish the natural speech and converted speech.

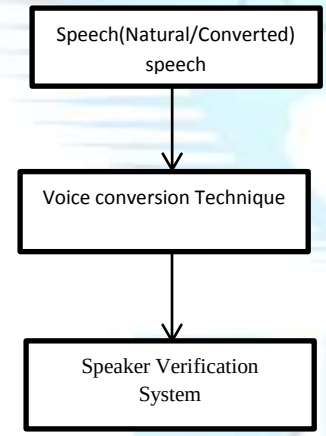
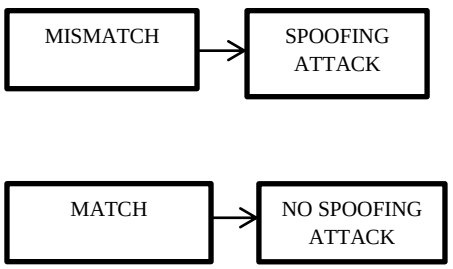


Fig1.Block diagram for detecting

the Natural/converted speech



4. Experimental Setup:

For different sound files the mean, covariance, log likelihood and the weight values are calculated and hence the variations in it are estimated using GMM. The sound variations are noticed as listed below

For had.wav,

MEAN	COVARIANCE	WEIGHT
1.03071754949 1 236e-04	0.0091	0.38837507 9055563
2.66247300175 4462e-04	4.3573e-005	0.61162492 0944436

For ate.wav,

MEAN	COVARIANCE	WEIGHT
4.14898985 4641171e-04	0.0038	0.2183954 10291125
-1.46999008 1295237e-04	3.3091e-005	0.7816045 89708876

5. RESULTS AND DISCUSSION:

The values thus obtained are compared and hence the variations in the values are of noted. A reference speech is compared with the same speech and also a different speech**.** The variation in the reference speech and the same speech is noted. If there results in mismatch means it tends to have spoofing attack. If the speech matches means there is no spoofing

attack thus the natural and the converted speech are of detected.

6. FUTURE WORK

In future work advanced algorithm will be used to improve the accuracy and a new speaker verification system will be implemented for more secure purpose.

REFERENCES

[1] J.P. Campbell Jr, "Speaker recognition: A tutorial," Proceedings of the IEEE, vol. 85, no. 9, pp. 1437–1462, 1997

[2] D.A. Reynolds, "An overview of automatic speaker recognition technology," in Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on. IEEE, 2002, vol. 4, pp. IV–4072.

[3] Y. Stylianou, O. Capp´e, and E. Moulines, "Continuous probabilistic transform for voice conversion," IEEE Trans. on Speech and Audio Processing, vol. 6, no. 2, pp. 131– 142, March 1998.

[4] A. Kain and M.W. Macon, "Spectral voice conversion for text-to-speech synthesis," in Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on. IEEE, 1998, vol. 1, pp. 285–288.

[5] T. Satoh, T. Masuko, T. Kobayashi, and K. Tokuda, "A robust speaker verification system against imposture using a HMM-based speech synthesis system," in Proc. 7th European Conference on Speech Communication and Technology (Eurospeech 2001), Aalborg, Denmark, September 2001, pp. 759–762.

[6] P. DeLeon, M. Pucher, and J. Yamagishi, "Evaluation of the vulnerability of speaker verification to synthetic speech," in Odyssey 2010: The Speaker and Language Recognition Workshop, Brno, Czech Republic, June 2010, pp. 151–158 (paper 28).

[7] J.F. Bonastre, D. Matrouf, and C. Fredouille, "Artificial

impostor voice transformation effects on false acceptance rates," in Eighth Annual Conference of the International Speech Communication Association, 2007.