ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

Detection of Voice Disguise by Various Disguising Factors

Abin Mathew George, Eva George

PG Scholar, Department of Communication Engineering, CAARMEL Engineering College, Perunad, Kerala, India

Assistant Professor, Department of E&C, CAARMEL Engineering College, Perunad, Kerala, India

ABSTRACT:Voice disguise has shown an increasing tendency for criminal activities like kidnapping, threatening calls. The most common way of disguising the voice adopted by the criminals is changing the pitch, so that speaker recognition will be difficult. In this paper, we focus on improving the detection performance of disguised voice, which is disguised by using software tools like Audacity, Cool Edit and Praat. First and foremost, we will be extracting mel-frequency cepstral coefficient as acoustic feature. Then we will be using Probabilistic Neural Network (PNN) classifier to detect whether the voice is disguised or not. Here we will be showing the output of detection performance of Support Vector Machine (SVM) classifier for various disguising factors, which is the existing system and also the output of PNN classifier for the disguising factor of -8 as of now.

KEYWORDS: electronic disguised voice; PNN; MFCC

I. INTRODUCTION

Voice disguise is being used for illegal purpose like kidnapping, fraud calls, emergency police calls and threatening calls. Voice disguise is a deliberate action to conceal one person's identity. This also affects the performance of speaker recognition system as it can be affected either by making variations in the communication channel or by making variations in a person's voice. There are two variations in communication channel: handset variations and environment variations. These variations were deeply studied by various researchers and they have proposed several normalizing techniques in order to counter these variations.

Variations in person's voice can be done in two ways: Deliberately and Non-deliberately. In deliberate variation of human voice, the speaker attempts to imitate other person's voice so as to falsify the listener. Non-deliberate variation of human voice comes due to emotion or physical condition such as cold, sore throat. This variations in human voice can be further divided into electronic and non-electronic disguised voice. Electronic voice disguise is disguising the voice electronically by using software tools like Audacity, Cool Edit and Praat, which is available in the internet. Non-electronic voice disguise is disguising the voice mechanically by placing an object between the mouth, by pinching nostrils etc. The feature which is widely used for identification of disguised voice and speaker recognition is MFCC. In sound processing, the mel-frequency cepstrum (MFC) is a representation of the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear mel scale of frequency. MFC is a collection of MFCC's. The scale of pitches judged by listeners to be equal in distance one from another is mel scale.

In previous work, we used SVM classifier to classify whether voice is disguised or not. A support vector machine constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification. Here, a good separation is achieved by the hyperplane that has the largest distance to the nearest training data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier.

In this paper, we proposed a new classifier called PNN classifier to classify whether the voice is disguised or not. A probabilistic neural network (PNN) is a feed forward network and predominantly a classifier to map any input pattern to a number of classifications.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

II. RELATED WORK

In [1] authors proposed a work which is divided into three parts: The proposed work is divided in three parts: the first one is a classification of the different options available for changing one's voice, the second one presents a review of the different techniques in the literature and the third one describes the main indicators proposed in the literature to distinguish a disguised voice from the original voice, and proposes some perspectives based on disordered and emotional speech. In [2] authors investigated the effect of intentional voice modifications on a state-of-the-art speaker recognition system and also showed that machine outperforms human in case of detecting whether the voice is disguised or not. In [3] authors focused on the classification of the speaker. The effect of 10 kinds of disguise voices on the performance of a Forensic Automatic Speaker Recognition System (FASRS) was studied in [4]. In [5] authors have presented the results of an ongoing study on the effects of common types of voice disguise, including increased voice pitch (even falsetto speech), lowered voice pitch and pinching the nose while speaking, on forensic speaker recognition (FSR) techniques. In [6] Blind detection of electronic disguised voice was introduced. Here, statistical moments of Mel frequency cepstrum coefficients (MFCC) are extracted as acoustic features of speech signals. Then an approach for detection of disguised voice based on the extracted features and Support Vector Machine (SVM) classifiers is proposed. The extensive experiments demonstrate that detection rates higher than 95 percent can be achieved, indicating that detection performance of the proposed approach is good. Detection performance of the proposed approach is demonstrated to be excellent, even when disguise methods used in training stage are different from the ones used in testing stage. However, when the disguising factor is +4 semitones, the detection performance is not good enough. In [7] authors investigated the principle of electronic voice transformation, and propose a blind detection approach using MFCC as the acoustic features and VQ-SVM (Vector Quantization-Support Vector Machine) as the classification method. By extensive experiments, it is demonstrated to have classification accuracy higher than 98 percent in most cases, indicating that the proposed approach has good performance and can be used in forensic applications.

III. PROBABILISTIC NEURAL NETWORK

A. *Introduction:*

It is closely related to Parzen window pdf estimator. It consists of several sub-networks in which each is a parzen window pdf estimator for each classes of its own. It consists of four layers. The first layer consist of input node which consists of a set of measurements, the second layer consists of Gaussian functions formed using the given data points as centers, the third layer performs the summation operation for the outputs of second layer of each class, the fourth layer selects the largest value from the outputs of the third layer. Then associated label of the class is determined.

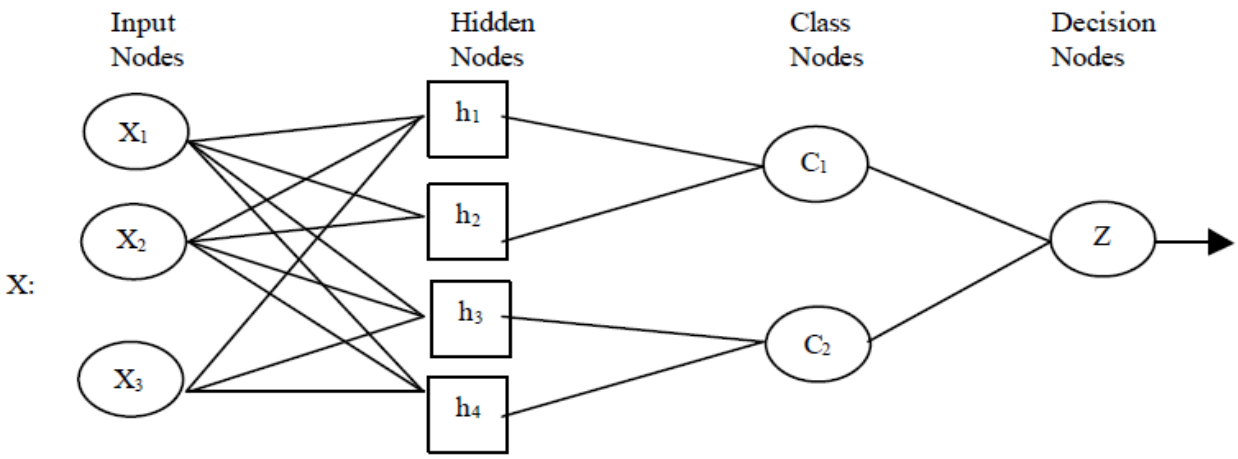

Fig.1. PNN Architecture

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

B. *Application of PNN:*

These are some of the applications of pnn shown below:

- It plays an important role in pattern recognition either in human face or speech recognition.
- It is used in classification of brain tissues in multiple sclerosis.
- It is used in classification of soil textures.
- It is used in classification of image patterns.

IV. MFCC EXTRACTION

The figure below shows the extraction of MFCC.

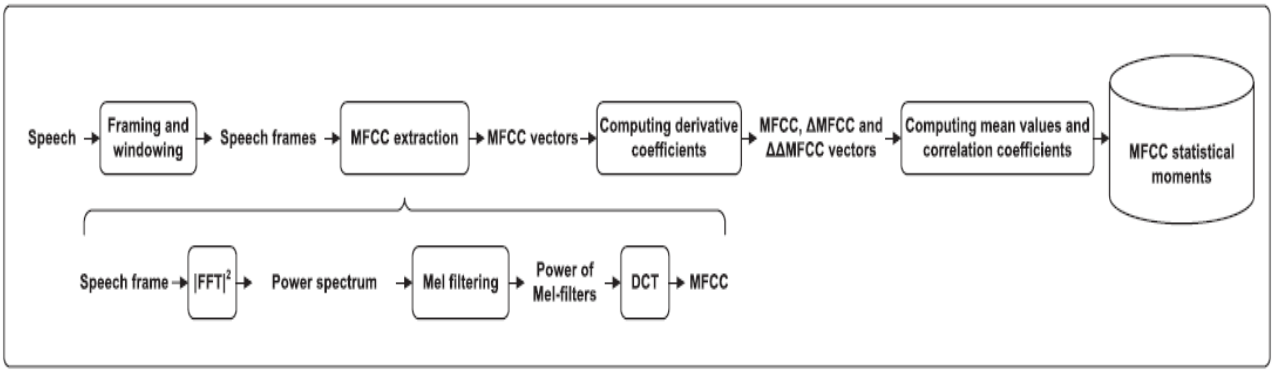

Fig.2. Extraction of MFCC

The following are the steps for the extraction of MFCC:

Step 1: Pre-emphasis

It is a technique used to enhance high frequencies of the signal in speech processing. Speaker information is contained more in the higher frequencies than in lower frequencies, so it increases the energy of signal at higher frequencies.

Step 2: Framing

Since speech is a continuous time varying signal, it is necessary to frame the signal.

Step 3: Windowing

The speech frames and window is being multiplied so as to minimize the spectral distortion both at the start and at the end of each frame. In this paper, Hamming window is used to obtain windowed frames. The equation for hamming window is shown in eq. (1). Here Z is the number of points in frame.

$$H(n) = 0.54 - 0.46cos\frac{2\pi n}{Z-1},\ \text{n} = 0,1,\dots,\text{Z}-1 \qquad \text{eq.(1)}$$

Step 4: Fast fourier transform

It is used to convert frames which consists of N samples from time domain to frequency domain.

Step 5: Mel-Frequency Warping

Here, a set of 20 triangular band pass filters are multiplied with magnitude frequency response, so as to get smooth magnitude spectrum. For a given frequency f the formula to calculate mel-frequency f_{mel} warping is given in eq.(2)

$$f_{Mel} = 1127\ln(1 + \frac{f}{700}) \qquad \text{eq.(2)}$$

Step 6: Discrete cosine transform

It is a compression step, which removes higher coefficients and keeps first coefficients.

Step 7: Calculation of mean and correlation coefficient

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

Consider a speech signal with N frames, assume V_{ij} to be the j^{th} component of the MFCC vector of the i^{th} frame and V_j to be the set of all the j^{th} components.

$$V_j = \{v_{1j}, v_{2j}, v_{3j}, \dots, v_{Nj}\}, \mathrm{j} = 1,2, \dots, \mathrm{L} \quad \text{eq.(3)}$$

where L is the dimension of MFCC vectors based on each frame.

The mean value of the speech signal can be calculated

$$E_j = E(V_j), \ \mathrm{j} = 1,2, \dots, \mathrm{L} \quad \text{eq.(4)}$$

The correlation coefficient of the speech signal can be calculated as shown in eq.(5).

$$CR_{jj'} = \frac{cov(V_j, V_{j'})}{\sqrt{VAR(V_j)}\sqrt{VAR(V_{j'})}}, 1 \leq \mathrm{j} < j' \leq L \quad \text{eq.(5)}$$

V. ALGORITHM FOR IDENTIFICATION OF DISGUISED VOICE

The proposed algorithm is based on MFCC and PNN classifier .Here audio wav file is given as an input, which can be either an original voice or disguised voice. In that way ,three databases are created, of which each contains 10 disguised voices and 10 original voices. Here disguising factor of -8 is used to disguise the voice. The features are extracted from each of these voices and is given as input to the PNN classifier in the form of vector and then classified into various classes. Then from the classes mentioned, PNN classifier will decide which class has maximum value and that particular value will be assigned to that input voice .These things will be done in training phase. An adaptive filter is used to remove the noise.

In testing phase, the tested voice will be passing through the PNN classifier and will be compared with all the voices in the database, whose features has already been extracted in training phase and assigned a value to each voice. It will decide which class should be assigned to the testing voice depending on the maximum value and we will come to know whether the tested voice is original or not. I have defined two classes named as 1 and 2 for original and disguised voice respectively. The features extracted are correlation MFCC, mean of MFCC, del-MFCC, mean of del MFCC, del-del MFCC and its mean. The figure for the identification of voice disguise is given in Fig.3.

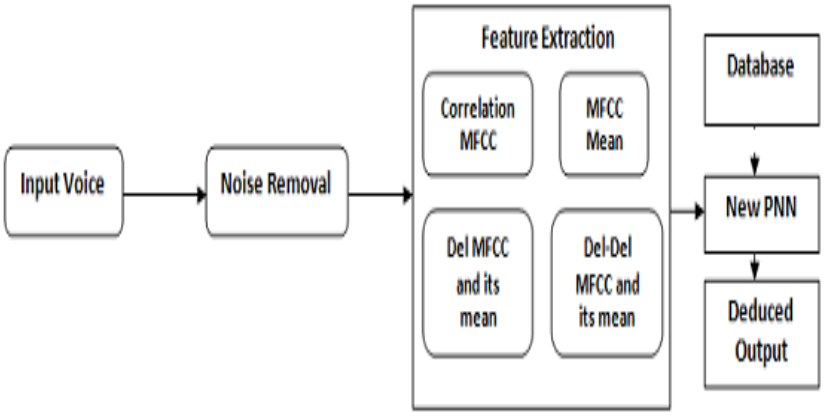

Fig.3. System block model

VI. SIMULATION RESULTS

The output of the existing system is shown in Fig.4. Here semitone is used as the disguising factor. In the existing system, the voices were disguised by disguising factor from -8 to +8. The procedure here was to create a database and

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

extract the features from the voices contained in the database and compare it with the features extracted from the testing voice using SVM classifier, so as to decide whether the voice is disguised or not. The voice was disguised using software tools like Audacity, Cool Edit, Praat and Rtisi. Fig.4. shows the detection performance of various tools used ,to detect disguised voice. Fig. 4. shows the detection rate of Rtisi is better than other software tools

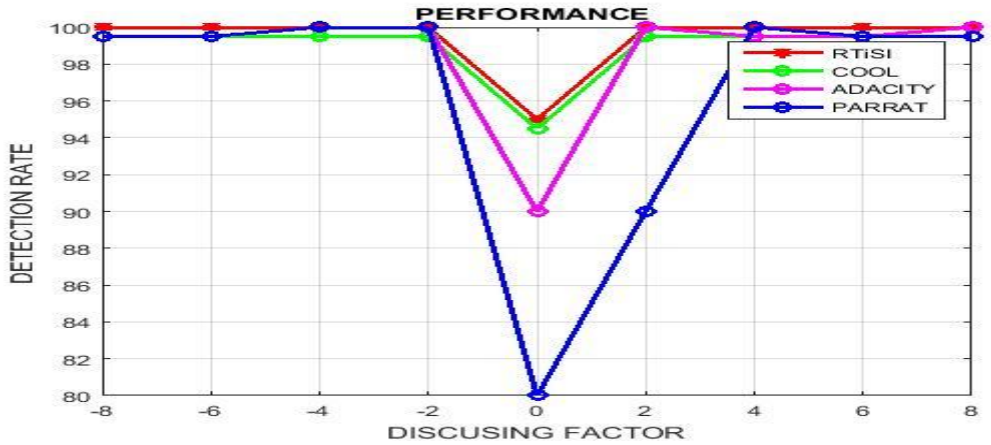

Fig.4. Detection Rate of Voice Disguise

Fig.5. shows the comparison of detection rate of disguised voice of existing and proposed system only for disguising factor -8. It shows that detection rate is high for proposed system than existing system. As shown in the figure, the detection rate of the proposed system is near to 99 percent, which shows PNN is good classifier. In the proposed system, we will be extracting six features from each voice. So in total, we will be obtaining 120 features from each database, as there are 20 voice samples in each database. In 120 features, we will be assigning value 1 for 60 features and value for the remaining using PNN classifier. These features will be compared with the features extracted from testing voice and selects the maximum value and decides whether voice is disguised or not. Here also voice is disguised using Audacity, Praat and Cool edit software tools. The voice is disguised by a factor of -8.

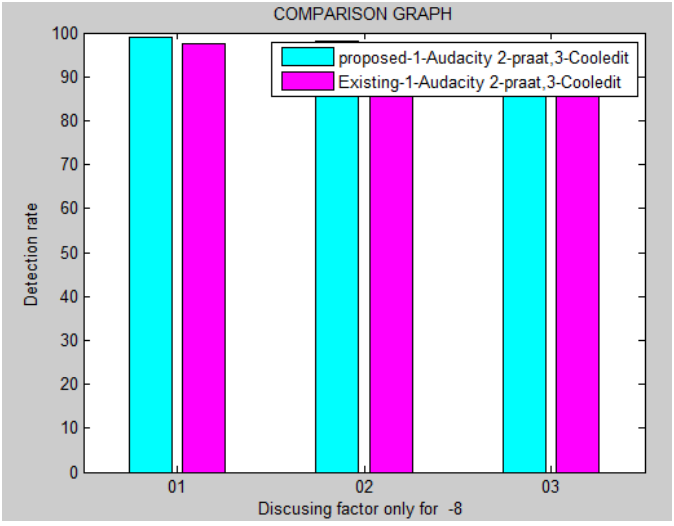

Fig.5. Comparison of Existing and Proposed System for -8 Disguising Factor

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 8, August 2015

VII. CONCLUSION AND FUTURE WORK

The simulation results showed that the proposed classifier with detection rate for disguising factor of -8 than that of the existing classifier. The proposed classifier PNN classifies the input vector more efficiently than that of SVM classifier. As the performance of the proposed algorithm is analyzed only for disguising factor of -8, we have to analyze it using various disguising factors.

REFERENCES

1. P. Perrot, G. Aversano, and G. Chollet, "Voice disguise and automatic detection: Review and perspectives," in Progress in Nonlinear Speech Processing (Lecture Notes in Computer Science). New York, NY, USA: Springer-Verlag, 2007, pp. 101–117.
2. S. S. Kajarekar, H. Bratt, E. Shriberg, and R. de Leon, "A study of intentional voice modifications for evading automatic speaker recognition," in Proc. IEEE Int. Workshop Speaker Lang. Recognit., Jun. 2006, pp. 1–6.
3. R. Rodman, "Speaker recognition of disguised voices: A program for research," in Proc. Consortium Speech Technol. Conjunct. Conf. Speaker Recognit. Man Mach., *Direct. Forensic Appl.*, 1998, pp. 9–22.
4. T. Tan, "The effect of voice disguise on automatic speaker recognition," in Proc. IEEE Int. CISP, vol. 8. Oct. 2010, pp. 3538–3541.
5. H. J. Künzel, J. Gonzalez-Rodriguez, and J. Ortega-García, "Effect of voice disguise on the performance of a forensic automatic speaker recognition system," in Proc. IEEE Int. Workshop Speaker Lang. Recognit., Jun. 2004, pp. 1–4.
6. Y. Wang, Y. Deng, H. Wu, and J. Huang, "Blind detection of electronic voice transformation with natural disguise," in Proc. Int. Workshop Digital Forensics Watermarking, 2012, LNCS 7809, pp. 336–343.
7. H. Wu, Y. Wang, and J. Huang, "Blind detection of electronic disguised voice," in Proc. IEEE ICASSP, vol. 1. Feb. 2013, pp. 3013–3017.

BIOGRAPHY

Abin Mathew George is a M.Tech student, specializing in Communication Engineering, who is presently undergoing his course in Believers Church Caarmel Engineering College, which is affiliated to M.G. University. He received Bachelor of Engineering (B.E.) degree from Cambridge Institute of Technology, Bangalore.

Eva George is an Assistant Professor of Electronics and Communication Department, who is presently teaching in Believers Church Caarmel Engineering College, which is affiliated to M.G. University.