ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

Identification of Voice Disguise for Various Disguising Factors using PNN

Abin Mathew George[1], Eva George[2]

PG Scholar, Dept. of Communication Engineering, Mahatma Gandhi University, Caarmel Engineering College, Pathanamthitta, Kerala, India[1]

Assistant Professor, Dept. of Communication Engineering, Mahatma Gandhi University, Caarmel Engineering College, Pathanamthitta, Kerala, India[2]

ABSTRACT: Voice disguise produces a negative impact on the forensic department, as it is difficult to analyse the voice as well as to identify the speaker or the criminal, who is doing such kind of criminal activity. In this paper, we will be extracting mel-frequency cepstral coefficient as acoustic feature, as it plays a very important role in voice detection and then we will be using probabilistic neural network classifier (PNN) to distinguish disguised voice from original voice. Then we will also show the results of comparison between the detection performance of support vector machine (SVM), which is the existing system. The voices are disguised using +2, +4, -2, -4 disguising factor as we use semitone as the disguising factor.

KEYWORDS: PNN, Electronic voice disguise, MFCC, Disguising factor.

1. INTRODUCTION

Voice disguise is done mostly for criminal activities like kidnapping, police threat calls, accessing important documents. It has imposed a serious challenge to the forensic department in order to identify the speaker. Voice disguise can also be said to conceal one's identity in short. Many options are available to the speaker in order to change its voice like changing the position of lips, keeping an object between the mouth etc. There are two kinds of disguised voices electronic disguised voice and non-electronic disguised voice. Identification of electronic disguising voice has caused a major threat to the society.

Disguising the voice can be done in two ways: Deliberately and Non-Deliberately. Deliberately disguising the voice is a kind of speaker trying to imitate the voice tone of another speaker. Non-deliberately disguising the voice is a kind of disguising the voice due to emotions and physical condition of the body like cold. Electronic disguised voice can be formed by using electronic scrambling devices. Non- electronic disguised voice is formed mechanically i.e. by keeping an object between the mouth, by closing your mouth, by pinching nostrils.

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.

II. SYSTEM MODEL

The speech signal is recorded in the form of .WAV format. The extraction of MFCC is shown in Fig.1. Then framing is done on the speech signal, as it is a continuous time varying signal, so it is required to frame the signal.

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

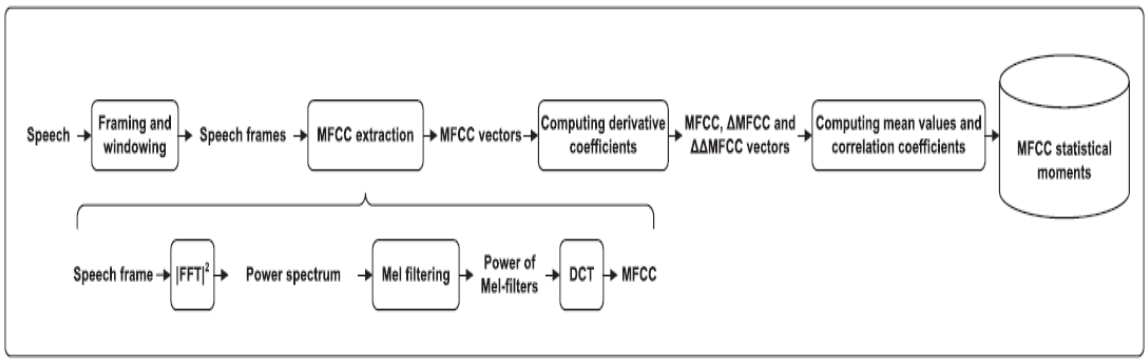

Fig. 1: MFCC Extraction

Then framed signal is windowed using hamming window so as to remove discontinuity at the start and the end of the frame. The equation for Hamming window is given by:

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

where Z is the number of points in the frame.

Then fast fourier transform is done so as to convert the frames which consist of N samples from time domain to frequency domain. After this transformation, by using 20 triangular band pass filters, we will get a smooth magnitude spectrum. The formula to calculate mel-frequency f_{mel} warping For a given frequency f is given by

$$f_{Mel} = 1127\ln\left(1 + \frac{f}{700}\right)$$

Then compression step is done using discrete cosine transform, which removes higher coefficients and keeps first few coefficients. Statistical moments of MFCC are also extracted i.e. mean and correlation coefficient.
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.

$$V_j = \{v_{1j}, v_{2j}, v_{3j}, \dots, v_{Nj}\}, j = 1,2,\dots,L$$

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

The mean value of the speech signal can be calculated

$$E_j = E(V_j), \qquad j = 1,2,\dots,L$$

The correlation coefficient of the speech signal can be calculated.

$$CR_{jj'} = \frac{cov(V_j, V_{j'})}{\sqrt{VAR(V_j)}\sqrt{VAR(V_{j'})}}, 1 \le j < j' \le L$$

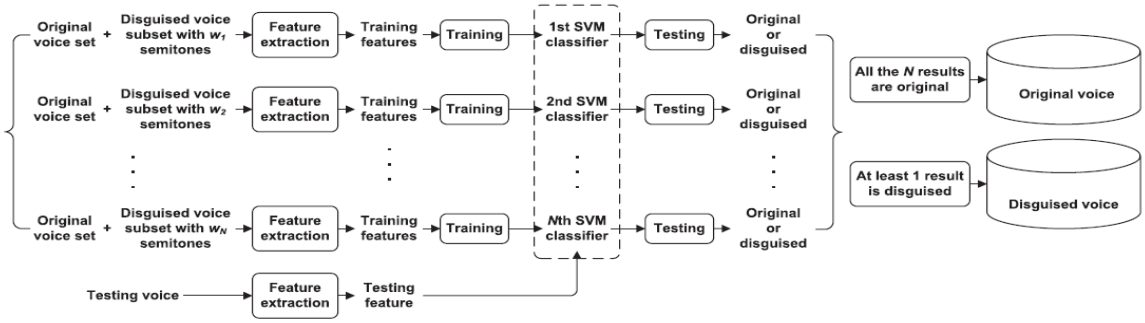

Fig. 2: Existing System Model

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

In the existing system, we used SVM classifier to detect whether the voice is original or disguised. The system model is shown in Fig.2. SVM classifier is used to compare the features extracted from the training database, which consists of disguised and original voices, and the features extracted from testing voice.

III. PROPOSED METHOD

We have proposed a new classifier called PNN classifier, which improves the detection performance of the disguised voice 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. The architecture of PNN is shown in Fig.3.

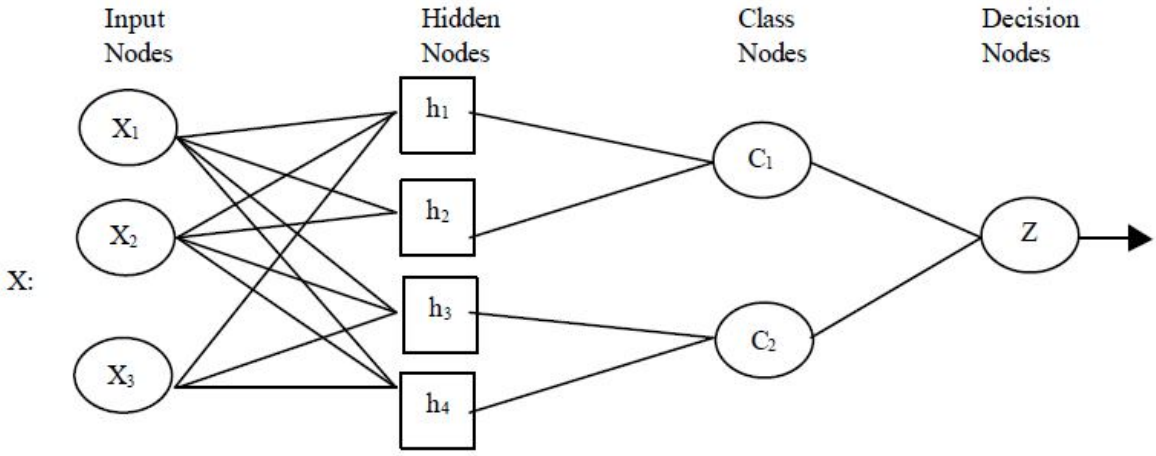

Fig. 3: Architecture of PNN

In similar fashion, the features extracted from each voice will be allocated a class i.e. 1 or 2. Class 1 represents original voice and class 2 represents disguised voice. 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. Correlation MFCC, mean of MFCC, del-MFCC, mean of del MFCC, del-del MFCC and its mean are the features extracted from the voice. The figure for the proposed system model is given in Figure 4.

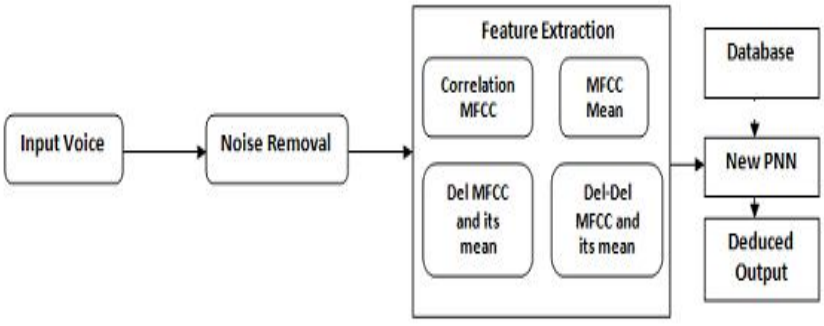

Fig. 4: Proposed System Model

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

IV. SIMULATION RESULTS

The simulation is done using MATLAB 2010. The voice is disguised using software tools like Audacity, Cool Edit, Praat and Rtisi. It is disguised using semitone as the disguising factor. In existing system, the procedure here was to create a database and 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. Figure 5 shows the detection rate of disguising voice for +2, -2, +4 and -4 disguising factor of existing system.

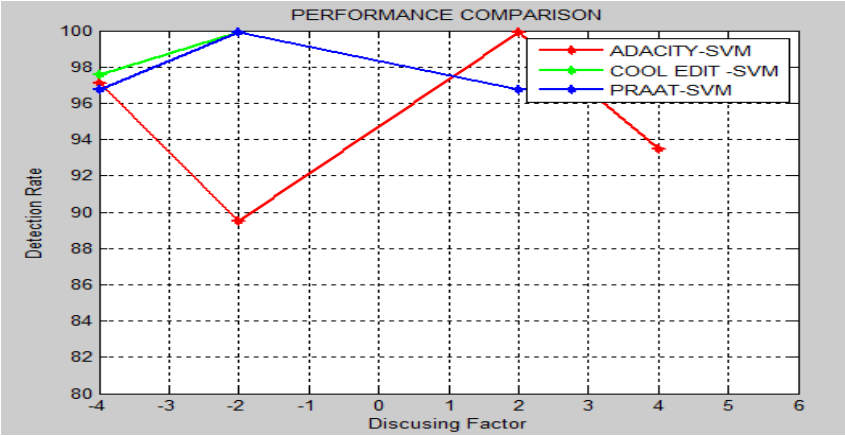

Fig. 5: Existing System Output

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. Fig.6. 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.

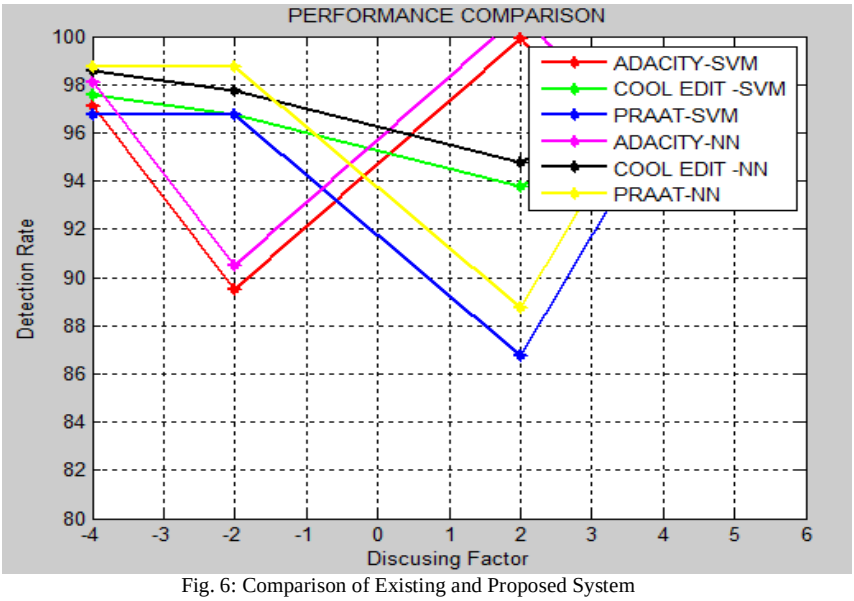

Fig. 6: Comparison of Existing and Proposed System

ISSN (Print) : 2320 – 3765
ISSN (Online): 2278 – 8875

International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 8, August 2015

V. CONCLUSION

In this paper, we have discussed about the extraction procedure of MFCC and also a brief introduction on PNN was given in section III. Detection performance of SVM classifier is also shown in the paper and also the comparison between the detection performance of SVM classifier and PNN classifier has also been carried out in the paper, which showed that detection rate of PNN classifier is greater than that of SVM classifier.

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.