ISSN (Online) 2394-6849

International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)
Vol 5, Issue 1, January 2018

Speaker Identification System Using Watermarking Technology For Spoofing Attack Of Voice Conversion: A Review

[1] Jui Trivedi, [2] Nikunj V. Tahilramani, [3] Ninad Bhatt Chandani, [4] D. Maheshwari
[1] PG student, [2] Assistant Prof. & Head, [3] Professor & Head, [4] Assistant Prof
Department of E&C
Silver Oak College of Engineering & Technology Ahmedabad, Gujarat, India.

Abstract: **- The task of Speaker Identification (SI) technique is to determine which authorized speaker provides an utterance. Voice conversion (VC) method pacts to hide the identity of the speaker. In the Spoofing attack, a manipulated voice is employed as the system input which is voice conversion. It is obligatory to protect speech samples from Spoofing attacks like one mimicry artist can mimic the voice of any person so at that time this system fails to provide security. The security check of watermarking pattern juxtaposes speech samples from the database to identify speaker at the end of the methodology. This paper addresses the review of various techniques of Speaker identification, spoofing attack of Voice Conversion and the techniques of embedding watermarking patterns on various speech samples to avoid the false identification of a subject in Speaker Identification System.**

Key words: **- Speaker Identification (SI); Voice Conversion(VC); Watermarking; Spoofing Attack.**

I. INTRODUCTION

In this technological epoch, security of any information is necessary. It is very onerous to identify which person is uttering from group of people. At several large organizations security system deteriorate due to speech interaction. An unrevealed speaker may endeavor to peer with any one of speaker which exists in database. The speaker identification is the method to figure out which person is speaking at that time. Many challenges affects openly or tortuously to the speaker identification system. Various Spoofing attacks that effect the system are impersonation, replay, speech synthesis and voice conversion. In Spoofing attack, manipulated voice is employed as the system input which is voice conversion[1].

Human mimicry is an example of impersonation. Replay is a speech played repeatedly. Speech synthesis is a method of effectuating expound language by machine on roots of written input. In Voice conversion, alteration of a person's original voice is software inaugurated. It is very difficult to distinguish natural and synthesized speech. The identification system is based on security check of watermarking pattern which is implanted on signal at transmitter side. That gives more desirable identification culmination. This paper is organized as succeeding: Section II outline speaker identification in speech. Watermarking in speech is discussed in section III. Section IV consists application of speech watermarking in spoofing attack.

II. SPEAKER IDENTIFICATION IN SPEECH

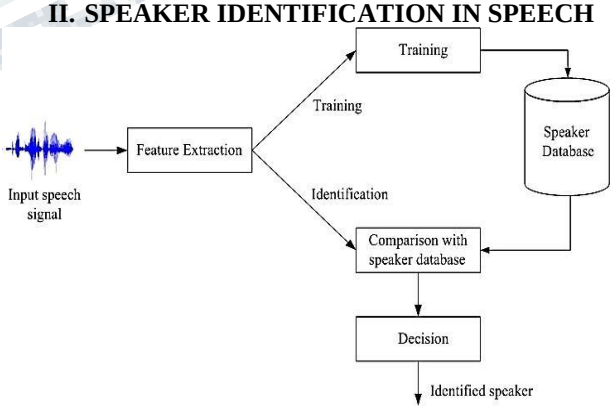

Fig. 1. Rudimentary Model of Speaker Identification System [2]

Speaker Identification system is based on feature extraction. Figure 1 displays rudimentary model of Speaker Identification system. Input speech signal is manoeuvre for feature extraction which is categorize as training and identification phases. Decision is based on collation with speaker database which is acquired from training phase and speaker is identified.

ISSN (Online) 2394-6849

International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)
Vol 5, Issue 1, January 2018

A. Pre-Processing

In speech dispensation short interval tranche is called frames and its size is heed as 10 to 40ms. Because of this variation in signal is distinguishable in short time [3]. Speech is segregated in number of frames. Short Time Energy and Zero Crossing Rates is restrained for each frame. If energy is lower than threshold then it is heed as noiseless period. Therefore the energy is extensively wield for the mensuration of starting and closure point of speech signal [4].Various pre-processing techniques are pre-emphasis, framing, silence removal, pre-quantization and windowing.

B. Feature Extraction Techniques

- Mel-Frequency Cepstral Coefficient

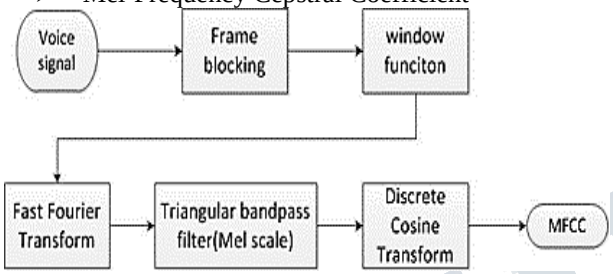

Fig. 2. Block Illustration of MFCC[5]

Framing is an approach of splitting the digitized speech samples into a small frames of span widespread from 20 to 40 ms. Several windows used for pre-processing are triangular, hamming, hanning window. Among them hamming window is best as it changes phase of voice signals to a designated range to make signal more continuous. Fast Fourier Transform transmute each frame contain samples from time to frequency domain. Dispensation Of Mel Filter Bank: The magnitude frequency which is equivalent to integrity at the center frequency and linearly it diminish response of each filter which is found to be triangular in outline and zero at middle frequency of very adjoining filters. Discrete Cosine Transform is functional to transfigure the logarithmic Mel spectrum into time dominion consuming DCT.

- Linear Predictive Cepstral Coefficients (LPCC)

LPC is utilized to evaluate spectrum of signal. It estimates speech models as a linear mixture of past samples[18]. This pattern diminishes the summation of squared difference among past samples and linearly predicted samples within limited interval. The distinctive set of predictor coefficients can be resolved by minimalizing such dissimilarity. The pre-emphasis is the first step to flatten spectrum of speech signal. Pre-emphasis increases the higher frequencies in the signal. The next stage is to edge the signal and take product of it with window function in order to decrease spectrum outflow in speech frame. In last phase, cepstrum is intended to cepstral analysis. Cepstral coefficients can also be considered from the LPC via set of recursive process. Various additional feature extraction systems are Log area ratio[6], Mean and Variance of the lasting phase[6], Perceptual Linear Prediction Coefficient, Mel filter bank slant structures[8], Delta and dual delta of MFCC characteristics[8], Cepstral Mean Subtraction[10].

- Perceptual Linear prediction (PLP)

PLP rejects inappropriate information of the speech and thus it increases speech recognition rate[19]. PLP is identical to LPC excluding that its spectral characteristics have been trans- formed to contest characteristics of human acoustic system. PLP approaches three perceptual aspects namely: Critical-band resolution curves, equal-loudness curve, and intensity-loudness power-law relation, which are known as the cubic-root.

C. Classification Techiques

Gaussian Mixture Model

In Gaussian mixture model, time consumed to establish using training phase is longer than the test phase. To speed up the process of obtaining the best parameters for Gaussian mixture model, it is obligatory to assess the initial parameters accurately. We use the K-means clustering to cluster the original feature vectors due to its fast, and worthy effect on clustering data. With K-means clustering resemble, all inceptive specification are clustered 128 groups as the 128 Gaussian models. Calculating weight for each group and combine these 128 Gaussian models as a Gaussian mixture model, and these 128 parameter groups are the inceptive parameters of the Gaussian mixture model. The EM algorithm is divided into two steps, as illustrated in Figure 5, calculating the E-step of the likelihood function and updating the parameters of the model in M-step.[5]

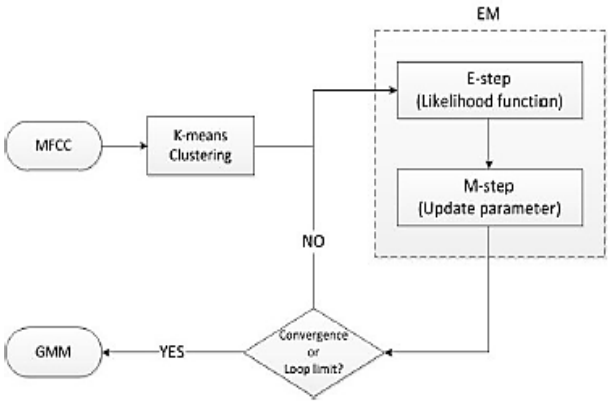

Fig. 3 Establishment of GMM Model [5]

ISSN (Online) 2394-6849

International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)
Vol 5, Issue 1, January 2018

Artificial Neural Network

An artificial neural network (ANN), habitually just known as neural network (NN), is an interrelated collection of artificial neurons that uses a statistical and computational model for information indulgence based on a connectionist advance to calculation[20]. Mainly an ANN is a system that fluctuate its configuration based on exterior or interior evidence that flows during the network. The model can have different forms, such as multi-layer perceptions. A MLP consist of one input layer, one hidden layer, and one output layer.

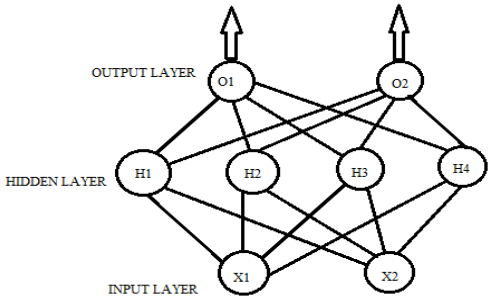

Fig. 4. Basic Model of ANN[20]

Various classification techniques are Polynomial Classifier [9], Vector quantization [11], Support Vector Machines [12], Hidden Markov MSodel[13].

III. SPOOFING ATTACK

In Spoofing attack, manipulated voice is employed as the system input which is voice conversion. Numerous types of spoofing attacks are speech synthesis, impersonation, replay, voice conversion. In impersonation subscriber's identity is changed. In speech synthesis spoken language is generated on base of written input. In replay valid data transmission is deceptively repeated.

Voice Conversion

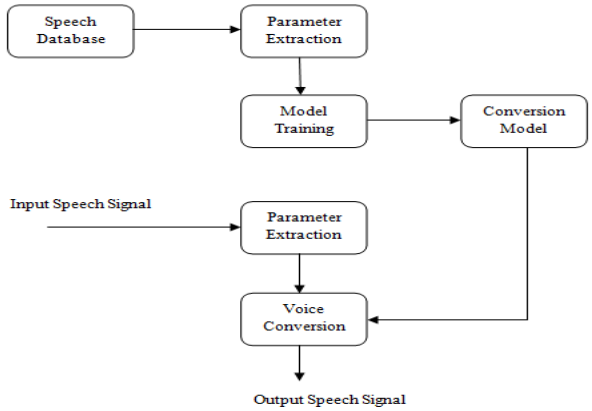

Fig. 5. Basic Block Diagram of Voice Conversion[17]

Voice conversion is constructed by establishing a speech record that reside speech signals both the source and target speaker's [17]. Various parameters extracted from these signals are pitch frequency, power values and mel-cepstrum. An erudition method is obligatory to construct a conversion model. Model must be able to capture original signals and transform it into target signal. Once model is trained, we test and install it by providing as input the source speaker speech. From this the characteristics are extracted. Subsequently using this model, the target speech is established and the output speech signal is generated. Various VC techniques are codebook mapping and VQ, HMM, GMM, ANN. Codebook mapping deliver relation between two spectra and in Vector Quantization mapping of two spectra is performed. In HMM, evolution to target signal is occurred via a sets of transition states. In GMM, alteration between the source and target signal is through the estimation of joint probabilities as well as exploiting a conditional probability whereas in ANN, a neural network is utilised to transform the speech signal. It can be used for learning and synthesis procedure.

IV. WATERMARKING IN SPEECH

The data which is inserted in the signal must be obtained and repel innumerable purposeful and accidental attacks.

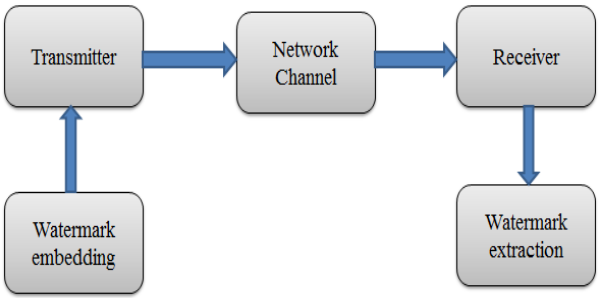

Fig. 6. Block Diagram of Watermarking in Speech[10]

Extraction section of digital speech watermarking are found in groups:

- In Blind speech watermarking, original signal is not required, logo or watermarked bits for watermark extraction.[14]
- In Non-blind speech watermarking, it requests the inventive signal and the watermarked signal for distinguishing watermark.

There is Steganography is also a possibility for data hiding in Audio and speech signal.

ISSN (Online) 2394-6849

International Journal of Engineering Research in Electronics and Communication Engineering (IJERECE)
Vol 5, Issue 1, January 2018

V. WATERMARKING FOR SPOOFING ATTACK

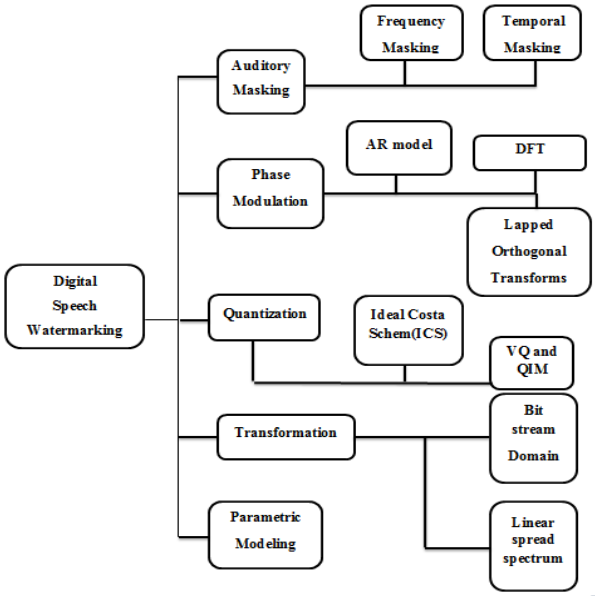

Fig. 7. Outline of techniques of Digital Speech Watermarking.[15]

Probability of spoofing attack in the speaker identification system is at the input side or sensor side. The appealed speaker data is already aware about the watermark of the system so that playback attack is possible in input side. Attack is also possible at the transmission side with replay attack or direct attack. To avoid the system from the attack watermark technology can be at both transmitter side as well as receiver side.

A. Phase Modulation

In phase modulation technique watermark bits are inserted by deploying phase of speech signal[21]. Advantage of phase modulation is that the watermarked and original speech has same power spectrum.

B. Quantization

Quantization techniques entrench the watermark in the tangential part of the speech[21]. This technique has enhanced the capacity of embedding the watermark in appropriate and inappropriate parts of the speech.

- Vector Quantizaton

In VQ, the original signals are segmented into non-overlapping frames and LPCs of every frame are intended as input vectors for VQ[21]. After receiving input vectors, they are associated with the adjacent code word in the codebook. By doing this we initiate output vector based on pre-defined formula. At the receiver side same catalogues are used to get preferred watermark.

- Quantization Index Modulation

Quantization Index Modulation (QIM) put on the accompanying quantizer for rounding off the original speech which entrenches the speech reliant on watermarks[21]. For digital watermarking, this technique can attain good stability between watermark embedding capacity and robustness.

C. Transformation

Transformation techniques gives attention on various techniques of speech signal such as creation, perception and bit rate.

- Linear Spread Spectrum

The spread spectrum (SS) technique endeavours to fleece or feast confidential information across the frequency spectrum of the speech signal which is self-regulating the actual signal. Hence, the closing signal inhabits a bandwidth which is more than required for transmission. This technique bids improved presentation, adequate data rate, high robustness but its restriction is that it announces noise in the sound file.

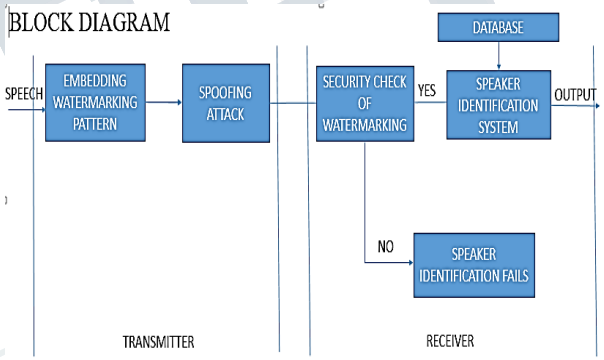

Fig. 8. Figure of Speaker Identification with Spoofing Attack Through Digital Speech Watermarking

Fig. 8 shows that at receiver side, initially features are removed from claimed speech, moreover this features are feed to classification model formerly judgement is prepared. Entire process is same as speaker identification process. Though once speaker is identified, it is tested for watermark convenience. If watermark is existing in appealed speech, it says that the font of the requested speaker is unaffected. Or else, spoofing attack is occurred. Occasionally, it is probable to way the voice conversion attack for making judgement and finding the source of attack. In this system watermarking pattern is embedded on transmitter side which is attacked by spoofing attack. Among various spoofing attack we are going to consider voice conversion as it is a voice manipulated by some person to achieve target speech. At receiver side detection of watermarking is occurred and if the system fails to detect the watermarking pattern the Speaker Identification fails. If it is detected then speech

signal are compared with already trained signal that are stored in the database, subsequently speaker is identified.

V. PERFORMANCE MEASURING PARAMETERS

- PARAMETERS FOR WATERMARKING DETECTION

A. Bit Error Rate

For sample, bits to be transferred are 11001100 and expected bits are 10000100. Compare both amount of bits. Therefore, the BER in this sample is 2/8*100 = 25 [16]

BER =E/ P x Q 100%

B. Signal to Watermark Ratio

Signal to watermark ratio is inspecting the effects of watermark on speaker identification system [1].

$$WSR = \frac{\text{Number Of Fram e Used For Watermarking In Speech}}{\text{Number Of Total Frame In Speech}}$$

- A. Percentage Of Identification Accuracy

$$PIA = \frac{\text{No. Of Correctly Identified Speakers}}{\text{Total No. Of Speakers}} \times 100$$

VI. CONCLUSION

Spoofing attacks is the leading intention to progress remote or online speaker identification system. Digital watermarking can effectively practice for numerous types of spoofing attack.it increases precision of speaker identification system in situation of insecure channels. ANN is better among other techniques but the accuracy for synthesizing signal was extraordinary only on deeper constructions. Hence it requires construction of several multiple layers for obtaining the output.

REFERENCES

[1] M.A Nematollahi, S.A.R Al-Haddad, Shyamala Doraisamy and M. Ranjbari, "Digital Speech Watermarking for Anti-Spoofing Attack in Speaker Recognition", IEEE Region 10 Symposium, 2014, pp. 476-479.

[2] Ronak Bajaj, "Features based on Fourier-Bessel Expansion for Application of Speaker Identification System", 2014.

[3] Kinnal Dhameliya and Ninad Bhatt, "Feature Extraction and Classification Techniques for Speaker Recognition: A Review", IEEE international Conference on Electrical, Electronics, Signals, Communication and Optimization (EESCO), Visakhapatnam, January 2015.

[4] Nidhi Desai, Kinnal Dhameliya and Vijayendra Desai, "Recognizing voice commands for robot using MFCC and DTW", International Journal of Advanced Research in Computer and Communication Engineering, Volume 3, Issue 5, May, 2014.

[5] Fang-Yie Leu, "An MFCC-Based Speaker Identification System,", Advanced Information Networking and Applications (AINA), 2017 IEEE 31st International Conference on, March 2017

[6] Jianglin Wang, An Ji and Michael T. Johnson, "Features for Phoneme Independent Speaker Identification", IEEE International Conference on Audio Language and Image Processing (ICALIP), Shanghai, July, 2012, pp. 1141-1145.

[7] Seiichi Nakagawa, Longbiao Wang and Shinji Ohtsuka, "Speaker Identification and Verification by Combining MFCC and Phase Information", IEEE transaction on audio, speech and language processing, Volume 20, Issue 4, May, 2012, pp. 1085-1095.

[8] Srikanth R Madikeri and Hema A Murthy, "Mel Filter Bank Energy- Based Slope Feature and Its Application to Speaker Recognition", IEEE National Conference on communication (NCC), Bangalore, January, 2011, pp. 1-4.

[9] Hemant A. Patil, Purushotam G. Radadia and T. K. Basu, "Combining Evidences from Mel Cepstral Features and Cepstral Mean Subtracted Features for Singer Identification", IEEE International Conference on Asian Language Processing, Hanoi, November, 2012, pp. 145-148.

[10] Nihalkumar G. Desai and Nikunj V. Tahilramani, "Speaker Recognition System Using Watermark Technology for Anti-Spoofing Attack: A Review", International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering,Volume 4, Issue 4, April, 2016, pp. 152-156.

[11] R. P. Ramachandran, K.R. Farrell, R. Ramachandran and R. J. Mammone, "Speaker Recognition—General Classifier Approaches and Data Fusion Methods", Pattern Recognition in Information Systems, Volume 35, Issue-12, December, 2002, pp. 2801-2821.

[12] C. J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition", Data Mining and Knowledge Discovery (Springer), Volume 2, Issue-2, June, 1998, pp. 121-167.

[13] Ghahramani Z., "An Introduction to Hidden Markov Models and Bayesian Networks", International Journal of

Pattern Recognition and Artificial Intelligence, Volume. 5, Issue-1, 2001, pp. 9-42.

[14] Nematollahi, Mohammad Ali, S. A. R. Al-Haddad, Shyamala Doraisamy, F. Zarafshan. "Blind Digital Speech watermarking Based on Eigen-value Quantization In DWT." Journal of King Saud University, December, 2014, pp. 58-67.

[15] Nematollahi, Mohammad Ali, and S. A. R. Al-Haddad. "An overview of digital speech watermarking." International Journal of Speech Technology, 2013, pp. 1-18.

[16] Seethal Paul , Sreelakshmi T.G. , "Performance Analysis and Study of Audio Watermarking Algorithms", International Journal Of Engineering And Computer Science ISSN:2319-7242 Volume - 3 Issue -8 August, 2014 Page No. 7540-7547

[17] Sathiarekha K, "A survey on the evolution of various voice conversion techniques", Communication Systems (ICACCS), 2016 3rd International Conference on, IEEE January 2016.

[18] Harshita Gupta, Divya Gupta, "LPC and LPCC method of feature extraction in Speech Recognition System", Cloud System and Big Engineering(Confluence),2016 6th International Conference, IEEE January 2016.

[19] Namrata Dave, "Feature Extraction Methods LPC, PLP and MFCC In Speech Recognition", IJARET, Volume 1, Issue VI, July 2013.

[20] Saravanan K. and S. Sasithra,"Review on Classification Based On Artificial Neural Networks", International Journal of Ambient System And Applications, Vol 2, No. 4, December 2014.

[21] Patel, Shruhad Kumar J., and Nikunj V. Tahilramani. "Information Hiding Techniques: Watermarking, Steganography: A Review."