International Journal of Engineering & Technology, 7 (2.16) (2018) 98-101

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research Paper

Acoustic comparison of electronics disguised voice using Different semitones

Mahesh K. Singh[1*], A. K. Singh[2], Narendra Singh[3]

[1,3]Jaypee University of Engineering & Technology, Guna (M.P.), India
[2]Thapar University, Patiala (Punjab), India
**Corresponding author E-mail: mahesh.092002.ece@gmail.com*

Abstract

This paper emphasizes an algorithm that is based on acoustic analysis of electronics disguised voice. Proposed work is given a comparative analysis of all acoustic feature and its statistical coefficients. Acoustic features are computed by Mel-frequency cepstral coefficients (MFCC) method and compare with a normal voice and disguised voice by different semitones. All acoustic features passed through the feature based classifier and detected the identification rate of all type of electronically disguised voice. There are two types of support vector machine (SVM) and decision tree (DT) classifiers are used for speaker identification in terms of classification efficiency of electronically disguised voice by different semitones.

Keywords: *Electronic disguised voice, Acoustic feature, Classifier, Speaker identification.*

1. Introduction

Under forensic casework speaker identification by electronically disguised voice is a major problem. This work emphasizes the effect of electronically disguised voice by different semitones [1], [2]. By using different semitones it's degraded or enhanced the voice quality. During the effect of electronic disguised voice pitch of speech signal has been changed by using different semitones [14], [15]. The disguised voice is mainly two types electronics disguised voice and non-electronic disguised voice. For electronics disguised voice the pitch of the speech signal may be changed by using different semitones [3]-[6]. In case of non-electronic disguised voice can be changed by physical deprivation of speech like raised pitch, lower pitch, pinched nostrils etc. In this proposed model emphasize speaker identification under electronic disguised method by using its acoustic feature analysis [20]. In this hierarchy analyzed that there are changing the speaker's normal voice by using different semitones [15], [16]. By using the different semitones it will be changed the pitch of the normal voice tones. The acoustic feature of normal and disguised voice by different semitones are extracted by using MFCC techniques. By using MFCC techniques easily extracted all the feature of speech [19]. In this paper acoustic feature analyzed by MFCC and calculate the mean value and correlation coefficients of the by normal voice and disguised voice using different semitones.

Presents voice modeling procedure for analyzing speaker identification system [17]. In this paper, techniques like speech spectral shaping, feature extraction, parametric mapping and statistical signal modeling are presented [7], [10]. That defines the role of feature extraction technique that acts as an efficient process for determining the information about the signal while discarding signals the other unwanted signal like noise [11]-[15]. This improves prediction rate of the system and provides fast and cost-effective speaker identification methods [2].These mathematical functions are representative of the various speech features. The module executes the functions and computations involved in calculating each of the mentioned features [9], [13]. The extracted features are either having a fixed dimension or multidimensional. A feature vector comprising all of these features is extracted for each frame. It improves the storage and processing efficiency of the speech signals. While feature based classification techniques used for correct identification of the speaker [5]-[8].

There are different feature classifications techniques are used for identification. Such feature based classification is DT and SVM. In this method derive a comparative analysis of speaker identification by SVM and DT. This work totally focused on the acoustic feature analysis and classification techniques used to calculate the efficiency of the speaker for correct identification rate [1].

2. Experimental Setup

Experimental classification in this work is classified in following different setup and explained in **Fig.1**.

A. Voice sample collection
B. Disguised voice by different semitones
C. Acoustic feature extraction
D. Classification and speaker identification

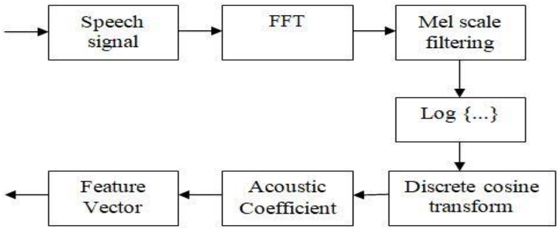

Fig 1: Speaker identification System for electronically disguised voice

A. Voice sample collection

For voice sample collection method there are 20 male students are selected. They speak the common accents in a normal voice. Their voice sample recorded by audacity recording tool with 32-bit quantization 8 KHz sampling rate in a closed room and noise free environment. All the recorded speech created a .wav file and put all file in a database [16].

Table 1: Speech collection details

Speech pattern/Setting	
Gender	**Male**
No. Of speakers	**20**
English	
Sampling Frequency	8 KHz
Recording Environment	Room
Utterance Recorded	“My research work in speech signal processing”
Recording duration	≥ 3.5 second
Disguised semitones	(-2,-4, +2, +4)
Total utterance	100

B. Disguised voice by different semitones

There is all the 20 speaker voice are disguised by different semitones for this observation normal voice electronically disguised by (-2,-4, +2, +4) semitones. For disguising the voice electronically change their pitch by (-2,-4, +2, +4) semitones. Change the pitch by -2 semitones are expanded by -10.91% change of voice (-10.91%) means to increase the pitch. Same as (-4) by (-20.41) as well as by (+2) semitones voice decrease by (12.24%) o pitch and changed by (+4) semitones decreased by (25.99%) voice pitch. All the electronic disguised voice using different semitones are done by changed the pitch by audacity tools.

C. Acoustic feature extraction

There is different feature extraction technique such as hidden Markov model, vector quantization, linear predictive cepstral coefficients & MFCC etc techniques used for feature extraction. Feature extraction techniques are used for extraction the different statistical feature of the speech signal [19]. In this work MFCC technique used for feature extraction in automatic speech recognition system. MFCC most commonly used for feature extraction by this feature extraction method the entire feature vector containing the linguistic message [14].

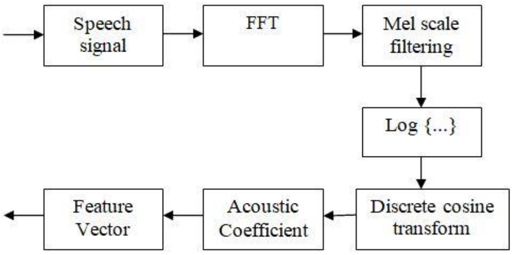

Fig 2: MFCC algorithm based acoustic feature

By using MFCC method derive entire acoustic coefficients by using the entire step shown in **Fig. 2.** Using this method derive the entire acoustic feature in term of mean and correlation coefficients [14], [15].

Here *x(n)* is a speech signal with *N* frames of MFCC vector. Suppose v_{mn} is the n^{th} value of MFCC dimensional vector of m^{th} frame, V_n is the vector of the n^{th}h component. Here V_n is given as:

$$V_n = \{v1n, v2n, \dots \dots, vNn\}; \ \ where\ n = 1,2, \dots \dots . L \qquad (1)$$

In this article, two types of acoustic feature are used. Firstly the mean E_n of each MFCC features V_n is extracted and then the correlation coefficients $CR_{nn'}$ among different MFCC features V_i and V_j are determined as given below –

$$E_n = E(V_n)\ ;\ n = 1,2, \dots \dots . L \qquad (2)$$

$$CR_{nn'} = \frac{cov(V_n, V_{n'})}{\sqrt{var(V_n)}\sqrt{var(V_{n'})}}\ ;\ 1 \le n < n' \le L \qquad (3)$$

The resulting values of meanE_n and correlation coefficients $CR_{nn'}$ are used together to create the statistical moments K_{MFCC} of MFCC vectors as described in below equation:

$$K_{MFCC} = (E_1, E_2 \dots . . E_L\ , CR_{12}\ , CR_{13}\ , \dots \dots CR_{L-1}) \qquad (4)$$

Similarly, the statistical moments of MFCC (K_{MFCC}) is determined. The text dependent speech signal has prior information about the text to be spoken by a speaker. In this algorithm, a text-dependent approach is used [10]. This can be described as the matching of the feature tendencies of normal voice acoustic feature with non-electronic disguised voice acoustic feature [7], [14].

D. Classification and speaker identification

This article two types of classifier SVM and DT are used for classifying the data for this study using 70% data for training and 30% data for testing purpose. All the training and testing data put in random size in nature. For finding the classification rate fold all training and testing data passed through each classifier and find the speaker identification rate.

SVM: SVM classifier is an algorithm that based on a nonlinear mapping to changed the original training data into its higher dimension. SVM classifiers are based on supervised learning that requires training and testing prior to classification. SVM receives a sequence of datasets and estimate each data inputs and further categorized it into two possible classes [1], [14], [15].

DT: Algorithm used in the decision tree is based on divides and conquer operation. This classifier has a tree-like structure, where all inner nodes represent a test on attributes, each branch denotes a result of the test, and leaf nodes correspond to classes or their distributions. It divided training and testing datasets into smaller training and testing dataset. It is derived by using entropy for different attributes:

$$E\ (N, L) = \sum\nolimits_{m\in 1} ((m) * (m)) \qquad (8)$$

Where (m) and (m) are the different datasets used in the classifier.

3. Acoustic feature analysis and speech classification

This section extracted the acoustic feature of electronically disguised voice by different semitones mean and correlation coefficients are compared. Subsequently, all the MFCC acoustic feature are classified using feature-based SVM and DT classifiers.

Acoustic Analysis- The acoustic mean value of normal voice and all types disguised voice of twenty speakers are shown in **Table 2.** From the acoustic feature that is shown electronic disguised by +2 and +4 semitones that have the greater mean value comparison to -2 and -4 semitones that is similar to the the acoustic mean value of normal voice.

Table 2: Acoustic mean value of normal voice and disguised voice

Voice/ Speaker	Mean Normal Voice	Mean Disguise Voice (-2)	Mean Disguise Voice (-4)	Mean Disguise Voice (+2)	Mean Disguise Voice (+4)
S1	6.55	5.59	5.83	6.79	6.36
S2	6.33	6.19	6.21	6.71	7.06
S3	4.85	4.90	4.85	4.58	4.75
S4	5.59	5.67	6.06	6.00	6.11
S5	5.59	5.59	6.39	5.86	5.66
S6	6.06	6.06	5.82	6.17	6.14
S7	5.48	5.26	5.76	5.95	6.66
S8	5.67	5.56	5.82	5.99	5.92
S9	6.24	6.48	6.58	5.65	5.63
S10	5.92	6.03	6.22	5.50	5.18
S11	5.20	5.20	6.12	5.66	5.94
S12	6.08	6.08	6.15	6.08	5.92
S13	5.55	5.85	6.04	5.97	6.19
S14	5.76	5.82	6.06	5.82	5.95
S15	6.27	6.27	6.30	5.84	5.24
S16	5.62	5.59	5.98	5.99	6.36
S17	5.79	5.99	6.56	5.87	5.90
S18	6.04	6.09	6.04	5.90	5.80
S19	5.02	5.18	5.76	5.50	6.42
S20	6.06	5.81	5.82	6.17	6.14

Table 3 represents that the average value of four types of electronically disguised voice. The table showed the average of table 1 for taking the average of four types of electronics disguised voice. By taking the average of all types of disguised voice that is nearer to normal voice. From average value of s6 is very close to the normal mean value. It is also shown the average acoustic mean value in the graphical representation in fig. 3. Graphical representation showed that the average value of s6,s8,s18 and s20 value is nearer to the normal mean value shown in **fig 3**.

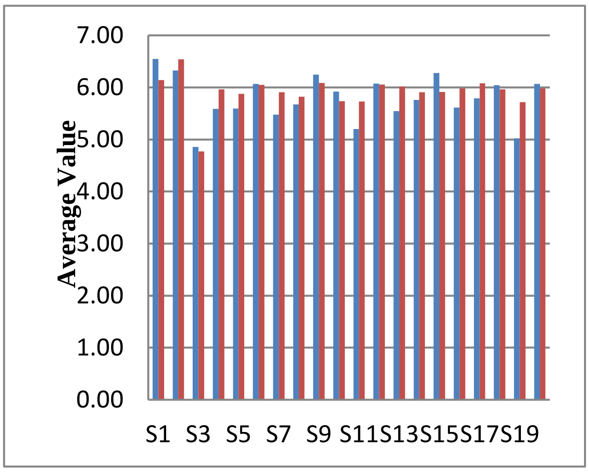

Fig.3: Average acoustic mean value of 20 speakers

The acoustic correlation value of normal voice and all types disguised voice of twenty speakers are shown in From the **correlation coefficients** acoustic feature that is shown electronic disguised by -2 and -4 semitones that have the greater mean value comparison to +2 and +4 semitones mean value that is similar to the acoustic mean value of normal voice.

Represents that the average correlation coefficients value of four types of electronically disguised voice and comparison result are shown. The table showed the average of table 5 for taking the average of four types of electronics disguised voice. By taking the average correlation coefficients value of all types of disguised voice that is closed to normal voice. From average value of s1 and s20, the disguised voice is very close to the normal mean value. It is also shown the average acoustic mean value in the graphical representation in **fig. 4**. Graphical representation showed that the average value of s3,s6, s11, and s20 value is nearer to the normal mean value shown in **fig 4.**

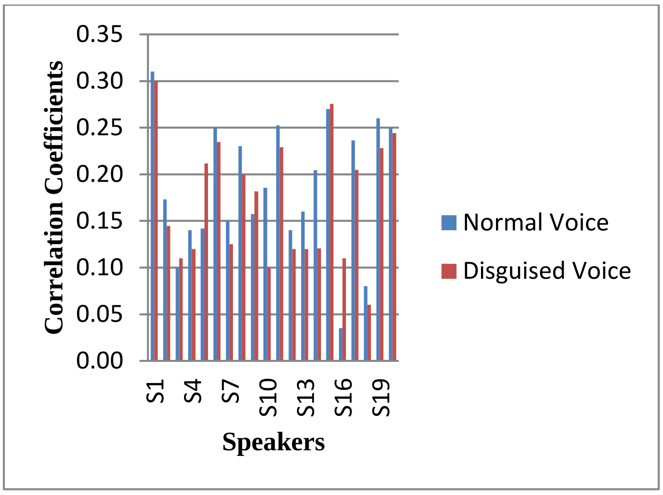

Fig.4: Average comparison of acoustic correlation coefficients of 20 speakers

The identification efficiency of speakers from electronically disguised voice using different semitones (-2, +2, -4, +4) it contains a training stage model and a testing model. For training model the speech signal, a speech database sets are created in this normal voice module and electronics disguised voice with above semitones are represented. There are analyzed the feature from the statistical value of the normal voice sets along with the disguised by different semitones used for testing and training purpose. A disguised voice from different semitones are used as the testing components to test from SVM and DT classifiers in classify the voice whether a training voice of disguised by different semitones of normal voice identify a particular speaker. The calculated database that is consisting of 100 voice samples including normal voice and disguised by (-2, +2, -4, +4) semitones with the duration of 3.5 sec from 20 speakers. The recorded file formats are taken as a the.WAV format with 8 kHz sampling rate, mono and 32-bit quantization [4]. Thereafter, the database is randomly divided into two adjacent parts: 20 normal voice sample segment of 20 speakers and 80 voice sample segment from disguised voice. Finally, 100 voice samples consist of disguised voice sample are considered. Out of 100 voice samples, 70 voice samples are used as the original set of voice for training purpose and 30 are used for testing purpose. For each disguising method, the performance of speech identification is evaluated under all the listed classifiers. The classification and average speaker identification results are shown below in Table 6.

Table 3: Speaker Identification Performance (%) of different classifier used in proposed model.

Classifier	SVM	DT
+2 (Semitone)	74.07	96.29
+4 (Semitone)	74.07	88.88
-2 (Semitone)	74.07	92.59
-4 (Semitone)	92.59	96.29
Average classifier efficiency	**78.70%**	**93.51%**

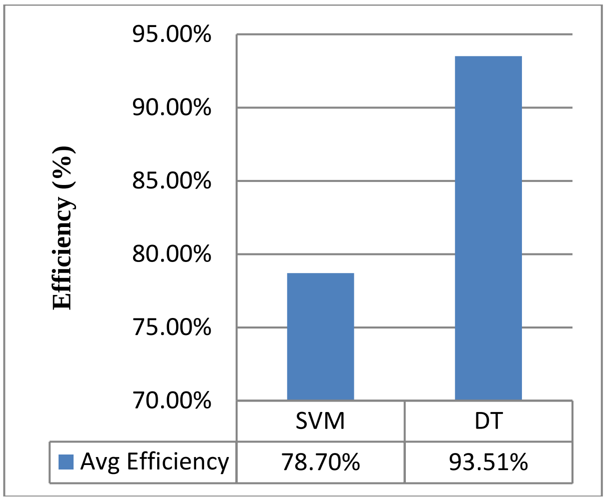

Fig 5: Speaker identification efficiency by classifier

Fig.5. shows that the classification through the SVM and DT classifiers, that gives the good result for speaker identification. Plot the result of disguised voice average classification efficiency of SVM classifier is (78.70%) with reference to normal voice as well as DT classifier result (93.51%) of speaker identification rate. DT has significantly higher efficiency or detection rates. Our results on SVM may be conflicting with some other evaluation relating SVM. Here superiority of DT is showed over other learning algorithms as it has achieved an excellent accuracy of DT 93.51% for disguised and 78.51% identification rate of SVM classifier.

4. Conclusion

This article shows the comparative analysis of MFCC based acoustic feature of voice disguised by different semitones. The resultant mean value disguised voice of speaker 6 is close to the mean value of normal voice. The speaker identification of feature-based classifier is 93.51% and 78.51% efficiently given the better result by these classification techniques. After all marginal speaker identification rate can be attaining nearer to 100%. From the acoustic feature of mean and correlation coefficients reveals the existing difference between normal voice and the disguised voice. Overall the speaker identification using DT classifier performs the better result over the SVM classifier for normal voice and disguised by different semitones.

References

[1]. A.F.Abdulwahab,S.A.Mohd and H.Husni, “Acoustic comparison of Malasian and Nigerian English accents,” Journal of telecommunication, electronics and computer engineering, vol. 9,No.3-5 pp.141-146,Nov.2017.

[2]. Marco Grimaldi and Fred Cummins, “Speaker identification using instantaneous frequencies,” *IEEE transactions on audio, speech, and language processing*, vol. 16, no. 6, pp.1097-1111, Aug. 2008.

[3]. Hamed Riazati Seresht, Seyed Mohammad Ahadi and Sanaz Seyedin, “Spectro-temporal power spectrum features for noise robust ASR,” *Circuits, Systems, and Signal Processing*, vol. 36, Issue 8, pp. 3222–3242, Aug. 2017.

[4]. C. Jingxu, Y. Hongchen, and S. Zhanjiang, “The speaker automatic identified system and its forensic application,” *in Proc. Int. Symp. Compute. Inf.*, vol. 1, pp. 96–100, 2004.

[5]. X. Zhu, G. Beauregard, and L. Wyse, “Real-time signal estimation from modified short-time Fourier transform magnitude spectra,” *IEEE Trans. Audio, Speech Lang. Process.*, vol. 15, no. 5, pp. 1645–1653, Aug. 2007.

[6]. R. Rodman, “Speaker recognition of disguised voices: A program for research,” *in Proc. Consortium Speech Technol. Conjunct. Conf. Speaker Recognition. Man Mach, Direct. Forensic*, pp. 9–22, Appl., 1998.

[7]. R. Hsiao, M. C. Fuhs, Y.-C. Tam, Q. Jin, and T. Schultz, “The CMUinterACT 2008 Mandarin transcription system,” in INTERSPEECH, 2008, pp. 1445-1448.

[8]. S. Amuda, H. Boril, A. Sangwan, and J. H. Hansen, “Limited resource speech recognition for Nigerian English,” in Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2010, pp. 5090-5093.

[9]. D. Vergyri, L. Lamel, and J.-L. Gauvain, “Automatic speech recognition of multiple accented English data,” in INTERSPEECH, 2010, pp. 1652-1655.

[10]. R. Soorajkumar, G. Girish, P. B. Ramteke, S. S. Joshi, and S. G. Koolagudi, “Text-Independent Automatic Accent Identification System for Kannada Language,” in Proceedings of the International Conference on Data Engineering and Communication Technology, 2017, pp. 411-418.

[11]. A. Rabiee and S. Setayeshi, “Persian accents identification using an adaptive neural network,” in Proceedings of the 2nd International Workshop on Education Technology and Computer Science, 2010, pp. 7-10.

[12]. L. M. Arslan and J. H. L. Hansen, “Language accent classification in American English,” Speech Communication, vol. 18, pp. 353-367, 1996.

[13]. A. Hanani, M. J. Russell, and M. J. Carey, “Human and computer recognition of regional accents and ethnic groups from British English speech,” Computer Speech & Language, vol. 27, pp. 59-74, 2013.

[14]. Haojun Wu, Yong Wang and Jiwu Huang, “Identification of electronic disguised voices,” *IEEE transactions on information forensics and security*, vol. 9, no. 3, pp.489-500, March 2014.

[15]. Haojun Wu, Yong Wang and Jiwu Huang, “Blind detection of electronic disguised voice,” *IEEE international conference on acoustics, speech and signal processing* (ICASSP), pp.3016-3017 May.2013.

[16]. Audacity: free audio editor and recorder [online]" in http://audacity.sourceforge.net.

[17]. T. Tan, “The effect of voice disguise on automatic speaker recognition*,” IEEE Int. CISP*, vol. 8. pp. 3538–3541. Oct. 2010.

[18]. Reza Entezari-Maleki, Arash Rezaei and Behrouz Minaei-Bidgoli, “Comparison of Classification Methods Based on the Type of Attributes and Sample Size,” *Journal of Convergence Information Technology*, Vol. 4, no.3, pp.09-17, Sept. 2009.

[19]. Rajeev Ranjan and Rajesh K. Dubey, “Isolated word recognition using HMM for Maithili dialect,”*IEEE*, International conference on signal processing and communication, pp. 32-328, Dec. 2016.

[20]. Cuiling Zhang, Tiejun Tan, “Voice disguise and automatic speaker recognition,” .*Elsevier: Science Direct. Forensic Science International*, vol.175, issues2-3, pp. 118–122, March 2008.