Original Article

Identifying speaker from disguised speech using aural perception and Mel-frequency cepstral coefficient

J. Praveena, Y. Krishna[1]

Department of ENT, Jawaharlal Institute of Postgraduate Medical Education and Research, Puducherry, [1]Department of Speech and Hearing, School of Allied Health Sciences, Manipal University, Manipal, Karnataka, India

Abstract

Objective: The present study intended to compare the accuracy of speaker identification using aural perception and semiautomatic method (Mel –Frequency Cepstral Coefficient; MFCC), when the speech is in disguise condition by using the handkerchief during recording and to check the percentage of correct identification in the semiautomatic method when the vowel and consonant segments were used for analysis. **Methods:** Thirty speaker's single sentence speech sample was recorded in undisguised and disguised conditions were randomly paired into the sets of one undisguised followed by five disguised samples for the task of speaker identification. In aural perceptual method the five judges listened to the samples and made a decision on the match. In MFCC method, from /ðə/ segment, ten coefficient values were extracted. The coefficient values were manually averaged and the pair that obtained the lowest value of Euclidean distance was determined to be the sample of the same speaker. The Kappa agreement was used to find the agreement between the two methods in speaker identification and the percentage of correct identification was calculated for the vowel and consonant segment analysis. **Results:** The results revealed the kappa value to be negative ($k < 0$) indicating no agreement between the two methods. The percentage of correct identification using aural perception ranged from 56.7% - 80% and for MFCC under whole word, consonant segment and vowel segment analysis were 46.7%, 26.7% and 53.33% respectively. **Conclusion:** The aural perception method had a greater percentage of correct identification than MFCC though it was not statistically significant for speaker identification from disguised speech.

Key words: Disguised speech, Mel-frequency cepstral coefficient, speaker identification

Introduction

"Speaker identification refers to any decision-making process that uses some feature of the speech signal to determine if a particular person is the speaker of a given utterance" (Atal, 1974). In speaker identification, the experimenter has to determine which one sample from the group of known voice would best match with the input voice sample. It involves the comparison of one or more samples of unknown voices (sometimes called the questioned samples) with one or more samples of

Address for correspondence:
Ms. J. Praveena, Department of ENT, Jawaharlal Institute of Postgraduate Medical Education and Research, Puducherry, India.
E-mail: praveenaind@gmail.com

Access this article online	
Quick Response Code:	Website: www.jisha.org
	DOI: 10.4103/0974-2131.185974

How to cite this article: Praveena J, Krishna Y. Identifying speaker from disguised speech using aural perception and Mel-frequency cepstral coefficient. J Indian Speech Language Hearing Assoc 2015;29:28-34.

known voice. One factor that may seriously affect both expert and lay speaker identification is vocal disguise. "Disguise refers to any alteration, distortion or deviation from the normal voice, irrespective of the cause."[1] Disguise may take many forms, including mimicry, adoption of a different accent, the use of external objects to affect vocal tract dynamics or the use of electronic devices.[2] Disguise can be deliberate or nondeliberate,[1] and the deliberate disguise can further be of electronic disguise (electronic scrambling) or nonelectronic disguise (online modification of speech) types. The other types of disguise also include change in pitch falsetto, pertinent creaky voice, whispering, faking a foreign accent, and pinching one's nose.[3] Thus, disguise can be made possible with deliberate, nondeliberate way, or with the use of external device.

Hecker in 1971[4] classified the methods for speaker identification into three types as following, speaker identification by listening or aural perceptual method (subjective method), speaker identification by visually inspecting the spectrogram of the speech samples (subjective method), and speaker identification by machines (objective method). The listening method or the aural-perceptual speaker identification (APSID) is the subjective method based on human auditory perception. The trained phoneticians carefully listen to the recordings and identifies if the samples are of the same speaker or different speakers. In the identification process, the factors such as listener's ability, voice characteristics of the speaker, and the nature of environment in which utterances are produced are to be considered.

In the automatic methods, the speaker identification is done using special algorithms by computer systems, and the participation of the investigator is very minimal in the process. In semiautomatic method, the computer processes the samples; it extracts parameters and analyses them according to a preset program. The examiner makes the interpretation based on closeness of the parameters between the samples. In semiautomatic methods of speaker identification, F1 and F2;[5-10] higher formants,[11] fundamental frequency,[12] linear prediction coefficients,[13] cepstral coefficients and Mel-frequency cepstral coefficient (MFCC),[14-17] long-term average spectrum,[18] and cepstrum;[15,19-24] have been used in the past.

A report published by the National Crime Records Bureau compared crime rate from 1953 to 2006 in India. The report noted that kidnapping has increased by 47.80% (from 5261, a rate of 1.40/100,000 in 1953 to 23,991, a rate of 2.07/100,000 in 2006).[25] In such instances, voice disguise can be observed to be a significant proportion of offense. This warrants the need for speaker identification to play an important role, in identifying the suspect in spite of their disguised speech. Speaker identification also extends the boundary for the scope of speech language pathologist (SLP) to serve as witness. It is important to know which of the speaker identification methods will be appropriate to correctly identify the speaker even if the speech is disguised. Thus, the present study aims at investigating speaker identification accuracy using APSID and MFCC for disguised speech using handkerchief and to compare the accuracy of the two methods in speaker identification and to check the percentage of correct identification in the semiautomatic method when the vowel and consonant segments were used for analysis.

Methods

Participants

Thirty adult males who were native Malayalam speakers within the age range of 18–23 years with perceptually normal voice and articulation participated in the study.

Materials

The standardized sentence from Consensus Auditory-Perceptual Evaluation of Voice, 2002, signifying the vowel production "The blue spot is on the key again" was chosen for the participants to say, for the purpose of analysis in both aural perception and MFCC methods. The whole sentence was presented for aural perception, and only/ðə/segment from the initial part of the sentence was chosen for the MFCC analysis.

Procedure

The participants were explained about the study and the purpose of recording their speech sample. The sentence "The blue spot is on the key again" written in upper case was given for the participants to read first. Two to three trials were given before recording to each participant to say the sentence in the form of a statement without stressing the words at a normal conversation level. During the trial, it was made sure that the rate of speech was normal and there were no pauses between the words. The participants were made to sit erect in front of the condenser microphone placed on the stand at approximately 10 cm away from the mouth of the participants. Recording was done in the Computerized Speech Lab (CSL model 4500) and saved in wave format (*.wav). First, the sentence was recorded in undisguised manner (normal situation). Recordings were repeated if the participants made errors or if they had any interruptions or change in intonation

or change in their rate of speech. The disguised speech sample was then recorded by covering the microphone with handkerchief of four folds. The same recording environment and handkerchief were used to record all samples.

The recorded samples were assigned random number as Q1–Q30 for the undisguised samples and S1–S30 for the disguised samples by an SLP to blindfold the experimenter from the names of the sample. Lottery method was carried out to randomize the disguised samples into, 30 sets each containing five samples. SLP was asked to choose one of the undisguised samples from the speaker's sample, and it was made sure that the disguised sample of the same speaker was present in the set. Thus, a group of six samples (one undisguised followed by five disguised) was present in each of the 30 sets.

Semiautomatic method

For semiautomatic approach of speaker identification, from the recorded sentence samples, the most frequently occurring word in spoken English "The"[26] at initial position of the sentence was chosen for MFCC extraction. The sentence was displayed in WAV format in CSL, from which /ðə/ segment was chosen. The cursor was placed at 10 different points within the portion of/ðə/utterance from waveform display. Each cursor point were subjected to fast Fourier transform in the Mel scaling and was further subjected to cepstral analysis from which MFCC of the peak was obtained for 10 different peak points in /ðə/ segment in CSL for all the 30 undisguised and the disguised samples. The MFCC was obtained by noting the quefrency (X-axis) and amplitude (Y-axis) of the peak. The 10 quefrency and the amplitude values for each sample were tabulated in Excel sheet and their average were obtained manually. The points were then sorted in ascending order with respective to the quefrency values, and the average of the first five and the next five were taken separately. Thus, for each of the sample, there were three averages, the whole word 10 points; from the ascending ordered quefrency values, the first five points were considered to be from the initial consonant segments and the later five points were from the vowel segments. This was done to check if there was variation in correct identification and if the predominant consonant segment (initial five values) and vowel segment (later five values) were only considered for Euclidean distance calculation.

In analysis of MFCC method, the tabulated quefrency (X-axis) and amplitude (Y-axis) of whole word 10 points, initial consonant segment five points, and later vowel segment five points for each of the sample were averaged separately in excel. The Euclidean distance between the undisguised sample and five disguised samples in each set for all the three types of averages was calculated manually in excel using the formula (1):

$$\text{Euclidean distance} = \sqrt{\left(X_2 - X_1\right)^2 + \left(Y_2 - Y_1\right)^2} \qquad (1)$$

where (X_1, Y_1) and (X_2, Y_2) are the averaged MFCC of undisguised sample and disguised sample, respectively.

The disguised sample which got the lowest value of Euclidean distance within the set was considered the sample of the same speaker. After similar calculation for the entire 30 sets of samples, the experimenter was given the numerical key for the each undisguised and disguised samples for each of the speaker. The key was then compared with the sample pair that obtained lowest Euclidean distance value. If the lowest Euclidean distance sample was from the sample of the same speaker, then it was considered as correct identification.

Aural perceptual method

The samples from each of the saved folder were edited using Adobe Audition software version 1.0 (Adobe systems), for them to be played in a sequence of one undisguised sample followed by the five disguised samples. The same random ordering of samples was used for both the methods. Thus, a set of 30 in the sequence of 6 samples (one undisguised followed by 5 disguised samples) were played for the aural perceptual judgment.

Five SLPs were involved for the aural perception judgment. They were instructed that the first played sample should be compared to the following five samples and to indicate if they were of the same speaker or different speaker. The five SLPs were asked to mark the number of the disguised sample sequence which they perceived to be was similar to the undisguised sample. All the 30 sets of samples were presented through loudspeaker from the Adobe Audition software version 1.0 in a quite comfortable room. After the judgment for all the 30 sets, the number of correct identification was noted by cross checking with the key if the judges have paired the same speaker's disguised sample and undisguised sample if so it was noted as correct identification.

Analysis

The statistical agreement between the aural perception and MFCC methods was analyzed using kappa measure. The number of correct identification for each sample by the 5 judges and MFCC method were noted down. Using Statistical Package for Social Sciences version 6 (International Business Machine Corporation [IBM] Licensor, Illinois, Chicago, United States 1989, 2011),

the percentage of positive proportion version (number of correct identification) and negative proportion (the number of incorrect identification) and the kappa agreement between the two methods were calculated.

Results and Discussion

Speaker identification is a decision-making process if the particular person is the speaker of the given sample.[15] The speaker identification process can be carried out using any of the three methods (a) aural perceptual, (b) spectrographic analysis, and (c) automatic or semiautomatic methods.[4] The aural perception is the technique of listening to the speaker samples and judging the similarity by the trained experimenter. In the semiautomatic method, the interpretation is made by the experimenter based on the features extracted. In our present study, we intended to compare the accuracy of speaker identification by using the aural perceptual and the semiautomatic method (MFCC) and to check the percentage of correct identification in the semiautomatic method when the whole word, and vowel and consonant segments were used for analysis under disguised condition brought about by using handkerchief during recording.

Thirty speakers speech sample in undisguised and disguised condition was randomly paired into the sets of one undisguised followed by five disguised samples. In the aural perceptual method, the task of the five judges was to listen to the disguised samples and to make decision which of the sample was similar to the undisguised sample. In the semiautomatic method of MFCC, /ðð/ segment was chosen from the speech sample; 10 coefficient values were extracted from the whole of /ðð/segment, and five MFCC were extracted from initial consonant segment including the transition duration. Another five points were extracted from the later steady portion of the word consisting of the vowel. The averages of each sample's MFCC were taken, and Euclidean distance was calculated between the undisguised and disguised samples within each of the sets. The pair that obtained the lowest value of Euclidean distance was determined as the sample of the same speaker. The lowest Euclidean distance and correct sample Euclidean distance are given in Table 1. The percentage of correct identification in both the method was considered; the kappa agreement value and the proportion of positive and negative percentage between the methods were also obtained in the process of analysis.

Aural perception method

In aural perceptual method from the total of 30 speakers, the number of correct identification varied among the five judges. Judge I and II matched the 22 number of speakers correctly, Judge III matched 20 speakers correctly; Judge IV matched 24 which was the maximum number of correct identification among all the judges, and Judge V matched only 17 speaker's samples correctly which was the least among the judges. The number and percentage of the correct identification by each judge are shown in Table 2.

The highest percentage of correct identification obtained using aural perceptual method was 80% while the lowest being 56.7%. These results are in accordance with the various other studies, which showed the range of percentage of correct identification ranging between 34%–75%[27] and 56%–79.4%[28] under normal condition recorded samples. While the recognition of samples from varied conditions can obtain error rate up to 30%,[29] correct identification for unfamiliar speakers ranged from 44.67% to 59%.[28]

Semiautomatic method

In MFCC method, the number of correct identification for the whole word, consonant segment, and vowel segment analysis were 14, 8, and 16, respectively. The results of the number and percentage of the correct identification using MFCC method for the whole word, consonant segment, and vowel segment analysis are given in the Table 3.

The highest percentage of correct identification was obtained for vowel segment analysis (53.3%), and the lowest for the consonant segment analysis (26.7%). These results are in positive statement with the other studies which used vowel segment for MFCC extraction in speaker identification.[15,23,30]

Comparison of the methods

The kappa agreement was used to compare the agreement between the two methods for the correct and incorrect speaker identification. The proportion of correct and incorrect identification (percentages of positive and negative proportions) was calculated between the results of five judges and the three MFCC analyses. The positive proportion or the correct identification proportion between the aural perception and MFCC analysis varied across judges and MFCC analysis. Tables 4 and 5 show the percentage of the positive and negative proportions of correctly identified samples between the aural perception and MFCC, respectively.

The percentage of positive proportion was higher than the negative proportion between the methods. The range of percentage of positive proportions was

Table 1: Lowest Euclidean distance sample and same speaker undisguised and disguised sample Euclidean distance

Undisguised speech sample	Lowest Euclidean distance sample		Disguised sample of the same speaker	
	Sample name	Euclidean distance value	Sample name	Euclidean distance value
Q21	S30	7.521708	S22	11.20588
Q20	S18	2.2491	S28	6.134795
Q25	S29	7.447003	S29	7.447003
Q9	S11	3.517863	S2	12.58392
Q19	S2	1.392819	S20	3.339049
Q29	S12	2.97748	S13	11.99307
Q22	S3	16.19527	S3	16.19527
Q12	S23	3.474158	S25	28.93901
Q27	S12	0.909874	S16	17.90962
Q11	S27	10.38048	S27	10.38048
Q1	S5	0.816253	S5	0.816253
Q28	S7	16.72479	S8	23.09938
Q12	S21	1.366649	S21	1.366649
Q24	S24	22.04016	S18	5.434613
Q4	S1	1.713043	S1	1.713043
Q5	S30	4.021556	S30	4.021556
Q13	S15	1.660981	S15	1.660981
Q8	S5	3.744008	S11	13.85103
Q18	S30	2.867612	S10	7.346007
S15	S9	2.64869	S9	2.64869
S10	S29	1.1374	S29	1.1374
Q17	S17	2.722957	S4	8.566038
Q7	S23	13.05564	S23	13.05564
Q3	S20	17.98402	S14	26.55436
Q30	S20	4.856009	S26	18.67191
Q16	S2	2.010203	S6	4.576957
Q14	S12	23.88462	S12	23.88462
Q26	S19	2.903207	S14	5.326363
Q23	S18	4.989739	S18	4.989739
Q6	S16	1.342615	S17	12.62007

Table 2: The number and percentage of correct identification by each of the judges in aural perceptual method

Judges	Number of correct identification	Percentage of correct identification (%)
Judge I	22	73.33
Judge II	22	73.33
Judge III	20	66.7
Judge IV	24	80
Judge V	17	56.7

Table 3: The number and percentage of correct identification using Mel-frequency cepstral coefficient method for the whole word and consonant and vowel segment analysis

Segments	Number of correct identification	Percentage of correct identification (%)
Whole word	14	46.7
Consonant	8	26.7
Vowel	16	53.33

larger between the aural perception and the consonant segment analysis having 37.5% as the minimum and 100% as the maximum. However, the range of the positive proportion for the MFCC whole word and vowel segment analysis with aural perception was restricted to 50–81.2%. Indicating there is a lesser reliability in the correlation of the aural perception and MFCC method, if the constant segment is involved in the MFCC extraction than when more points of MFCC (whole word) and the predominant vowel segment is considered;[28] it have also pointed that the interspeaker cepstral distance is reliable when vowel segment was used than the consonant segment for MFCC extraction.

As there was greater variation between the positive and negative proportions for identification, there was no kappa agreement observed between the methods and the values were negative. There was no significant correlation between the MFCC analysis and the five judges' correct identification as $P > 0.05$. The kappa

Table 4: The percentages of positive proportion between the aural perceptual judgment and Mel-frequency cepstral coefficient method

Segments	Judge I (%)	Judge II (%)	Judge III (%)	Judge IV (%)	Judge V (%)
Whole word	50	81.2	62.5	68.8	62.5
Consonant	37.5	100	62.5	62.5	87.5
Vowel	62.5	68.8	62.5	81.2	50

Table 5: The percentages of negative proportion between the aural perceptual judgment and Mel-frequency cepstral coefficient method

Segments	Judge I (%)	Judge II (%)	Judge III (%)	Judge IV (%)	Judge V (%)
Whole word	25	18.8	25	31.2	37.5
Consonant	31.8	22.7	31.8	27.3	36.4
Vowel	14.3	21.4	28.6	21.4	35.7

value (k) and significant *P* value between the MFCC and five judges correct identification are given in Table 6.

We also checked for the variation in number of correct identification and the agreement between the methods if/ðð/segment was analyzed as whole, initial consonant segment, and later vowel segment when subjected to Euclidean distance calculation. In our present study, the percentage of correct identification was only 46.7% for whole word analysis, 26.7% for the initial consonant segment, and 53.3% for the later vowel segment MFCC extraction. Under undisguised condition, the literature reports that MFCC yields the accuracy of 90% and above[15,23,31] in speaker identification. Under the mismatched sample comparison such as presence of noise and speech sample recorded from mobile, the performance of the objective method in speaker identification is reported to be poorer than the aural perception.[29]

The lesser percentage of correct speaker identification in MFCC method could be attributed to three factors: (i) Disguised condition, (ii) large sample size, and (iii) segment chosen for MFCC extraction. There are no other studies which have compared the methods of aural perception and MFCC in an externally induced disguised condition using handkerchief while recording. In the presence of disguise such as use of handkerchief externally on the microphone would tend to alter the amplitude of the speaker's voice and also use of handkerchief on the microphone induces distortion to the spectrum of the speech sample recorded and this would vary depending on the factors such as thickness and material of the kerchief used while recording. During analysis, it was noted that in the calculation of Euclidean distance, the amplitude (Y-axis value) was one of the contributing units which increased the distance between the MFCC of the undisguised and disguised sample of the same speaker as the handkerchief reduced the amplitude of the speech during disguised condition recording, which lead to incorrect identification. Sample size in our study was 30; it has been noted that speaker recognition systems based on pitch extraction perform well when the numbers of speakers are small[15] and the performance significantly decreases when the number of speakers increases.[32] This could have been another contributing factor for the lesser identification accuracy. The consonants are acoustically unstable than the vowels and thus, consonants will have greater variability in the same speakers for the second utterance; this can be observed in the present study that when later segment consisting of predominantly the vowel segment, the percentage of correct identification was 53.3% which was higher than the whole word (46.7%) and initial consonant segment (26.7%) analysis.

Table 6: The kappa value (κ) and significant value (*P*) between the Mel-frequency cepstral coefficient method and five judges correct identification

	Kappa value (κ)	Significance value (*P*)
MFCC and Judge I	−0.67	0.657
MFCC and Judge II	−0.192	0.201
MFCC and Judge III	−0.085	0.602
MFCC and Judge IV	−0.198	0.141
MFCC and Judge V	−0.179	0.310

All the kappa value is negative and not significant. MFCC: Mel-frequency cepstral coefficient

Based on the above results, it can be concluded that aural perception is having greater percentage of correct identification than MFCC method, in speaker identification from disguised speech. Although aural perception was having better percentage of correct identification than MFCC, there is no agreement between the methods (k < 0). MFCC extraction

from vowel segment has better accuracy for speaker identification.

Conclusion

The aural perception is having greater percentage of correct identification than MFCC method in speaker identification from disguised speech. There is no agreement between the methods in correct speaker identification. MFCC extraction from vowel segment has better accuracy for speaker identification. Future implication is to focus on speaker identification in different disguised conditions.

Financial support and sponsorship
Nil.

Conflicts of interest
There are no conflicts of interest.

References

1. Rodman R. Speaker Recognition of Disguised Voices: A Program for Research. In Proceedings of the Consortium on Speech Technology in Conjunction with the Conference on Speaker Recognition by Man and Machine: Directions for Forensic Applications, Ankara, Turkey: COST250 Publishing Arm; 1998. p. 9-22.
2. Clark J, Foulkes P. Identification of voices in electronically disguised speech. Int J Speech Lang Law 2007;14:195-221.
3. Künzel HJ, Gonzalez-Rodriguez J, Ortega-García J. Effect of Voice Disguise on the Performance of a Forensic Automatic Speaker Recognition System. In ODYSSEY04-The Speaker and Language Recognition Workshop; 2004.
4. Hecker MH. Speaker Recognition: An Interpretive Survey of the Literature. American Speech and Hearing Association; January, 1971.
5. Stevens KN, Williams CE, Carbonell JR, Woods B. Speaker authentication and identification: A comparison of spectrographic and auditory presentations of speech material. J Acoust Soc Am 1968;44:1596-607.
6. Atal BS. Automatic speaker recognition based on pitch contours. J Acoust Soc Am 1969;45:309.
7. Noll AM, Spectrum ST. Cepstrum techniques for vocal-pitch detection. J Acoust Soc Am 1964;36:296-302.
8. Hollien H. The acoustics of crime: The new science of forensic phonetics. J Acoust Soc Am 2002;90:1703-4.
9. Kuwabara H, Sagisak Y. Acoustic characteristics of speaker individuality: Control and conversion. Speech Commun 1995;16:165-73.
10. Lakshmi P, Savithri SR. Benchmark for Speaker Identification Using Vector F1 & F2. Proceedings of the International Symposium, Frontiers of Research on Speech & Music. FRSM; 2009. p. 38-41.
11. Wolf JJ. Efficient acoustic parameters for speaker recognition. J Acoust Soc Am 1972;51:2044-56.
12. Atkinson JE. Inter- and intraspeaker variability in fundamental voice frequency. J Acoust Soc Am 1976;60:440-6.
13. Markel JD, Davis SB. Text-independent speaker recognition from a large linguistically unconstrained time-spaced data base. IEEE Trans Acoust Speech 1979;27:74-82.
14. Fakotakis N, Tsopanoglou A, Kokkinakis G. A text-independent speaker recognition system based on vowel spotting. Speech Commun 1993;12:57-68.
15. Atal BS. Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification. J Acoust Soc Am 1974;55:1304-22.
16. Reynolds DA. Speaker identification and verification using Gaussian mixture speaker models. Speech Commun 1995;17:91-108.
17. Rabiner LR, Juang BH. Fundamentals of Speech Recognition (Prentice Hall PTR. Englewood Cliffs, New Jersey; 1993.
18. Kiukaanniemi H, Siponen P, Mattila P. Individual differences in the long-term speech spectrum. Folia Phoniatr (Basel) 1982;34:21-8.
19. Luck JE. Automatic speaker verification using cepstral measurements. J Acoust Soc Am 1969;46:1026-32.
20. Furui S. Cepstrsl analysis techniques for automatic speaker verification. IEEE Trans Acoust Speech 1981;29:254-72.
21. Li KP, Wrench EH Jr. Text-independent speaker recognition with short utterances. J Acoust Soc Am 1982;72:S29-30.
22. Higgins AL, Wohlford R. A New Method of Text-independent Speaker Recognition. From Proceedings of the Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'86; 11 April, 1986. p. 869-72.
23. Che C, Lin Q. Speaker Recognition Using HMM with Experiments on the YOHO Database. Eurospeech; 1995. p. 625-8.
24. Jakkar SS. Benchmark for Speaker Identification using Cepstrum [Unpublished Independent Diploma Project]. University of Mysore; 2009.
25. Crime Rate in India. National Crime Bureau Available from: http://www.en.wikipedia.org/wiki/Crime_in_India. [Last updated on 29 May 2016; Last cited on 2016 Feb 8].
26. The Oxford Corpus 2006. Available from: https://en.wikipedia.org/wiki/Most_common_words_in_English. [Last Cited on 2011 Dec 20].
27. Amino K, Arai T, Sugawara T. Effects of the phonological contents on perceptual speaker identification. In: Springer Berlin Heidelberg: Speaker Classification II; 2007. p. 83-92.
28. Amino K, Arai T. Effects of linguistic contents on perceptual speaker identification: Comparison of familiar and unknown speaker identifications. Acoust Sci Technol 2009;30:89-99.
29. Alexander A, Botti F, Dessimoz D, Drygajlo AN. The effect of mismatched recording conditions on human and automatic speaker recognition in forensic applications. Forensic Sci Int 2004;146:S95-9.
30. Hasan MR, Jamil M, Rahman MG. Speaker Identification Using Mel Frequency Cepstral Coefficients Variations; December, 2004. p. 1-4.
31. Barry RM, Savithir SR. Speaker Identification from Electronically Disguised Voice. Proceedings of the 45th ISHA CON Chennai: Indian Speech and Hearing Association-Tamil Nadu. 1st-3rd February, 2013. p. 308-9.
32. Ezzaidi H, Rouat J, O'Shaughnessy D. Towards Combining Pitch and MFCC for Speaker Identification Systems. Proceedings of the Seventh European Conference on Speech Communication and Technology (Eurospeech 2001), Aalborg, Denmark; September, 2001. p. 2825-8.