Auditory and Acoustic Features from Clue-Words Sets for Forensic Speaker Identification and its Correlation with Probability Scales

Babita Bhall[1*], Singh CP[2], Rakesh Dhar[3] and Rajesh Soni[1]

[1]Physics Division, Forensic Science Laboratory, Madhuban, Karnal, Haryana, India

[2]Physics Division, State Forensic Science Laboratory, Delhi, India

[3]Department of Applied Physics, Guru Jambeshwar University of Science & Technology, Hisar, India

[*]**Corresponding author:** Babita Bhall, Forensic Science Laboratory Physics Forensic Lab, Madhuban 132037, Karnal, Haryana, India, Tel: 7206690795; E-mail: babitabhall@gmail.com

Abstract

An experiment carried out on speaker identification by semi-automatic measurement of parameters with the goal of collaborating numerical data as well as descriptive data with that of probability scales. The 15 sets of speech samples of 15 speakers selected randomly from 100 actual crime cases, in Hindi utterances for purpose of speaker identification test subjected to spectrographic analysis. Speaker specific acoustic parameters, namely 1st formant frequency at a particular location (F1); 2nd formant frequency at a particular location (F2); and 3rd formant frequency at a particular location (F3) measured for the set of speech samples for all the 15 speakers. Also, the auditory analysis based on the linguistic features and phonetic features noted of each of the 15 sets of speech samples. We developed software to calculate the similarity percentage for the numerical data measured on the basis of acoustic analysis and numerical values assigned to auditory parameters on the basis of auditory analysis computed according to one of the nine probability scales. Most of the existing methods take only acoustic features to obtain numerical results for the purposes of speaker identification.

Keywords: Speaker identification; Spectrographic analysis; Acoustic parameters; Frequency; Auditory parameters; Subjective probability; Objective probability

Introduction

The fast increasing development and penetration of communication technology surely helped humankind in better, accessible and efficient communication but it is not without its ill consequences. Information and Communication technology (ICT) has also helped anti-social elements in committing more organized and white collar crimes, and in turn, law enforcement agencies should be better equipped with advanced technology to counter or deal with such crimes. Speaker Identification technology is one of the many tools which our law enforcement agencies could rely upon and it is also a popular identification technique used for monitoring and authenticating human subjects using their speech signal. But the question often arises about how to involve auditory parameters in numerical terms in the results of speaker identification [1,2].

Kersta [3] postulated that spectrograms of the given individual were unique, individualistic and permanent. Kinston, et al. [4] discussed the various aspects of the utilization of statistics in criminalistics. Tosi, et al. [5] in one of their experiment professed that the differences between inter-speaker and intra-speaker variability, which according to them, stems mainly from the anatomical differences in vocal tract but these did not correlate or quantified with specific acoustic parameters. Koenig [6] in 1986 conducted a survey with FBI to determine the error rate of spectrographic identification of voice and compared spectral pattern of two voice samples by comparing formant shaping, beginning, mean and end formant frequency, timing, pitch etc. of each and every individual word [7]. Meuwly et al. [8] used likelihood ratio to suggest the worth of evidence of questioned recording and measured that how this recording scores for suspected speaker models when compared to relevant non-suspected speaker models. Kinoshita [9] investigates about the possibility to perform forensic speaker identification under forensically realistic data, with traditional acoustic parameters [10] and Bayesian Likelihood Ratio and concluded that the likelihood ratio based discriminant test was found to be one of the effective ways in evaluating speech data. Aitken [11] used Bayesian approach to combine objective and subjective probabilities in one formula. Singh [12] in 2005 studied isolated spoken words with similar vowel quality as syllabic nuclei preceded by consonants of similar place of articulation as clue-words for forensic importance. As published by Singh [13] in 2007 in "An Introduction to Forensic Speaker Identification Procedure", the verbal probability scales are positive identification, identification with high probability, probable identification, possible identification, no opinion, possible elimination, probable elimination, elimination with high probability elimination and positive elimination. Becker, at al. [14] applied UBM-GMM verification system to semi-automatically extracted formant features and concluded two important advantages:

- Dimensionality of feature vector space was small and,
- There was direct relationship between the configuration of vocal tract not only as an average, but also with the speaker-specific variations which were expressed in the entire distribution.

Nowadays, mostly numerical data obtained with the help of acoustic features is computed against the probability scales; surely, it is of immense help to the courts of law. The scope of the present study is to combine the acoustic features with that of the auditory parameters so that one new paradigm may be established and the law enforcement agencies may do away with separately looking subjective and the

objective data. In this paper, an experiment of 15 actual crime cases conducted on speaker identification by using semi-automatic method of measurement of acoustic parameters [15].

Experimental Methods

Sampling of speech material: A set of clue-words collected from the database of clue-words of actual crime samples accessed from State Forensic Laboratories of Haryana and Delhi (India). These sets of clue-words were collected without noting the particulars of the cases. Clue-words are the words that are similar having same vowel quality selected from questioned as well as specimen speech sample from the set of verbatim/non-verbatim words. The set of clue-words of a speaker consists of different vowels, namely, /ʌ/, /ɛ/, /ɑ/, /æ/, /a/, /u/, /i/, /ɔ/, /o/, /e/ & /ə/ preceded and succeeded by different consonants uttered at similar places of articulation. The selection of sets of clue-words for an informant should be in such a manner that the vowels as a nuclei must be present in different words. The clue-words selected specifically, used to study various acoustic, linguistic and phonetic features from contextual text. These sets of clue-words from questioned and the specimen speech samples for speakers selected from among the randomly selected 15 informants (accused/ suspect/ complainant) in Hindi/ English/ Punjabi utterances having male and female ratio as 12:3 in varying age groups from 15-60 years mostly taken from the northern parts of India. The questioned and specimen speech samples

considered for this study for the purposes of speaker identification taken on one-to-one basis in the closed sets with options of acceptance and rejection. These sets of speech samples selected for speakers whose questioned recording had been made through mobile network, landline and also from direct recording. The speech samples of questioned as well as specimen samples digitised at sampling rate of 22050 Hz and 16 bit quantization in mono signed.

Experiment

The sets of clue-words of 15 speakers for questioned as well as of specimen speech samples subjected to spectrographic analysis using computerized speech lab (CSL). The formant frequencies (F1, F2 and F3) were analyzed using LPC (Linear predictive Coding) at a particular location of vowel nuclei. The result of this analysis found to be tabulated for one speaker in Table 1. The linguistic and phonetic features tabulated for one speaker as in Table 2. The auditory analysis contributed the phonetic and linguistic features of the speaker's specific characteristic and spectrographic analysis provided the acoustic features of the speaker's utterances. These features subjected to the software specifically developed for statistical evaluation and a correlation of phonetic, linguistic and acoustic events to that of probability scales used in the procedure for speaker identification test [16].

English Transcription of Hindi words	Word	Nuclei vowel	Questioned			Specimen		
			F1(Hz)	F2(Hz)	F3(Hz)	F1(Hz)	F2(Hz)	F3(Hz)
thik	tik	/i/	300	2253	2872	300	2253	2872
hai	jɑ	/ɑ/	522	1567	3801	522	1567	3801
kal	kʌl	/ʌ/	406	1431	2369	406	1431	2369
nai	nʌːi	/ʌ/	600	1779	2611	600	1779	2514
nai	nʌːi	/i/	290	1248	2273	290	1248	2273
paise	pʌɛ	/ʌ/	542	1741	2476	542	1741	2476
paise	pʌɛ	/ɛ/	493	1847	2611	493	1847	2611
paise	sɛ	/ɛ/	503	1712	2311	503	1712	2311
dal	dɑl	/ɑ/	648	1547	2514	648	1547	2514
paya	pɑ	/ɑ/	590	1431	2573	609	1431	2573
paya	jɑ	/ɑ/	571	1547	2631	571	1547	2631
koi	kɔi	/i/	426	1499	3191	426	1499	3191
bat	bɑt	/ɑ/	696	1402	2485	696	1402	2485
bahut	bʌhut	/ʌ/	455	1006	2505	455	1006	2505
bahut	bʌhut	/u/	416	1344	2389	416	1344	2389
inquiry	ink	/i/	542	2050	2466	542	2050	2466
inquiry	kvɑjri	/ɑ/	629	1470	2476	629	1470	2476
inquiry	kvɑjri	/i/	309	2292	2582	309	2292	2582
tum	tum	/u/	387	1286	2553	387	1286	2553

kahan	kɑhɑn	/ɑ/	600	1335	2543	600	1335	2543
kahan	kɑhɑn	/ɑ/	464	1625	2698	464	1625	2698
ghar	ghʌr	/ʌ/	667	1499	3791	667	1499	3791
subah	su	/u/	348	967	2282	329	948	2282
subah	bʌh	/ʌ/	522	1277	2456	522	1277	2456
pura	pu	/u/	358	1064	2689	358	1064	2689
pura	rɑ	/ɑ/	803	1538	2543	803	1538	2543
are	ʌːre	/ ʌ/	619	1576	3791	619	1576	3791
are	ʌːre	/e/	445	1828	2573	445	1828	2573
office	ɔːfis	/i/	406	1944	2524	377	1944	2543
kam	kɑm	/ɑ/	629	1325	2466	629	1325	2466
chori	tsɔri	/i/	319	2137	2514	319	2137	2514
aap	ɑːp	/ɑ/	735	1257	2534	735	1257	2534

Table 1: Features extracted for a set of clue-words for one speaker.

Questioned	Specimen
Stylistic Impression-Normal	Stylistic Impression-Normal
Delivery of speech-Medium	Delivery of speech-Medium
Phonation- Medium	Phonation-Medium
Physiological pitch level-Medium	Physiological pitch level-Medium
Flow of speech (qualitative)-Easy	Flow of speech (qualitative)-Easy
Flow of speech (quantitative)-Normal Fluent	Flow of speech (quantitative)-Very Fluent
Plosive Formation-Medium	Plosive Formation-Medium
Nasality-Normal	Nasality-Normal
Intonation pattern-Level	Intonation pattern- Level
Dynamic of Loudness-Medium	Dynamic of Loudness- Medium
Speech Rate-Medium	Speech Rate- Medium
Speech Variation-Medium	Speech Variation- Medium
Striking time features-Compression of words/ Compression of Statement	Striking time features- Compression of words/ Compression of Statement
Pauses-Normal	Pauses-Normal

Table 2: Linguistic and phonetic features noted for one speaker.

The software for statistical evaluation and correlation of verbal probability scales designed in such a manner that the auditory characteristics and acoustic features got combined in statistical operation using Bayes method i.e. after taking values of formant frequencies (F1, F2 and F3) i.e. acoustic features from questioned as well as from the specimen speech samples and comparing them for similarities only, a similarity percentage calculated and the same thing done in the case of linguistic and phonetic features [17]. These numerical results obtained from objective as well as subjective data got combined by providing different weightages to each of them in such a manner that objective data gets more credential in comparison to the subjective one and then computed against one of the verbal probabilities scales according to the criteria.

Results and Discussions

Figures 1 and 2 show waveform with phonetic transcript of words / tik/, / jɑ/, /kʌl/ and /nʌːi/; their respective spectrogram with formant

marking and their respective LPCs [18]. Formant frequencies (F1, F2 and F3) measured for all the 15 speakers separately. For each set of clue-words, with the help of formant frequencies (F1, F2 and F3), percentage of similarity/dissimilarity calculated using the software.
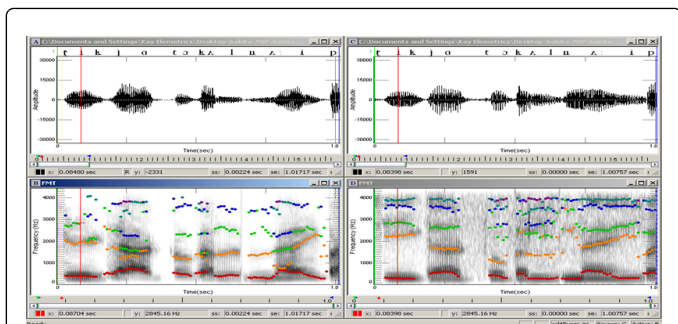

Figure 1: Waveform with phonetic transcript of words /tik/, / jɑ/, / kʌl/ & /nʌ:i/ in window A and B, their respective spectrogram with formant marking in windows C and D.

Weightage values assigned to the linguistic and phonetic features also determined and percentage of similarity/dissimilarity based on that value calculated using the software modules. Then the combination of objective and subjective probabilities of acoustic and auditory parameters calculated.

The results of all the percentages tabulated in Table 3. Similarity percentage of more than 90% having more than 3 usable formants and more than 20 matching clue-words/word segments yields positive identification. Whereas, similarity percentage of more than 80% having 2 or more usable formants and more than 15 matching clue-words/word segment produces probable identification. In the result, tabulated in Table 3, 12 out of 15 produces probable identification and 3 out of 15 yields positive identification [19]. Figure 3 shows the output of the software for one speaker.
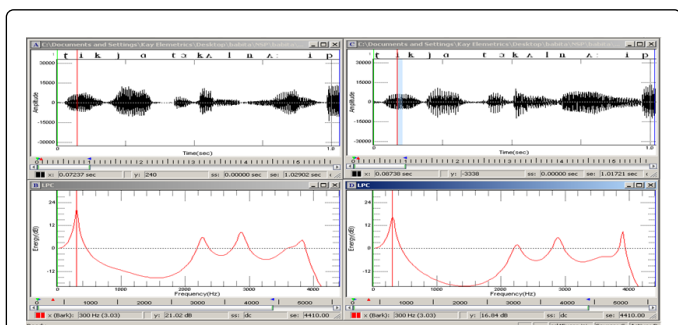

Figure 2: Waveform with phonetic transcript of words /tik/, / jɑ/, / kʌl/ & /nʌ:i/ in window A and B, their respective LPC in windows C and D.

The probability scale derived verbally as per the criteria based on number of clue-words, number of formants and the percentage of matching of both acoustic and auditory features used in most of the methods of the speaker identification tests by semi-automatic measurement of acoustic parameters [20,21].
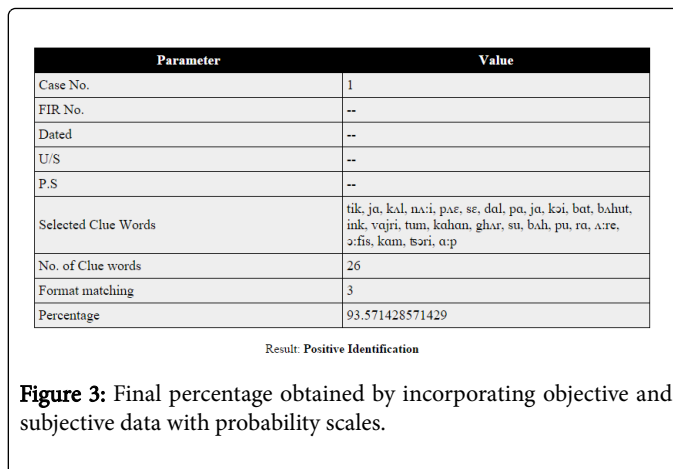
Parameter	Value
Case No.	1
FIR No.	--
Dated	--
U/S	--
P.S	--
Selected Clue Words	tik, jɑ, kʌl, nʌ:i, pʌɛ, se, dɑl, pɑ, jɑ, kɔi, bɑt, bʌhut, ink, vɑjri, tum, kɑhɑn, ghʌr, su, bʌh, pu, rɑ, ʌ:re, ɔ:fis, kɑm, tsɔri, ɑ:p
No. of Clue words	26
Format matching	3
Percentage	93.571428571429

Result: **Positive Identification**

Figure 3: Final percentage obtained by incorporating objective and subjective data with probability scales.

Speaker	Number of Clue Words/ Word Segments	Number of Formants	Percentage (%)	Probability
1	26	3	93.57	Positive Identification
2	27	3	82.51	Probable Identification
3	29	3	89.42	Probable Identification
4	33	3	90.42	Positive Identification
5	38	3	84.3	Probable Identification
6	38	3	86.66	Probable Identification
7	39	3	83.06	Probable Identification
8	21	3	86.19	Probable Identification
9	18	3	82.27	Probable Identification
10	19	3	82.59	Probable Identification
11	41	3	85.33	Probable Identification
12	32	3	91.37	Positive Identification
13	37	3	85.56	Probable Identification
14	31	3	83.08	Probable Identification
15	33	3	86.47	Probable Identification

Table 3: Results of all the 15 speakers with similarity percentage after combining objective and subjective similarity percentage.

Unlike the verbal probability scales derived based on the criteria, the results of the present study which is based on by combining the

auditory features and the acoustic parameters is found promising and robust in the sense that the correlation of statistical probability by a combination of objective and subjective probabilities provided the percentage of similarity. May the threshold value is possible to be fixed for percentage of similarity on the basis of the results of the combined objective and subjective probabilities obtained from this study for the larger database.

Conclusion

New methods like automatic speaker identification, text independent speaker identification found to be developed in the recent past. However, these methods of automatic speaker identification did not achieve the desired acceptability in/by scientific community for application in criminal justice system. The semi-automatic measurement of acoustic parameters combined with auditory features continues to be an acceptable technique for speaker identification by experts. The decision drawn based on the criteria of the number of clue-words, number of formants and percentage of matching conclude the verbal probability scales.

The result of the present study is promising as well as its importance lies in for deriving a numerical probability by combining the objective with the subjective probabilities. Combining these probabilities and its presentation in the courts of law will be more feasible than verbal probability. Extending this study with all the 100 speakers will be helpful in coming out with a threshold value of probability in the verbal probability scales.

References

1. Smrkovski L (1975) Study of Speaker Identification by Aural and Visual Examination of Non-Contemporary Speech Samples. J Assoc Off Anal Chem 59: 927-937.

2. Jurafsky D, Martin JH (2014) Speech and Language Processing, Pearson.

3. Kersta LG (1978) Voice Print Identification infallibility. J Acoust Soc Am 34.

4. Kirk PL, Kingston CR (1964) Evidence Evaluation and Problems in General Criminalistics. J Forensic Sci 9: 434-444.

5. Black J, Lashbrook W, Nash E, Oyer H, Pedry C, et al. (1972) Reply to Speaker Identification by Speech Spectrograms: Some Further Observations. J Acoust Soc Am 54: 535.

6. Koenig B (1986) Spectrographic Voice Identification. A forensic survey. J Acoust Soc Am 79: 2088-2090.

7. Haji Hosseini H, Amir Fattahi R, Gazor S, Soltanian-Zadeh H (2016) Robust Estimation and Tracking of Pitch Period using an Efficient Bayesian Filter. IEEE Transactions on Audio, Speech and Language Processing 24: 1219-1229.

8. Meuwly D, Drygazlo A (2001) Forensic Speaker Recognition based on Bayesian Framework and Gaussian Mixture Modelling (GMM). The Speaker Recognition Workshop Crete, Greece.

9. Kinoshita Y (2002) Use of Likelihood Ratio and Bayesian Approach in Forensic Speaker Identification. Proceedings of the 9th Australian International Conference on Speech Science & Technology, Melbourne, Australia.

10. Wolf JJ (1972) Efficient Acoustic Parameters for Speaker Recognition. J Acoust Soc Am 51: 2044-2057.

11. Kirk PL, Kingston CR (1964) Evidence Evaluation and Problems in General Criminalistics. J Forensic Sci 9: 434-444.

12. Singh CP, Manisha K (2005) Syllabic Nuclei of Similar Vowel Quality as a Clue for Forensic Speaker Identification – A Study on the isolated spoken words. The Indian Police Journal.

13. Singh CP (2007) An Introduction to Forensic Speaker Identification Procedure. Advance Interactive Training Course on Forensic Speaker Recognition, Directorate of Forensic Science, Ministry of Home Affairs, Govt of India.

14. Becker T, Jessen M, Grigoras C (2008) Forensic Speaker verification using Formant Features and Guassian Mixture Models. INTERSPEECH 2008, 9th Annual Conference of the International Speech Communication Association, Brisbane, Australia.

15. Hazen B (1973) Effects of Differing Phonetic Contexts on Talker Identification. J Acoust Soc Am 54: 650-658.

16. Singh CP, Singh SR, Singh R (1997) Voice Spectrographic Studies of Disguise in Hindi of Kashmiri Speaker and the effect on Bilabial, Nasal, Dental/Alveolar, fricatives and the following Vowels, CBI Bulletin.

17. Sambur MR (1975) Selection of Acoustic features for Speaker Identification. IEEE Trans on Acoustic, Speech and signal processing 23: 176-182.

18. Garner PN (2011) Cepstral normalisation and the signal to noise ratio spectrum in automatic speech recognition. Speech Communications 53: 991-1001.

19. Rose P (2005) Forensic Speaker Recognition at the Beginning of the Twenty· First Century. An Overview and a Demonstration. Aust J Forensic Sci 37: 47-92.

20. Anguera X, Bozonnet S, Evans N, Fredouille C, Friedland G, et al. (2012) Speaker Diarization: A Review of Recent Research. IEEE Transactions on Audio, Speech and Language Processing 20: 1-15.

21. Besson O, Dobigeon N, Tourneret JY (2014) Joint Bayesian Estimation of Closed Subspaces from Noisy Measurements. IEEE Signal Processing Letters 21: 1-10.