Research Article

Combining Acoustic Parameters and Auditory Features Using Bayes Theorem and Establishing Their Correspondence with the Probability Scales by Semi-Automatic Methods in Forensic Speaker Identification

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

[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, Physics Division, Forensic Science Laboratory, Madhuban, Karnal, Haryana, India. Tel: +917206690795; Email: babitabhall@gmail.com

Abstract

Evolution has resulted in humans and non-humans creating a wide range of sounds used to warn of danger, find mates, and communicate. Only humans are able to produce set of unique distinguishable sounds called, phonemes. One emerging field of forensic investigation is using acoustic parameters and auditory features to conduct speaker identification between known patterns to unknown samples. Then using the method of Bayes' Theorem, determines the probability statements of similarity. In this paper, we consider two sets of speech samples, questioned and the other is a known specimen speech sample obtained randomly from the actual crime cases. The two speech samples underwent to spectrographic analysis and statistically compared using the Formant Frequencies (F1, F2 & F3) at particular locations. The percentage of similarities between the unknown (Questioned) and the known specimen were ascertained by formant frequencies (acoustic parameters), and for numerical values assigned to the descriptive data (auditory features). Bayes' Theorem was used to combine objective probability obtained from the acoustic parameters and subjective probability obtained from the auditory features. These values computed against one of the nine probability scales with the help of the software developed by the author. The study showed that this method can be used to compare unknown voice samples to known samples and assign probability statements of similarity.

Introduction

As a consequence of evolution, humans, animals and birds have all developed the ability to produce different types of sounds which allows them to understand signs of danger, locate a mate, and communicate higher forms of thought. Primates had an advanced system of communication that includes vocalization, hand gestures and body language. Humans unlike other primates have the ability to articulate sounds to produce a set of distinguish sounds, called phonemes and this has led to a development of language. This ability of humans separates us from our less evolved cousins which are unable to articulate though they can produce vowel like sounds.

Human and a large number of animals have the ability to identify others by listening to the sounds of their voice. The degree of accuracy with which identification is performed under all sorts of conditions still remains under question especially under the purview of Forensic Science, which deploys the relative usefulness of spectrograms as a supplement to careful listening. There are two basic methods of speaker identification, one is a subjective method

where identification of the individual making the sounds is made by the human mind and the second one is an objective method where the identification is determined by mechanical or electronic means. The most general acoustic parameters of speech include (1) time, (2) formant frequencies and (3) intensity distribution within all bands of frequency simultaneously present in the instantaneous speaker output; in other words, the parameters portrayed by a spectrogram. Comparisons of these general or derived spectral/temporal parameters are the basis of all speaker identification systems, both subjective and objective. One source of variation of these spectral parameters depends on phonetic content; in systems of Speaker identification, in which it is desirable to minimize the phonetic source of variability. Comparisons of a known vocal pattern to unknown voices are performed by using similar sounded words produced by the speakers. The problem is that, even maintaining a similar set text, values of the selected acoustic parameters will differ not only among different talkers called interspeaker variability, but also vary within the same talker called intraspeaker variability if different utterances of the same words are compared. Researchers have found different parameters from the sets of clue-words obtained from questioned as well as the specimen speech samples i.e., parameters that convey the least intraspeaker variability and the most interspeaker variability possible in all conditions that may occur in normal or even in disguised speech [1-4]. Various studies have been conducted on speaker dependent parameters are described in the literatures [5-7]. Different studies have been conducted regarding the statistical interpretation of the evidence obtained during the course of a criminal investigation, using the Bayes' theorem [8-11]. Comprehensive studies for speaker identification procedures, methods and linking the statistical results to a probability scales was conducted in 2002, 2005 and 2016 [12-14].

In this paper, a comparative study was conducted comparing a questioned (unknown) speech sample with that of a known sample using formant frequencies (F1, F2 & F3), also known as acoustic parameters, and auditory features and then combining them both using auditory features (subjective probability) and acoustic features (objective probability) to calculate the final similarity probability. The author and her team developed a new method using Bayes' Theorem and utilizing new software for the purposes of calculating probability value of similarly between the two voice patterns using the 1-9 probability scales.

Experimental Methods

Sampling of Speech Material: A set of clue-words for questioned as well as specimen sample were extracted and prepared from text uttered by the suspect while asking for bribery (as this is the text dependent technique). The sets of clue-words contained different type of vowels, namely, /æ/, /i/, /ɑ/, /o/, /u/, /ʌ/, /ɔ/ and /ɛ/ which is either preceded or succeeded by the consonants CVC, VC, or CV uttered at similar places of articulation. Selected clue-words are used to extract and study the acoustic parameters i.e. first Formant Frequency (F1) at particular location; second Formant Frequency (F2) at particular location; third Formant Frequency (F3) at particular location and a number of auditory features. This particular speaker was selected randomly from among the data base of actual crime samples. Questioned speech sample has been prepared from the recording present in the mobile and specimen speech sample has been prepared from the direct recording in the laboratory. Both these samples are digitized at sampling rate of 22050 Hz and 16 bit quantization in mono signed.

Experiment: A Set of clue-words were subjected to a spectrographic analysis using the Computerised Speech Lab (CSL-4500). The auditory parameters (F1, F2 & F3) at particular location of vowel nuclei were measured. Auditory features comprised of linguistic and phonetic features were collected. The data was entered into the software developed by the authors which calculate their similarity percentages and weighing objective and subjective data differently using Bayes' theorem.

Results and Discussions

The results of the acoustic parameters (F1, F2 & F3) at particular location of vowel nuclei are tabulated in Table 1. Auditory features comprised of linguistic and phonetic features are shown in the observation sheet in Figure 2. Figure 1 shows the intonation pattern with formant markings of the words /kæsis/, /mʌin/, /ho/ & /dʒɑtɑ/ and LPC of the vowel /æ/ showing the value of its First Formant Frequency (F1 = 503 Hz). Similarly, values of Second Formant Frequency (F2) and Third Formant Frequency (F3) were also measured. Values for Formant Frequencies (F1, F2 & F3) were measured and calculated for other vowels and their values were measured for questioned as well as specimen samples.

English Transcription of Hindi words	Word	Nuclei vowel	QUESTIONED			SPECIMEN		
			F1(Hz)	F2(Hz)	F3(Hz)	F1(Hz)	F2(Hz)	F3(Hz)
cases	kæsis	/æ/	503	2079	3511	503	2079	4023
cases	kæsis	/i/	464	1721	2273	464	1721	2273
main	mʌin	/ʌ/	522	1683	2224	522	1683	2224
main	mʌin	/i/	503	2021	3878	503	2021	3878
ho	ho	/o/	455	1683	2379	455	1683	2379

jata	ʤata	/ɑ/	729	1586	3772	729	1586	3772
jata	ʤata	/ɑ/	619	1828	2514	619	1692	2514
depend	dipɛnd	/i/	464	1625	2215	464	1625	2215
depend	dipɛnd	/ɛ/	580	1896	2418	580	1896	2418
karta	kʌrta	/ʌ/	619	1712	2843	619	1712	2398
karta	kʌrta	/ɑ/	716	1625	2398	716	1625	2398
hai	hʌi	/ʌ/	590	1896	2398	590	1896	2398
hai	hʌi	/i/	522	2002	2340	522	2002	2340
do	dɔ	/ɔ/	445	2311	3182	445	2311	3182
teen	tin	/i/	416	2195	2689	416	2195	2863
se	sɛ	/ɛ/	455	1654	2437	455	1422	2437
upar	upʌr	/u/	416	1044	2602	416	1044	2602
upar	upʌr	/ʌ/	493	1470	2456	493	1470	2456
namaskar	nʌmʌʃkar	/ʌ/	522	1238	2273	522	1238	2273
namaskar	nʌmʌʃkar	/ʌ/	522	1344	3482	522	1344	3482
namaskar	nʌmʌʃkar	/ɑ/	542	1586	3714	542	1586	3714
uncle	unkʌl	/u/	1663	2592	3598	1663	2592	3598
uncle	unkʌl	/ʌ/	513	1576	2408	513	1576	2408
ji	ʤi	/i/	377	2485	3849	377	2137	3849
ai	ɑi	/ɑ/	638	1499	3830	638	1683	3830
ai	ɑi	/i/	551	1808	3791	551	2050	4043
ar	ɑr	/ɑ/	542	2021	3704	542	2021	3984
ho	ho	/o/	484	1634	2776	484	1799	2776
jaaegi	ʤajɛgi	/ɑ/	493	1857	3810	493	1857	3810
jaaegi	ʤajɛgi	/i/	426	2331	3994	426	2331	3994

Table 1: Features extracted for a set of clue-words for one speaker.
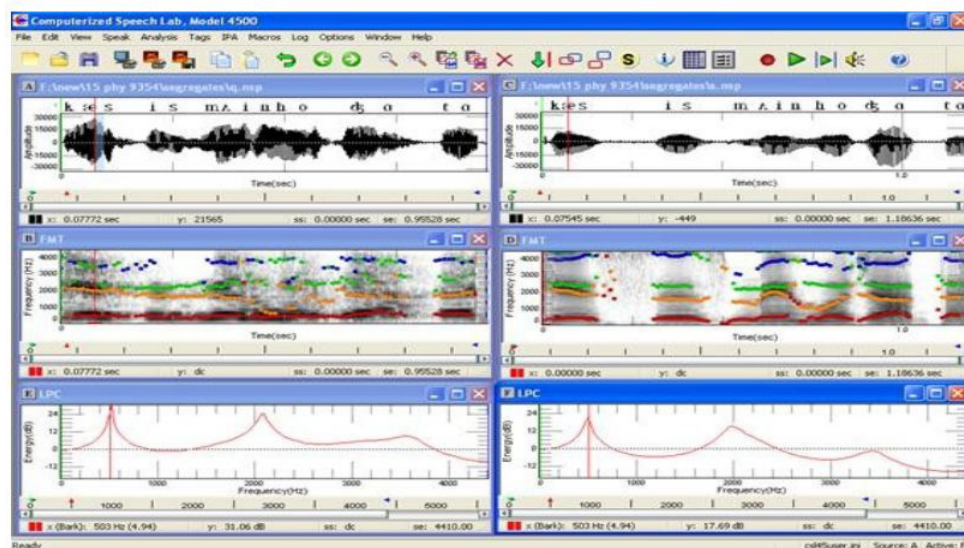
Figure 1: Waveform with phonetic transcript of words /kæsis/, /mʌin/, /ho/ & /ʤɑtɑ/ in window A and C; their respective spectrogram with formant marking in windows B and D &; their respective LPC in windows E and F

Figure 2 shows the final observation sheet with the auditory features for the questioned as well as specimen samples; duration of both samples, clue-words selected for the spectrographic analysis, their final percentage after combining acoustic and auditory parameters by using Bayes' Theorem, number of formants used and the final take on the probability scale.

Figure 2: Observation sheet showing auditory features, duration, selected clue-words, number of formants used of questioned as well as specimen speech sample, final percentage and its correlation on the probability scale.

The probability scale has been calculated by evaluating the final percentage that is composed of a combination of (1) acoustic features and auditory parameters, (2) the number of formants used, (3) and the number of clue-words selected. The software weighs these three factors in calculating final probability of similarity between an unknown sample to a known sample. In this case, the evaluation concluded a 90.2% match and therefore, a Positive Identification.

Conclusion

Based on the result of this study, an unknown voice samples can be compared with known specimen samples to determine the percentage of similarity by combining both the acoustic parameters and auditory features, individually as well as in combination of both using Bayes' Theorem. This method offers promising application in the field of forensic and law enforcement. The current method incorporates subjective probability which has not been used to date. The method provides probability statements for that of the voice of the offender matches with that of a suspect. Once this method has been reproduced by others and determined relatable, it will greatly assist law enforcement agencies and the courts.

The author has plans for future large scale studies consisting of 100 speech samples of questioned as well as specimen samples selected randomly from the data base.

References

1. Endress W, Bambach W, Flosser G (1971) Voice Identification as a function of Age, Voice Disguise and Voice Imitation. J Acoust Soc Amer 49: 1842-1848.

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

3. Holmgren GL (1967) Physcial and Psychological Correlates of Speaker Recognition. Journal of Speech, Language, and Hearing Research 10: 57-66.

4. Mathur S, Chaudhary SK, Vyas JM (2016) Effect of Disguise on Fundamental Frequency of Voice. Journal of Forensic Research 7: 2157-7145.

5. Samber MR (1975) Selection of Acoustic Features for Speaker Identification IEEE Trans n Acoustic, Speech and Signal Processing 23: 176-182.

6. Tosi O, Oyer H, Lashbrock W, Pedey C, Nical J, et al. (1972) Experiment on Voice Identification. Journal of Acoustical Society of America 51: 2030-2043.

7. Wolf JJ (1972) Efficient acoustic parameters for speaker recognition. Journal of Acoustical Society of America 51: 2044 2057.

8. Aitken CGG (2000) Statistical Interpretation of Evidence/Bayesian Analysis. University of Edinburgh, Edinburgh UK: 717-724.

9. Kinoshita Y (2002) Use of Likelihood Ratio and Bayesian Approach in Forensic Speaker Identification. Ratio and Bayesian Approach: 297-302.

10. Meuwly D, Drygazlo A (2001) Forensic Speaker Recognition based on Bayesian Framework and Gaussian Mixture Modelling (GMM). ISCA Archive: 1-6.

11. Besson O, Dobigeon N, Tourneret J-Y (2014) Joint Bayesian Estimation of Closed Subspaces from Noisy Measurements. OATAO 21: 1-4.

12. An Introduction to Forensic Speaker Identification Procedure Advance Interactive Training Course on Forensic Speaker Recognition, CBI Bulletin, Directorate of Forensic Science, Ministry of Home Affairs, Govt. of India. Vol.XIII, No.1, January2005.

13. Kinoshita Y (2002) Use of Likelihood Ratio and Bayesian Approach in Forensic Speaker Identification, School of Languages and International Education, University of Canberra.

14. Bhall B, Singh CP, Dhar R, Soni R (2016) Auditory and Acoustic Features from Clue-Words Sets for Forensic Speaker Identification and its Correlation with Probability Scales. Journal of Forensic Research 7: 1-5.