Research Journal of Forensic Sciences ISSN 2321–1792
Vol. **6(3),** 1-6, April **(2018)** *Res. J. Forensic Sci.*

Effect of combining auditory features with acoustic parameters on the probability scales in forensic speech recognition

Babita Bhall[1*], C.P. Singh[2] and Rakesh Dhar[3]
[1]Physics Division, Forensic Science Laboratory, Madhuban, Karnal, Haryana, India
[2]Physics Division, State Forensic Science Laboratory, Delhi, India
[3]Dept. of Applied Physics, Guru Jambeshwar University of Science and Technology, Hisar, India
babitabhall@gmail.com

Available online at: www.isca.in, www.isca.me
Received 11th January 2018, revised 13th April 2018, accepted 24th April 2018

Abstract

Attempts to understand the phenomenon and mechanism of speech sounds led humans to discover visual representation of them in terms of frequency-time graphs which helps them to understand acoustic parameters, which give the voice of humans 'uniqueness'. One of the emerging field of forensic science is using acoustic parameters and auditory features to perform speaker identification test by comparing known to unknown samples. In this paper, we consider two sets of speech samples, questioned and known specimen speech sample data base obtained from the actual crime cases. The two speech samples underwent to the method of auditory analysis and spectrographic analysis. The percentage of similarities between the unknown sample (Questioned) and the known sample were ascertained by formant frequencies, and for numerical values assigned to the 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 to correlate with one of the nine probability scales with the help of the software programs developed by the authors. This study reveals how the resultant probability changes, if auditory features were also taken into account along with that of the acoustic parameters while calculating the final similarity percentage.

Keywords: Spectrographic analysis, acoustic parameters, formant frequency, auditory parameters, Bayes' theorem.

Introduction

Several efforts have been made since last two centuries with an aim to understand human vocal tract, its functioning and mechanism so that the phenomenon of speech sounds can be apprehended and put to its suitable use. This phenomenon is wuite helpful in cases in which the recorded conversation is the only available evidence present.

Alexander Melville Bell in 1867 was a pioneer in understanding and recognizing the process of speech sounds when he developed a form of visible speech using a type of phonetic symbols.

Latest developments started with the experiments of Grey and Kopp, 1944 when they developed the sound spectrograph as a means of visualizing speech signals. This process was made more robust with the experiments of Kersta in 1962, when he coined the term 'voiceprint'.

The most general acoustic parameters of speech include i. time; ii. formant frequencies; and iii. intensity distribution within all bands of frequency simultaneously present in the instantaneous speaker output. Formant Frequencies are produced of acoustic event in vocal tract, which can be easily altered by the pharyngeal, laryngeal and oral cavity musculature. Comparisons of these general or derived spectral/temporal parameters are the basis of all speaker identification systems. One source of variation of these spectral parameters depends on phonetic context, in which it is desirable to minimize the phonetic source of variability[1-4]. Studies have been conducted on speaker dependent parameters are described in the literatures[5-7]. Various studies have been performed regarding statistical interpretation of the evidence obtained during the course of a criminal investigation and subsequently incorporating that evidence later in the final results, 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 set of questioned speech sample with that of a known speech sample using formant frequencies (F1, F2 and F3) and auditory features.

Combination of auditory features and acoustic parameters to calculate resultant probability was attempted in this study. The effect of combining auditory features with that of the probability derived from the acoustic parameters is the subject matter of this study and is finally correlated with any one of the nine probability scales in Forensic Speaker Recognition.

A new approach was attempted in this work; by using Bayes' Theorem and utilizing the developed program to calculate the resultant probability obtained after combining auditory features and acoustic parameters.

Materials and methods

Sampling of Speech Material: A set of clue-words from questioned as well as specimen sample were obtained and prepared from text uttered by the suspect. 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 as in CVC, VC, or CV uttered auditory similarly. Selected clue-words from questioned and specimen samples are used to extract the frequencies 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 case samples.

Questioned speech sample has been prepared from the recording present in the mobile and specimen speech sample has been chosen from the direct recording. Both of these samples are digitized at the 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 and 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 weighed objective and subjective data differently in the final score using Bayes' Theorem.

Results and discussion

The results of the acoustic parameters (F1, F2 and 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-3. Figure-1 shows the intonation pattern with formant markings of the clue-words. Figure-2 shows Linear Prediction Coding (LPC) of the vowel /e/ showing the value of its First Formant Frequency (F1 = 774 Hz). Similarly, values of Second Formant Frequency (F2) and Third Formant Frequency (F3) were also measured. Values for Formant Frequencies (F1, F2 and F3) were measured for other vowels in the similar manner for questioned as well as specimen speech sample.

Figure-3 shows the final observation sheet with the auditory features for both the questioned and specimen samples; duration of both samples, clue-words selected for the spectrographic analysis, their final percentage obtained after combining two types of values, i.e. i. those values of formant frequencies (F1, F2 and F3) which are similar for questioned and specimen speech sample. ii. and those auditory features which have similar values for questioned and specimen speech sample by using Bayes' Theorem, number of formants used and the final take on the probability scale.

The probability scale has been identified with the help of the software after careful consideration of the i. final percentage as shown in the observation sheet; ii. the number of formants used; iii. and the number of clue-words selected. This is the criterion which is deployed by the researchers to calculate the resultant probability in India. The software developed by the authors, weighs these three factors in calculating the resultant probability. In this case, the final similarity percentage comes out to be 87.71%, numbers of formant frequencies used are three and 23 clue-words are taken, therefore, as per the criteria, the resultant probability is the Probable Identification.

Conclusion

Standard or traditional criteria used by the scientific fraternity/researchers to calculate the final probability is based on the following points: i. number of formant frequencies used in the experiment; ii. number of clue-words used in the experiment; and iii. similarity percentage of the similar vowels obtained after comparing questioned and the specimen speech samples i.e. acoustic (objective) features.

All the above three factors were to take into account and considered while evaluating the resultant and final probability which is present among any one of the nine available verbal probability scales.

But in this study, the percentage value calculated for the acoustic parameters is 90%, taking into account only those values of formant frequencies for vowels whose values are same for questioned and the specimen speech sample.

Similarly, for the auditory features, this percentage is 78.57%. If we take into account the percentage value of acoustic parameters only, then we get positive identification by using the criteria, as in this case, we have three Formant Frequencies, namely F1, F2 and F3 and 23 clue-words.

But the final probability which we get is the probable identification; this is because of the low value of percentage of similar auditory features. This shows that even auditory features can change the final results, if they are also taken into account, like happened in this situation and the final probability we obtained is Probable instead of the positive identification. Similarly, vice versa can also happen, e.g., if the percentage of similar vowels in case of acoustic parameters comes out to be less than 90% but the similarity percentage of auditory features comes out to be more than 90%, keeping other parameters like number of clue-words and number of formants unchanged, then we can get Positive identification instead of the Probable one.

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

English Transcription of Hindi Words	Word	Nuclei vowel	Questioned			Specimen		
			F1(Hz)	F2(Hz)	F3(Hz)	F1(Hz)	F2(Hz)	F3(Hz)
Haan	H**ɑ**n	**/ɑ/**	774	1605	2282	774	1476	2282
Jee	ʤ**i**	/i/	438	1186	2063	438	1186	2063
Bhaiya	Bh**ʌ**jɑ	**/ʌ/**	851	1818	2560	851	1818	2560
Bhaiya	Bhʌj**ɑ**	**/ɑ/**	703	1341	2334	703	1341	2334
Bhai	Bh**ɑ**i	**/ɑ/**	651	1960	2573	651	1779	2573
Bhai	Bhɑ**i**	/i/	593	1947	2405	593	1947	2405
Saab	S**ɑ**b	**/ɑ/**	683	1528	2160	683	1657	2160
Ne	N**ɛ**	**/ɛ/**	580	1077	1773	580	1077	1773
Bataya	B**ɑ**tɑjɑ	**/ɑ/**	1019	1393	2418	1019	1393	2418
Bataya	Bɑt**ɑ**jɑ	**/ɑ/**	1044	1399	2302	1044	1399	2302
Bataya	Bɑtɑj**ɑ**	**/ɑ/**	1032	1328	2244	1032	1328	2244
Dusri	D**u**siri	/u/	967	1928	2618	967	1928	2463
Dusri	Dus**i**ri	/i/	438	1167	1792	438	1167	1792
Dusri	Dusir**i**	**/i/**	967	1709	2437	967	1709	2437
Party	P**ɑ**rti	**/ɑ/**	677	1141	1573	677	1141	1573
Party	Pɑrt**i**	/i/	490	1019	2379	490	1199	2379
Dilli	D**i**lie	/i/	432	1167	1850	432	1167	1850
Dilli	Dil**i**e	/i/	451	967	2186	451	967	2186
Dilli	Dili**e**	/e/	664	1141	1573	664	1141	1573
Approach	Pr**o**ʧ	/o/	625	1051	2701	625	1051	2701
Karne	Kʌrn**ɛ**	**/ɛ/**	522	1006	2224	522	1006	2224
Ki	**Ki**	/i/	909	1122	1702	909	1257	1999
Koshish	K**o**ʃiʃ	**/o/**	542	1122	2830	542	974	2830
Koshish	Koʃ**i**ʃ	**/o/**	471	1825	2824	471	1825	2824
Rahe	R**ʌ**hɛ	**/ʌ/**	567	1199	1580	567	1199	1580
Rahe	Rʌh**ɛ**	**/ɛ/**	529	1077	2263	529	1077	2263
He	H**ɛ**	**/ɛ/**	709	1696	2205	709	1696	2108
IAS	**ʌ**i	**/ʌ/**	696	1425	2205	696	1425	2205
IAS	ʌ**i**	/i/	613	1328	2147	613	1328	2147
IAS	**E**s	**/e/**	477	1199	1728	477	1199	1728

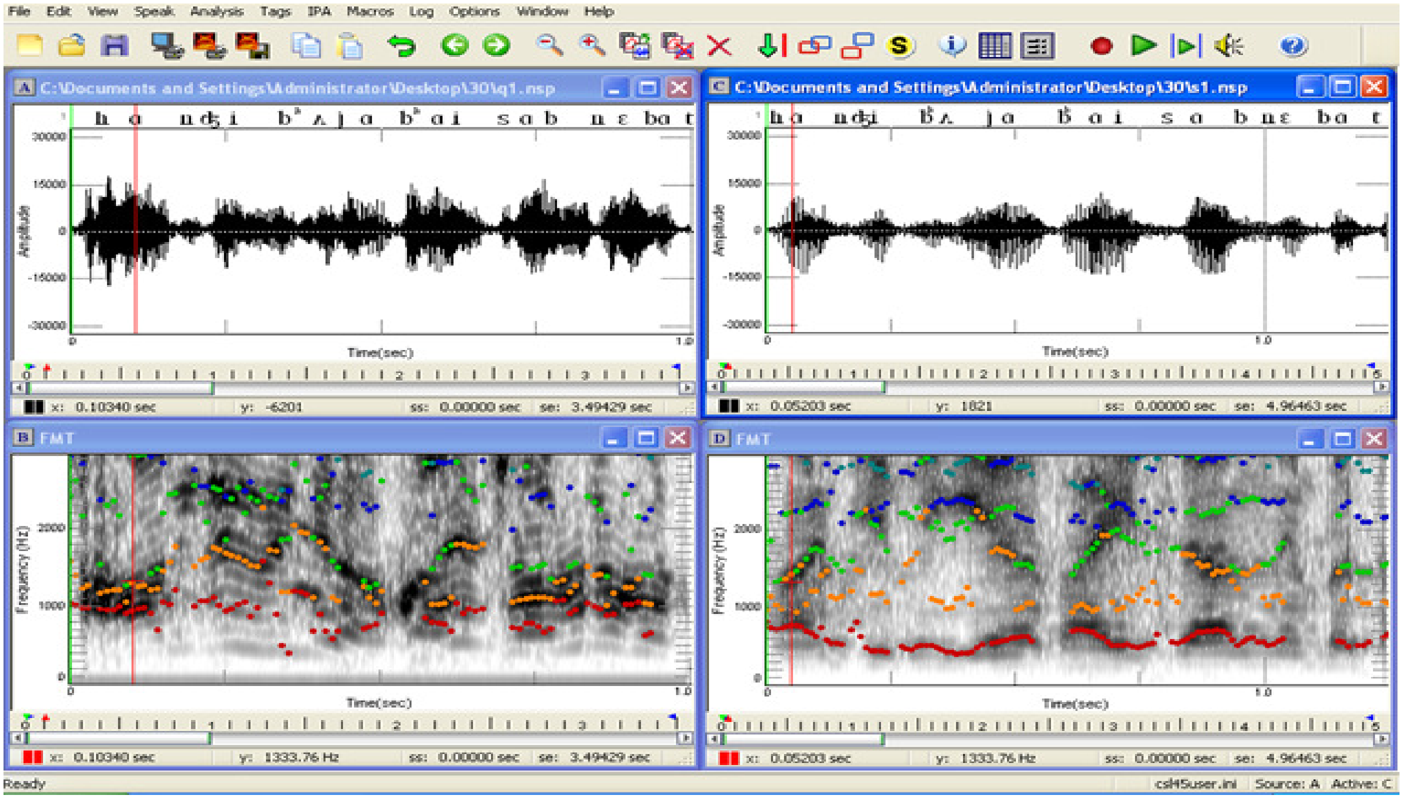

File Edit View Speak Analysis Tags IPA Macros Log Options Window Help
A C:\Documents and Settings\Administrator\Desktop\30\q1.nsp
C C:\Documents and Settings\Administrator\Desktop\30\s1.nsp
Amplitude
Time(sec)
x: 0.10340 sec y: -6201 ss: 0.00000 sec se: 3.49429 sec
x: 0.05203 sec y: 1821 ss: 0.00000 sec se: 4.96463 sec
B FMT
D FMT
Frequency (Hz)
x: 0.10340 sec y: 1333.76 Hz ss: 0.00000 sec se: 3.49429 sec
x: 0.05203 sec y: 1333.76 Hz ss: 0.00000 sec se: 4.96463 sec
Ready
csl45user.ini Source: A Active: C

Figure-1: Waveform with phonetic transcript of words /hɑn/, /ʤi/, /bhʌjɑ/,/bhɑi/,/sɑb/ and /nɛ/ in window A and C; their respective spectrogram with formant marking in windows B and D.

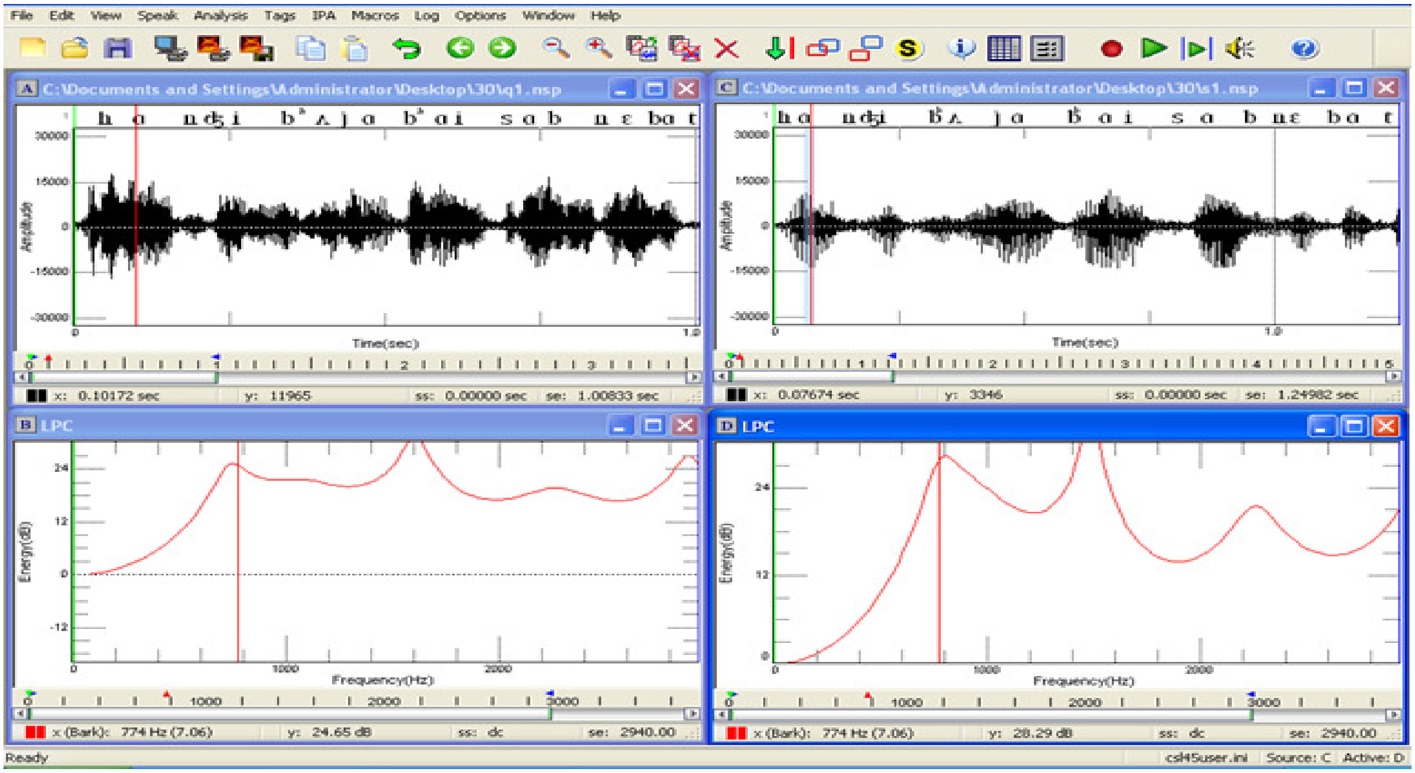
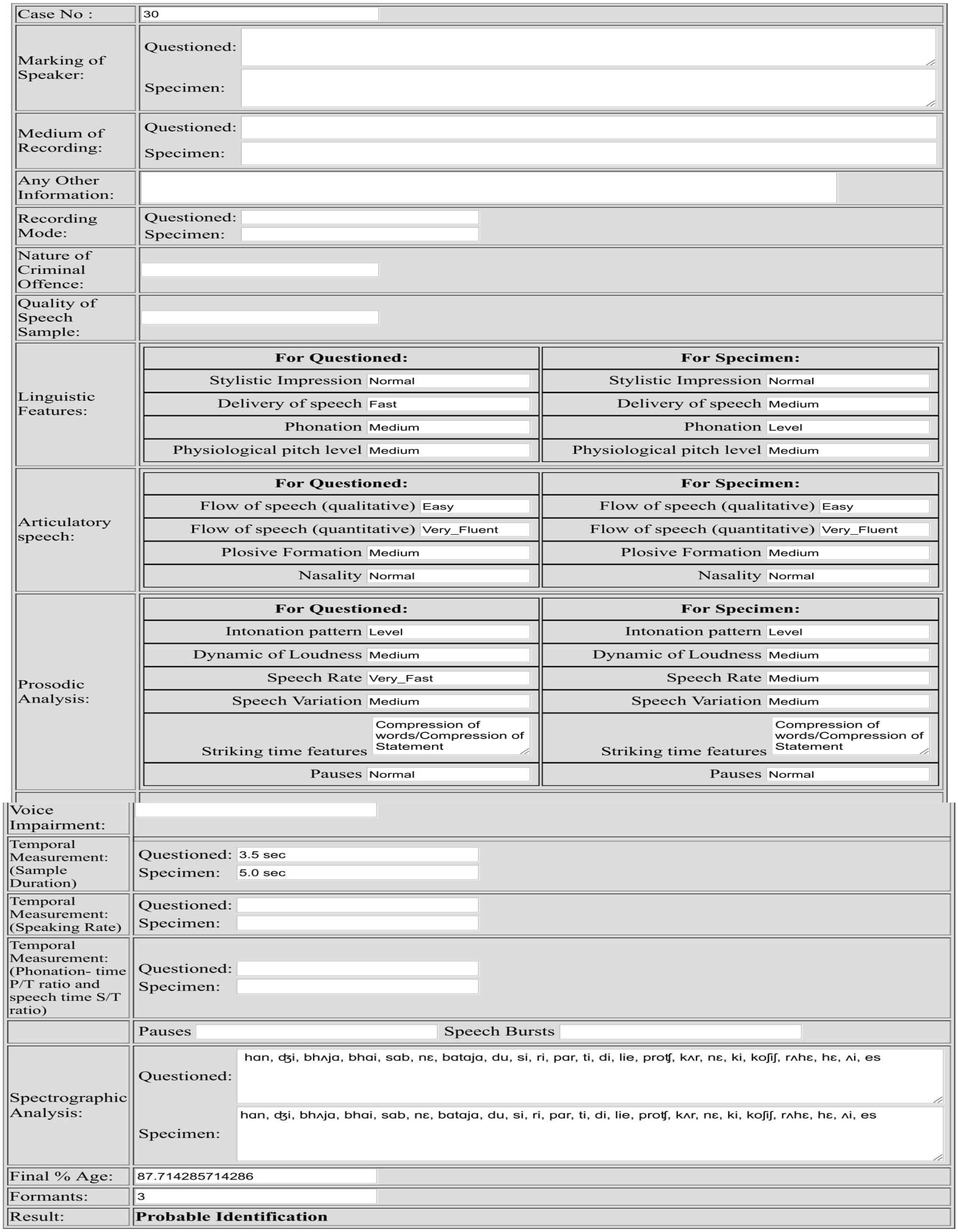

File Edit View Speak Analysis Tags IPA Macros Log Options Window Help
A C:\Documents and Settings\Administrator\Desktop\30\q1.nsp
C C:\Documents and Settings\Administrator\Desktop\30\s1.nsp
Amplitude
Time(sec)
x: 0.10172 sec y: 11965 ss: 0.00000 sec se: 1.00833 sec
x: 0.07674 sec y: 3346 ss: 0.00000 sec se: 1.24982 sec
B LPC
D LPC
Energy(dB)
Frequency(Hz)
x (Bark): 774 Hz (7.06) y: 24.65 dB ss: dc se: 2940.00
x (Bark): 774 Hz (7.06) y: 28.29 dB ss: dc se: 2940.00
Ready
csl45user.ini Source: C Active: D

Figure-2: Waveform with phonetic transcript of words /hɑn/, /ʤi/, /bhʌjɑ/,/bhɑi/,/sɑb/ and /nɛ/ in window A and C; their respective LPC in windows B and D.



Case No :	30		
Marking of Speaker:	Questioned:		
	Specimen:		
Medium of Recording:	Questioned:		
	Specimen:		
Any Other Information:			
Recording Mode:	Questioned:		
	Specimen:		
Nature of Criminal Offence:			
Quality of Speech Sample:			

Linguistic Features:	**For Questioned:**		**For Specimen:**	
	Stylistic Impression	Normal	Stylistic Impression	Normal
	Delivery of speech	Fast	Delivery of speech	Medium
	Phonation	Medium	Phonation	Level
	Physiological pitch level	Medium	Physiological pitch level	Medium
Articulatory speech:	**For Questioned:**		**For Specimen:**	
	Flow of speech (qualitative)	Easy	Flow of speech (qualitative)	Easy
	Flow of speech (quantitative)	Very_Fluent	Flow of speech (quantitative)	Very_Fluent
	Plosive Formation	Medium	Plosive Formation	Medium
	Nasality	Normal	Nasality	Normal
Prosodic Analysis:	**For Questioned:**		**For Specimen:**	
	Intonation pattern	Level	Intonation pattern	Level
	Dynamic of Loudness	Medium	Dynamic of Loudness	Medium
	Speech Rate	Very_Fast	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

Voice Impairment:				
Temporal Measurement: (Sample Duration)	Questioned:	3.5 sec		
	Specimen:	5.0 sec		
Temporal Measurement: (Speaking Rate)	Questioned:			
	Specimen:			
Temporal Measurement: (Phonation- time P/T ratio and speech time S/T ratio)	Questioned:			
	Specimen:			
	Pauses		Speech Bursts	
Spectrographic Analysis:	Questioned:	han, ʤi, bhʌja, bhai, sab, nɛ, bataja, du, si, ri, par, ti, di, lie, proʧ, kʌr, nɛ, ki, koʃiʃ, rʌhɛ, hɛ, ʌi, es		
	Specimen:	han, ʤi, bhʌja, bhai, sab, nɛ, bataja, du, si, ri, par, ti, di, lie, proʧ, kʌr, nɛ, ki, koʃiʃ, rʌhɛ, hɛ, ʌi, es		
Final % Age:	87.714285714286			
Formants:	3			
Result:	**Probable Identification**			

Figure-3: 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.

References

1. Holmgren G.L. (1967). Physial and Psychological Correlates of Speaker Recognition. *Journal of Speech, Language, and Hearing Research*, 10, 57-66. doi:10.1044/jshr.1001.57.
2. Endress W., Bambach W. and Flosser G. (1971). Voice Spectrograms as a function of Age, Voice Disguise and Voice Imitation. *Journal of Acoustical Society of America*, 49, 1842-1848. https://doi.org/10.1121/1.1912589.
3. Tosi O., Oyer M., Lashbrock W., Pedey C., Nicol J. and Nash E. (1972). Experiment on Voice Identification. *Journal of Acoustical Society of America*, 51, 2030-2043.https://doi.org/10.1121/1.1913064.
4. Wolf J.J. (1972). Efficient acoustic parameters for speaker recognition. *Journal of Acoustical Society of America*, 51(6), 2044-2057. https://doi.org/10.1121/1.1913065
5. Hazen B. (1973). Effects of differing phonetic contexts on spectrographic speaker identification. *The Journal of the Acoustical Society of America*, 54(3), 650-660. https://doi.org/10.1121/1.1913645.
6. Samber M.R. (1975). Selection of Acoustic Features for Speaker Identification. *IEEE Transactions on Acoustic, Speech and Signal Processing*, 23(2), 176-182. 10.1109/TASSP.1975.1162664.
7. Aitken C.G.G. (2013). Statistical Interpretation of Evidence/Bayesian Analysis. University of Edinburgh, Edinburgh, UK, 173-179. ISBN: 978-0-12-800647-4.
8. Meuwly D. and Drygazlo A. (2001). Forensic Speaker Recognition based on Bayesian Framework and Gaussian Mixture Modelling (GMM). *The Speaker Recognition Workshop Crete, Greece*, 18-22, 145-150.
9. Kinoshita Y. (2002). Use of Likelihood Ratio and Bayesian Approach in Forensic Speaker Identification. School of Languages and International Education, University of Canberra. Australian Speech Science and Technology Association Inc., 297-302.
10. An Introduction to Forensic Speaker Identification Procedure (2005). Advance Interactive Training Course on Forensic Speaker Recognition. CBI Bulletin, Directorate of Forensic Science, Ministry of Home Affairs, Govt. of India, XIII(1).
11. Besson O., Dobigeon N. and Tourneret J.Y. (2014). Joint Bayesian estimation of close subspaces from noisy measurements. *IEEE Signal Processing Letters*, 21(2), 168-171. 10.1109/LSP.2013.2296138.
12. Mathu R.S., Chaudhary S.K. and Vyas J.M. (2016). Effect of Disguise on Fundamental Frequency of Voice. *Journal of Forensic Research: Open Access*, 7(3). ISSN: 2157-7145 JFR, doi:10.4172/2157-7145.1000327.
13. Bhall B., Singh C.P., Dhar R. and 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: Open Access*, 7. ISSN: 2157-7145 JFR, doi:10.4172/2157-7145.1000338.