Role of Second Formant Frequency(F2) in Forensic Speaker Identification

NIKITA SUTHAR[1], *Jawaharlal Nehru University,*

ABSTRACT

Identification of a speaker specific feature is an almost impossible task because of the enormous number of features which can affect a single voice utterance.F2, (marker of vowel position) plays an important role in vowel identification. The present study is an statistical analysis conducted on the 16 speakers' F2, F2- F1 and F3-F2 values for the vowels; "/a:/, /i:/ and /u:/. Vowel /a/ showed considerable height difference in Known sample (where the researcher knew the Speaker) and Unknown sample (someone else collected the sample).

1 Introduction

Using formants to analyze speech signals in a forensic enquiry has been around and accepted, since the beginning of the field. There are infinite number of variables which can affect any forensic enquiry. Formant analysis is the most prominent method to analyze sound. F0 is an accepted and reliable source of speaker distinction in a Forensic enquiry. It helps in distinguishing gender, tone and many other emotional and psychological factors, from a speaker's voice. F1 shows the height of the vowel and F2 shows the position of vowel sound in the mouth. The paper is going to focus on the importance of F2 in identifying an individual speaker out of many suspect samples.

Stevens (1972b) introduced the term quantal factor with respect to the acoustic signals used for vocal communications. Stevens (1972b) stated that certain speech sounds are more distinctly recognizable than the others. He focused on three major vowels, [i], [u] and [a]. Acoustically speaking for [i] the second and the third formants, F2 and F3 are both high; [a] has low , F2 and F1 is high whereas [u] has a low , F1 and F2 . Articulatory and Acoustic analysis have shown that these vowels are the limiting articulations of a vowel triangle which constitutes a Language-universal.

All these three vowels have well defined spectral properties. A central spectral peak occurs at the above 1KHz for[a], a High frequency spectral peak occurs for [i] and a low frequency spectral peak occurs for [u].

[1]nickybhadrecha@gmail.com

2 Data and Variables

The study was conducted on 32 Marwari Speakers. 16 belonged to Bishnoi Community of the small city of Bikaner, and 16 were from Brahmin community. Both varieties contained 8 male and 8 female speakers. Two types of recordings were taken, which were Known sample and Unknown samples. Known Samples were termed as Suspect Sample (SS) and Unknown samples were marked as Question Sample(QS). Each QS was taken separately and then the researcher looked for the closest SS to it. For the purpose of this study only F2 was identified for each sample and analysed. Three major tests, T-testing, Standard Deviation and mean were marked down. The final step was to mark the accuracy of the final outcomes of these tests. Three long vowels were used for the following study. ‘a:’, ‘i:’ and ‘u:’. The vowels very recorded by the researcher herself. For any kind of recording Bernard’s method was used, i.e. Recording a vowel in a h_d format. But since there are no matching words in Marwari for both varieties, which can fulfil these conditions exactly, the next alternative was to retrieve vowels in a CVC format. So, for SS, CVC format was used, where both ‘C’s were stops. The words which were chosen for the recording of suspect sample, were common for both varieties. These words were specifically selected in order to see the difference between the pronunciation of two varieties, when they have the same vocabulary. Words for the unknown sample recordings were taken from the conversations researchers had with the participants. The same pattern was used for this selection also.ie. CVC. Finding optimum words from a QS is always difficult. In this study, the major problem which arrived was that finding the exact same word in both varieties as a QS was not possible. CVC pattern was strictly followed and both consonants were stops, but they were not common in both varieties. For example, a Bishnoi speaker when asked what a child is called, always replied , “ ta:b r” which was also true for a Brahmin speaker. But, the same speaker repeatedly used word, “ ting r” instead of “ ta:b r” while having a conversation with the researcher, whereas it remained “ tab r” for Brahmin speaker even in a conversation. This made this study more realistic, because every forensic enquirer knows that, acquiring a question sample is never ideal. Sometimes, the words are not audible, sometimes they are audible, but suspects become adamant to mispronounce them, while giving a SS. And sometimes the position of vowels is difficult to assess because of the high word speed and low amplitude of the recorded samples. In forensics, stops have always been chosen over any other consonant to follow or precede the vowel in questions, because of their minimum effect on the vowel. By choosing the same method researcher fixed few variables for the enquiry.

The next step was done by another researcher. That was she jumbled all the available question sample’s numbering. The purpose of this task was to keep the anonymity of the sample and see if the researcher will be able to match out the real QS to its intended SS.

3 Auditory Analysis

An auditory analysis was conducted on the acquired data, for both SS and QS. During the auditory analysis, all the known samples i.e. the suspect samples were listed carefully and remarks were made by the listener. Since the researcher knew all the suspects personally, these remarks were conducted by two other different linguists, in order to avoid any predetermined biases. These biases can affect the remarks on suspects, in a way that the suspect might be more talkative or outspoken but while giving the recording he or she became more conscious. This kind of situation might result to researcher writing his or her own remarks, rather than being true to the recording. This problem was solved, because the analysis was conducted by other linguists.

The recordings were played for the researchers and they were asked to assign the auditory details given in the table for each participant. This analysis was conducted for both suspect sample and question sample.

The auditory analysis was divided into four tables. The first table represents the auditory analysis of known Brahmin speakers. The second table represents the Auditory Analysis of known Bishnoi Speakers. Both of these tables contain data from sixteen speakers from each variety. The data is divided into ten categories, the sex of the speaker, Speaking Mode, Stylistic Feature, Respiratory Form, Pitch Level, Striking Features of Voice, Fluency, Understandability, Loudness, Speech Rate. The table is arranged in such a way that, the first eight speakers are male and the next eight speakers are females, for each variety.

The next two tables contained the question samples obtained by random conversations from each speaker. The third table, has all the question samples for the male participants, the forth table has the auditory analysis of question samples of all the female speakers.

The parameters designed by the CFSL, gives an extensive overview of auditory features of the speakers. Below is one such table obtained during the analysis.

S.No.	**PARAMETERS**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**
1.	**Sex**	M	M	M	M	M	M	M	M	F	F	F	F	F	F	F	F
2.	**Speaking Mode** **i. Disguised** **ii. Normal**	NORMAL	NORMAL	NORMAL	NORMAL	NORMAL	NORMAL	NORMAL	NORMAL	NORMAL	NORMAL	NORMAL	NORMAL	NORMAL	NORMAL	NORMAL	NORMAL
3.	**Stylistic Feature** **i. Simple** **ii. Polite** **iii. Emotional** **iv. Expressive** **v. Ironical**	SIMPLE	EXPRESSIVE	EXPRESSIVE	POLITE	EXPRESSIVE	EXPRESSIVE	SIMPLE	IRONICAL	EXPRESSIVE	EXPRESSIVE	SIMPLE	POLITE	SIMPLE	SIMPLE	EXPRESSIVE	POLITE

4.	**Respiratory Form** **i. Nasal** **ii. Oral**	ORAL	NASAL	NASAL	NASAL	ORAL	ORAL	ORAL	NASAL	ORAL	NASAL	NASAL	ORAL	ORAL	ORAL	ORAL	ORAL
5.	**Pitch Level** **i. High** **ii. Normal** **iii. Low**	HIGH	NORMAL	NORMAL	NORMAL	NORMAL	NORMAL	NORMAL	HIGH	HIGH	NORMAL	NORMAL	NORMAL	NORMAL	NORMAL	NORMAL	NORMAL
6.	**Striking Features of Voice** **vi. Coarse** **i. Hoarse** **ii. Creak** **iii. Pressed voice** **iv. Normal voice**	NORMAL	CREAK	CREAK	CREAK	NORMAL	NORMAL	HOARSE	NORMAL	NORMAL	NORMAL	NORMAL	NORMAL	NORMAL	NORMAL	NORMAL	PRESSED
7.	**Fluency** **i. Hasty** **ii. Very fluent** **iii. Normal fluent** **iv. Sluggish** **v. Stopping**	NORMAL	NORMAL	NORMAL	FLUENT	NORMAL	NORMAL	NORMAL	HASTY	NORMAL	NORMAL	NORMAL	FLUENT	NORMAL	NORMAL	FLUENT	FLUENT
8.	**Understandability** **i. Easily** **ii. Hard** **iii. Hardly**	EASILY	EASILY	EASILY	EASILY	EASILY	EASILY	EASILY	EASILY	EASILY	EASILY	EASILY	EASILY	EASILY	EASILY	EASILY	EASILY
9.	**Loudness** **i. Very loud** **ii. Loud** **iii. Medium** **iv. Soft**	VERY LOUD	LOUD	MEDIUM	MEDIUM	MEDIUM	MEDIUM	LOUD	LOUD	VERY LOUD	MEDIUM	MEDIUM	MEDIUM	LOUD	LOUD	MEDIUM	MEDIUM
10.	**Speech Rate** **i. Very fast** **ii. Fast** **iii. Medium** **iv. Slow** **v. Very Slow**	MEDIUM	MEDIUM`	MEDIUM	FAST	FAST	MEDIUM	MEDIUM	MEDIUM	MEDIUM	MEDIUM	MEDIUM	MEDIUM	MEDIUM	MEDIUM	MEDIUM	MEDIUM

Table 1: Auditory Analysis of Known Brahmin Speakers (Brahmin Suspect Samples)

3.1 Observations

The most important feature in this variety is that, many speakers had a nasalized voice. This feature is prominent in many Brahmin variety across the Indo- Aryan language families. Bishnoi speaker, female speakers' voices were identified as loud, one tempo higher than Brahmin females.

4 Acoustic Analysis

Three values, which were important for this study were, F1, F2 and F3. As mentioned earlier, this study's main focus was on role of F2, and whether it is an important factor in determining a speaker specific feature. Acoustic space charts were created for all the suspect samples and they were paired to the nearest possible match in the question sample. On the basis of these charts two separate lists were created. On containing all the Suspect Sample values and one with all the question sample values. In order to determine the role of F2 here, F2-F1 and F3- F2 were chosen. This gave us 6 separate value for each vowel. SF2, QF2, S (F2-F1), Q(F2-F1), S (F3- F2) and Q (F3- F2). For each list, mean and standard deviation was calculated and finally, to verify the significance of the results, t-distribution and p-values were analyzed.

Bishnoi Female ii						
	SF2	QF2	S(F2-F1)	Q(F2-F1)	S(F3-F2)	Q(F3-F2)
t	3.1187		3.1422		-2.388	
df	13.493		12.821		13.99	
p-value	0.007838		0.007905		0.0316	
alternative hypothesis	TRUE		TRUE		TRUE	
95 confidence Lower	103.0474		106.2587		-391.82727	
Upper	562.1089		576.0051		-21.01256	
mean of x	2289.641		1827.43		595.4648	
mean of Y	1957.063		1486.298		801.8847	

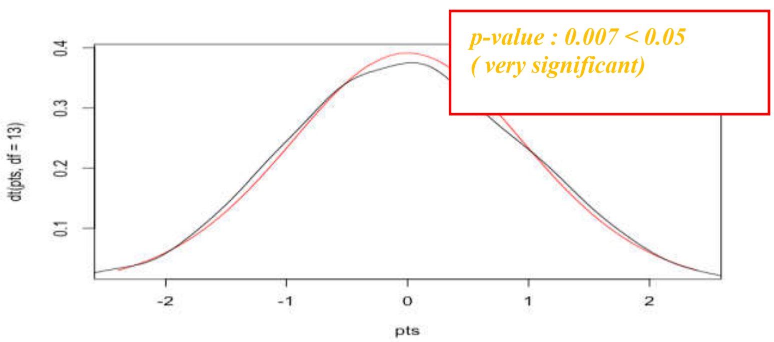

Figure 1 Bishnoi Females sound /i:/ SF2 vs QF2 , PAIRED T-TEST

4.1 Conclusion

In all three categories degrees of freedom were almost 13, which itself is a significant number. Then the p-value in all three is > 0.05. these points prove that the null hypothesis, i.e. F2 is not an important factor in determining speaker specific feature, has been proven wrong.

The other alternative hypothesis was;'*F2 is significant*'

The strong df and lower p-value proves the alternative hypothesis true.

Brahmin male a:						
	SF2	QF2	S(F2-F1)	Q(F2-F1)	S(F3-F2)	Q(F3-F2)
t	-2.3886		-0.037722		0.21035	
df	10.339		10.883		13.998	
p-value	0.03729		0.9706		0.8364	
alternative hypothesis	TRUE		TRUE		TRUE	
95 confidence Lower	-221.41777		-111.8466		-184.5411	
Upper	-8.18501		108.0823		224.6736	
mean of x	1274.916		603.1728		1368.469	
mean of Y	1276.798		717.9742		1348.403	

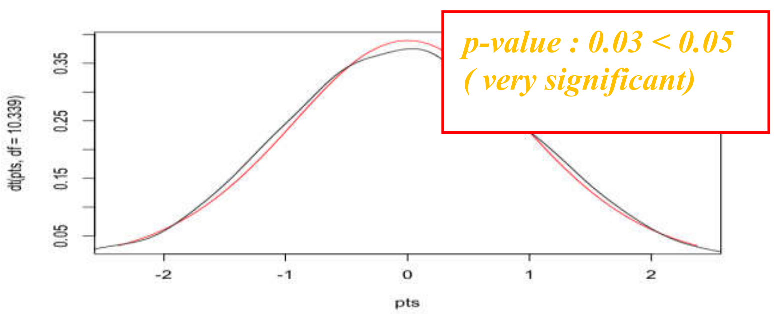

Figure 2 Brahmin male sound /a:/ SF2-F1 vs QF2-F1, Paired t-test

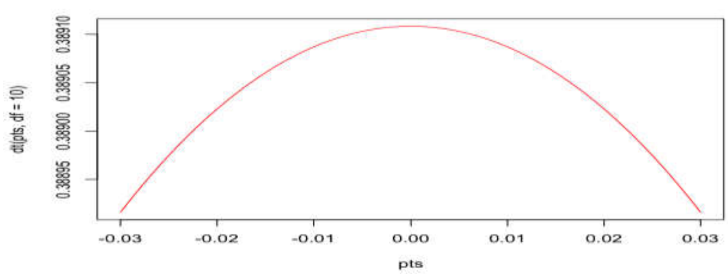

Figure 3Brahmin male sound /a:/ SF2 vs QF2, Paired t-test

4.2 Results

In the first categories degrees of freedom were 10. which if analysed alone, is not a significant number. Then the p-value of F2 separately gave us the result which is less than > 0.05. But the significance of the p-value is questionable for the last two categories. This is the most important finding of this paper. It shows that F2 alone has strong evidence against the null hypothesis, but it might get affected if analysed with F1 or F3.

4.3 Conclusion

The overall conclusion from the current study is that the speakers of the Brahmin variety had a general nasalized tone for almost 8 speakers. The Bishnoi speakers on the hand had a loud voice. Bishnoi speakers' loud and fast voice output created a higher mean pitch for the variety than that of Brahmin speakers.

Nasalization affected the F1 values of Brahmin speaker a lot. But in all these cases, the F2 remained somewhat unchangeable and provided positive results. In forensic speaker identification, for a nasalized voice it is better to look for the F2, which depicts the position of vowel in the mouth, i.e. how back or front the vowel is. So, we can say that through the help of the current study, it can be shown that F2 shows the highest level of accuracy for a nasalized voice and for a non-nasal oral voice, Even though both F2 and F1 should be taken into consideration and to achieve more accurate results, F2 plays a significantly better role while analyzing a nasal sound. The analysis should contain both auditory and acoustic analysis.

References

Baldwin, John R., and Peter French. 1990. *Forensic phonetics.* s.l.:Pinter Publishers.

Becker, Timo, Michael Jessen, and Catalin Grigoras. 2008. Forensic speaker verification using formant features and Gaussian mixture models. In *Ninth Annual Conference of the International Speech Communication Association*.

Foulkes, Paul, and Peter French. 2012. Forensic speaker comparison: A linguistic–acoustic perspective. In: L. &. P. Tiersma, ed. *The Oxford Handbook of Language and Law.* s.l.:Oxford: Oxford University Press.

Fujimura, Osamu, 1967. On the second spectral peak of front vowels: a perceptual study of the role of the second and third formants. *Language and Speech* 10(3): 181-193.

Fujisaki, Hiroya, and Takako Kawashima. 1968. The roles of pitch and higher formants in the perception of vowels. *IEEE transactions on audio and electroacoustics* 16(1): 73-77.

Halberstam, Benjamin, and Lawrence J. Raphael. 2004. Vowel normalization: the role of fundamental frequency and upper formants. *Journal of Phonetics* 32(3): 423-434.

Hirahara, Tatsuya, and Hiroaki Kato. 1992. The effect of F0 on Vowel Identification. *Speech perception, production, and linguistic structure*: 89-112.

Hollien, Harry. 2013. *The acoustics of crime: The new science of forensic phonetics*. Springer Science & Business Media.

Lindblom, Björn, Diehl, Randy and Creeger, Carl. 2009. Do 'Dominant Frequencies' explain the listener's response to formant and spectrum shape variations?. *Speech communication*, 51(7): 622-629.

Marrero, Victoria, Elena Battaner, Juana Gil, Joaquim Llisterri, María Machuca, Montserrat Marquina, Carme De La Mota, and Antonio Rios. 2008. Identifying speaker-dependent acoustic parameters in Spanish vowels. *Journal of the Acoustical Society of America* 123(5): 3877-3877.

Rodman, Robert, D. McAllister, D. Bitzer, L. Cepeda, and Pamela Abbitt. 2002. Forensic speaker identification based on spectral moments. *Forensic Linguistics* 9: 22-43.

Rose, Phil. 2002. *Forensic speaker identification*. cRc Press.

Rose, Phil. 2006. Technical forensic speaker recognition: Evaluation, types and testing of evidence. *Computer Speech & Language* 20(2-3): 159-191.