

# Imitation of the Voice – An Acoustic Study

Babi Duli

Ph.D Research Scholar, Dept. of Linguistics and Phonetics, School of Phonetics and Spoken English  
EFL University, Hyderabad – 500605 Email:- bobbyciefl@gmail.com Ph: +91 9494006604

**Abstract:** *The present study presented here is an imitation of a voice sample imitated the south Indian film star Mahesh Babu by a professional imitator/mimicry artist in a TV show. The imitator tried to imitate the dialogue of the hero from his block buster hit movie titled Dookudu ([dʊ:ku/ʊ]). The aim of the study is to investigate how closely the imitations matched in the selected acoustic parameters of the original voice. It was found that he was able to imitate dialogue very closely, but timing at the segmental level showed little change in the direction of the targets. Mean formant frequency did not match the target voice. It is too distant in the timing, mean of formant frequencies and in pitch*

**Keywords:** imitation, fundamental frequency, formant frequency, pitch, timing, forensic, phoneticians, culprits, acoustics, YouTube

## 1. Introduction

Imitation is an act of attempting others voice which is unique by birth. It can be done either with the knowledge of the actual person or in his/her absence. If it is done in his/her presence, chances are very limited to make the actual person as culprit or to indulge in any of the immoral activities, but in absence it will go to any extent (either entertain or not). If it entertains, the imitator gets the awards, if not; only the original speaker can get the repression. But according to law 100 culprits can escape from the punishment but one good person should not be caught in injustice. Here the forensic experts trained a lot especially to trace out the culprits and save the innocent. Even phoneticians also trained in this connection to connect to the culprits' voices from a collection of compendious voices to do the justice to the common compatriot.

## 2. Earlier research review

Only a handful of studies exist in this area, and to our knowledge only two that deal with acoustic parameters in some detail. Bessler [2] has studied a caricatured impersonation of de Gaulle. The most relevant finding with respect to the present study was the fact that the impersonator exaggerated both mean fundamental frequency level and range. Stressed syllable durations were also exaggerated. The impersonation was primarily meant for entertainment purposes and not accuracy of imitation. The generalisability of the results may therefore be questioned. In the other study [3], vowel formant frequencies and fundamental frequencies in imitations were compared with the corresponding values for the original voices. Although the imitators managed to change their formant and fundamental frequencies in the direction of the target values "they were not able to adapt these parameters to match or even be similar to those of the imitated persons." [3, p. 1842]

## 3. Methodology

The target voice sample is collected from the YouTube from a popular Telugu comedy show in E-TV titled *Jabardasth*. The source voice sample is collected from the super hit movie titled *Dookudu*. The video files are converted into

.wav files by using 'video pad' and 'wave pad'. Praat software is used for segmenting the voice samples and for analyzing.

## 4. Acoustic Analysis

The speech materials are digitized and used for fundamental frequency level and pitch level for the acoustic analyses. All the speech files are labelled at word level. The mean of the Fundamental frequency, formant frequencies and the mean of the Pitch are then computed. First three formant frequencies are calculated for all the words which are presented in the material. All tokens are numbered consecutively to ensure that comparisons between tokens in the original, and imitation of the voice were made between in identical contexts.

## 5. Results

### 5.1 Timing

Imitator reached for maintain the closest timing at the initial word in the continuous speech but he failed after that. Initially the difference between imitator and Mahesh is a very close but after it is moved to the falling stage and acquired negative results. This is one of the key factors for identifying the culprit. It is proved that voice is unique in nature. The following table1 will give the clarity regarding the timing of the voice samples.

**Table1:** Total timing of the utterance

Token	I – Total (in Ms*)	M- Total (in MS*)	Dif
/bHaja:nike/	595	567	28
/mi:niN/	319	431	-112
/tʃelijani/	232	312	-80
/bla rana:dʃi/	686	804	-118

\***Millisecons I** stands for imitator voice **M** stands for Mahesh voice **Dif** stands for difference of the voice

A more detailed analysis of the timing showed on a line diagram. At the initial stage it is very close to the original voice but later the imitator's voice reached to the falling stage and original voice reached to the raising stage. This is clearly seen in the following Figure 1

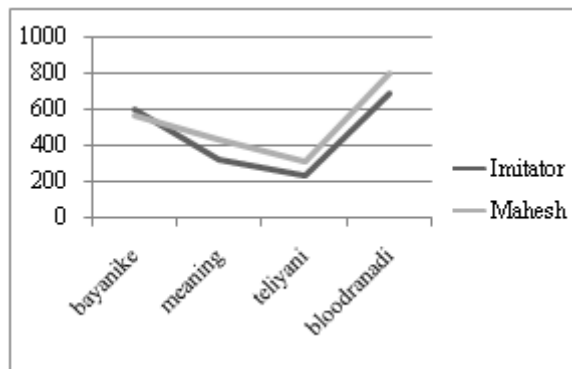


Figure 1: Timing of the utterance

## 5.2 Formant frequency

The second best way of examining voice samples is that get the mean of the formant frequency. In F1, F2, F3 only Mahesh has the dominating voice than the imitator. Imitator failed to reach the original voice samples. He attempted but failed due to less preparation. Even this is not possible to anybody to imitate the original voice samples alike because as it stated above voice is unique in nature. By examining the mean of the formant frequencies also forensic experts and phoneticians identify the culprits or voices. The following table2 explains about the mean of the formant frequency

Table 2 : Mean of the formant frequencies

Token	M-F1	I-F1	Dif	M-F2	I-F2	Dif	M-F3	I-F3	Dif
/bHaja:nike/	718.39	564.51	153.88	1704.26	1603.07	101.19	2535.72	2449.05	86.67
/mi:niN/	389.56	376.47	13.09	2006.89	1788.94	217.95	2590.42	2357	233.42
/t5elijani/	562.91	471.42	91.49	1882.86	1697.43	185.43	2575.49	2362.59	212.9
/bla rana:d5i/	672.87	582.28	90.59	1673.09	1614.60	58.49	2712.80	2549.11	163.69

F stands for frequency

For the table 2 the following bar chart (figure2) represents the results of the mean of the formant frequencies.

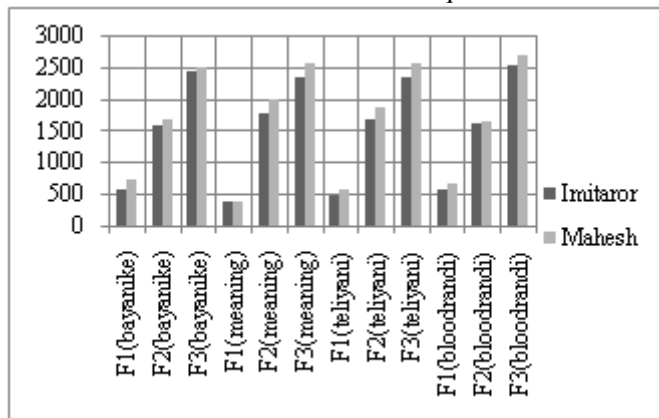


Figure2: Mean of the formant frequency

## 5.3. Pitch

In any voice pitch plays a crucial role. The following table3 is clearly showing that the imitator has the pitch in between 148Hz to 160 Hz but Mahesh has 100Hz to 125Hz. It means where Mahesh pitch levels end, there the imitators pitch levels starts. Based on this assumption, Pitch plays a very crucial role in any individual's voice.

Table3: Pitch of the voice

Token	I-Pitch	M-pitch	Dif
/bHaja:nike/	156.07	121.78	34.29
/mi:niN/	158.71	123.25	35.46
/t5elijani/	152.78	109.38	43.4
/bla rana:d5i/	148.67	100.13	48.54

Imitator has the high pitch, but Mahesh has the low Pitch. Of course the pitch levels are merely close when it is to a naked ear, but they are too distant when it is acoustically analyzed. The following line diagram (Figure3) represents how the voices are distant.

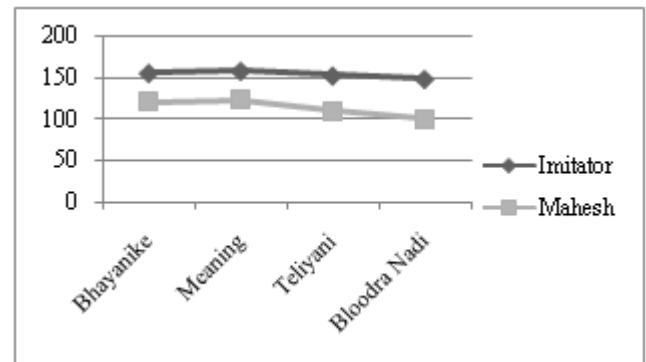


Figure3: Pitch of the voice

## 6. Conclusion

Based on the present study the author found that the voice is unique in nature. It is highly impossible to copy. Though it is very close to the original voice to a naked ear but when it is analyzed acoustically, then there are so many factors reveal that no two voices are alike. It is proved based on this study. It is too distant in the timing, mean of the formant frequencies, and the pitch. There are other voices and imitations also to work on like (the famous actors/artists in the Tollywood) NTR (Nandamuri Taraka Ramarao), ANR (Akkineni Nageswara Rao), Krishna, Chiranjeevi, etc.,

## References

- [1] Anders Eriksson and Pär Wretling. How flexible is the human voice? – A case study of mimicry *Umeå University, S-901 87 Umeå, Sweden.*
- [2] Bessler, P. (1991) La caricature de de Gaulle par Tissot: Étude phonostylistique. In: *Information/Communication*, 12, 19–32. Canadian Scholars' Press.
- [3] Endres, W., W. Bambach & G. Flösser. (1971) Voice spectrograms as a function of age, voice disguise, and voice imitation. *J. Acoust. Soc. Am.*, 49, 1842–1848.

- [4] Philip Rose.( 2002) *Forensic Speaker Identification*: Taylor And Francis Series. London
- [5] Traunmüller, H. & A. Eriksson. (1995) The perceptual evaluation of F0 excursions in speech as evidenced in liveliness estimations. *J. Acoust. Soc. Am.*, **97**, 1905–1915.

### Author Profile



**Mr. Babi Duli**, a Ph.D research scholar at the EFL University, from the Department of Linguistics and Phonetics, School of Phonetics and Spoken English, continuing his research in the area of Forensic Phonetics. He had a master degree from Adikavi Nannaya University, Rajahmundry. He was having experience in teaching Pronunciation. He worked for several prestigious institutions, and trained many students in the area of pronunciation.