Xournals

Academic Journal of Forensic Science

ISSN: 2581-4273 | Volume 06 | Issue 01 | April-2023

Voice Impersonation Examination by Spectrographic Analysis: A Voice Comparative Study

Ananya Jain[1], Riya Bansal[2]

Available online at: www.xournals.com

Received 09th January 2023 | Revised 16th January 2023 | Accepted 30th January 2023

Abstract:

The most popular form of communication is audio voice, however occasionally it is employed improperly or illegally. The Forensic Speaker Identification System faces a potential challenge from voice imitation, one of the main disguises that is on the rise. There are numerous ways to conceal a human voice such as Self-disguise, impersonating someone else, or stealing their identity are all examples of this. Here, identifying impersonation using a person's real voice is crucial for establishing ownership. It is crucial to determine whether a voice is being impersonated or belongs to the actual speaker when a person disputes ownership of voice evidence that sounds like them. This research presents a novel two-stage verification approach for the mimicry voice signal. The first stage involves comparing intonation patterns using spectrograms of voice of original artists and their respective mimicry artists, while the second stage is based on differences in fundamental frequencies and pitch.

Keywords: *Speaker identification, voice disguise, spectrogram, pitch, fundamental frequency.*

Authors:

1. *B.Sc, Forensic Science, Garden City University, INDIA*
2. *Junior Scientific Officer, Sherlock Institute of Forensic Science, INDIA*

Introduction

The human voice is thought of as one of our most personal characteristics. People tend to converge towards the language they observe around them, whether it's copying word choices, mirroring sentence structures or mimicking pronunciations. In the Odyssey (Homer, 850 BC), Helen of Troy is said to have circled the wooden horse while yelling to different Greek soldiers in the voices of their wives and sweethearts because she suspected betrayal. This tale is noteworthy for two different reasons. That is one of the first examples of voice mimicry that has been documented. It may also be the earliest instance of vocal mimicry used for deception **(Singh *et al.*, 2017).**

This research study is primarily concerned with the identification of speakers whose voices have been disguised and focuses on speaker verification and recognition on the basis of the intonation pattern **(www.research.csc.ncsu.edu).** The intonation model is typically evaluated based solely on the similarities and differences between several samples, without any explicit reference. Mimicry voice is using synthetic speech against speaker verification based on the spectrum and pitch analysis **(Hautamäki *et al.*, 2017).** Pitch refers to the highness or lowness of the voice and articulation is the way you pronounce individual sounds. While the generation of speech sounds is the emphasis of articulation, whether you pronounce them correctly is the subject of pronunciation. Both impressions of believability and intelligibility are influenced by the sound characteristics of articulation and pronunciation. During person-to-person interactions, your speech frequently has distinct, strong tones. Articulatory phonetics refers to the equipment used to produce speech sounds as well as the cognitive and physical parameters that specify the range of potential speech sounds and sound patterns. The size and shape of the speaker's vocal tract have an impact on the variety of sounds that can be produced by a human. The oral and nasal cavities, the glottis, the tongue, the velum or soft palate, the hard palate, the teeth, and the lips make up the vocal tract **(Delvaux *et al.*, 2017).** With so many factors influencing human speech, it becomes simple for us to identify a person just by their voice. So, these speech traits effectively don't alter even when someone tries to impersonate another person, making them useful for detecting disguised voices.

Additionally, the person trying to mimic the voice of another person has their own speech traits, which will occasionally come through in the imitation. Another method of verification is carried out using the PRAAT software where we compare the spectrograms of the two speakers and look for words that both speakers use frequently to determine whether or not voice impersonation has been done.

Obviously, the situation can change. A single person or two distinct people cannot pronounce the same word or sentence with the same intonation. An expert impersonator or an artist impersonating someone else will also have some distinctions that can identify disguise **(Latorre *et al.*, 2014).**

Another type of speech disguise is self-disguise, which involves concealing one's identity to avoid detection when one of one's voice recordings is being examined in court **(Hautamäki *et al.*, 2017).**

In this paper, we attempt to provide a strategy for avoiding the mimicking voice, which could pose a security problem. Humans have a tendency to mimic the speaking style of some of the famous personalities. But, from a security perspective, using a mimicked voice for any voice recognition system as a stand-in for an existing voice model is a difficult problem. Mimicry voices are highly vulnerable for any speaker recognition system. Voice dialling, banking over a telephone network, database access services, security control for secret information, and remote access to computers are all areas where a mimicking attack is likely to occur. So, to verify the claim speaker, one must choose the speaker's speech file from among the various speaker models already in use in the system **(Kanrar & Mandal, 2015).**

Objectives

The main objective of this research is:

- To compare and analyse voices of mimic artists with the voice of the original artists.
- To extract information about variations in fundamental frequency i.e., Pitch, Minimum and maximum pitch & pitch count.
- To determine instances of voice impersonation and analyse the effects of speech disguise.

Methodology

Here, we demonstrate that even the most skilled impersonators are unable to accurately imitate certain patterns used by the target speaker.

- The investigation primarily involved comparing the intonation patterns of the words delivered by the mimic artist and the genuine speaker. PRAAT is the software that is used for the study.
- The examination process involved extracting audio samples of the mimic artist and the genuine

speaker speaking the same phrase, dialogue, or word from various sources.

- The audio file was then converted into the FLAC file format using Any Audio Convertor, opened in PRAAT, the channels were changed to mono, and then chose the view and edit option to play the audios.
- The spectrogram of each audio file was displayed. Furthermore, each word that appeared often in both mimicked and original audio files was separated, and the intonation patterns of those words were examined and snapshots were.
- The variations in the harmonic composition of the frequency spectrum and differences in the pitch count is used to illustrate voice concealment.
- The artists such as Mr. Shahrukh Khan, Mr. Aamir Khan, and Late Mr. Irrfan Khan, as well as their impersonators Mr. JayVijay Sachan, Mr. Sumedh Shinde, and Mr. Sunil Pal, respectively, were selected for the original and mimicked audio samples.

Results and Discussion

The first stage of examination of voice impersonation is spectrographic comparison testing of intonation pattern. The spectrogram is the bottom half of the Sound Editor window which displays and provides the information of the acoustic characteristics of speech such as formants, pitch contour, duration and intensity. Voice bar is the dark bar in the spectrogram and shows the intensity of the sound **(www.corpus.eduhk.hk).**

Two speaker voices were taken into consideration, one from the original artist's category and the other one from the mimicked artist's category. We have selected three pairs of original and mimicry artist for this research. First is Mr. Shahrukh Khan whose voice is marked as 'O- **(www.youtube.com)** and his mimicry artist, Mr. JayVijay Sachan's voice as 'M-1' **(www.youtube.com).** Similarly, voice of Mr. Aamir Khan is marked as 'O-2' **(www.youtube.com)** and his mimicry artist, Mr. Sumedh Shinde as 'M-2' **(www.youtube.com)** & of Late Mr. Irrfan Khan as 'O-3' **(www.youtube.com)** and his impersonator, Mr. Sunil Pal as 'M-3' **(www.youtube.com).**

Figure 1 & 2 represent the wave form, spectrogram of the two incoming voices 'O-1' & 'M-1' in the very compact form. The first row presents the spectrogram and the second row presents the acoustic signal of the speaker. We select the word spoken by both the speaker which contains at least one vowel and in the same language. Here we manually selected segment portion of a dialogue from movie 'Jab Tak Hai Jaan'. The spoken- words 'ज़ुल्फ़ों' and 'नफ़रत' of both the speaker were taken into consideration (refer Table No. 2) and represented by Fig.1, 2 & 3 in which the upper spectrogram is the original voice 'O-1' and the lower spectrogram is the mimicked voice 'M-1'.

Table No. 1: Marking of original and mimic artists

Original Artists	**Marked as**	**Mimic Artists**	**Marked as**
Shahrukh Khan	O-1	JayVijay Sachan	M-1
Aamir Khan	O-2	Sumedh Shinde	M-2
Mr. Irrfan Khan	O-3	Sunil Pal	M-3

The intonation pattern of Shahrukh Khan spoken words largely differs with that of JayVijay Sachan spoken word.

Table No. 2: Clue words taken from 'O-1' & 'M1'

Voice Samples	Common words from samples
'O-1' & 'M-1'	'ज़ुल्फ़ों' & 'नफ़रत'

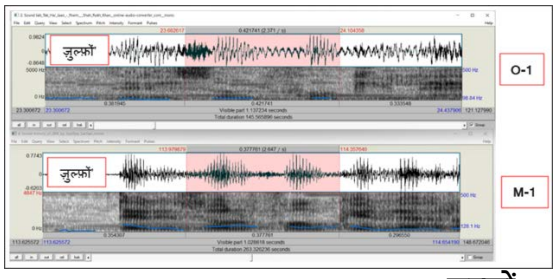

Figure No. 1: Spectrogram of word 'ज़ुल्फ़ों'

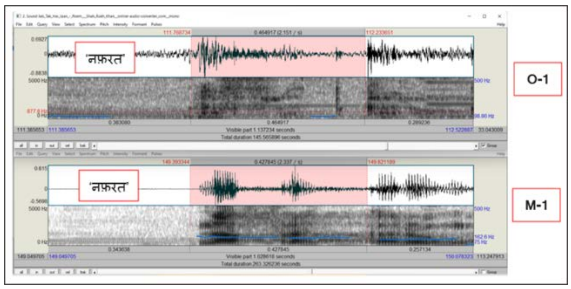

Figure No. 2: Spectrogram of word 'नफ़रत'

When the intonation pattern of the common words spoken by 'O-2' and 'M-2' were compared, the difference was clearly visible. Figs. 3 and 4 represent

the intonation pattern of words 'Very Good' & 'अच्छा' (Table No. 3).

Table No. 3: Clue words taken from 'O-2' and 'M-2'

Voice Samples	Common words from samples
'O-2' & 'M-2'	'Very Good' & 'अच्छा'

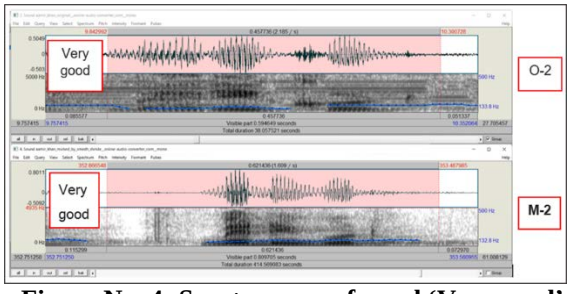

Figure No. 4: Spectrogram of word 'Very good'

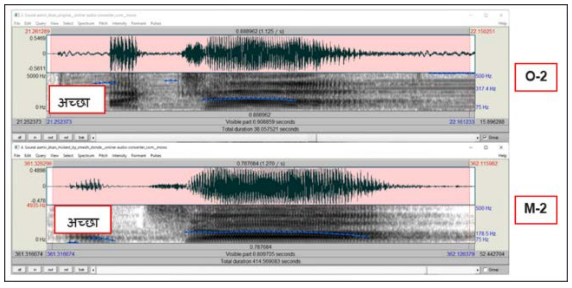

Figure No. 5: Spectrogram of word 'अच्छा'

The words 'दुनिया' and 'शराफ़त' (refer Table 4) pronounced by Late Mr. Irrfan Khan (O-3) and his impersonator Mr. Sunil Pal (M-3) are taken into account for the comparison of the intonation pattern represented by Figure No. 5 & 6.

Table No. 4: Clue words taken from 'O-3' and 'M-3'

Voice Samples	Common words from samples
'O-3' & 'M-3'	'दुनिया' and 'शराफ़त'

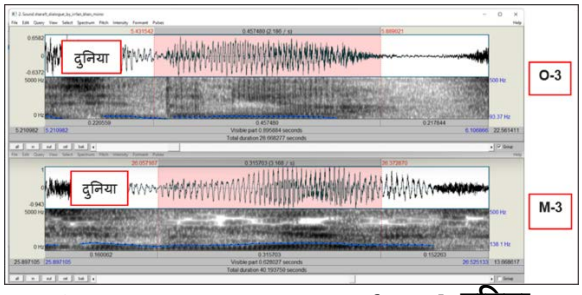

Figure No. 6: Spectrogram of word 'दुनिया'

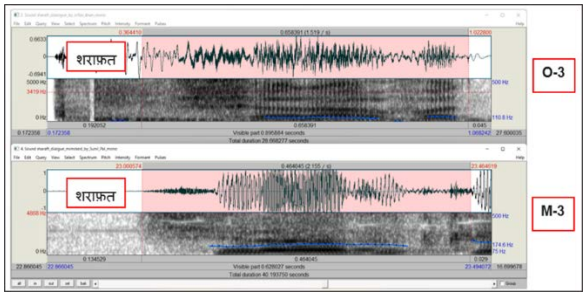

Figure No. 7: Spectrogram of word 'शराफ़त'

Second stage of verification involves comparison of common words spoken by the original artists and the mimicry artists on the basis of differences in the pitch count and Fundamental frequency (F0). Pitch is the variation in the fundamental frequency, F0 which serves as an important acoustic cue for tone, lexical stress, and intonation.

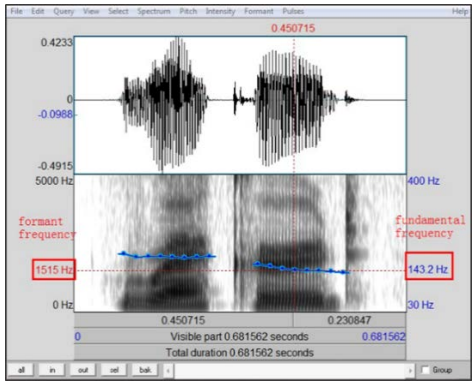

Figure No. 8: Fundamental Frequency & Pitch Count (www.corpus.eduhk.hk)

It is marked in blue on the right side of the window. In the spectrogram, the blue line stands for the pitch's rising and falling (Figure No. 7). It can be used to see the stress, tone of word, and intonation of the sentence. The intensity is marked by darkness of the bands in the waveform, and marked as a yellow line in the spectrogram **(www.corpus.eduhk.hk).**

Here we manually selected segment portion of a dialogue from movie 'Jab Tak Hai Jaan'. The spoken-words 'ज़ुल्फ़ों' and 'नफ़रत' of both the speaker were taken into consideration (refer Table No. 2) and represented by Fig.1, 2 & 3 in which the upper spectrogram is the original voice 'O-1' and the lower spectrogram is the mimicked voice 'M-1'.

From Table No. 1, it is showed that the fundamental frequency and pitch count is almost same for each individual considered in this study but difference is observed when the same word is spoken by the original speaker and the mimicry artist.

Table No. 5: Comparison of F0 and Pitch Count

Artists	Words	F0 (Hz)	Pitch count
O-1	'ज़ुल्फ़ों'	98.51	17
M-1		128.3	25
O-1	'नफ़रत'	98.89	12
M-1		163.31	28
O-2	'Very good'	133.8	37
M-2		132.8	40
O-2	'अच्छा'	317.4	28
M-2		178.5	43
O-3	'दुनिया'	93.37	32
M-3		138.1	40
O-3	'शराफ़त'	110.8	28
M-3		174.6	34

The Figure No. 1 showing the intonation pattern of the word- 'ज़ुल्फ़ों' spoken by Mr. Shahrukh Khan varies greatly with the same word spoken by Mr. JayVijay Sachan The intonation pattern also makes a very apparent distinction in pitch and intensity. From Table No. 1, the fundamental frequency varies greatly as 'O-1' has F0 value of 98.51Hz and that of 'M-1' is 128.3Hz. The number of pitch count for the speaker changes, 17 for Mr. Khan and 25 for Mr. Sachan. For the second word- 'नफ़रत', the F0 of 'O1' was observed to be 98.89Hz and pitch count was 25 whereas for 'M-1', F0 was 163.31 and pitch count, 28. The F0 range and pitch count is almost same for all the words spoken by 'O-1' but it varied greatly for 'M-1' as he was trying to imitate the voice of the original artist. This shows that even though 'M-1' spoke the same words with almost same style and accent of 'O-1', the difference in F0 and pitch count clearly revealed the case of voice impersonation, even if the intonation pattern showed slight similarity. If we compare and contrasted different words, we might find similar differences. Each person has a unique fundamental frequency value and pitch count, which can change to some extent but not as greatly as was seen in the case of voice concealment.

The comparison of fundamental frequency, pitch count, and the range of maximum and minimum frequency that a person can perceive, for example, can be used to identify self-disguise. Referring Table 1, when the word 'Very good' & 'अच्छा' are considered, the F0 value, 133.8Hz to 317.4Hz and pitch count, 37 and 28, respectively are obtained. Now, even if the 'O-2' tries to conceal his voice, the intonation pattern, F0 value and pitch count will be nearly identical, thus resolving the question of voice self-disguise.

The comparative study of the intonation patterns and the differences in the fundamental frequency and pitch count for all the words that were considered in this study here results that leads to the conclusion that even though the voice of mimicry artist has spoken the same context with the same speaking style of their respective original artist, the difference in articulation is observed.

Conclusion

This paper describes an experiment that addresses voice denial, the claim made by the speaker heard in a voice recording that the recording actually belongs to an impostor. This is a significant issue in forensic research.

The approach taken in this article comprised of examining the voice disguise in two stages: the intonation pattern examination of audio sample of the mimicry artists with that of the original speaker. Pitch count variation and Fundamental Frequency (F0) variation made up the second stage of verification. By carrying out this research, we came to the conclusion that no matter how skillfully the impersonation is done, the results showed that no one can totally conceal the voice of another individual. There will always be some variation in the frequency, pitch, intensity, and intonation pattern that aids in the research and confirmation of voice concealment cases.

References:

1.4. Analysis of speech in PRAAT (no date) Phonetics and Phonology. Available at: https://corpus.eduhk.hk/english_pronunciation/index.php/1-4-analysis-of-speech-in-praat/.

Achha: PK dialogue promo 4: Aamir Khan & Anushka Sharma: In cinemas now (2014) YouTube. Available at: https://youtu.be/UWnuOqE4waw.

Best mimicry of Shahrukh Khan by Jayvijay Sachan (2016) YouTube. YouTube. Available at: https://www.youtube.com/watch?v=Lkzhmv773RE.

Delvaux, Véronique, et al. "Voice Disguise vs. Impersonation: Acoustic and Perceptual Measurements of Vocal Flexibility in Non Experts." *Conference of the International Speech Communication Association*, 2017, https://doi.org/10.21437/interspeech.2017-1080.

Evolution of Aamir Khan - Sumedh Shinde (2018) YouTube. Available at: https://youtu.be/MlALvE6vSuo.

Hautamäki, Rosa González, et al. "Acoustical and Perceptual Study of Voice Disguise by Age Modification in Speaker Verification." *Speech Communication*, vol. 95, Elsevier BV, Dec. 2017, pp. 1–15. https://doi.org/10.1016/j.specom.2017.10.002.

Irrfan Khan V/S Nawazuddin Siddiqui best comedy by Sunil Pal (2017) YouTube. Available at: https://youtu.be/Gvtup-BZoS8.

Jab Tak Hai Jaan - Poem: Shah Rukh Khan (2012) YouTube. Available at: https://youtu.be/cx_R9BKosAE.

Kanrar, Soumen, and Prasenjit Mandal. "Detect Mimicry by Enhancing the Speaker Recognition System." Advances in Intelligent Systems and Computing, Springer Nature, 2015, pp. 21–31. https://doi.org/10.1007/978-81-322-2250-7_3.

Latorre, Javier, et al. "Speech Intonation for TTS: Study on Evaluation Methodology." Conference of the International Speech Communication Association, 2014, https://doi.org/10.21437/interspeech.2014-204.

Rodman, R.D. (2000) *Speaker Recognition of Disguised Voices: A Program for Research*. rep. Available at: https://research.csc.ncsu.edu/speakerrecognition/Papers/AnkaraPaper.pdf (Accessed: June 9, 2000).

Sharafat ki Duniya Ka Kissa Hi Khatam Best Whatsapp Status (2018) YouTube. Available at: https://youtu.be/OnarqzQCIlU.

Singh, Rita, et al. "Voice Disguise by Mimicry: Deriving Statistical Articulometric Evidence to Evaluate Claimed Impersonation." *IET Biometrics*, vol. 6, no. 4, Institution of Engineering and Technology, Feb. 2017, pp. 282–89. https://doi.org/10.1049/iet-bmt.2016.0126.