Xournals

Academic Journal of Forensic Science

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

Spectrographic and Statistical Analysis of Speech Recorded Through Different Recording Devices

Moulya. B. P[1], Geetam Shukla[2]

Available online at: www.xournals.com

Received 24th January 2023 | Revised 07th February 2023 | Accepted 24th February 2023

Abstract:

Speaker Identification is the identification of the speaker speaking the current utterance and Speaker Verification is the verification from the utterance of whether the speaker is who he claims to be. Since people are more likely to deny one's voice in many situations during ongoing crimes. This technique aids in the resolution of such cases and in the identification of the guilty. This research paper involves recognizing and verifying the speakers based on the intonation pattern of the words they speak. It is a sample-based investigation. Here, ten people's specimens are taken for analysis, with five from northern part of India and five from southern part of India. This report emphasizes the variations in intonation patterns resulting from different recording devices. Audio recording devices include a handset, a recorder, and a laptop. The difference is noted, and the extent to which the differences can be considered is also inspected.

Keywords: *Speaker Recognition, Intonation Patterns, Intensity, Pitch, Formants, Standard deviation.*

Authors:

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

Introduction

The term Speaker Recognition consists of Speaker Identification which is the identification of the speaker speaking the current utterance and Speaker Verification that refers to the verification from the utterance of whether the speaker is who he claims to be **(Almaadeed *et al.*, 2015).** Unlike to other biometric characteristics like fingerprints and faces, a human voice is a biometric characteristic that is not yet frequently employed for person identification. A system uses a recording of a speaker's speech to verify or ascertain the speaker's identification in automatic voice recognition, also referred to as speaker recognition **(Mokgonyane *et al.*, 2021).**

There are two types of speaker recognition: text-dependent and text-independent. The process is known as text-dependent speaker verification when the lexicon of the spoken utterances is limited to a single word or phrase across all speakers, as opposed to text-independent speaker verification, where speakers are free to say whatever they want without having their utterance constrained.

In this paper the approach is based on text-dependent verification for the comparison of intonation pattern of speech signals. The basic idea is taken from the fact that the vocal tract of a speaker is what distinguishes the voice from each other. Each person's vocal tract is unique in terms of size and shape which creates differences in their pitch frequency, intensity and formant frequency **(Chaubey *et al.*, 2022)**.

A man produces sound waves whenever he speaks, when the vocal folds come together and vibrate due to passing of air through them during exhalation from the lungs, they produce sound. This vibration produces the sound wave. Sound waves are characterised on the basis of frequency, amplitude and wavelength. The frequency, also known as pitch, is the number of times per second that a sound pressure wave repeats itself. The (quasi-)periodic structure of voiced speech signals are approximated by the fundamental frequency of a speech signal, which is sometimes indicated by F0. The vocal folds, when appropriately tensed, create an oscillation in the airflow. The mean size of oscillations per second, measured in Hertz, is the fundamental frequency. Formant frequencies, also known as F1, F2, and F3, are created when this fundamental frequency is amplified or attenuated by different parts of the resonating body **(Ali *et al.*, 2006; Magdin *et al.*, 2019).**

Intensity is the energy of a sound wave degree of loudness associated with it. It is measured in decibels. The intensity helps us in knowing how a person talks certain words. The amplitude of the sound wave is nothing but its height or the maximum distance that a medium's vibrating particles are moved from their average position during sound production. Due to variations in amplitude and pitch in a speaker's voice over different recordings, no two digital signals are identical, even when the same words are spoken by the same speaker **(Almaadeed *et al.*, 2015).**

The intonation patterns are evaluated in the software called Praat. Praat converts speech to digital signals. It consists of waveform view and a spectrogram view of a sound. The waveform view gives amplitude information over time whereas the spectrogram shows the frequency information over time, basically the amplitude is shown through the shadings. It enables us to determine a sound's pitch, intensity, and formant values. In this paper the examination of the intonation pattern of the words is done using these parameters. The findings and observations are listed in the table below, along with the conclusion **(Bhrati and Bansal, 2025; Wirdayanthi, 2022)**.

Objectives

- To collect voice samples of ten people.
- To compare the intonation patterns of collected voice samples recorded through different recording devices such as a recorder, mobile phone and laptop

Methodology

The collection of samples for the present study was done according to the requirements of the objective.

1. The Dataset - Voice samples of ten people were recorded thrice using different recording devices. The dataset contains 78 seconds of recordings from 10 speakers. Five of them were from the northern part of India which included states of Delhi, Haryana, Rajasthan and Uttar Pradesh, while the other five were from the southern part of India which included states of Karnataka and Tamil Nadu.

2. The sample audio file was opened in Praat (Open → Read from file….). Then the voice sample was converted to mono channel and it was analysed by clicking on “view & edit” option for viewing the intonation pattern. This opens a new tab of intonation pattern. In this tab there are options for getting the pitch, intensity and formant values. All

of the values were recorded in a tabular column and thoroughly examined.

Table No. 1: Description of collected data

S. No.	Description of Data	Value
1.	No. of speakers	10
2.	No. of samples per speaker	3
3.	Total Duration	78 sec
4.	Total Size	11.037 MB

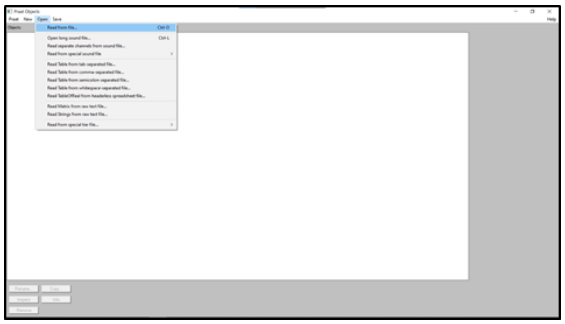

Figure No. 1: Praat Window

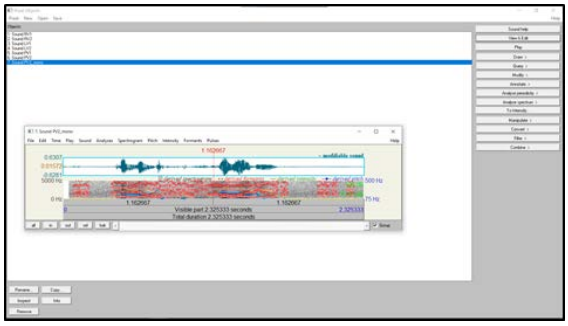

Figure No. 2: Praat Sound Window

3. Devices used for the recording were a handset, a recorder and a laptop. The specifications for the same is given below in Figure No. 3. Clue words were picked from the sentence and analysed in Praat software for pitch, intensity, and formants. Every speaker was asked to utter a sentence "Hello! How are you" in all the recording devices considered in this paper. Each person is given a marking P1 to P10 and the clue word "Hello" is taken from the sentence spoken by the speakers. (Table No. 2).

S. No.	Recording Devices	Images	Specifications	Marked as
1.	Voice Recorder		Sony Stereo IC Recorder ICD-PX470	D-1
2.	Handset		iPhone SE 2 (Version: iOS 14.7.1)	D-2
3.	Laptop		HP LAPTOP-MIS234QO Microsoft Windows 10 Home Single Language Version-10.0.19045	D-3

Figure No. 3: Table showing details of recording devices

Table No. 2: Marking of the person and clue word taken

S. No.	Person Marked as	Voice through recording devices	Clue word Taken
1.	P1	D1, D2, D3	Hello
2.	P2	D1, D2, D3	Hello
3.	P3	D1, D2, D3	Hello
4.	P4	D1, D2, D3	Hello
5.	P5	D1, D2, D3	Hello
6.	P6	D1, D2, D3	Hello
7.	P7	D1, D2, D3	Hello
8.	P8	D1, D2, D3	Hello
9.	P9	D1, D2, D3	Hello
10.	P10	D1, D2, D3	Hello

Observations

Intonation pattern of each person in three recording devices were viewed together at once in Praat software. All the three patterns were arranged one after the other for analysis and screenshot of the same is captured and pasted below. The voice sample recorded in voice recorder, handset and laptop is marked as D-1, D-2 and D-3 respectively.

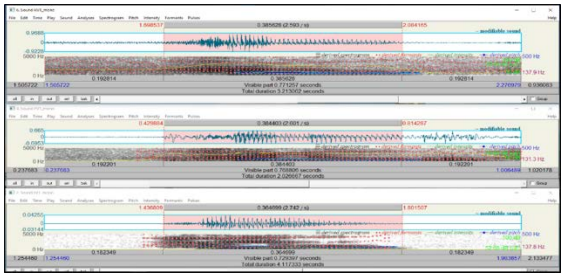

Figure No. 4: Intonation patterns of person 'P1' in three devices D1, D2 & D3

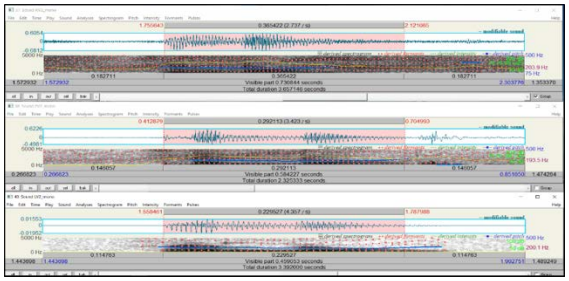

Figure No. 5: Intonation patterns of person 'P2' in three devices D1, D2 & D3

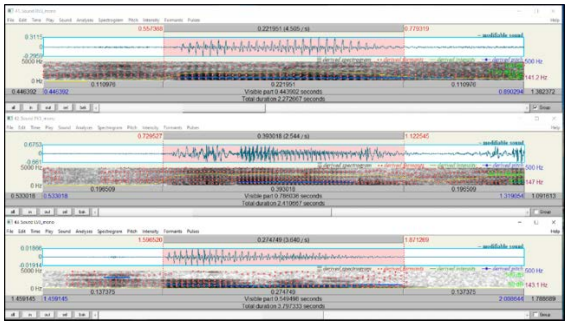

Figure No. 6: Intonation patterns of person 'P3' in three devices D1, D2 & D3

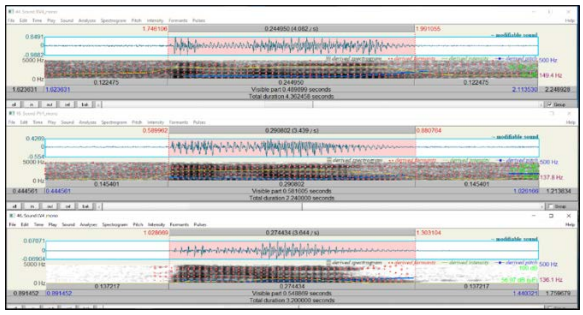

Figure No. 7: Intonation patterns of person 'P4' in three devices D1, D2 & D3

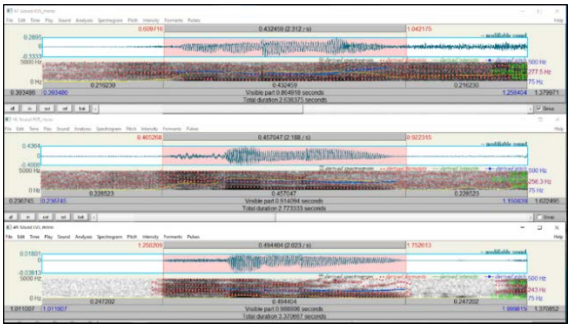

Figure No. 8: Intonation patterns of person 'P5' in three devices D1, D2 & D3

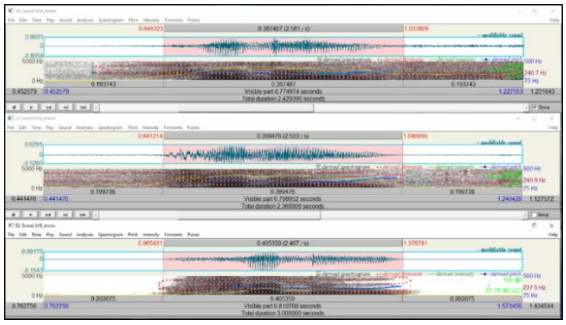

Figure No. 9: Intonation patterns of person 'P6' in three devices D1, D2 & D3

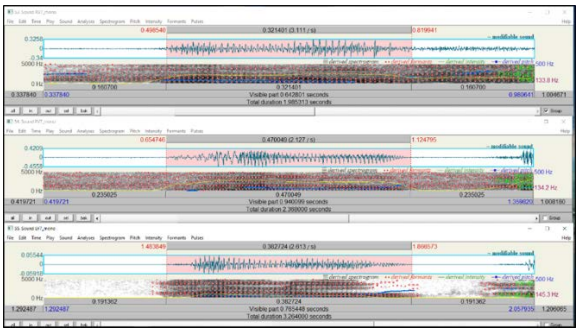

Figure No. 10: Intonation patterns of person 'P7' in three devices D1, D2 & D3

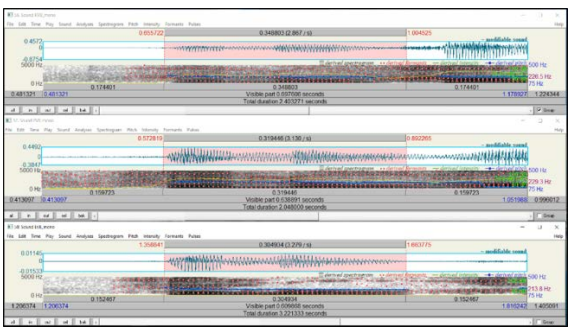

Figure No. 11: Intonation patterns of person 'P8' in three devices D1, D2 & D3

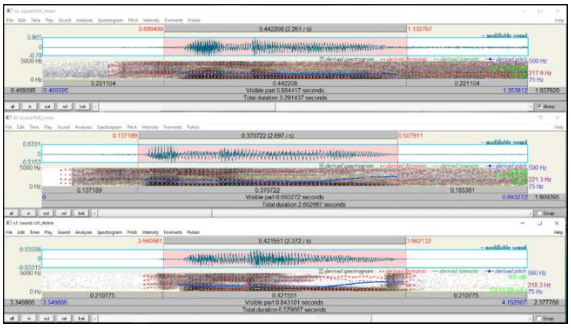

Figure No. 12: Intonation patterns of person 'P9' in three devices D1, D2 & D3

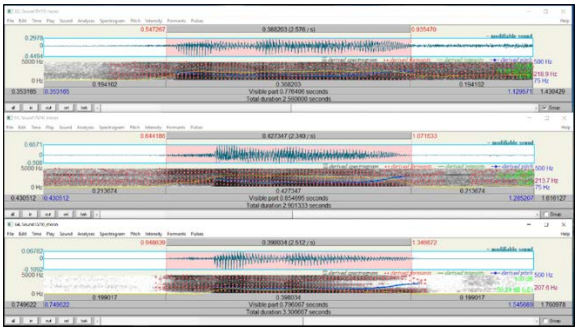

Figure No. 13: Intonation patterns of person 'P10' in three devices D1, D2 & D3

Results and Discussion

The values of Pitch (P), Intensity (I), Formant1 (F1) and Formant 2 (F2) are noted in a tabular column below. The unit of intensity is decibels (dB) and the unit of pitch and formants is hertz (Hz).

Figure No. 14: Table showing values of intensity and pitch of the sound recorded in recording devices D1, D2 & D3

S. No.	Person	D1		D2		D3	
		I (dB)	P (Hz)	I (dB)	P (Hz)	I (dB)	P (Hz)
1.	P1	77.77	137.58	79.5	131.36	52.64	137.6
2.	P2	79.3	205.16	76.9	191.9	47.28	204.02
3.	P3	71.93	141.2	80.51	147.03	46.48	143.12
4.	P4	81.9	149.36	77	137.18	56.96	136.12
5.	P5	73.26	277.4	75.91	256.29	51.65	243.04
6.	P6	78.95	240.71	79.03	240.9	61.7	237.45
7.	P7	65.9	115.7	75.45	134.15	56.75	145.32
8.	P8	74.13	226.5	76.54	229.3	45.68	204.7
9.	P9	80.71	211.8	77.78	221.3	53.52	218.27
10.	P10	74.3	218.8	78.48	213.73	59.23	207.5

Standard deviation for values of intensity and pitch for each person in different recording devices was calculated using the following formula:

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(xi - \bar{x})^2}{N}}$$

σ – Standard deviation
xi – Data Value
$\bar{x}$ – Mean
N – Total population

Figure No. 15: Table showing values of F1 and F2 of voice samples recorded in devices D1, D2 & D3

S. No.	Person	D1		D2		D3	
		F1 (Hz)	F2 (Hz)	F1 (Hz)	F2 (Hz)	F1 (Hz)	F2 (Hz)
1	P1	646.74	1433.37	506.13	1395.8	554.58	1357.96
2	P2	674.1	1784.8	443.37	1697.4	485.6	1110.86
3	P3	539.17	1414.77	504.42	1320.14	478.97	1025.02
4	P4	497.16	1411.11	438.11	1411.8	415.2	1356.02
5	P5	531.27	1646.82	652.03	1834.14	559.27	1514.46
6	P6	677.16	1591.9	623.5	1590.6	618.83	1564.69
7	P7	450.3	1301.67	455.48	1408.4	463.54	1196.58
8	P8	573.8	1518.23	444.8	1374.04	465.09	1263.98
9	P9	654.1	1480.59	490.93	1412.18	532.8	1324.42
10	P10	564.4	1578.21	636.62	1424.62	621.6	1400.7

Fundamental frequency (F0) is the sound wave produced from vocal cord. It is intensified or damped by numerous parts of the resonation body. Resulting frequencies are multiple of fundamental frequency (F0). These are called formants. Formants are different for different vowels and consonants; this distinguishes how the person talks and the manner in which he or she talks is also determined. As you can observe from the table above, these frequencies are unique to each person. This helps us in the identification of the person. These determine a person's accent and dialect. Variations in digital signals of voice samples has occurred for a variety of reasons. The reasons could include natural variations due to environmental factors, which include background noise, due to type of location like open or closed, angle of the recorder while recording, movements while talking and also the model of the recording devices. The other reason is individual factor which includes the state of mind of the speaker, his emotions like fear, anger, happiness, sad etc. and also the variations in the vocal tract of each individual. These factors in turn affect the Pitch, Intensity, Base and other properties of the speech signal. After the calculations, it is found that the variations seen in intensity is 11.3dB and the variation seen in pitch is 7.93Hz.

Conclusion

Every person is distinctive in their own particular manner. This uniqueness has paved way for the forensic scientists to differentiate people from each other and provide their valuable researches. Speaker Identification is growing and is a futuristic subject that we need to dwell on. This paper has tried to analyse this uniqueness in a person's voice and the factors causing disturbances while analysing for the speaker's identity. This paper uses statistical study to come to conclusion for the average variations created in pitch and intensity of same voice signal in different recording devices. First the values of the pitch and intensities for the same voice in different recording devices were noted in a tabular column, then the standard deviation in the values from each set of the same person is calculated. Later average for ten people were calculated to know the average variations created when same voice is recorded in different recording devices.

References:

Ali, Ahmed, et al. Formants Based Analysis for Speech Recognition. 2006, https://doi.org/10.1109/iceis.2006.1703179.

Almaadeed, Noor, et al. "Text-Independent Speaker Identification Using Vowel Formants." Journal of Signal Processing Systems, vol. 82, no. 3, 2015, pp. 345–356., https://doi.org/10.1007/s11265-015-1005-5.

Bharti, Roma, and Priyanka Bansal. "Real Time Speaker Recognition System Using MFCC and Vector Quantization Technique." International Journal of Computer Applications, vol. 117, no. 1, 2015, pp. 25–31., https://doi.org/10.5120/20520-2361.

Chaubey, Ashutosh, et al. "Improved Relation Networks for End-to-End Speaker Verification and Identification." Interspeech 2022, 2022, https://doi.org/10.21437/interspeech.2022-10064.

Magdin, Martin, et al. "Voice Analysis Using PRAAT Software and Classification of User Emotional State." International Journal of Interactive Multimedia and Artificial Intelligence, vol. 5, no. 6, 2019, p. 33., https://doi.org/10.9781/ijimai.2019.03.004.

Mokgonyane, Tumisho Billson, et al. "A Cross-Platform Interface for Automatic Speaker Identification and Verification." 2021 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (IcABCD), 2021, https://doi.org/10.1109/icabcd51485.2021.9519322.

Wirdayanthi, A.A. Istri. "UTILIZATION OF PRAAT IN DETERMINING THE AUTHENTICITY OF VOICE." IJFL (International Journal of Forensic Linguistic), vol. 3, no. 1, Apr. 2022, pp. 81–89., https://doi.org/https://www.ejournal.warmadewa.ac.id/index.php/ijfl/index.