

■ REVIEW ARTICLE

Study on Disguised Voice Recording using Voice Changer Apps

Aditi¹, Monika Chauhan², Akbar Ali³

ABSTRACT

Authors' Affiliations:

¹Student

²Associate Professor, School of
Basic and Applied Sciences,
Galgotias University, Greater
Noida, Uttar Pradesh

³Senior Scientific Assistant
Forensic Science Laboratory,
Patna, Bihar 800023, India.

Corresponding Author:

Monika Chauhan

Associate Professor, School of
Basic and Applied Sciences,
Galgotias University,
Greater Noida, Uttar Pradesh.

Email:

monika0283@gmail.com

Speech science has a storied and illustrious past. Many studies are conducted for research on speech and voice acoustics, which focuses on speech. The human voice is our most basic and self contained instrument. When it comes to voice processing, the source filter paradigm may be used. It's common to think of the voice as having two distinct processes: one that produces the initial tone and another that produces the final tone. This is done in order to create a speaker's identity by providing an accurate phonetic profile of that individual. Other related fields include improving low quality audio files and verifying disputed recordings to see whether they have been manipulated, edited, or otherwise tampered with. In contrast, a digital recording samples the transduced waveform and transforms each sample into a binary number code similar to that used by computers. Speech analysis was also used in the investigation of recordings allegedly made by Osama bin Laden, the world's most wanted terrorist. This research will discuss if voice changer apps can be used to mask one's voice. There have been several instances where the perpetrator has used these apps to alter their speech. This research looks at how the characteristics of any voice can be altered or not by the use of these apps, how they function, and how they affect voice recognition. This paper also examines whether these apps really assist the perpetrator in eluding voice recognition experts.

KEYWORDS | Forensic speech analysis, voice changer apps, Audio analysis

INTRODUCTION

SPEECH SCIENCE HAS A LONG AND glorious history. Many studies, including mine, which explores the speech, is conducting research on speech and voice acoustics.¹ Chatting, singing, laughing, moaning, crying, or shouting are examples of sounds produced by a person using the vocal tract. Human voice pitch is generated primarily by the vocal cords. (Other sound-generating mechanisms include unvoiced consonants, taps, whistling, and humming, which all come from the same general area of the body.) The human voice is our most basic and self-contained

instrument.²

Indian languages represent the diverse layers of social order and caste. Individuals' speech repertoire includes a range of modes and dialects that are appropriate for various social situations. The better a speaker's rank is, the more speech forms he or she has at his or her fingertips.³ To signify the speaker's social standing and the social context in which they speak, speech is altered in a number of ways.

When it comes to speech synthesis, the source filter paradigm may be used. It's common to consider the voice to have



How to cite this article

Aditi. Study on Disguised Voice
Recording using Voice Changer Apps.
Indian J Forensic Med Pathol.
2021;14(3 Special):316-321.

two distinct processes: one that produces the initial sound and another that modifies it. For example, the larynx (also known as the “speech box”) emits a sound with a wide variety of frequencies. The root is the vibrating vocal folds in the larynx.⁴ The filter is the vocal tract, which runs from the vocal folds to the lips or nostrils.

The larynx is a device located at the top of trachea. The vocal folds are two vibrating flaps of flesh in the voice. In a resonant system, standing waves can form. When the sound is up at the vocal folds and low at the mouth, standing vibrations, or resonances, occur in the vocal tract.⁴

Exhalation, phonation, and articulation are the three stages of speech development at the periphery level.

Formants are the more noticeable frequencies that differentiate distinct vowel sounds.

Voice Changer

A voice modulator (also called a voice changer) is a computer software programme that changes the human voice in real time. Not all voice changers are electronic devices of their own. Many of these voice changers are software programmes that can alter large swathes of vocabulary.⁵

Voice changers are also used as speech disguisers. When a voice changer is used, the accent varies so dramatically that even close friends and family members are unable to say who is speaking to, based solely on speech. They largely act by altering the user’s voice pitch.⁶

A defendant was found guilty of delivering a bomb message over the internet in the case of United States v. Gilbert. According to the district court, the defendant purchased a toy voice changer on the day the phone bomb threat was made. To disguise the speaker’s voice, this voice changer may alter the sound of the speaker’s voice up or down.⁷

The defendant made some offensive calls by using the voice changer. Since adjusting the sound of the audio conversations so that the defendant’s own voice could be heard, an investigator was able to prove that the defendant was the one making the calls. The majority of voice changers alter the tone of a person’s natural voice.⁸

In Gilbert, the court dealt with the testimony of a voice changer by making an investigator testify to the defendant’s name. Although an expert may

be able to analyse metadata associated with a voice file, the required analysis, would be challenging.⁹

MATERIALS AND METHODS

For this research voice samples of 10 person (5 male & 5 female) with five filters including original in three voice changer apps has been calculated. The age group of the voice sample is 20-55 years.

Details of Different Apps

The transcript of the voice samples is in Devanagari script. The script is in Devanagari because Hindi is our mother tongue and it is easy to read by everyone and also easy to analyse. For analysis, different tools like Hash Cal, Gold Wave, Adobe Audition, Praat were used.

Hash Cal is used for Hash Calculation. This is essentially a set of numbers resulting from a complicated mathematical theorem. Gold wave is used for waveform Analysis. Adobe Audition is used for Spectral Analysis. Praat helps in Spectrographic Analysis.

RESULTS AND DISCUSSION

All the samples were properly arranged in a database with all the details including file name, file size, sampling rate, duration, format, mode of sample, details of recorder, hash value. Having a proper database concluded different variations in the samples.

Every filter is changing the file size, file duration, hash value of the original sample. The hash value is calculated by the software hash calculator.

In Hash cal the readings of SHA1 and MD5 is observed.

After having the details of each and every sample, it is observed that these apps are changing the audio properties of the original file.

When the details of each file are calculated, there is difference in every file. This shows that this app changes the details of original file. Here one original file’s hash calculation and one modulated file’s calculation is attached.

Then the file duration and sampling rate is notified from the Gold Wave app [10]. All conversions are performed by Gold Wave using Windows speech tools, so the consistency of the voice or the precision of the identification is solely dependent on that software.

S.N	NAME	SAMPLE NAME	MALE/FEMALE	AGE
1.	Person 1	P1	Female	29
2.	Person 2	P2	Female	21
3.	Person 3	P3	Male	32
4.	Person 4	P4	Male	25
5.	Person 5	P5	Male	25
6.	Person 6	P6	Male	36
7.	Person 7	P7	Male	52
8.	Person 8	P8	Female	48
9.	Person 9	P9	Female	20
10.	Person 10	P10	Female	22

Table 1: Details of the persons

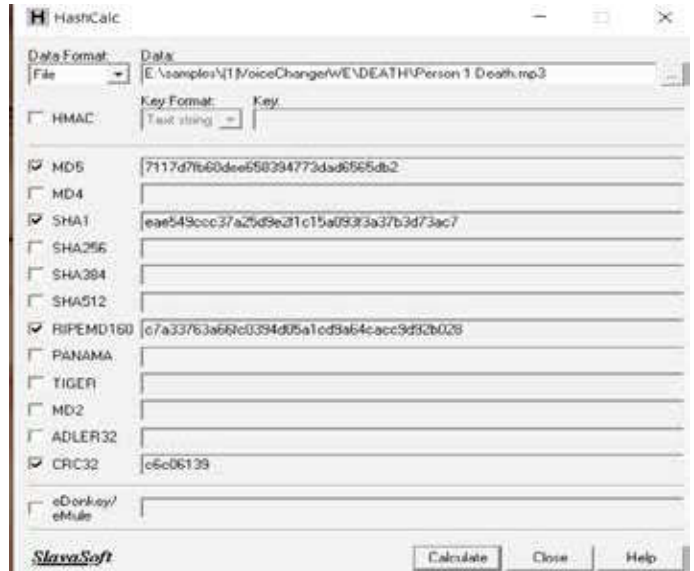


Figure 1: Representation of Hash calculations

 APP 1: Voice Changer WE Powered by FMOD studio by firelight Technologies	File name APP 1: Voice Changer WE	Filters Original Death Deep voice Robot Villain Zombie
 APP 2 (MyVoiceChanger) Powered by FMOD studio by firelight Technologies	File name APP 2 (MyVoiceChanger)	Filters Original Chipmunk Girl man multiple Slow Down
 APP 3: Voice Changer Studio Powered by Ngwyen Van Kim, Hanoi, Vietnam	File name APP 3: Voice Changer Studio	Filters Original Bee Devil Gaint Helium Hexa Fluoride

Table 2: Voice changer Apps details

File name	Samples	Filters	Samples Rates
APP 1: Voice Changer WE Powered by FMOD studio by firelight Technologies	51	Original	16000Hz, 64kpbs, Mono
	52	Death	16000Hz, 64kpbs, Mono
	53	Deep voice	16000Hz, 64kpbs, Mono
	54	Robot	16000Hz, 64kpbs, Mono
	55	Villain	16000Hz, 64kpbs, Mono
	56	Zombie	16000Hz, 64kpbs, Mono
APP 2 (MyVoiceChanger) Powered by FMOD studio by firelight Technologies	51	Original	24000Hz, 112kbs, Joint stereo
	52	Chipmunk	24000Hz, 56kbs, Joint stereo
	53	Girl	24000Hz, 96kbs, Joint stereo, VBR
	54	man	24000Hz, 64kbs, Joint stereo
	55	multiple	24000Hz, 96kbs, Joint stereo, VBR
	56	Slow Down	24000Hz, 96kbs, Joint stereo, VBR
APP 3: Voice Changer Studio Powered by Ngwyen Van Kim, Hanoi, Vietnam	51	Original	44100Hz, 320kbs, Joint stereo
	52	Bee	44100Hz, 192kbs, Joint stereo
	53	Devil	44100Hz, 192kbs, Joint stereo
	54	Gaint	44100Hz, 192kbs, Joint stereo
	55	Helium	44100Hz, 192kbs, Joint stereo
	56	Hexa Fluoride	44100Hz, 192kbs, Joint stereo

Figure 2: Difference in Sampling Rate

FILE NAME	SAMPLE	FILTERS	SIZE	DURATION	LATENCY TIMES (ms)
Person 6	S 31	Original	172kb	21.927	Start- 0.450 End- 0.146
	S 32	Death	234kb	29.887	Start- 0.874 End- 0.060
	S 33	Deep voice	214kb	27.363	Start- 0.759 End- 0.132
	S 34	Robot	150kb	19.119	Start- 0.390 End- 0.619
	S 35	Villain	228kb	29.127	Start- 0.579 End- 0.039
	S 36	Zombie	379kb	48.459	Start- 0.922 End- 0.084

Table 3: Difference in latency time

This result shows that there is no change in the sampling rate of the files of same voice changer app. There are changes in the sampling rate of different apps but this varies due to the app's coding.

Latency in Adobe Audition refers to the time between which an audio signal reaches a device and when it returns (usually measured in milliseconds).

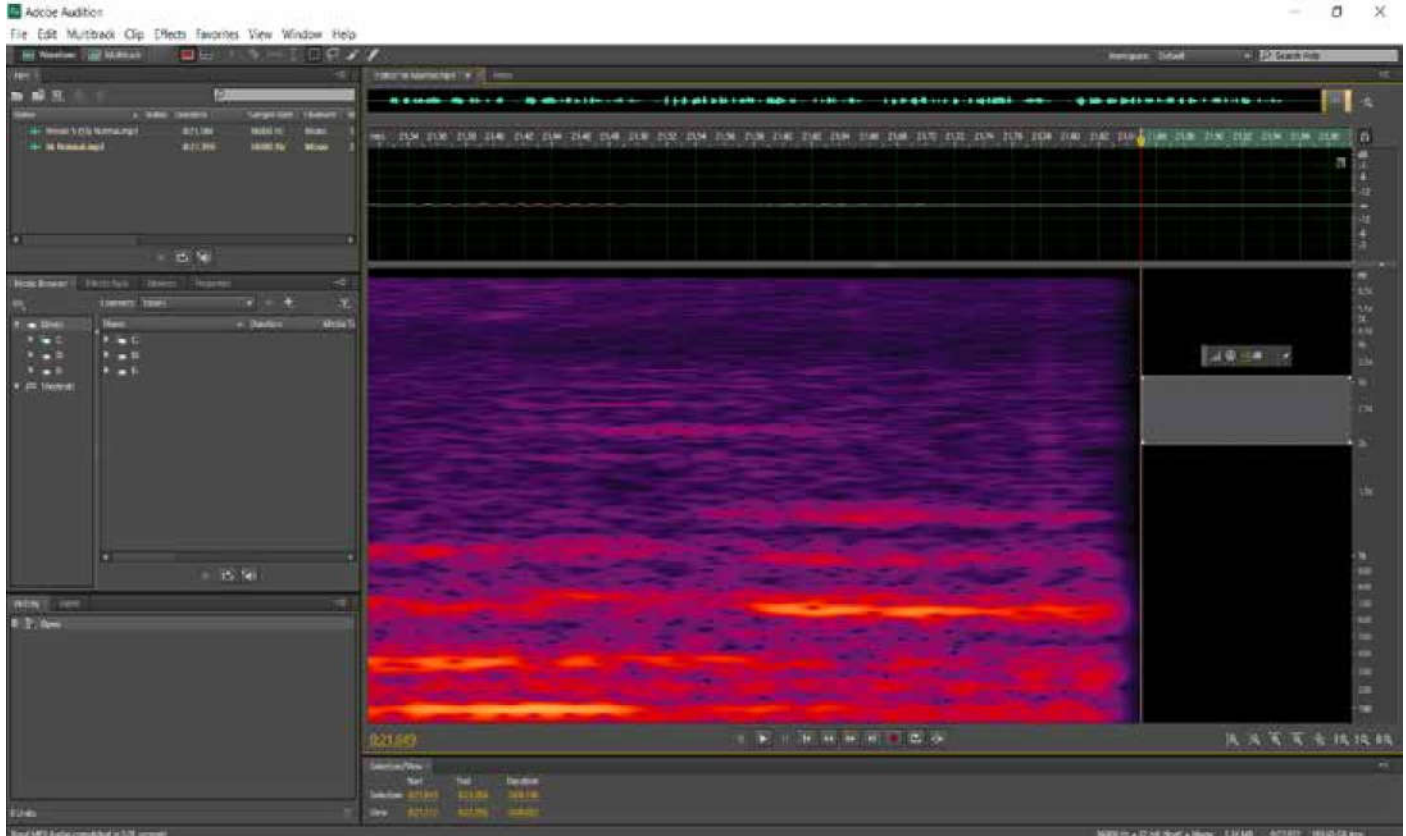
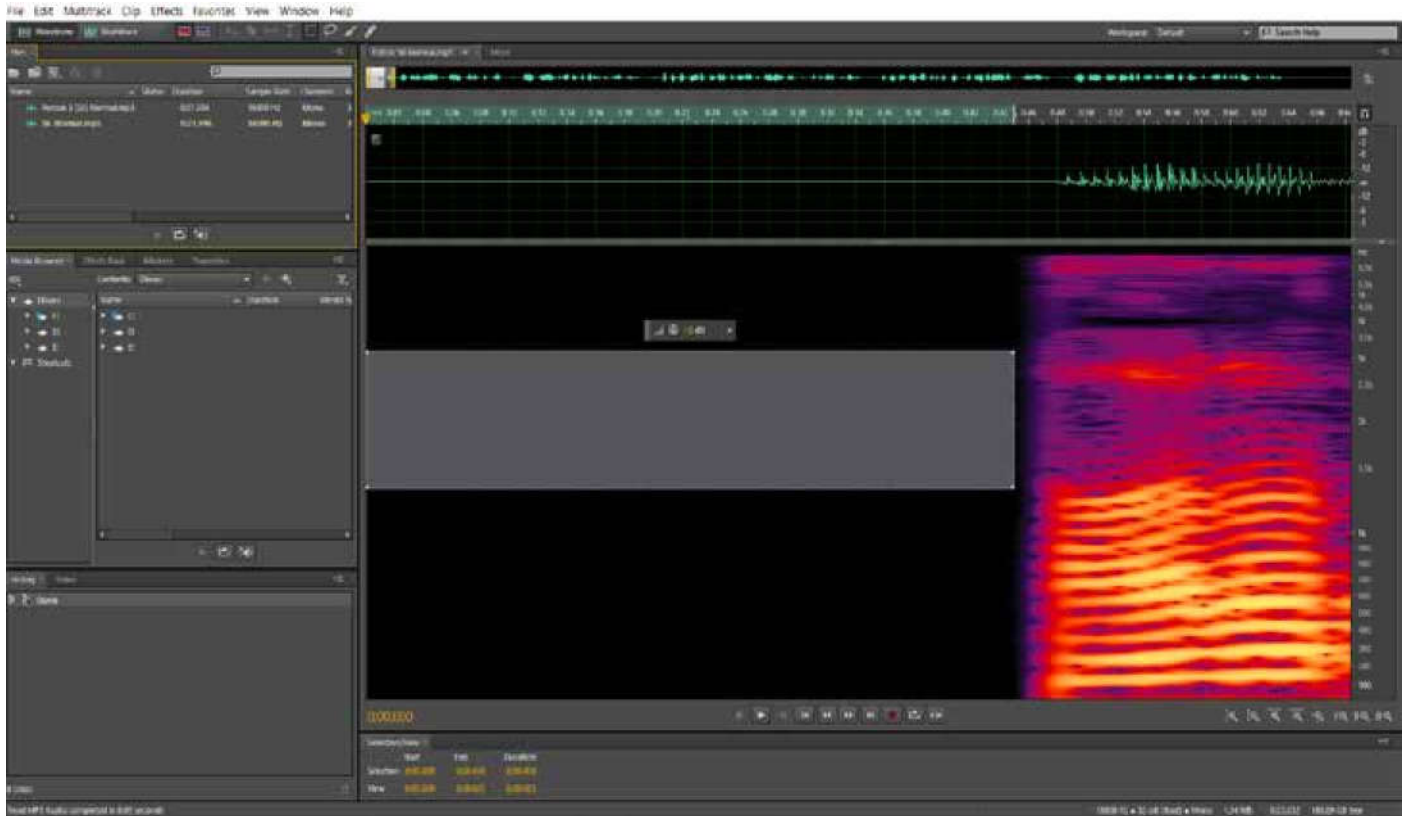


Figure 3: Calculation of Latency time.



Figure 3: Spectral analysis of Samples.

Analog-to-digital switching, buffering, digital signal processing, transfer time, digital-to-analog conversion, and the speed of sound in the transmission medium are all potential contributors to delay in an audio system.^{11,12} After analysis it is found that the latency of each sample differs from another sample. This shows that phonetic features changes.

Praat: Spectral analysis is the process of measuring or calculating the amount of acoustic energy present in a signal at various frequencies.^{13,14} Note that this will change at any time, so specifying that spectral analysis applies to a certain point in time is normally necessary.

The spectral analysis shows the features of any

individual's voice. The observation shows lack of differences in both of the file. This seems that identity cannot be concealed through these apps.

The results of the study are presented and discussed with reference to the aim of the study, which is to determine disguise on voice recording using voice changer app. The objectives of this paper is discussed above so here results are presented. The first objective of the research is about the features get changed through these apps. So, as per results of study, only phonetic features can be changed by the filters of this app. This means that the duration of the file, size of the file, sound effects is changed by these apps. Original features of the person is not able to get changed.

As the samples are observed under praat software, it is clearly visible, the original features is not disguised through these apps.

The second objective of paper is that it is possible for any perpetrator to hide his/her identity through these voice changer apps. Results of my study shows that it is not possible to hide one's identity through these apps. It is found from analysis that no original features of the voice is changed through these apps.

CONCLUSION

A voice changer is a software application that modifies human speech in real time. Many of these voice changers are computer software apps that can alter large chunks of language. Phone texts, voice mail, ransom demands, prank calls, and calls to emergency or police numbers are all examples of recorded fragments that may be investigated. Forensic experts can restore, recover, optimise, and interpret audio recordings using a variety of scientific instruments and procedures. For my study, recordings of the voices of ten people, using five filters, including the original, in three voice changer applications, have been collected. A variety of methods for analysis, including hash cal, goldwave, Adobe Audition, and Praat has been used.

According to the findings of this research, the filters in this app will only alter phonetic functions. These applications adjust the length of the file, the scale of the file, and the sound effects. When the samples are examined with Praat tools, it is apparent that the original features are not obscured by these applications.

Acknowledgement:

The author would like to thank Dr. A.K. Jain, for being a source of inspiration and advice. He thankful for his guidance, full support and care. He also expresses his sincere gratitude and thanks to all the faculties and staff members of division of forensic science for their kind guidance and valuable suggestions during the completion of my paper. He is also thankful to all who gave him the opportunity to do work on this dissertation and their valuable suggestions and encouragement.

Conflict of Interest:

The authors declare that there is no conflict of interest in this article.

Source of Funding: Nil

REFERENCES

1. **Kabari, L. G.** Speaker Recognition System Using Signal Processing Tool In Matlab. *International Journal of Advanced Research in Computer Science*, 2019; 10(4), 5-8.
2. **No authors listed.** Indian Languages: Classification of Indian Languages. (2013, December 17). Retrieved from <https://www.yourarticlelibrary.com/language/indian-languages-classification-of-Indian-languages/19813>.
3. **Zhang C, van de Weijer J, Cui J.** Intra- and inter-speaker variations of formant pattern for lateral syllables in Standard Chinese. *Forensic Sci Int.* 2006 May 10; 158(2-3):117-24.
4. **Narayan, H.** (2021, April 27). India, a land of many tongues. Retrieved from <https://www.thehindu.com/thread/arts-culture-society/india-a-land-of-many-tongues/article19445187.ece>
5. **No authors listed.** Voice Analysis. (n.d.). Retrieved from <https://what-when-how.com/forensic-sciences/voice-analysis/#:~:text=Everyone knows of forensic speaker, authenticity or integrity examination.>
6. **Catanzaro, M.** (2017, January 25). Voice Analysis Should Be Used with Caution in Court. Retrieved from <https://www.scientificamerican.com/article/voice-analysis-should-be-used-with-caution-in-court/>
7. **No authors listed.** Forensic Audio and Video Analysis. (n.d.). Retrieved from <http://www.forensicsciencesimplified.org/av/index.htm>
8. **Cognola, F., Baronchelli, I., & Molinari, E.** (2019, June 17). Inter- vs. Intra-Speaker Variation in Mixed Heritage Syntax: A Statistical Analysis. Retrieved from <https://www.frontiersin.org/articles/10.3389/fpsyg.2019.01528/full>
9. **No authors listed.** New Indian Xpress. (2019, February 21). Man uses voice changer app, cheats men. Retrieved from <https://www.newindianexpress.com/states/karnataka/2019/feb/21/man-uses-voice-changer-app-cheats-men-1941572.html>
10. **Ponterio, R.** Retrieved from <https://web.cortland.edu/flteach/mm-course/sound.html>
11. **No Authors listed.** GoldWave Audio Editor Software - Audio & Video Editing ... (n.d.). Retrieved from <http://goldwave.com/goldwave.php>
12. **No authors listed.** Audition User Guide. (n.d.). Retrieved from <https://helpx.adobe.com/audition/user-guide.html/audition/system-requirements.ug.html>
13. **J. Spragens, Xrickrockx, RonNovy, Steve_C2, Tilted_lab.** (2012, May 15). Audition CS 5.5 and the future of Audition. Retrieved from <https://community.adobe.com/t5/audition/audition-cs-5-5-and-the-future-of-audition/td-p/3418780>
14. **No authors listed.** Spectral analysis Lu. Retrieved from <https://person2.sol.lu.se/SidneyWood/praaate/spect.html>