July - December 2022 Volume 31 Issue 2 PRINT ISSN: 2277-1867
ONLINE ISSN: 2277-8853

JOURNAL OF FORENSIC MEDICINE SCIENCE AND LAW

Official Publication of Medicolegal Association of Maharashtra

MULTISPECIALITY, MULTIDISCIPLINARY, NATIONAL
PEER REVIEWED, OPEN ACCESS, MLAM (SOCIETY) JOURNAL
Indexed with Scopus (Elsevier) & Index Copernicus (Poland)

Editorial Office Address
Department of Forensic Medicine & Toxicology, Third Floor, Library Building, Seth G S Medical College & KEM Hospital, Parel, Mumbai, Maharashtra, India. Pin-400 012. Email id: mlameditor@gmail.com Phone: 022-24107620 Mobile No. +91-9423016325.

JOURNAL OF FORENSIC MEDICINE SCIENCE AND LAW
(Official Publication of Medicolegal Association of Maharashtra)
Email.id: mlameditor@gmail.com

PRINT ISSN: 2277-1867
ONLINE ISSN: 2277-8853

Original Research Article

Speaker identification of the Electronically Disguised Voices

Palak Aneja[a], Sumit Kumar Choudhary[b], Surbhi Mathur[c]

[a]Research Scholar; [b]Senior Assistant Professor, School of Forensic Science, Risk Management & National Security, Rashtriya Raksha University Gandhinagar, Gujarat, India. [c]Senior Assistant. Professor, School of Forensic Science, National Forensic Science University, Gandhinagar, Gujarat, India.

Article Info

Received on: 13.09.2021
Accepted on: 09.02.2022

Key words
Acoustic,
Disguise,
Forensic,
Speaker,
Identification.

Abstract

Background: Voice is said to be unique for every individual and can be used for the identification. Significant intra variation in Speech samples can be seen with the changes in the physical and mental conditions. Analysis of disguised voice samples is challenging to the Forensic Speaker Identification Expert. Speech processing softwares which are freely available and can be used to change voice samples. These softwares add some background variations along with the alter in the frequency distribution in the voice. **Methodology:** This paper focuses on the examination of the voice samples which are processed by a freely available android application which changes the female voice samples to that of male. The examination was done on the OT- Expert 6.0. The Examination of the voices was done by the Aural-Acoustics method. **Results:** For the examination of the voice samples on acoustics parameters such as fundamental frequency(F0) and Formant Frequencies (F1, F2, F3, F4) for 5 vowels (/ʌ/, /ɔ/, /i/, /I/, /u/) compared with their control recordings showed prominent and noteworthy differences. Aural examination parameters such as relative pauses, background variations, and linguistic features such as delivery of speech, specific pronunciations for words remained similar. **Conclusion:** Some specific formant frequencies for 5 vowels (/ʌ /, /ɔ /, /i /, /I /, /u/) combined with auditory parameters especially the degradation of background of the audio for suspecting a case of electronically disguised voice and then compared it with a possible list of suspects for speaker identification. Therefore, based on some specific aural-acoustics features electronically disguised voices can be identified amongst the suspected voice sample.

1. Introduction

Speaker recognition is the technique used for the identification of the speaker using speaker-specific information from the speech signals. Speaker identification is the process of determining from which of the registered speakers a given utterance comes;

How to cite this article: Aneja P, Choudhary SK, Mathur S. Speaker identification of the Electronically Disguised Voices. J For Med Sci Law 2022;31(2):56-60.

Corresponding author: Palak Aneja, Research Scholar, School of Forensic Science, Risk Management & National Security, Rashtriya Raksha University, Gandhinagar, Gujarat, India. Email id: palakaneja78083@gmail.com (M): +91-9999832247.

speaker verification is the process of accepting or rejecting the identity claimed by a speaker[1] Some applications include secure access control by voice, customizing services, or information to individuals by voice, indexing or labelling speakers in recorded conversations or dialogues, surveillances.[2]

In voice disputed cases encountered in Forensic Science both the scenarios speaker identification and verification are very significant. Identification is correlated with the physiological and behavioural characteristics of an individual's speech production system.[3] To hide their identity criminals, use various disguise methods like keeping a handkerchief over the phone's mic or producing nasal sound, doing mimicry to cheat with the system etc.[1] All these methods open new challenges to the forensic speaker identification expert. The last decade's development in the software technology has provided more options in disguising the identity. There are multiple android and windows operating applications that change the given voice to that of alien, robot, other gender, child etc. Forensic Speaker identification in India and some other countries majorly depends on the Formant Frequency, Pitch, and Auditory methods for analysis.[4]

This research has reported a voice conversion method based on analysis and transformation of the characteristics that define a speaker's voice.[5] This manuscript gave an overview of real-world applications by extensively studying existing systems of voice conversion, and discussed remaining challenges in the voice conversion system.[6] In one of the study Voice feminization therapies were exploited for male to female transgender women by increasing the fundamental frequency and increasing the formant frequency, the result suggested that voice was convincing as not male.[7] Research conducted to investigate the contribution of formant frequencies for gender identification found that female speaker was perceived as a female even at a speaking fundamental frequency in the typical male range, whereas for male speaker's gender perception was less accurate at speaking fundamental frequencies at 165 Hz and higher. Even if there was considerable overlap between genders, significant differences in formant frequencies of male and female were seen.[8]

2. Experimental Procedures

To conduct this study, voice samples of 50 females from age 20- 35 years were collected. All the recordings were collected in a controlled environment at the Acoustics Laboratory at the School of Forensic Science & Risk Management, using Motu Audio System. Hindi Transcript was given to all the speakers. Each speaker was asked to recite the given script for five times. To make them familiarize with the script, it was given 30 minutes before the recording to each speaker. For creating the electronically disguised sample every fourth recording of each speaker was converted into the male voices, using a free android application. The selected android application was highly rated and have multiple changing parameters.

Changing the voice samples into the Child, old-age, alien, Robot, etc. For this study, the recoded female voice samples were converted to Male voices. After processing each recording, Speaker Identification was done using the Semi-Automatic method. Selected parameters for auditory analysis were quality of speech, delivery of the speech, nature of pause, background variation, and speech rate. In Spectrographic analysis Pitch and Formant frequencies (F1, F2, F3, F4) were studied. Five vowels studied are **/ʌ/, /ɔ/, /i/, /I/, /u/**. Wilcoxon Signed Rank Test was used as a statistical procedure to determine whether the mean difference between two sets of observations is significant ($p<0.05$) or not.

3. Results

On Auditory Analysis, parameters such as **Voice quality, Delivery of Speech, Nature of Pauses, Background Variation, and Speech Rate** in the controlled audio was found to be good, clear, natural pauses, no background noise, and speech rate was also observed to be normal. The audios were recorded in the control situation, to study the effect of electronic disguising software on these parameters. In the electronically disguised voice, the **Voice Quality** was degraded in 42%, 50% had normal quality of speech and in the remaining 8% of the audios the quality of voice was good. **Delivery of Speech** was degraded in all the audios but in 20% of the audios, the delivery was degraded to an extent that it was difficult to understand the content of the audios. While examining **Nature of Pauses,** in 50 % of the audios the duration of pauses was increased, in 40% the duration of pauses was decreased, and only in 10% audio the pauses remained the same. **Background Variation** was added through reverberations were added in the audios. **Speech Rate** showed degradation and high variation from their original voices. After the auditory analysis of the audios, both control and disguised audios were run through the Acu- Expert 6.0 for **Spectrographic analysis**. The fundamental frequency of the audio and formant frequencies (F1, F2, F3, F4) for the five vowels **/ʌ/, /ɔ/, /i/, /I/, /u/** were marked. The values were then compared among themselves using Wilcoxon Signed Rank Test. The results of Wilcoxon Signed Rank Test were tabulated and shown below.

Figure 1: Distribution of Voice Quality in electronically disguised voice.

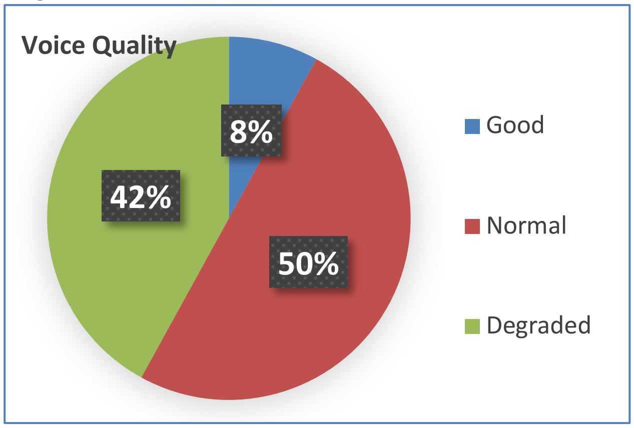

Figure 2: Distribution of Nature of Pauses in electronically disguised voice.

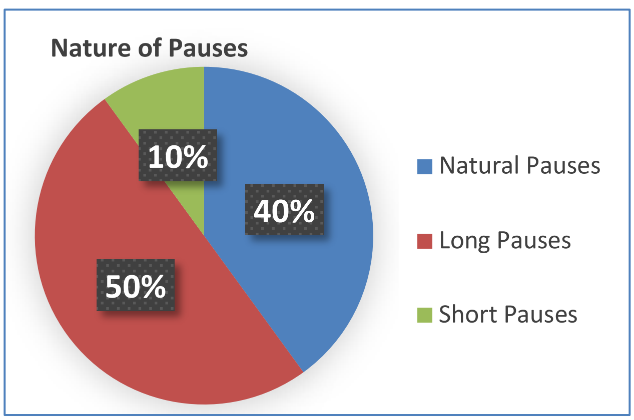

Figure 3: Distribution of Delivery of Speech in electronically disguised voice.

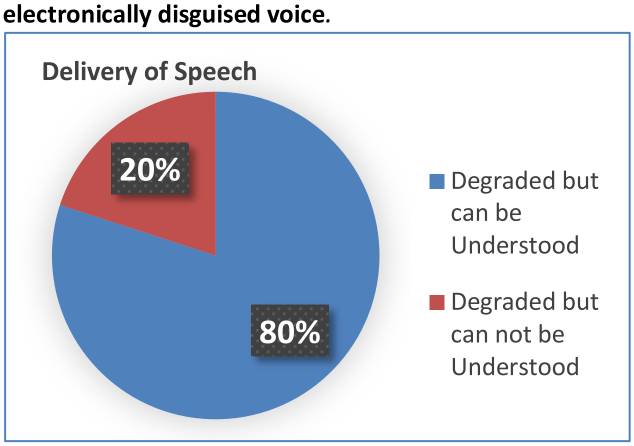

Figure 4: Distribution of Background Variation in electronically disguised voice.

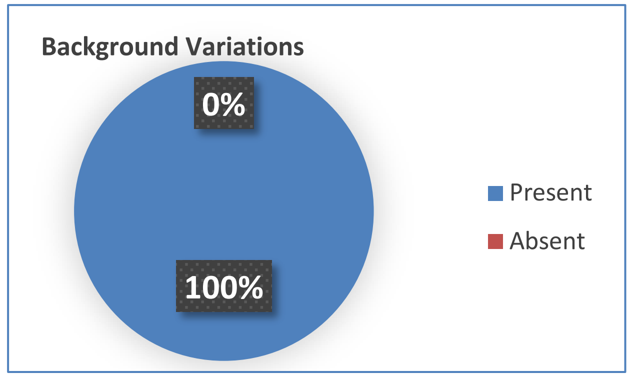

Table 1: Statistical Results of the fundamental frequency of Female Samples Compared with Disguised Male Sample- *Fundamental Frequency (F0).*

	EF0 - CF0
Z	-6.155
Asymp. Sig. (2-tailed)	.000

Table 1 above indicates that the fundamental frequency was highly significant ($p<0.005$) therefore, there is significant positive difference in the fundamental frequency of the control and electronically disguised voice.

Table 2: Statistical Results of Formant Frequency of Female samples Compared with Morphed Male for vowel /ʌ / - Formant Difference for vowel / ʌ /

	EF1 - CF1	EF2 - CF2	EF3 - CF3	EF4 - CF4
Z	-1.368	-3.546	-1.577	-5.148
Asymp. Sig. (2-tailed)	.171	.000	.115	.000

Table 2 above indicates that the fundamental frequencies (F2 & F4) was highly significant ($p<0.005$) therefore, there is significant negative difference in the formant frequencies (F2 & F4) of the control and electronically disguised voice. Also, formant frequencies (F1 & F3) were non-significant ($p>0.005$).

Table 3: Statistical Results of Formant Frequency of Female samples compared with Disguised Male for vowel /ɔ/ - Formant Difference for vowel / ɔ/

	EF1 - CF1	EF2 - CF2	EF3 - CF3	EF4 - CF4
Z	-6.073	-3.183	-2.074	-4.372
Asymp. Sig. (2-tailed)	.000	.001	.038	.000

Table 3 above indicates that the fundamental frequencies (F1, F2 & F4) was highly significant ($p<0.005$) therefore, there is significant positive difference in the formant frequencies (F1, F2 & F4) of the control and electronically disguised voice. Also,

formant frequency (F3) was non-significant ($p>0.005$).

Table 4: Statistical Results of Formant Frequency of Female samples compared with Disguised Male sample for vowel /i/ - *Formant Difference for vowel /i/*

	EF1 - CF1	EF2 - CF2	EF3 - CF3	EF4 - CF4
Z	-6.093	-2.999	-4.372	-.527
Asymp. Sig. (2-tailed)	.000	.003	.000	.598

Table 4 above indicates that the fundamental frequencies (F1, F2 & F3) was highly significant ($p<0.005$) therefore, there is significant positive difference in the formant frequencies (F1 & F3) and negative difference in formant frequency (F2) of the control and electronically disguised voice. Also, formant frequency (F4) was non-significant ($p>0.005$).

Table 5: Statistical Results of Formant Frequency of Female samples compared with Disguised Male sample for vowel /I/ - *Formant Difference for vowel /I/*

	EF1 - CF1	EF2 - CF2	EF3 - CF3	EF4 - CF4
Z	-6.093	-3.278	-4.382	-3.183
Asymp. Sig. (2-tailed)	.000	.001	.000	.001

Table 5 above indicates that the fundamental frequencies (F1, F2, F3 & F4) was highly significant ($p<0.005$) therefore, there is significant positive difference in the formant frequencies (F1, F3 & F4) and negative difference in formant frequency (F2) of the control and electronically disguised voice.

Table 6: Statistical Results of Formant Frequency of Female samples compared Disguised Male sample for vowel /u/ -*Formant Difference for vowel /u/*

	MF1 - CF1	MF2 - CF2	MF3 - CF3	MF4 - CF4
Z	-6.083	-.343	-3.069	-3.790
Asymp. Sig. (2 tailed)	.000	.731	.002	.000

Table 6 above indicates that the fundamental frequencies (F1, F3 & F4) was highly significant ($p<0.005$) therefore, there is significant positive difference in the formant frequencies (F1, F3 & F4) of the control and electronically disguised voice. Also, Formant Frequency (F2) is non-significant with negative difference in formant frequency (F2) of the control and electronically disguised voice.

4. Discussion

After processing the audio in software, the output audio was altered with different properties to that of control recording. In auditory analysis of the control and electronically disguised voices the parameter which showed maximum variation was the background variation. Background of the audio is referred to as the non-voiced regions of the audio signal. On listening to the electronically disguised male voices there were significant changes in the background of the audio. The Voice of the female samples was converted in such a manner that the voice was heavy, reverberating and recorded in a closed empty area. In some of the audio's articulation of words was not very clear, researchers had to listen twice or thrice to understand the words.

This could be because the formant frequencies showed significant changes. The values of formant frequency F1, F2, F3, F4 for 5 selected vowels were decreased for all vowels in disguised voice sample to that of values of control recording. On analyzing the average pitch of the control and morphed voice it can be noted that the average pitch of the electronically disguised male was in the range of 105- 175 Hz whereas the average pitch of the male voice lies in the 85 to 155 Hz.[9] The parameters selected for the analysis are limited, and includes only those which are commonly used and accepted in the identification of the speaker in the criminal investigation.[4] There can be many more changes in the signal of the audio which could be analyzed by the speech signal processing experts. Forensic Speaker Identification expert together with signal processing experts can explore this challenge and seek solution in the examination of the electronically disguised voice, so that the changed signal can either be linked to the original signal or it can reprocess and converted to the original audio.

5. Conclusion

The study clearly highlighted that auditory analysis of all the aural parameters of the electronically disguised voice samples with respect to their control voice had shown degradation and remarkable deviations in the voice conditions. But identification of disguised voice only on the basis of auditory parameters of speech can be tricky and challenging for the expert. This is because disguising softwares are altering the vocal parameters on the basis of predefined algorithms.

From spectrographic parameters of electronically disguised voices along with their original voices, it can be concluded that parameters such as pitch of the voices shows comprehensive variations, thus cannot be reliable parameter for identification. However, respective formant frequencies for the given vowel are the suitable parameter for speaker identification especially in samples disguised into male.

Certainly, with limitations, some specific formant frequencies indicative of the morphing can be combined with auditory parameters for suspecting a case of morphing and can be compared with possible list of suspects for speaker identification based on spectrographic analysis.

Contributor ship of Author: All authors equally contributed.

Ethical Clearance: Yes.

Conflict of interest: None to declare.

Source of funding: None to declare.

Acknowledgments

Authors would like to acknowledge Acoustics Laboratory, School of Forensic Science and Risk Management. A special thanks Dr. Gaurav Singh Kushwaha, Assistant professor, School of Physical Education and Sports; Dr. Parvesh Sharma, Assistant Professor, School of Forensic Science and Risk Management; Mr. Sudhanshu Shekhar Tiwari, Assistant Professor, School of Forensic Science and Risk Management Rashtriya Raksha University for their continuous support during the study. Also, we are very much thankful to all those who have contributed their time and effort for providing samples for our research work without whose co-operation and trust this project might not have been possible.

List of Abbreviations:

F0 -Fundamental Frequency
F1- Formant Frequency 1
F2- Formant Frequency 2
F3- Formant Frequency 3
F4- Formant Frequency 4
CF0- Control Fundamental Frequency
CF1- Control Formant Frequency 1
CF2- Control Formant Frequency 2
CF3- Control Formant Frequency 3
CF4- Control Formant Frequency 4
EF0- Electronically Disguised Fundamental Frequency
EF1- Electronically Disguised Formant Frequency 1
EF2- Electronically Disguised Formant Frequency 2
EF3- Electronically Disguised Formant Frequency 3
EF4- Electronically Disguised Formant Frequency 4

References

1. Rose P. Forensic Speaker Identification. 1st ed. London: CRC Press; 2002. p.381.
2. Furui S. Chapter 7 - Speaker Recognition in Smart Environments. In: Aghajan H, Delgado RLC, Augusto JC, editors. Human-Centric Interfaces for Ambient Intelligence. USA: Elsevier, Academic Press; 2010. p. 163–84.
3. Kent RD, Kent RA, Read C. The Acoustic Analysis of Speech. Albany, New York, USA: Thomson Learning; 2002. p. 324.
4. Eriksson A. Aural/Acoustic vs. Automatic Methods in Forensic phonetic case work. In Neustein A, Patil HA, editors. Forensic Speaker recognition: Law Enforcement and counter terrorism. New York (USA): Springer; 2012. p. 41-69.
5. Rentzos D, Vaseghi S, Turajlic E, Yan Q, Ho CH. Transformation of speaker characteristics for voice conversion. In 2003. p. 706–11.
6. Mohammadi SH, Kain A. An overview of voice conversion systems. Speech Commun. 2017; 88:65–82.
7. Gallena SJK, Stickels B, Stickels E. Gender Perception after Raising Vowel Fundamental and Formant Frequencies: Considerations for Oral Resonance Research. J Voice. 2018; 32(5):592–601.
8. Gelfer MP, Bennett QE. Speaking Fundamental Frequency and Vowel Formant Frequencies: Effects on Perception of Gender. J Voice. 2013; 27(5):556–66.
9. Titze IR. Principles of voice production.1st ed. Englewood Cliffs, N.J: Prentice Hall; 1994; p. 354.