

FORMANT FREQUENCY DIFFERENCES IN VOWELS BETWEEN NORMAL AND ELECTRONICALLY DISGUISED MALE VOICE: COMPARATIVE ANALYSIS

PRASOON, P.^{1*} – VERMA, R.² – RAMAKRISHNAN, P. N.¹

¹ *Physics Division, Central Forensic Science Laboratory, Telangana, India.*

² *Document Division, Central Forensic Science Laboratory, Chandigarh, India.*

**Corresponding author
e-mail: prakhar.prasoon01[at]gmail.com*

(Received 14th May 2024; Revised 23rd July 2024; accepted 01st August 2024)

Abstract. Voice disguise involves the deliberate alteration of one's voice for deceptive purposes and is commonly employed in criminal activities such as ransom or threatening calls. This manipulation may include non-electronic methods, such as accent adoption and pitch adjustment, as well as electronic methods like phone voice changers and software. In this study, we compared the variation in formant frequency values of five vowels in normal and electronically disguised male voices using Consonant-Vowel-Consonant (CVC) and Consonant-Vowel-Vowel-Consonant (CVVC) words. Five normal male voice samples were recorded in the laboratory, and the same samples were electronically disguised with a Voice Changer app. These disguised samples were then analyzed using Multi Speech Model 3700 software. Through statistical analysis, including t-tests and p-value calculations, it was revealed that only vowels /i/ and /o/ exhibited significant formant frequency variation in all formants, example F1, F2, and F3. Moreover, it was noted that only the F2 and F3 frequencies were associated with more significant or nearly significant results. This highlights the intricate impact of electronic disguises on vocal characteristics. These findings enhance our understanding of the acoustic shifts induced by voice disguises, which is crucial for forensic voice analysis and speaker recognition applications. The results presented in this paper contribute to the comprehension of acoustic shifts induced by voice disguises and can be beneficial for applications in forensic voice analysis and speaker recognition.

Keywords: *voice, voice disguise, formants, vowels, electronic disguised voice*

Introduction

Advancements in technology have led forensic investigations to increasingly rely on biometrics, like voice analysis, facial recognition, and eye-based identification. Voice identification has become crucial in cases involving serious crimes such as murder, kidnapping, and assault. However, merely comparing voices isn't enough; determining their typicality is also crucial. A measure called the likelihood ratio helps gauge the strength of forensic evidence, and recent developments in Automatic Speaker Recognition have simplified this estimation. Within disguised voices research, various studies have delved into particular features and types of disguise. Voice disguise involves intentionally altering one's voice to conceal identity, commonly used in criminal activities like ransom, hoax, or threatening calls (Didla, 2020; Jia et al., 2008; Clark and Foulkes, 2007; Eklund and Traunmüller, 1997; Endres et al., 1971). In forensic science, one study indicates that voice disguise is employed in 52% of cases involving offenders who believe their voice is being recorded (Hautamäki et al., 2017; Figueiredo and Souza Britto, 2013; Masthoff, 2013; Eriksson et al., 2010; Farrús Cabeceran et al., 2010). Disguised voices fall into two categories: non-electronic and electronic. Non-electronic methods involve simple techniques like covering the mouth

with cloth, adopting particular accents, adjusting pitch to imitate, using external objects to alter vocal dynamics, and changing the positioning of articulators like lips or tongue. On the other hand, electronic disguise involves technologies such as phone voice changers, voice changer apps, and software or algorithms for voice conversion (Chen et al., 2016).

Research from previous years has highlighted various ways in which the voice can be intentionally altered or disguised. These include changes in the voice source, such as using falsetto, altering pitch, employing vocal fry, or speaking softly. Resonance adjustments like nasal or hyper nasal tones, altering the vocal tract, or speaking with items in the mouth also play a role. Additionally, modifications in language, such as using different dialects or foreign accents, and adjustments in speaking style, like varied tempos, stress patterns, or monotonous speech, are common methods observed (Rodman and Powell, 2000). In the contemporary landscape, a multitude of easily accessible applications and software enables the effortless alteration of voices, leading many individuals to utilize these tools for sending disguised voice messages without disclosing their true identities. Various voice-changing devices can also be employed during phone calls to modify one's voice. Consequently, this trend has given rise to a series of legal complexities. As a result, the identification and comparison of electronically disguised voices produced by specific applications or software with the original voices have become arduous tasks in the present circumstances (Chen et al., 2016).

Within the realm of Forensic Phonetics, the process of Forensic Speaker Identification encounters numerous obstacles, with the analysis and comparison of disguised voices against their authentic counterparts emerging as a particularly formidable challenge. Voice disguise represents a relatively recent area in forensic phonetics, with limited studies delving into the different forms of voice alteration, their implications, and their effects on investigative procedures.

Materials and Methods

To examine the relationship between the normal and disguised voices, the present study compared the value of formants for each of the five vowels among the normal and electronically disguised male voices with Consonant-Vowel-Consonant (CVC) and Consonant-Vowel-Vowel-Consonant (CVVC) words. For this, 5 normal male voices were recorded and then disguised with Voice Changer application. Following that, the GoldWave V6.77 software was used to extract the vowel segments from each word in the sample. After that, the spectrogram and waveform of the vowels were acquired using Multi Speech, Model 3700 software, and the values of the first three formants, i.e. F1, F2, and F3, were recorded for each vowel segments. The relation between the formant frequencies of the normal and disguised voice was then identified statistically. *Figure 1* illustrates all the outlined procedures

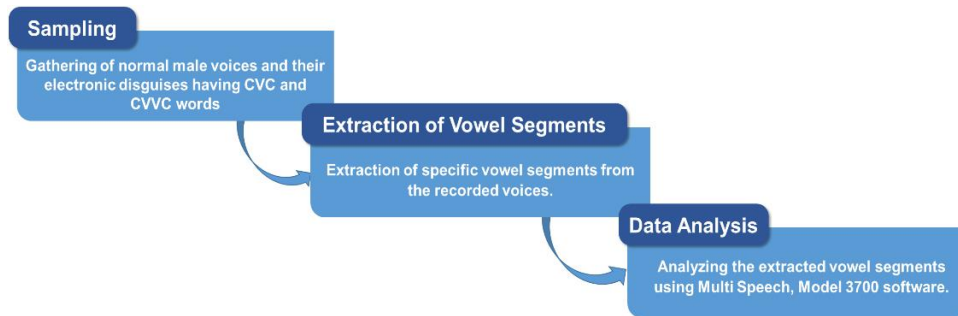
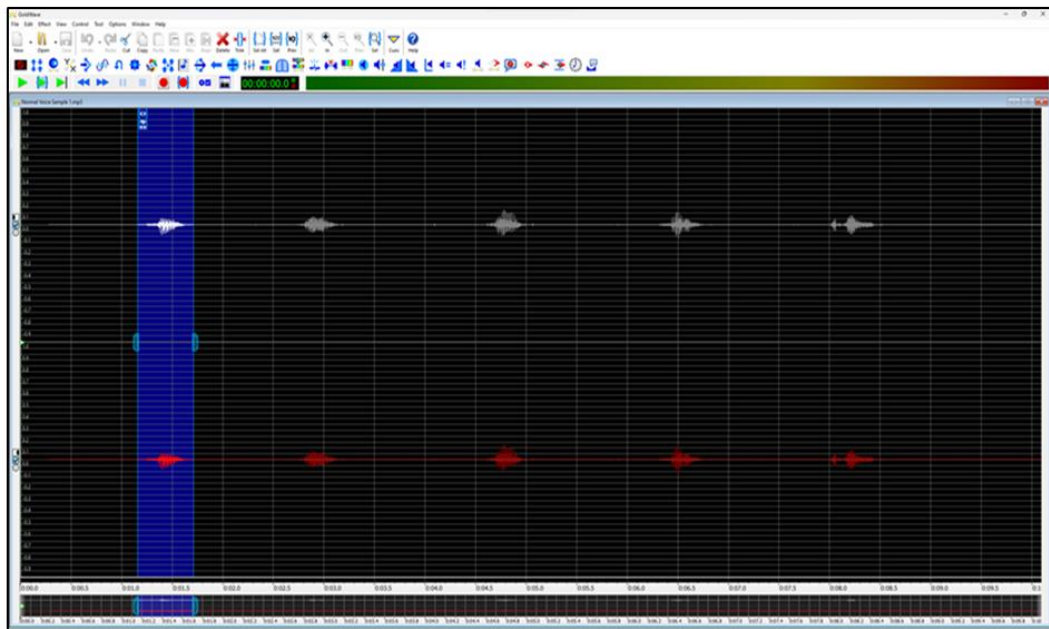
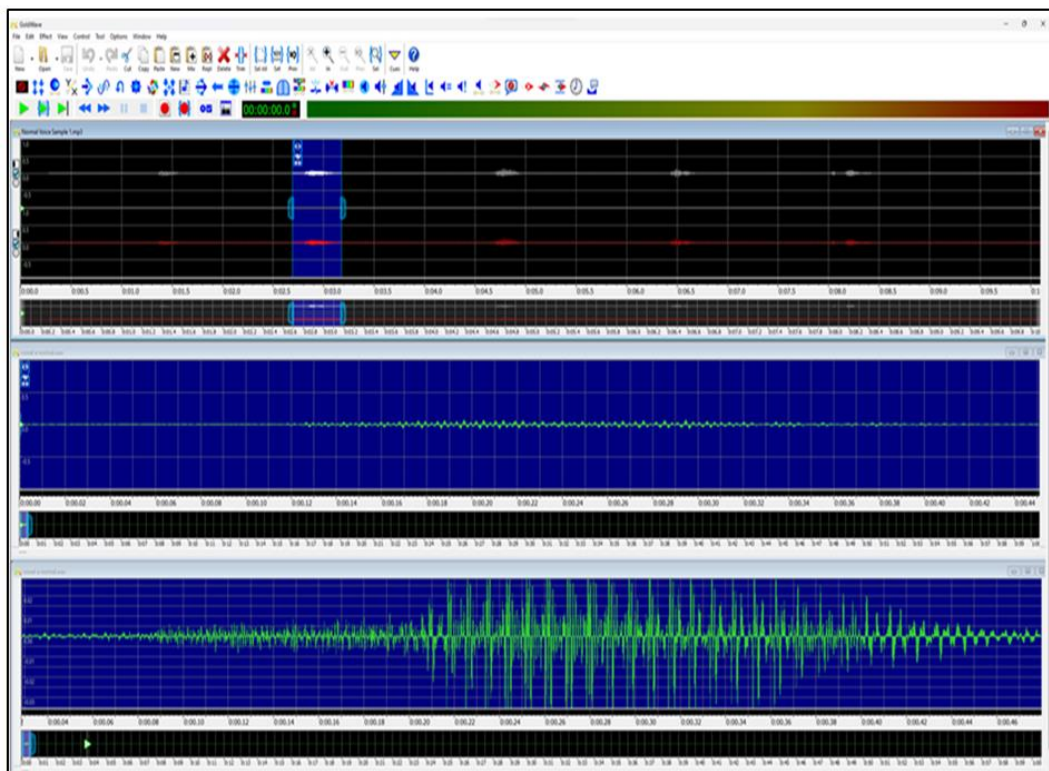


Figure 1. A flowchart's insight into outlined processes.

For conducting the analysis, voice samples were obtained from 5 male speakers between the ages of 20 and 25 years, i.e., all samples were taken from different individuals. The suspects' digital recordings were taken on a 'Fifine Ampligame AV6 Desktop Microphone' and were in Waveform Audio File Format (.wave). The voice samples were recorded in a silent atmosphere with little reverberation. The suspects were given a written document containing a set of five words-four consisting of Consonant-Vowel-Consonant (CVC) patterns and one featuring a Consonant-Vowel-Vowel-Consonant (CVVC) structure ("had," "need," "nice," "mode," and "rude")-while their voices were being recorded for the study. All 5 voice samples were then disguised using 'Optimus Robot Sound effect' in the Voice Changer Application. The extraction of vowels from the normal as well as disguised voice is performed using GoldWave Software version 6.77 (Figure 2). Using Multi-Speech, Model 3700, the separated vowels of all normal and disguised speech samples were analyzed, and the formant values were recorded (Figure 3). Each extracted vowel from the electronically disguised and normal voices is likewise subjected to the Linear Predictive Coding (LPC) analysis, as shown in Figure 4. The disparity in formant values for each vowel across individual samples has been computed, aiming to assess the frequency alterations following voice disguise. To determine statistically significant differences between the two groups, data was analyzed using the Student's t-test. All hypotheses were tested at a 0.05 level of significance.



(a)



(b)

Figure 2. Extraction of vowels from the normal and disguised voice using GoldWave Software: (a) Selection of words containing vowels from normal and electronically disguised voice samples; (b) Extraction of vowels from the normal and electronically disguised voice samples.

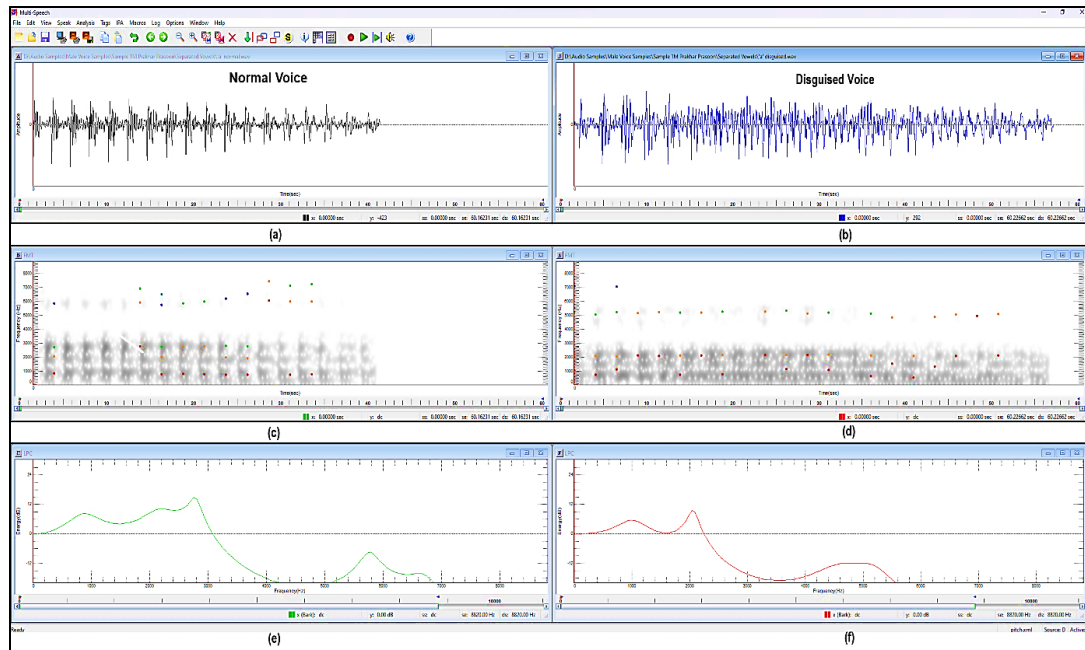


Figure 3. Analysis of Vowels of Normal and Electronic Disguised Voice using Multi-Speech, Model 3700: (a) Amplitude vs. Time graph for vowel /a/ of normal voice; (b) Amplitude vs Time graph for vowel /a/ of disguised voice; (c) Frequency vs Time graph for vowel /a/ of normal voice; (d) Frequency vs Time graph for vowel /a/ of electronic disguised voice; (e) LPC for vowel /a/ of normal voice; 3(f) LPC for vowel /a/ of electronic disguised voice

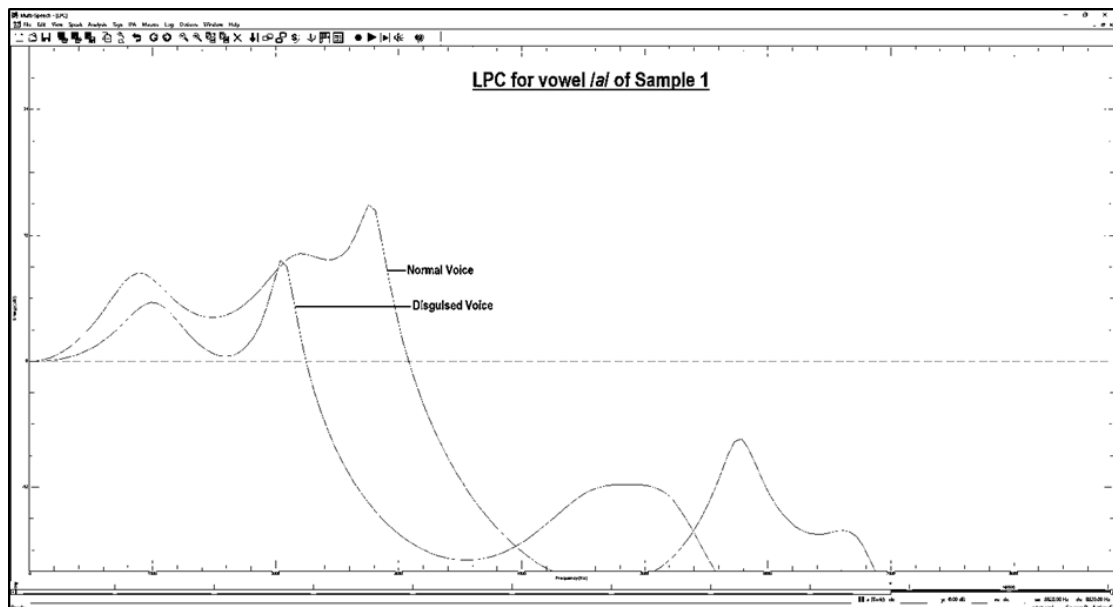


Figure 4. Comparison of Linear Predictive Coding (LPC) for Vowel /a/ of normal vs. electronic disguised voice.

Results and Discussion

Table 1 and Table 2 display the first three formant values for both normal and disguised male voice samples from five male speakers. To assess the shifts in formant frequency values subsequent to the use of voice disguise, the study computed the differences in formant values among distinct samples for every vowel shown in Table 3.

To mitigate the presence of negative values, a square root transformation method was employed as part of the analytical approach. Statistical metrics, specifically the mean, median, and standard deviation, were computed across all samples. A two-tailed t-test was conducted for the statistical measurement of the differences between normal and electronic disguised voice. A comparative analysis of formant values across all vowels between normal and electronically disguised voices within all the samples has been depicted in the accompanying bar graphs (*Figure 5*).

Table 1. Formant values for five vowels from normal voice samples.

Formants	/a/			/e/			/i/			/o/			/u/		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
Sample 1	871.52	2148	2777	340.9	2509	3023	905.5	1032	2642	984.3	2967	5676	361.2	1629	2838
Sample 2	494.89	1750	4318	265.6	3137	5848	134.3	1457	4516	543.7	3033	5569	327.9	2810	5457
Sample 3	731.13	1688	3848	3316	3715	7099	1145	4061	6305	770.6	3484	5052	356.5	2592	6232
Sample 4	1012.4	1654	2476	163.3	2627	2956	1306	2306	2864	640.6	1975	3342	513.4	2447	6005
Sample 5	692.71	1841	3924	383	3154	3446	688.6	1721	3654	808.7	937.7	3701	422.7	1358	3727

Table 2. Formant values for five vowels from electronic disguised voice samples.

Formants	/a/			/e/			/i/			/o/			/u/		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
Sample 1	1008.87	2057	4546	244.2	2127	3062	799.9	2109	4868	564	2242	4883	282.9	2232	3172
Sample 2	458.62	1591	4268	139.5	2603	3547	1192	4273	4690	468	2609	5255	1058	4442	4763
Sample 3	1257.42	3109	4735	161.9	2341	3620	709.8	3148	3801	563.2	1960	4482	277	2129	4825
Sample 4	820.06	1920	4914	2149	2332	5464	570.1	1545	2379	526.8	1713	2475	481.2	1846	4280
Sample 5	520.17	1468	3163	340.1	2586	3693	464.3	1406	3670	944.9	2903	4518	669.3	1915	3235

Table 3. Differences in the formant frequencies of the Normal and Disguised Voice Samples, with their respective mean values, median, and standard deviation for every vowel in each sample.

Formants	/a/			/e/			/i/			/o/			/u/		
	1	2	3	1	2	3	1	2	3	1	2	3	1	2	3
Sample 1	137.35	91.2	1770	96.66	382.1	38.6	105.6	1077	2226	420.3	725.2	792.9	78.34	603.5	333.5
Sample 2	36.27	159.2	49.13	126.1	533.8	2301	1057	2816	174.7	75.69	424.6	314.2	730.2	1632	693.6
Sample 3	526.29	1421	886.9	3154	1374	3479	435.1	913.6	2504	207.4	1524	570.3	79.55	463.8	1407
Sample 4	192.34	266.2	2439	1986	295	2509	735.7	760.9	485.5	113.8	262	866.5	32.12	601	1725
Sample 5	172.54	373.8	761.4	42.98	567.5	246.3	224.3	315.1	16.74	136.2	1965	817	246.6	557.9	491.5
Mean	212.96	462.2	1181	1081	630.6	1715	511.6	1177	1081	190.7	980.3	672.2	233.4	771.6	930.1
Median	172.54	266.2	886.9	126.1	533.8	2301	435.1	913.6	485.5	136.2	725.2	792.9	79.55	601	693.6
Standard Deviation	185.18	546.4	931.6	1421	430.4	1504	387.7	959.4	1188	137	734.1	230.1	289.5	484.2	604.8

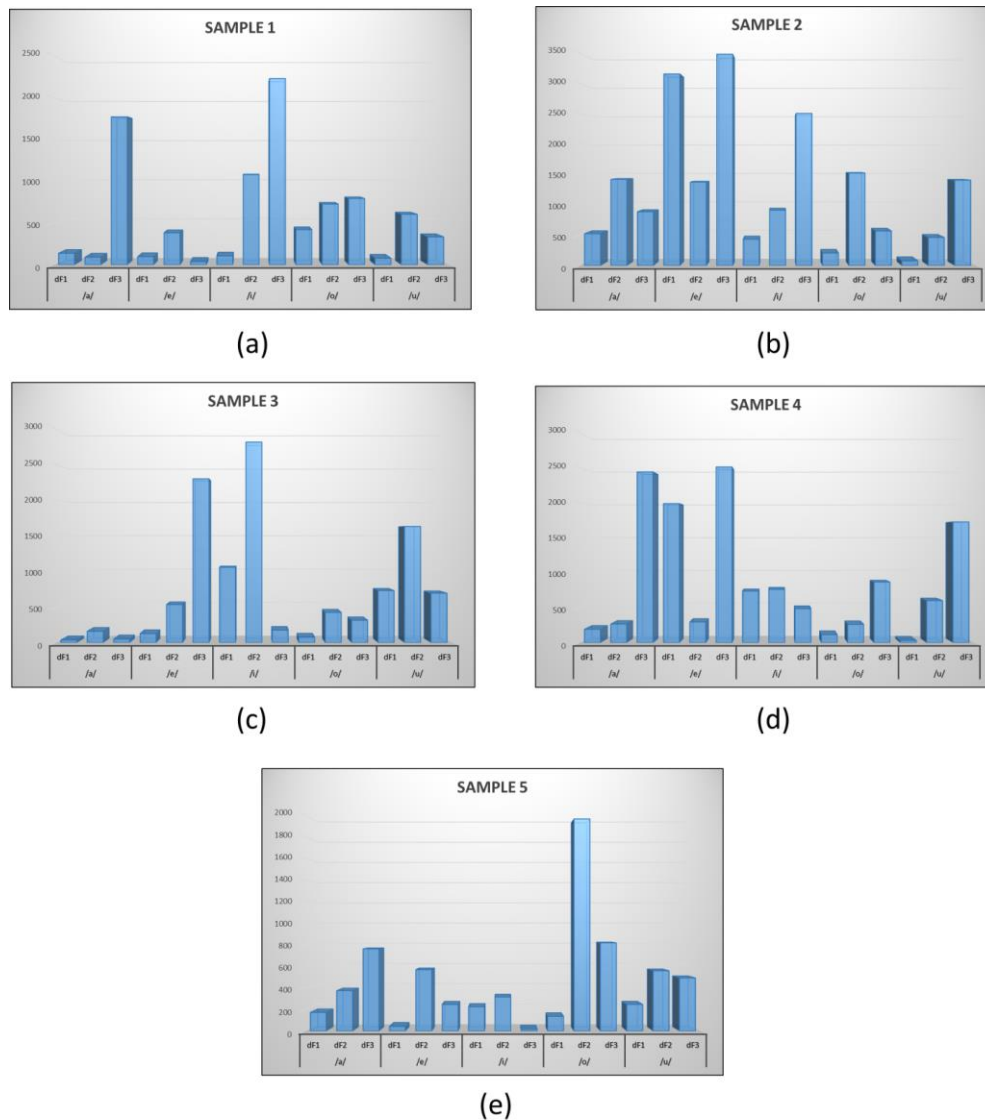


Figure 5. Formant value comparison of voice samples with normal and electronically disguised voices across each vowel.

From the presented *Table 4*, *Table 5* and *Table 6*, displaying p-values calculated using Student's t-test for formants F1, F2, and F3 across various vowels, it can be inferred that the variations in formants F1, F2, and F3 demonstrate differing significance levels across vowels. Vowels /i/ and /o/ consistently exhibit significant differences in multiple formants, suggesting distinctive acoustic characteristics compared to other vowels. Conversely, vowels /e/ and /u/ generally show non-significant differences in formants. The interpretation of these findings should consider the level of significance ($p < 0.05$) and the specific formant being analyzed

Table 4. p-values for formant F1 across vowels calculated using a t-test.

Vowel	Formant	p-value	Sig. $p < 0.05$	Level of significant
a	F1	0.056	Not significant	0.05
e		0.164	Not significant	
i		0.041	Significant	

o	0.035	Significant
u	0.145	Not significant

Table 5. *p-values for formant F2 across vowels calculated using a t-test.*

Vowel	Formant	p-value	Sig. p<0.05	Level of significant
a	F2	0.131	Not significant	0.05
e		0.030	Significant	
i		0.051	Not significant	
o		0.040	Significant	
u		0.023	Significant	

Table 6. *p-values for formant F3 across vowels calculated using a t-test.*

Vowel	Formant	p-value	Sig. p<0.05	Level of significant
a	F3	0.047	Significant	0.05
e		0.063	Not significant	
i		0.115	Not significant	
o		0.002	Significant	
u		0.026	Significant	

The comprehensive study into formant values of normal and electronically disguised male voice samples from five speakers yielded valuable insights into the impact of voice disguise on acoustic characteristics. The study focused on formants F1, F2, and F3, conducting a meticulous analysis and employing statistical measures to discern significant differences across vowels. The following conclusions can be drawn from the presented tables and analyses:

Formant F1

The statistical evaluation revealed that for vowel /a/, the difference in formant F1 was not statistically significant at the 0.05 level. This implies that the use of electronic disguise did not lead to a substantial alteration in the F1 frequency for the vowel /a/. Similarly, for vowels /e/ and /u/, the differences in formant F1 were also not significant, suggesting that electronic disguise had a minimal effect on F1 frequencies for these vowels. In contrast, significant differences were observed for vowels /i/ and /o/, with p-values of 0.041 and 0.035, respectively. This indicates that the F1 frequencies for /i/ and /o/ underwent notable alterations when subjected to electronic disguise.

Formant F2

Significant differences were noted for vowels /e/, /i/, /o/ and /u/ with p-values of 0.030, 0.051, 0.040, and 0.023, respectively. These findings suggest that electronic disguise significantly influenced the F2 frequencies for these vowels. For vowel /a/, the difference in formant F2 was not significant, indicating that the F2 frequency remained relatively unaffected by the disguise.

Formant F3

Significant differences were observed for vowels /o/ and /u/, with p-values of 0.002 and 0.026, respectively. This implies that electronic disguise had a discernible impact on the F3 frequencies for these vowels. For vowel /a/, a significant difference in formant F3 was noted (p-value of 0.047), indicating an effect of electronic disguise on

F3 frequencies. However, for vowels /e/ and /i/, the differences in formant F3 were not statistically significant.

Conclusion

The comprehensive study comparing formant values between normal and electronically disguised male voice samples from five speakers provided valuable insights into the impact of voice disguise on acoustic characteristics. The investigation, with a specific focus on formants F1, F2, and F3, conducted meticulous analyses and employed statistical measures to discern significant differences across various vowels. For Formant F1, electronic disguise had negligible effects on /a/, /e/, and /u/, while /i/ and /o/ exhibited significant changes. In terms of Formant F2, significant differences were observed for /e/, /i/, /o/, and /u/. However, /a/ showed minimal impact on F2 frequency. Formant F3 analysis revealed notable effects on /o/ and /u/, indicating a discernible impact of electronic disguise. Vowel /a/ also showed a significant difference, while /e/ and /i/ had minimal impact on F3 frequencies. In conclusion, the study elucidates that electronic disguise can result in varying effects on formant frequencies across different vowels. While some vowels exhibit significant alterations, others show minimal changes. These nuanced findings contribute to a deeper understanding of the acoustic consequences of voice disguise, with potential applications in fields such as forensic voice analysis and speaker recognition. The study underscores the importance of considering vowel-specific variations when assessing the impact of electronic disguise on formant frequencies, paving the way for advancements in voice analysis and related technologies.

Acknowledgement

This research is self-funded.

Conflict of interest

The authors confirm that there is no conflict of interest involve with any parties in this research study.

REFERENCES

- [1] Chen, Q., Li, J., Li, Y. (2016): Forensic identification for electronic disguised voice based on supervector and statistical analysis. – In 2016 Conference of The Oriental Chapter of International Committee for Coordination and Standardization of Speech Databases and Assessment Techniques (O-COCOSDA), IEEE 4p.
- [2] Clark, J., Foulkes, P. (2007): Identification of voices in electronically disguised speech. – International Journal of Speech Languages and the Law 14(2): 195-221.
- [3] Didla, G.S. (2020): A review of voice disguise in a forensic phonetic context. – International Journal of English Literature and Social Sciences 5(3): 721-725.
- [4] Eklund, I., Traunmüller, H. (1997): Comparative study of male and female whispered and phonated versions of the long vowels of Swedish. – *Phonetica* 54(1): 1-21.

- [5] Endres, W., Bambach, W., Flösser, G. (1971): Voice spectrograms as a function of age, voice disguise, and voice imitation. – *The Journal of the Acoustical Society of America* 49(6B): 1842-1848.
- [6] Eriksson, A., Llamas, C., Watt, D. (2010): The disguised voice: imitating accents or speech styles and impersonating individuals. – *Language and Identities* 8: 86-96.
- [7] Farrús Cabeceran, M., Wagner, M., Erro Eslava, D., Hernando Pericás, F.J. (2010): Automatic speaker recognition as a measurement of voice imitation and conversion. – *The International Journal of Speech, Language and the Law* 1(17): 119-142.
- [8] Figueiredo, R.M.D., Souza Britto, H. (2013): A report on the acoustic effects of one type of disguise. – *International Journal of Speech Language and the Law* 3(1): 168-175.
- [9] Hautamäki, R.G., Sahidullah, M., Hautamäki, V., Kinnunen, T. (2017): Acoustical and perceptual study of voice disguise by age modification in speaker verification. – *Speech Communication* 95: 1-15.
- [10] Jin, Q., Toth, A.R., Black, A.W., Schultz, T. (2008): Is voice transformation a threat to speaker identification? – In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE 4p.
- [11] Masthoff, H. (2013): A report on a voice disguise experiment. – *International Journal of Speech Language and the Law* 3(1): 160-167.
- [12] Rodman, R., Powell, M. (2000): Computer recognition of speakers who disguise their voice. – In *Proceedings of the International Conference on Signal Processing Applications and Technology (ICSPAT2000)* 5p.