

Vowel Space in Disguise Voice

Lakshmi Prasanna P*

Associate Professor in Speech-Language Pathology, Helen Keller's Institute of Research and Rehabilitation for the Disabled Children, RK Puram, Secunderabad-06, Telangana, India

Abstract

The acoustic characteristics of vowels provide vital information on frequency, intensity, formant frequencies etc. The vocal tract is responsible for frequency shaping of the signal from the vocal folds, which is crucial for formant frequencies. Using the first two formant frequencies, the current study was aimed to check the vowel space in various disguise voice. A 35-year-old Telugu native speaker participated in the study. Two types of recordings were made one is by using PRAAT and the other one is by using voice changer app. The participant was asked to phonate three vowels (a, i and u). Formants (F1 and F2) were measured by using PRAAT software. The obtained data was used for further analysis. The paired sample t test reveals that there is a high significant difference between original with women & girl type voices for the first formant. Second formant showed significant difference between original and women type voice. Both F1 and F2 comparisons had a negative mean score, which might be attributed to the original voice's lower formant frequencies versus the disguised voices higher formants. Overall, the presents study found that the formant frequencies for "man" and "boy" voice types were higher after using the voice changer app, indicating that they were more similar to the original female voice. The vowel space varies and is higher in disguised voices than in normal sounds, despite the fact that all the voices are from the same person. When comparing the regular voice to other disguised voices, the current study findings, according to the author, results can aid in understanding the formants. The formant frequency range is also changed by the gadgets and numerous voice changer apps, which makes it higher than usual.

Keywords: PRAAT; Formant frequencies; Disguised voices; Vowel space and voice changer app

Corresponding author:

Lakshmi Prasanna P

✉ lakshmiprasannaspeechpathologist@yahoo.com

Tel: +918977034115

Associate Professor in Speech-Language Pathology, Helen Keller's Institute of Research and Rehabilitation for the Disabled Children, RK Puram, Secunderabad-06, Telangana, India

Citation: Prasanna LP (2022) Vowel Space in Disguise Voice. J Biomed Sci, Vol. 11 No. 8: 76

Received: 29-July-2022, Manuscript No. IPJBS-22-12919; **Editor assigned:** 01-Aug-2022, PreQC No. IPJBS-22-12919(PQ); **Reviewed:** 15-Aug-2022, QC No. IPJBS-22-12919; **Revised:** 22-Aug-2022, Manuscript No. IPJBS-22-12919(R); **Published:** 29-Aug-2022, DOI: 10.36648/2254-609X.11.8.76

Introduction

Voice is a complicated process where an individual can change in many ways to fulfil communication. The speaker can be identified by voice characteristics and qualities either perceptually or acoustically. Individually, the person can also be identified by the suprasegmental aspects. The acoustic characteristics of voice consist of fundamental frequency (F0), intensity, jitter, shimmer, harmonics to noise ratio (HNR), and formant frequencies, etc. The formant frequencies are the resonant frequencies that occur when the signal passes from the vocal fold through the oral tract into the vocal tract. The resonant frequencies are altered by the articulators in the vocal tract. The formant frequencies are the steady state acoustic energy which is present in the

vowels. Vowels are produced without any constrictions in the oral tract. The relationship between the tongue position and formant frequencies is considered, with tongue height and tongue advancement being related to the formants. As stated by many researchers, the first formant frequency (F1) increases as the constricted area becomes larger, and the second formant frequency (F2) increases as the constricted area is located more anteriorly. Vowel space is defined as an acoustic measure to get the size of the vowel articulatory space which is constructed by using F1 and F2 of vowels. EMA measurements and functional data analysis can show the articulatory movements made when a voice is being disguised [1].

Studied the relationship between formants and tongue positions,

finding that F1 varied considerably with differences in the y dimension of tongue position whereas F2 was associated in a similar way with variations in the x and y positions [2]. Unfortunately, there is no a principle, obvious way to identify the speakers for whom the covariation is not significant. Instead, analyses conducted within-speaker are likely to make reasonable assumptions about the size of the tongue working space based on the size of the acoustic vowel space for the majority of speakers. Also stated that the variance in F1 reflects tongue height, whereas F2 is a far more complex reflection of tongue variation in both dimensions.

A speaker has a wide range of options available to him for altering his voice and deceiving either a human or an automatic system [3]. The formant frequencies of a voice can be altered by electronic scrambling or, more simply, by utilising intra-speaker variability. This can be done by altering the pitch or the location of the articulators, such as the lips or tongue, which affects the formant frequencies.

To establish a benchmark for speaker identification by using $F_2 \approx F_1$, results showed that /i:/ has benchmarking of 75% (better) compared with vowels /a:/ and /u:/. Vowel /u:/ had poor benchmarking. Authors stated that the results indicated that vowel /i:/ had 75% benchmarking $F_2 \approx F_1$ of vowel /i:/ could be used for speaker identification [4]. Investigated vowel space in 72 Telugu normal speakers across age, gender, and dialects (three regionals: Andhra Pradesh, Rayalaseema, and Telangana), finding that vowel space decreases with age; females have larger vowel space than males; and speakers from the Coastal region have larger vowel space than those from the Telangana and Rayalaseema regions [5]. Stated that the first two (F1 and F2) formant frequencies of the vowels are lower in the males than females [6].

The length of the pharyngealoral tract, the constriction of the vocal tract, and the degree of constriction narrowness influence the formant frequencies of vowels. Disguised voice has a significant impact on speaker recognition. If advanced electronic manipulations [7], such as vocoders or communicating via voice synthesis, were utilised, claim that speaker identification would be exceedingly difficult, if not impossible, in many circumstances [8]. Studied on 30 participants to compare un disguised and disguised conditions by using aural perception and Mel frequency methods in which they found that aural perception showed 56.7 to 80 % and where as 46.7%, 26.7% and 53.3% for Mel frequencies in whole word, consonant and vowel segmental analysis [9]. Authors concluded that aural perceptual methods had higher percentage of speaker identification in disguised condition than Mel frequency. The results revealed the kappa value to be negative ($k < 0$) indicating no agreement between the two methods. The percentage of correct identification using aural perception ranged from 56.7% - 80% and for MFCC under whole word, consonant segment and vowel segment analysis were 46.7%, 26.7% and 53.33% respectively. The aural perception method had a greater percentage of correct identification than Mel frequency method. Stated that the perception of a speaker's identity depends on a wide range of characteristics, including both naturally occurring ones like dialect and familiarity with the spoken language and different

means of disguising one's identity through methods like imitation and impersonation [10]. The use of an automatic speaker verification method to recognise modal to disguised voice utterances suggests that speaker pairs classified as simple by this method were also simple for the average listener [11]. In the difficult trials, listeners also made more mistakes. The listening results show that native and non-native groups performed similarly for non-target pairs, but that target (same speaker) trials were more tough for the non-native group. The results of a study to compare the effects of natural voice and the three types of voice disguise show that the reference populations that contain speech data with the same type of disguise very slightly affect the system's performance when the three types of voice disguise are used [12]. However, if the reference group is created using just regular speech, the effects are typically more severe and distinct for the three forms of disguise being assessed.

The Long-Term Average Spectrum was utilised by the authors to compare the male and female reding samples [13]. Results revealed that the female voice had higher levels of aspiration noise in the spectral regions corresponding to the third formant than the male voice, giving the female voice a "breathier" character. Another effect of this increased aspiration noise presence is the lower spectral tilt in the female voices. Compared the normal and high-pitched disguise voice in males using LTAS [14]. The results indicated that Skewness and Kurtosis decreased in disguised condition compared to normal speech.

Trained two groups using the perceptual approach to identify speakers, and the results revealed that both groups of listeners were able to distinguish speakers with a reasonably high degree of accuracy (92% correct) when both components of the stimulus pair were unmasked [15]. The presence of a disguised speech sample in the stimulus pair had a substantial negative impact on listener performance (59%—81% correct, depending on the disguise). Automatic speaker verification methods have been used to identify disguise voice [16].

In the current scenario, technology has improved to the point where everything is possible with a single touch on smart phones. Based on the requirements, various apps are also available on the Google Play store. If the technology is used for any bad deeds, then a problem arises. From this perspective, the present study was aimed at studying the vowel space by measuring formant frequencies and comparing them between the original voice and various disguised voices.

Method and Procedure

A Telugu native female aged 35 years participated in the study. The participant is healthy and has no history of any neurological, speech, or hearing problems.

Stimuli

Three vowels /a/, /i/ and /u/

Soft wares used

1. Voice changer app
2. PRAAT soft ware

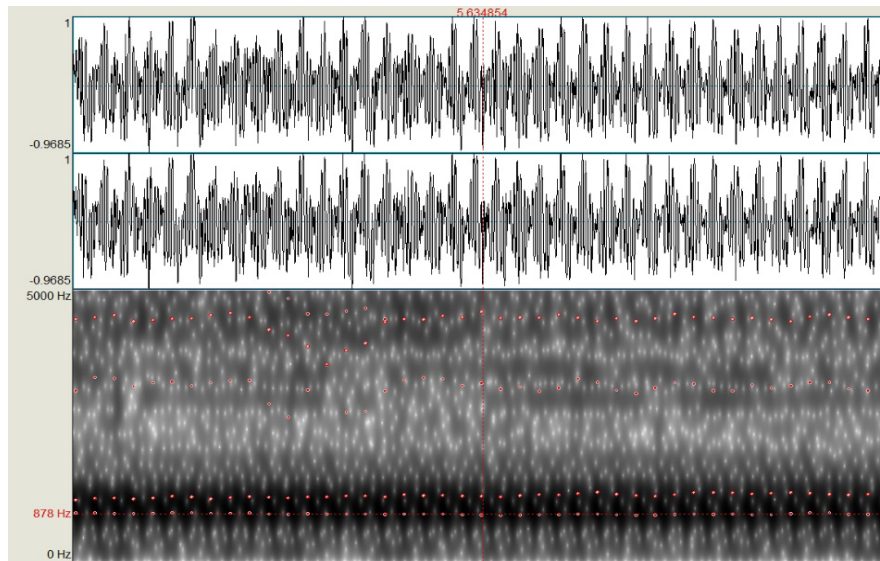


Figure 1 Showing spectrum for vowel /a/. Arrows represents F1 and F2 formants.

Table 1. Shows the formants (F1 and F2) of all the types of voices for all the vowels.

Vowels	Original		Original app		Male		Boy		Women		Girl	
	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
a:	789	1379	1049	1401	789	1162	853	1103	773	1690	784	1919
i:	336	1532	373	2057	254	2289	302	2017	515	1822	619	2361
u:	402	844	406	1064	493	1690	622	1959	530	1370	646	1609

Procedure

The study was carried out by using the Voice Changer app, which was downloaded from the Google Play store on a Samsung A30S hand set. In all the options provided, the examiner chooses a boy, a man, a girl, a woman. The participant was asked to record all the vowels in sustained phonation by using the voice changer app and they were saved in different modes like boy, man, girl, woman, and original voice in the app. A separate record of original voices was made by using the PRAAT 6.02.06 version¹.

Analysis

All the voice recordings were analysed by using PRAAT software. Formants (F1 and F2) were analysed by keeping cursor on the three consecutive dots showed on the spectrum and saved for further analysis. The spectrum was showed in the (Figure 1). All the formants obtained from the various voices were measured and stored for further analysis.

Statistical analysis

Mean, SD were calculated and paired sample t test was done to compare the data and check the significant difference.

Results and Discussion

The first two formants were analysed in all the vowels in various types of voices; the formant values were given in (Table 1). The original voice (recorded from PRAAT) showed lesser formants for all the vowels when compared to the original app (recorded from

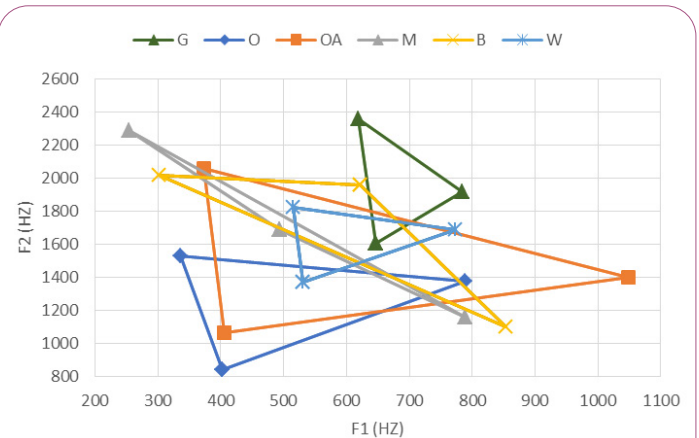
the voice changer app). The woman and girl type disguised voices showed higher formants in all the vowels when compared to the original voice, except for the first formant of "/a:/", which showed in the formant range, i.e., 773 Hz (W) and 784 Hz (G) respectively. This indicates that the higher formants were obtained due to the algorithms and functions used in the voice changer app to change the voice being different when compared to other laboratory or field devices.

This should be noted further that in any kind of disguised voice, higher formants will be shown when compared to the reference voice, in particular with the speaker identification analysis. Vowel/u/showed higher formants in man and boy type disguised voices than the original female voice and the original app voice, which showed that even though the voice is perceived as low pitch due to the effect used in the application, the formants remain higher in the acoustic analysis. This indicates that the disguised voices show higher formants in the back high vowel (/u/). The "man" type of voice showed a higher second formant for the front high vowel (i) when compared to other types, except for the "girl" type of voice. This clearly shows that the disguised voice may have some features of the original voice where it is nearly showing the high formant frequency, which is close to female voices. The score was incorporated in the table 1.

For all vowels /a/, /i/, and /u/ in different voices, the mean and SD were determined. When compared to other forms of voices, the mean formant frequencies of the original voice have been shown to be lower. When compared to the original PRAAT

Table 2. Mean and SD scores of F1 and F2 for all the vowels in various types of voices.

Types of voices	F1 (Hz)		F2 (Hz)	
	Mean	SD	Mean	SD
Original	509.00	244.72	1251.66	361.24
Original App	609.33	381.11	1507.33	504.96
Male	512.00	268.00	1713.66	563.87
Boy	592.33	276.69	1593.00	460.56
Women	606.00	144.82	1627.33	232.42
Girl	683.00	88.50	1963.00	377.92

**Figure 4** Shows vowel space in various types of voices.

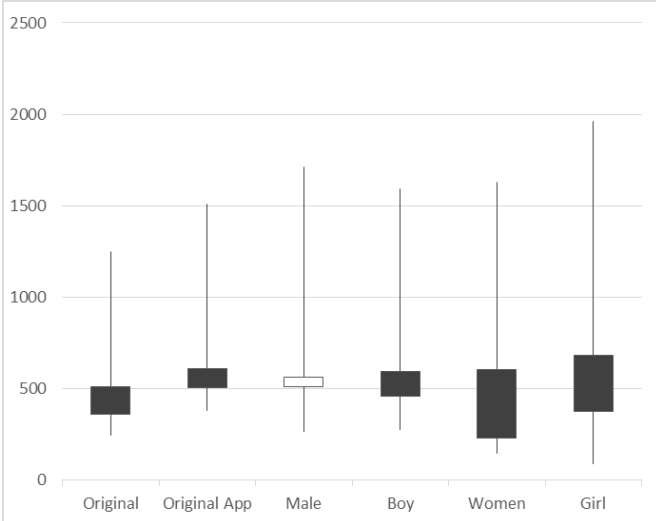
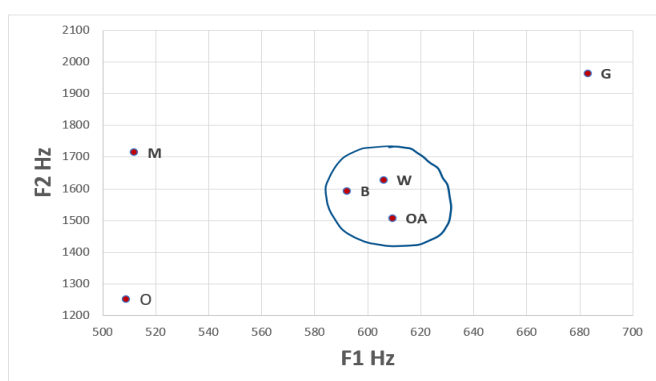
The mean scores of the original app, women, and girl disguised voices show little variation, with perhaps an octave increase in these voices when compared to the original PRAAT recording voice. These findings should be considered in the analysis of speaker identification or any other acoustic measures. **Figure 2** shows the mean and SD scores of the formants in various types of voices. It is observed that among all the voice types, the "girl" voice has higher formants and all the disguised voices are higher than the normal voice. Strong evidence was observed in the "male" voice type as the formants were higher than in the other types of voices. **Figure 3** shows the position of the formants of all the type voices where the formants are in higher position than the original voice. Interestingly it was observed that 'boy', 'women' and 'original app' voices found in slightly same region where other voices are scattered and plotted in different regions.

Vowel space

The first two formants (F1 and F2) of the three vowels in one original voice and five disguised voices were analysed. The obtained values of the formants were plotted in a graph where the X-axis (F1 Hz) and Y-axis (F2 Hz) were considered to get a vowel triangle. The vowel triangle gives the information related to the space/area of a particular person's speaking vowels. The present study showed vowel space in different voices of the same person, which gives a clear idea of how the formants vary. The data is summarised in the table and depicted in the (**Figure 4**).

The vowel space for all types of voices varies, with the original app voice having a greater vowel space. It was interesting to see that the vowel space for the original voice (PRAAT recorded) was at the bottom of the graph, whereas the vowel space for the other disguised voices was at the top. The vowel spacing of the girl voice type was higher than the other voice types. When compared to other voices, the women's type of vowel space is in the middle of the graph and has a small area.

Overall, the findings of this study show that vowel space varies between different types of disguised voices, as well as between the recorded original (PRAAT) and the voice changer app. This means that when compared to the original (PRAAT) voice, the formant values of various disguised voices increase. Since PRAAT was used for the majority of the acoustical study, the

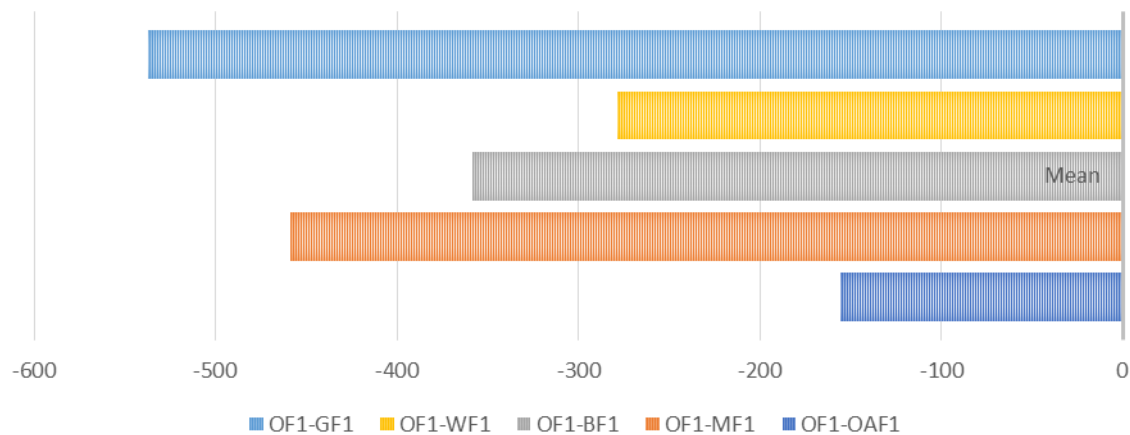
**Figure 2** Shows the mean and SD of all the formants in various types of voices.**Figure 3** Shows the position of the means of formants for all the types of voices.

recording voice, the values revealed greater means in all forms of disguised voice as well as original app voice. This could be due to the algorithms, or it could be due to different settings in that specific device. Typically, the male voice should have lower formant frequencies when compared to the female voice, but in the present study it shows that the male and boy types of disguised voices are present with higher formants than the original female voice. The obtained scores were given in (**Table 2**).

Table 3. Shows the mean and SD scores for first formant (F1).

Pairs	Mean	SD	t	Sig
OF1-OAF1	-155.33	366.78	-.734	.540
OF1-MF1	-459	586.93	-1.35	.308
OF1-BF1	-358	633.04	-.980	.431
OF1-WF1	-278.66	149.48	-3.22	.084
OF1-GF1	-537.33	14.15	-65.75	.000*

*p<0.01 (high significance)

**Figure 5** Shows the mean scores of F1 comparisons.

author emphasises that the type of device, recording mode, and technology can all get an impact on the formant frequencies.

Statistical analysis

First formant (F1)

The original (PRAAT) recorded voice was compared to other disguised voices recorded from the voice changer app by using a Paired Sampled T test, which showed that the mean scores obtained are in negative range. Higher mean scores were observed in the combination of original voice and girl type voice. There is a highly significant difference between the original voice (PRAAT recorded) and the girl type voice. The original voice belongs to the female voice, where the typical formants lie in the higher formant region. The voice changer app recorded voice shows higher formants, including male and boy voice types. Statistically, there is no significant difference found between the original voice and the original app voice accompanied by various disguise voices. Due to the increase in formant variation in the voice changer app, there was a high significance found between original and girl type voices. The data are summarized in the (Table 3) and depicted in the (Figure 5).

The Paired sample correlation test reveals that there is a positive correlation between the original and other types of voices, where there is a significant difference found between the original voice and women's and girl's voices, with a strong positive correlation. This happens due to the increase in formants of women's and girl's voices, which are heard higher than the original voice type. The other pairs showed no significant difference, with high

Table 4. Shows the correlation and significance scores for the first formant (F1).

Pairs	N	Correlation	Sig
OF1 & OAF1	3	.91	.259
OF1 & MF1	3	.76	.449
OF1 & BF1	3	.55	.625
OF1 & WF1	3	.99	.020*
OF1 & GF1	3	1.0	.016*

*p<0.05 (significance)

Table 5. Shows the mean and SD scores for second formant (F2).

Pairs	Mean	SD	t	Sig
OF2-OAF2	-255.66	253.38	-1.74	.223
OF2-MF2	-462	589.71	-1.35	.308
OF2-BF2	-341.33	559.50	-1.05	.401
OF2-WF2	-375.66	130.61	-4.98	.038*
OF2-GF2	-711.33	151.79	-8.11	.015*

*p<0.05 (significance)

Table 6. Shows the correlation and significance scores for the second formant (F2).

Pairs	N	Correlation	Sig
OF2 & OAF2	3	.881	.314
OF2 & MF2	3	.247	.841
OF2 & BF2	3	.089	.943
OF2 & WF2	3	.997	.047*
OF2 & GF2	3	.917	.262

*p<0.05 (significance)

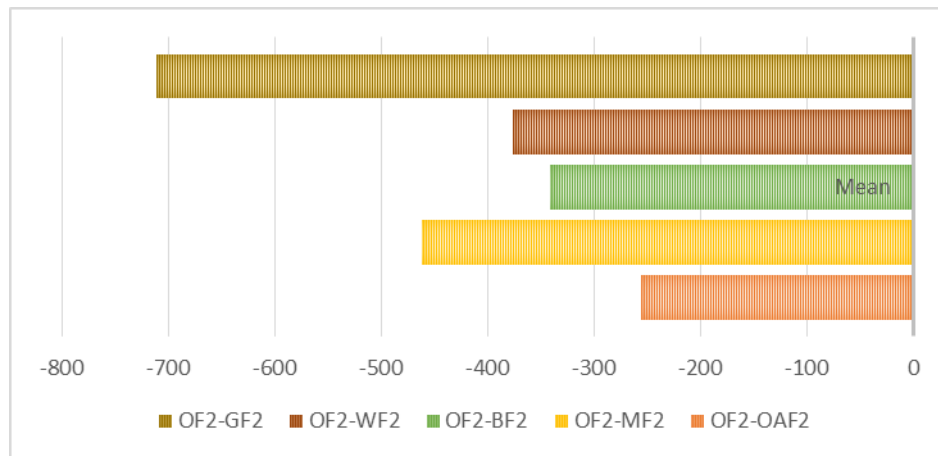


Figure 6 Shows the mean scores of F2 comparisons.

correlation and high significant values, which occurred due to the limited sample size. The data is given in the (Table 4).

Second formant (F2)

A paired sample t test reveals that the mean scores are in a negative range, which could be the reason why some of the original formants are less than others. When the original voice was combined with the girl type voice, higher means were observed. There is a significant difference found between women and girls of type with the original voice. There is a strong correlation observed between all the types of voices, with a significant difference found only with women's voices when compared with the original voice. This indicates that the original voice (female) had a higher formant and, due to the filters used by the voice changer app, the formant frequencies for all the types of voices showed a higher formant range. Hence, there is no significant difference found between the original voice and boy or man-type voices. Perhaps there is some significance found between the girl and woman types. This could be due to the higher formants in both the types than in the original voice. The data is summarised in the following (Tables 5 and 6) and depicted in the (Figure 6).

Conclusion

The present study was done to check the vowel space in a disguised voice by comparing it with the normal voice. Vowel space was larger in the original app voice than in the normal and other disguised voices. The author also stated that the formant frequencies were increased in the disguised voice with the effect of the voice changer app. The male and boy disguised voices also show higher formant frequencies, which is close to the normal recorded (female) voice. Hence, there is no significant difference between the normal voice and the other disguised voices (original app, male and boy). There is a highly significant difference found between the original voice and girl type voices in F1 and a strong correlation observed between original voices with other disguised voices, accompanied by the significant difference between women and girl voices when compared with the normal voice. There is a significant difference found between women and girls' types in F2 with a strong correlation between other types,

accompanied by a significant difference between women's types in voice when compared with normal voice. This could be due to the increase of formants in women's and girl's voices. The increase in formant frequency in male and boy disguised voices was also noticed where they occurred in the typical female formant range, which is caused by the voice changer app filters. As a result, the vowel space increased in the disguised voice when compared to the normal voice. The present study author also agreed with [17] Author reviewed various articles and stated that the voice of the perpetrator is disguised, and fake documentation is produced. This will have a negative impact on the evidence's authenticity. Manipulation of voices is thus inevitable in the field of audio forensics. When identifying speakers, the first stage in figuring out whether or not the voice test is hidden may be the detection of disguised voices. Similarly, [18] Showed that when speakers tried to sound younger, they would increase their fundamental frequency (f_0) and speech rate, and when they tried to seem older, they would speak slower. Implies that the study of imitation to find mimicry- or spoofing-invariant traits may be a useful technique for biometric analysis, particularly of masked voices [19]. More research is required to consider nonlinear phenomena that occur during speech generation and outcomes from the field of natural language processing [20] in order to significantly improve the quality of voice transformation systems. Finally, current study author concluded that advanced technology [21] can cause serious effects in the identification of speakers, especially in forensic voice analysis.

Conflict of Interest

The author declared that there are no conflicts of interest.

Acknowledgement

Sincere thanks to the participant and Helen Keller Institute for allowing the author to complete the study.

Source of Funding

This study was done under the part of Research at Helen Kellers Institute, Secunderabad, INDIA.

References

- 1 Tavi L, Kinnunen T, Meister E, González-Hautamäki R, Malmi A (2021) Articulation During Voice Disguise: A Pilot Study. In International Conference on Speech and Computer Springer, Cham 680-691.
- 2 Lee J, Shaiman S, Weismer G (2016) Relationships between the tongue positions and formant frequencies in female speakers. *J Acoust Soc Am* 139: 426-440.
- 3 Perrot P, Aversano G, Chollet G (2007) Voice disguise and automatic detection: Review and perspectives. *Progress in Nonlinear Speech Processing* 101-117.
- 4 Prasanna LP, Savithri SR (2009) Benchmark for speaker identification using vector F1~F2. *Proceedings of the international symposium on speech and music* 38-41.
- 5 Krishna Y, Rajashekhar B (2012) Vowel space areas across age, gender and dialects in telugu. *Language in India* 12: 357-369.
- 6 Man CY (2007) An Acoustical analysis of the vowels, diphthongs and triphthongs in Hakka Chinese: Paper presented at the 16th International Congress of Phonetic Sciences.
- 7 Pickett JM (1996) *The Sounds of Speech Communication - A Primer of Acoustic Phonetics and Speech Perception*.
- 8 Clark J, Foulkes P (2007) Identification of voices in disguised speech. *Int J Speech Lang Law* 14: 195-221.
- 9 Praveena J, Krishna Y (2015) Identifying speaker from disguised speech using aural perception and Mel-frequency cepstral coefficient. *J Indian Speech Language Hearing Assoc* 29: 28.
- 10 Eriksson A, Llamas C, Watt D (2010) The disguised voice: imitating accents or speech styles and impersonating individuals. *Lang identities* 8: 86-96.
- 11 Hautamäki RG, Sahidullah M, Hautamäki V, Kinnunen T (2017) Acoustical and perceptual study of voice disguise by age modification in speaker verification. *Speech Commun* 95: 1-15.
- 12 Künzel H, Gonzalez-Rodriguez J, Ortega-García J (2004) Effect of voice disguise on the performance of a forensic automatic speaker recognition system. In ODYSSEY04-The Speaker and Language Recognition Workshop.
- 13 Mendoza E, Valencia N, Muñoz J, Trujillo H (1996) Differences in voice quality between men and women: use of the long-term average spectrum (LTAS). *Journal of voice: official journal of the Voice Foundation*, 10: 59-66.
- 14 Pickett JM (1996) *The Sounds of Speech Communication - A Primer of Acoustic Phonetics and Speech Perception*.
- 15 Reich AR, Duke JE (1979) Effects of selected vocal disguises upon speaker identification by listening. *J Acoust Soc Am* 66: 1023-1028.
- 16 Zhang C, Tan T (2008) Voice disguise and automatic speaker recognition. *Forensic Sci Int* 175: 118-122.
- 17 Kaur HP, Kaur R (2021) Identification and comparison of disguised voices with the genuine voice under various circumstances using spectrographically analysis: A review study. *Int J of Adv Trends Com App (IJATCA)* 8: 54-58.
- 18 Waller SS, Eriksson M (2016) Vocal age disguise: The role of fundamental frequency and speech rate and its perceived effects. *Front Psychol* 7: 1814.
- 19 Singh R, Gencaga D, Raj B (2016) Formant manipulations in voice disguise by mimicry. *International Workshop on Biometrics and Forensics (IWBF)* 1-6.
- 20 Stylianou Y (2009) Voice transformation: A survey. *Acoustics, Speech, and Signal Processing, IEEE International Conference on* 3585-3588.
- 21 Prasanna LP, Sangeetha G (2011) Use of Long-Term Average Spectrum (LTAS) In Speaker Identification in High Pitched Disguise Condition. *Proceedings of the international symposium on speech and music* 113-116.