

# Acoustic Analysis for Comparison and Identification of Normal and Disguised Speech of Individuals

Mathur S<sup>1</sup> and Vyas JM<sup>2</sup>

<sup>1</sup>Assistant Professor, Gujarat Forensic Sciences University, Gujarat, India

<sup>2</sup>Director General, Gujarat Forensic Sciences University; Gandhinagar, Gujarat, India

\*Corresponding author: Mathur S, Assistant Professor, Gujarat Forensic Sciences University, Gujarat, India, Tel: 079-23977171, E-mail: surbhi.mathur@gfsu.edu.in

Citation: Mathur S, Vyas JM (2016) Acoustic Analysis for Comparison and Identification of Normal and Disguised Speech of Individuals. J Forensic Sci Criminol 4(4): 403. doi: 10.15744/2348-9804.4.403

Received Date: March 28, 2016 Accepted Date: August 30, 2016 Published Date: August 31, 2016

## Abstract

Although the rapid development of forensic speaker recognition technology has been conducted, there are still many problems to be solved. The biggest problem arises when the cases involving disguised voice samples come across for the purpose of examination and identification. Such type of voice samples of anonymous callers are frequently encountered in crimes involving kidnapping, blackmailing, hoax extortion and many more, where the speaker makes a deliberate effort to manipulate their natural voice in order to conceal their identity due to the fear of being caught. Voice disguise causes serious damage to the natural vocal parameters of the speakers and thus complicates the process of identification. The sole objective of this doctoral project is to find out the possibility of rendering definite opinions in cases involving disguised speech by experimentally determining the effects of different disguise forms on personal identification and percentage rate of speaker recognition for various voice disguise techniques such as raised pitch, lower pitch, increased nasality, covering the mouth, constricting tract, obstacle in mouth etc by analyzing and comparing the amount of phonetic and acoustic variation in of artificial (disguised) and natural sample of an individual, by auditory as well as spectrographic analysis.

**Keywords:** Forensic; Speaker recognition; Voice; Speech; Disguise; Identification

## Introduction

The science of crime investigation relies on one basic principle known as the “Principle of exchange”. According to this principle when two objects come in contact with one another, there will be transfer of substance between the two [1-4]. Similarly for the crime site it is believed that no matter where a criminal goes or what a criminal does, by coming into contact with things, a criminal can leave all sorts of evidence, including DNA, fingerprints, footprints, hair, skin cells, blood, bodily fluids, pieces of clothing, fibers and more. At the same time, they will also take something away from the scene with them. The scenario reverses in the situation where there is no immediate crime scene like cases involving blackmailing, kidnapping, extortion, threatening, anonymous calls, ransom calls, hoax calls, obscene calls, harassment calls, match fixing etc, where the criminals resort to the aid of telephones and mobiles in order to maintain their anonymity for fear of detection [5-7]. In these circumstances, the voice of an individual is an important clue for identification.

Forensic speaker identification is the application of science to solve the problems related to identification of the unknown speaker in criminal investigation. Voice is the unique sound produced by vocal organs of living organism [8]. Although DNA is considered as the most important evidence for criminal investigation, but the truth is that DNA can't talk [9,10]. The voice of a person can be successfully used as a biometric feature as it is well accepted by the users and can be easily recorded using microphones and hardware of low costs [5,11,12]. It can provide an unconventional and more secure means of permitting entry without any need of remembering a password, lock combination etc or the use of keys, magnetic card or any other fallible device which can be easily stolen [7,13,14].

In the present era, widely available facilities of telephones, mobiles and tape recorders results in the misuse of the device and thus, making them an efficient tool in commission of criminal offences, where the criminals frequently misuse these modes of communication, believing that they will remain incognito, and nobody would recognize them. It is fortunately no longer true [15-17]. The voice of an individual can successfully recognize him and pin the crime on him [18-20].

With the advancement of crime, the criminals are now capable of imparting deliberate change in their voice characteristics to prevent recognition and for misleading the investigation [21,22]. For example: “A criminal make use of a simple handkerchief over

the speaker with the intention to modify his voice". This is the biggest limitation faced by the voice experts all over India [23]. This study aims to solve problems occurring in the speech of individual due to different forms of disguise and assist the experts while examination of such challenging voice exhibits [24-26].

Therefore, the sole objective of this project is to find out the possibility of rendering definite opinions in cases involving disguised speech by experimentally determining the effects of different disguise forms on personal identification and percentage rate of speaker recognition for various voice disguise techniques such as raised pitch, lower pitch, increased nasality, covering the mouth, constricting tract, obstacle in mouth etc by analyzing and comparing the amount of phonetic and acoustic variation in of artificial (disguised) and natural sample of an individual, by auditory as well as spectrographic analysis [27,28].

## Materials and Methods

This research was conducted at Voice Division of Directorate of Forensic Science, Gandhinagar and Institute of Forensic Science, Gujarat Forensic Sciences University, Gandhinagar. The study included disguise samples and control samples of 200 individuals of different sex, religion and age groups, mostly of Gujarat origin. Out of 200, 102 samples were collected from male speakers and 98 from female speakers of age group 20 to 60 years. Most of the speakers both male and females were in the age group of 25 to 35 years. All the voices samples were collected using high quality Digital recorder. The disguise voice samples were carefully collected from each speaker under distinctive condition which imposes certain variations in the acoustic and perceptual parameters of recorded voice sample. Besides this three control samples (routine voice sample) were also collected from each individual, in order to study the degree of variations among disguise voice and natural voice of a person. The disguise conditions on which we focused were:

- a. Keeping hand/cloth on mouth
- b. Variations in the vocal pitch
- c. Simulating anger
- d. Condition of extreme cold
- e. Condition of bad throat
- f. Chewing pan or tobacco
- g. Constriction of vocal tract
- h. Pinching nostrils
- i. Pulling cheeks
- j. Changing the accent and talking style
- k. Mimicry

### Materials required

1. Go Gear Philips digital recorder
2. High quality head phones
3. Data Cable
4. Gold wave software
5. Computerized speech lab model- 4500
6. SIS Software
7. Voice Net automatic software

### Steps for sample collection

1. A transcript was prepared with the contents designed to simulate a blackmailing call of approximately 2 min duration, which was presented to each individual for collection of their voice sample.
2. The recording were conducted in sound proof recording room of DFS, Gandhinagar
3. While collecting voice samples, all the speakers were asked to recite the same transcript four times in same session i.e. one in disguised state (with his/her choice) and three in control state. Therefore, a total four samples were collected from 200 different speakers.
4. All the Voice samples were collected on Go Gear mix Philips Digital recorder at a distance of approximately 40 cm from the mouth of the speaker.
5. A duly filled consent form from each speaker was collected along with their voice samples. Also a declaration was provided to each speaker to ensure the secrecy and usability of their voice samples.
6. The detailed records of name, age, sex, concerned guardian, geographical origin and educational background of each speaker was maintained properly along with their samples.

## Examination and analysis of voice samples

All the disguised and control speech samples of each individual were then subjected to different softwares for comparison in order to determine the similarities and dissimilarities in their auditory and spectrographic parameters. Almost 22 acoustic parameters were compared for identification of disguised speakers including:

- Auditory features: quality of speech sample, delivery of speech, frequently used words, pronunciation, accent, talking style, dialect used, flow of speech, degree of phonation, nature & degree of pauses, nasality and speech time (S/T) rate.
- Spectrographic parameters: fundamental frequency; formant frequency, formant patterns, amplitude, energy patterns, pitch patterns, loudness, transitional characteristics, bandwidth [29-31].

The results for the analysis were recorded and were statistically evaluated to frame the final conclusions. The statistics applied include:

1. Pearson correlation in order to measure the association between ideal and disguised speech parameters.
2. Chi-square test for assessing the dependency between a set of observed values (disguised speech parameters) and those expected (Control speech parameters).
3. Z-test for measuring the amount and nature of variations between disguised and control voice samples of individuals.

## Preparing Files for Analysis

Each recording device has its individual format of recording the voice file. The files with inappropriate format do not suit for spectrographic analysis, therefore, it is recommended to convert the file into the accepted format:

- Sampling rate: 11025 Hz
- Bit rate : 172 Kbps
- Bit depth: 16 bits
- Channel: Mono
- File Format: Wave with the help of Goldwave Software and saved.

## Results and Conclusions

The subjects were asked to give one of the voice samples by doing some modifications in their original voice. Among the 200 different subjects, the following disguise techniques were adopted including constriction of tract (6%), lowering of pitch (6%), pinching nostrils (9%), pulling cheeks (3%), raising pitch (10%), changing tone/accnt (1%), covering mouth (34%), simulating anger (5%), state of cold (2%), mimicry (3%), with some obstacle in mouth (9%), protruding lips (3%), throat infection (3%) and whispering (6%) (Figure 1).

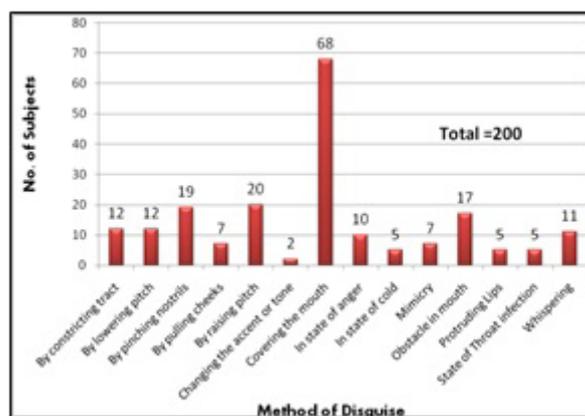


Figure 1: Chart showing the disguise techniques preferred by the different speakers (Total, N=200)

## Aural parameters in disguised and control/normal speech

While examination, analysis and comparison of disguised speech sample (n=200) with their respective controls (n=200), it was observed that, the disguising of the voice leads to the degradation of aural parameters with respect to the normal voice conditions. The auditory analysis was carried out in presence of 3 expert listeners in the age group of 25-40 years, using high quality headphones. The disguised and control speech samples of each individual were listened again and again in order to determine the amount of similarities and dissimilarities between them. The results of the analysis were recorded in the proper format (Figure 2).

The parameters like quality of speech, delivery of speech, flow of speech, speech rate and dynamic loudness degrades at higher degree in the disguised conditions as compared to their respective control samples.

**1. Demographic Information:**

QUESTIONED SAMPLE: Male/female, Age, Height, Weight, etc.

SPECIMEN SAMPLE: Male/female, Age, Height, Weight, etc.

CONTROL SAMPLE: Male/female, Age, Height, Weight, etc.

**2. Acoustic Parameters:**

ACUSTIC PARAMETER	QUESTIONED SAMPLE	SPECIMEN SAMPLE	CONTROL SAMPLE
Duration of speech	Medium	Medium	Medium
Pitch	Medium	Medium	Medium
Intensity	Normal	Normal	Normal
Timbre	Normal	Normal	Normal
Speech Rate	Normal	Normal	Normal
Number & Length of Pauses	Normal (short duration)	Normal (short duration)	Normal (short duration)

**3. Perceptual/Subjective Parameters:**

ACUSTIC PARAMETER	QUESTIONED SAMPLE	SPECIMEN SAMPLE	CONTROL SAMPLE
Duration of speech	Normal	Normal	Normal
Pitch	Normal	Normal	Normal
Intensity	Normal	Normal	Normal
Timbre	Normal	Normal	Normal
Speech Rate	Normal	Normal	Normal
Number & Length of Pauses	Normal (short duration)	Normal (short duration)	Normal (short duration)

Figure 2: Observation sheet maintained for recording the results of auditory analysis

Degree of phonation in disguised voice samples showed moderate variations as compared to that in their control samples. The parameters like nasality and nature of pauses were found to be consistent with that in their control samples.

Most of the aural parameters of voice samples disguised by constricting tract, pinching nostrils, covering mouth, obstacle in mouth, in state of cold, in state of throat infection and whispering showed higher deviations from that in their respective control samples. On other hand the voice samples disguised by simulating anger, pulling cheeks and changing accent/tone showed high consistency and similarity in aural parameters with that of their respective control counterparts.

The variations in the aural parameters significantly depend upon the type of speech sample, and were found to be independent of sex of the speaker.

**Quality of speech:** Voice quality is derived from a variety of laryngeal and supralaryngeal features, running continuously through the individual’s speech. Speech quality degrades at higher level with the condition of voice disguise. About 61% of the total disguised speech samples collected from 200 subjects were having low quality of speech, while the percentage of low quality of speech in control voice samples was only 12% (5 times less than that in disguised samples).

A strong negative correlation in speech quality was observed between the samples disguised by constricting tract, lowering pitch, pinching nostrils, raising pitch, covering mouth, in state of cold, mimicry, obstacle in mouth, protruding lips, throat infection and whispering when compared to speech quality in their respective control samples, indicating significant variations (at alpha=0.05) between two samples (Table 1 and 2).

Method of Disguise	Correlation Coefficient With Control Samples
Constricting tract	-0.94
Lowering pitch	-0.33
Pinching nostrils	-0.73
Pulling cheeks	+0.94
Raising pitch	-0.99
Simulating anger	+0.987
Changing accent/Tone	+0.99
Covering the mouth	-0.95
State of cold	-0.693
Mimicry	-0.629
Obstacle in mouth	-0.896
Protruding Lips	-0.693
State of Throat infection	-0.944
Whispering	-0.949

Table 1: Pearson correlation coefficient for speech quality between disguised & control voice samples of both males and females (TOTAL, N=200)

On other hand, high consistency and a strong positive correlation was observed in speech quality of the voice samples disguised by pulling cheeks, simulation of anger and changing of accent/tone and their control counterparts, indicating non-significant variations (at  $\alpha=0.05$ ) between two samples (Table 1 and 2).

Disguise type	Chi-Square value	p-value
By constricting tract	12	0.0025
By lowering pitch	14.4	0.0007
By pinching nostrils	30.76	<0.0001
By pulling cheeks	0.44	0.8025
By raising pitch	7.51	0.0234
Covering the mouth	63.06	<0.0001
Simulating anger	0.4	0.8187
In state of cold	4.8	0.0907
Mimicry	8.57	0.0138
Obstacle in mouth	23.19	<0.0001
Protruding Lips	4.8	0.0907
State of Throat infection	4.33	0.1147
Whispering	15.16	0.0005

**Table 2:** p-values for chi-square test for speech quality of disguised and control voice samples (TOTAL, N=200; df=2)

**Delivery of speech:** As voice imitation involves the manipulation of articulators to deliver a sound more close to the model voice, majority of about 65% the total disguised speech samples collected from 200 subjects showed low speech delivery, while the percentage of low speech delivery in control voice samples was 25% (2.6 times less than that in disguised samples).

A strong negative correlation in speech delivery was observed between the samples disguised by constricting tract, pinching nostrils, raising pitch, covering mouth, in state of cold, mimicry, obstacle in mouth, throat infection and whispering when compared to speech delivery in their respective control samples, indicating significant variations (at  $\alpha=0.05$ ) between two samples (Table 3 and 4).

On other hand, high consistency and a strong positive correlation was observed in speech delivery of the voice samples disguised by lowering pitch, pulling cheeks, simulation of anger, protruding lips and changing of accent/tone and their control counterparts, indicating non-significant variations (at  $\alpha=0.05$ ) between two samples (Table 3 and 4).

Method of Disguise	Correlation Coefficient With Control Samples
Constricting tract	-0.99
Lowering pitch	+0.98
Pinching nostrils	-0.987
Pulling cheeks	+0.99
Raising pitch	-0.99
Simulating anger	+0.989
Changing accent/Tone	+0.99
Covering the mouth	-0.988
State of cold	-0.99
Mimicry	-0.989
Obstacle in mouth	-0.989
Protruding Lips	+0.99
State of Throat infection	-0.978
Whispering	-0.988

**Table 3:** Pearson correlation coefficient for speech delivery between disguised & control voice samples of both males and females (TOTAL, N=200)

Disguise type	Chi-Square value	p-value
By constricting tract	10.74	0.001
By lowering pitch	0.22	0.639
By pinching nostrils	18.24	<0.0001

Disguise type	Chi-Square value	p-value
By pulling cheeks	0.001	0.99
By raising pitch	6.83	0.009
Covering the mouth	7.57	0.0059
Simulating anger	0.001	0.99
In state of cold	11.1	0.0009
Mimicry	14.23	0.0002
Obstacle in mouth	9.56	0.002
Protruding Lips	0.005	0.98
State of Throat infection	9.2	0.0024
Whispering	7.54	0.006

**Table 4:** p-values for chi-square test for delivery of speech of disguised and control voice samples (TOTAL, N=200; df=1)

### Degree of phonation

25% of the total disguised speech samples of both males and females showed low degree of phonation. While none of total control samples of both males and females showed low degree of phonation. This was due to the fact that phonation occurs when the potential energy of the airstream compressed airstream below the larynx converts into the kinetic energy of egressive airflow producing audible sounds. Any form of constriction or modification of the laryngeal passage (in case of voice disguise) results in the turbulence in the airflow causing audible friction, degrading the degree of phonation of voice.

A strong negative correlation in degree of phonation was observed between the samples disguised by constricting tract, pinching nostrils, in state of cold, obstacle in mouth, protruding lips, throat infection and whispering when compared to that in their respective control samples, indicating significant variations (at  $\alpha=0.05$ ) between two samples (Table 5 and 6).

On other hand, high consistency and a strong positive correlation was observed in degree of phonation of the voice samples disguised by lowering pitch, pulling cheeks, raising pitch, simulation of anger, protruding lips, mimicry and changing of accent/ tone and their control counterparts, indicating non-significant variations (at  $\alpha=0.05$ ) between two samples (Table 5 and 6).

Method of Disguise	Correlation Coefficient With Control Samples
Constricting tract	-0.989
Lowering pitch	+0.904
Pinching nostrils	-0.541
Pulling cheeks	+0.923
Raising pitch	+0.983
Simulating anger	+0.99
Changing accent/Tone	+0.989
Covering the mouth	+0.374
State of cold	-0.988
Mimicry	+0.987
Obstacle in mouth	-0.744
Protruding Lips	-0.5
State of Throat infection	-0.693
Whispering	-0.756

**Table 5:** Pearson correlation coefficient for degree of phonation between disguised & control voice samples of both males and females subjects (TOTAL, N=200)

Disguise type	Chi-Square value	p-value
By constricting tract	12.9	0.0016
By lowering pitch	2.27	0.321
By pinching nostrils	18.38	0.0001
By pulling cheeks	0.31	0.856
By raising pitch	1.23	0.541
Covering the mouth	24.74	<0.0001
Simulating anger	0.001	0.99

Disguise type	Chi-Square value	p-value
In state of cold	6.67	0.0356
Mimicry	0.001	0.99
Obstacle in mouth	13.52	0.0012
Protruding Lips	1.23	0.5406
State of Throat infection	10	0.0067
Whispering	12.33	0.0021

**Table 6:** p-values for chi-square test for degree of phonation of disguised and control voice samples (TOTAL, N=200; df=2)

**Flow of speech:** Flow of speech strongly degrades with voice disguise. About 62% the total disguised speech samples collected from 200 subjects including both males and females showed degraded and low flow of speech, because of the unnatural manipulation of the vocal tract. On other only 11% of control voice samples of both males and females were having low flow of speech.

A strong negative correlation in speech delivery was observed between the samples disguised by constricting tract, pinching nostrils, raising, changing accent/tone, covering mouth, in state of cold, mimicry, obstacle in mouth, protruding lips, throat infection and whispering when compared to speech delivery in their respective control samples, indicating significant variations (at  $\alpha=0.05$ ) between two samples (Table 7 and 8).

On other hand, high consistency and a strong positive correlation was observed in flow of speech of the voice samples disguised by lowering pitch, pulling cheeks and simulation of anger and their control counterparts, indicating non-significant variations (at  $\alpha=0.05$ ) between two samples (Table 7 and 8).

Method of Disguise	Correlation Coefficient With Control Samples
Constricting tract	-0.99
Lowering pitch	+0.989
Pinching nostrils	-0.978
Pulling cheeks	+0.99
Raising pitch	-0.988
Simulating anger	+0.99
Changing accent/Tone	-0.981
Covering the mouth	-0.989
State of cold	-0.99
Mimicry	-0.989
Obstacle in mouth	-0.99
Protruding Lips	-0.986
State of Throat infection	-0.991
Whispering	-0.978

**Table 7:** Pearson correlation coefficient for flow of speech between disguised & control voice samples of both males and females subjects (TOTAL, N=200)

Disguise type	Chi-Square value	p-value
By constricting tract	10.74	0.001
By lowering pitch	0.005	0.98
By pinching nostrils	13.06	0.0003
By pulling cheeks	0.001	0.99
By raising pitch	10.67	0.0011
Covering the mouth	35.75	<0.0001
Simulating anger	0.005	0.98
In state of cold	4.51	0.0337
Mimicry	5.16	0.0231
Obstacle in mouth	14.69	0.0001
Protruding Lips	6.76	0.0093
State of Throat infection	4.11	0.0426
Whispering	6.77	0.0093

**Table 8:** p-values for chi-square test for flow of speech of disguised and control voice samples (TOTAL, N=200; df=1)

**Speaking rate:** Speech rate strongly degrades with voice disguise. About 43% the total disguised speech samples collected from 200 subjects including both males and females, showed high variations in speech rate from their control counterparts, because voice disguise is a conscious effort where at each point the impersonator has to go slow to impart perfection in imitated voice model. Sometimes the mimicry artist has to impersonate the voice of person having high speaking rate than his normal capacity.

A weak correlation in speech rate was observed between the samples disguised by constricting tract, lowering pitch, pinching nostrils, pulling cheeks, raising pitch, changing accent/tone, covering mouth, in state of cold, mimicry, obstacle in mouth, protruding lips, throat infection and whispering when compared to speech rate in their respective control samples, indicating significant variations (at  $\alpha=0.05$ ) between two samples (Table 9 and 10).

On other hand, moderate correlation was observed in speech rate of the voice samples disguised by simulation of anger and their control counterparts, indicating non-significant variations (at  $\alpha=0.05$ ) between two samples (Table 9 and 10).

Method of Disguise	Correlation Coefficient With Control Samples
Constricting tract	-0.99
Lowering pitch	+0.217
Pinching nostrils	-0.678
Pulling cheeks	-0.803
Raising pitch	-0.216
Simulating anger	+0.693
Changing accent/Tone	-0.99
Covering the mouth	-0.962
State of cold	-0.693
Mimicry	-0.996
Obstacle in mouth	-0.945
Protruding Lips	-0.945
State of Throat infection	-0.5
Whispering	-0.782

**Table 9:** Pearson correlation coefficient for speech rate between disguised & control voice samples of both males and females (TOTAL, N=200)

Disguise type	Chi-Square value	p-value
By constricting tract	13	0.0015
By lowering pitch	8.19	0.0167
By pinching nostrils	15.69	0.0004
By pulling cheeks	8	0.0183
By raising pitch	7.46	0.024
Covering the mouth	45.43	<0.0001
Simulating anger	2.49	0.2879
In state of cold	6.65	0.0360
Mimicry	6	0.0498
Obstacle in mouth	21.4	<0.0001
Protruding Lips	10.01	0.0067
State of Throat infection	6.67	0.0356
Whispering	7.37	0.0251

**Table 10:** p-values for chi-square test for speech rate of disguised and control voice samples (TOTAL, N=200; df=2)

**Nasality:** The percentage of nasality and nonnasality in disguised samples of both males and females was found to be 12% and 88% respectively. 100% of the control voice samples of both males and females showed non nasal sounds. The chi-square value for nasality in all disguised and control voice samples was found to be 23.45 ( $p<0.0001$ ;df=1) which was found to be significant at  $\alpha=0.05$ , rejecting the null hypothesis and accepting the alternate hypothesis that the variations seen in nasality significantly depends upon the type of speech sample.

**Dynamic loudness:** About 40% the total disguised speech samples collected from 200 subjects including both males and females showed low loudness. Loudness varies under different disguise conditions depending upon how much kinetic energy is been delivered to egressive speech sound by the impersonator. On other hand only 10% of control voice samples of both male and female subjects were having low loudness.

A weak correlation in dynamic loudness was observed between the samples disguised by constricting tract, pinching nostrils, pulling cheeks, raising pitch, obstacle in mouth, simulating anger, in state of cold, mimicry, protruding lips, throat infection and whispering when compared to dynamic loudness in their respective control samples, indicating significant variations (at  $\alpha=0.05$ ) between two samples (Table 11 and 12).

On other hand, strong positive correlation was observed in dynamic loudness of the voice samples disguised by lowering pitch and their control counterparts, indicating non-significant variations (at  $\alpha=0.05$ ) between two samples (Table 11 and 12).

Method of Disguise	Correlation Coefficient With Control Samples
Constricting tract	-0.596
Lowering pitch	+0.885
Pinching nostrils	-0.649
Pulling cheeks	-0.189
Raising pitch	+0.189
Simulating anger	-0.277
Changing accent/Tone	+0.5
Covering the mouth	+0.536
State of cold	-0.5
Mimicry	-0.115
Obstacle in mouth	+0.104
Protruding Lips	-0.945
State of Throat infection	-0.99
Whispering	-0.61

**Table 11:** Pearson correlation coefficient for dynamic loudness between disguised & control voice samples of both male and female subjects (TOTAL, N=200)

Disguise type	Chi-Square value	p-value
By constricting tract	20.31	<0.0001
By lowering pitch	1.57	0.4561
By pinching nostrils	17.56	0.0002
By pulling cheeks	7.77	0.0205
By raising pitch	9.11	0.011
Covering the mouth	70.73	<0.0001
Simulating anger	10.04	0.0066
In state of cold	6.67	0.0356
Mimicry	7.65	0.0218
Obstacle in mouth	9.47	0.0088
Protruding Lips	6.33	0.0422
State of Throat infection	6.67	0.0356
Whispering	18.33	0.0001

**Table 12:** pp-values for chi-square test for dynamic loudness of disguised and control voice samples (TOTAL, N=200; df=2)

**Nature of pauses:** 95% of the disguised voice samples and 100% of the control voice samples showed normal pauses. Only the voice samples disguised in state of cold and throat infection, showed abnormal pauses (5%).

The chi-square value for nature of pauses in all disguised and control voice samples was found to be 8.31 ( $p=0.0039$ ;  $df=1$ ) which was found to be significant at  $\alpha=0.05$ , rejecting the null hypothesis and accepting the alternate hypothesis that the variations seen in nature of pauses significantly depends upon the type of speech sample.

### Spectrographic parameters in disguised and control/normal speech

Computerized speech lab model4500 was used for conducting the spectrographic analysis of voice samples. The spectrographic parameters like fundamental frequency, formant bands, formant frequencies, energy levels were found to be significantly more reliable in cases involving disguised speech samples than the aural parameters. The values of these voice parameters in disguised remained more consistent with that seen in their respective control samples as compared to aural parameters.

Third formant (F3) and Fourth formant (F4) were found to be more essential in identification of the disguised voice samples, followed by parameters like fundamental frequency (F0), first formant (F1), second formant (F2) and energy levels. Fifth formant (F5) found to be least important for comparison and identification of disguise voice samples.

**Fundamental Frequency (F0):** F0 was found to be crucial parameter for identification of voice samples disguised by constricting tract, lowering pitch (in male subjects), changing accent, pulling cheeks, in state of cold, simulating anger and covering mouth. The values of F0 in samples disguised by these techniques showed no significant variations from their control counterparts (at significance level of 0.05).

F0 does not found to be important for identification of voice samples disguised by lowering pitch (in female subjects), pinching nostrils, raising pitch, mimicry, obstacle in mouth, in state of throat infection and whispering. The values of F0 in samples disguised by these techniques showed significant variations from their control counterparts (at significance level of 0.05).

**First formant (F1):** F1 was found to be crucial parameter for identification of voice samples disguised by constricting tract, lowering pitch (in male subjects), pulling cheeks, raising pitch, in state of cold, simulating anger and covering mouth. The values of F1 in samples disguised by these techniques showed no significant variations from their control counterparts (at significance level of 0.05).

F1 does not found to be important for identification of voice samples disguised by lowering pitch (in female subjects), pinching nostrils, changing accent, mimicry, obstacle in mouth, in state of throat infection and whispering. The values of F1 in samples disguised by these techniques showed significant variations from their control counterparts (at significance level of 0.05).

**Second formant (F2):** F2 was found to be crucial parameter for identification of voice samples disguised by constricting tract, lowering pitch (in female subjects), mimicry obstacle in mouth, throat infection and whispering. The values of F2 in samples disguised by these techniques showed no significant variations from their control counterparts (at significance level of 0.05).

F2 does not found to be important for identification of voice samples disguised by lowering pitch (in male subjects), pinching nostrils, pulling cheeks, raising pitch, changing accent, covering mouth, simulating anger and in state of cold. The values of F2 in samples disguised by these techniques showed significant variations from their control counterparts (at significance level of 0.05).

**Third formant (F3):** F3 was found to be crucial parameter for identification of voice samples disguised by lowering pitch, pinching nostrils, pulling cheeks, raising pitch, changing accent, covering mouth, obstacle in mouth and whispering. The values of F3 in samples disguised by these techniques showed no significant variations from their control counterparts (at significance level of 0.05).

F3 does not found to be important for identification of voice samples disguised by constricting tract, simulating anger, in state of cold, mimicry and in state of throat infection. The values of F3 in samples disguised by these techniques showed significant variations from their control counterparts (at significance level of 0.05).

**Fourth formant (F4):** F4 was found to be crucial parameter for identification of voice samples disguised by lowering pitch, pinching nostrils, changing accent, covering mouth, simulating anger, in state of cold, mimicry, in state of throat infection and whispering. The values of F4 in samples disguised by these techniques showed no significant variations from their control counterparts (at significance level of 0.05).

F4 does not found to be important for identification of voice samples disguised by constricting tract, pulling cheeks, raising pitch and obstacle in mouth. The values of F4 in samples disguised by these techniques showed significant variations from their control counterparts (at significance level of 0.05).

**Fifth formant (F5):** F5 was found to be crucial parameter for identification of voice samples disguised in state of throat infection, state of cold, changing accent, pinching nostrils and constricting tract. The values of F5 in samples disguised by these techniques showed no significant variations from their control counterparts (at significance level of 0.05).

F5 does not found to be important for identification of voice samples disguised by whispering, obstacle in mouth, mimicry, simulation of anger, covering mouth, raising pitch, lowering pitch and pulling cheeks. The values of F5 in samples disguised by these techniques showed significant variations from their control counterparts (at significance level of 0.05).

**Energy contour:** Energy pattern was found to be crucial parameter for identification of voice samples disguised by lowering pitch, raising pitch, pulling cheeks, pinching nostrils, change of accent and simulation of anger. The values of energy in samples disguised by these techniques showed no significant variations from their control counterparts (at significance level of 0.05).

Energy pattern does not found to be important for identification of voice samples disguised by constricting tract, covering mouth, in state of cold, mimicry, obstacle in mouth, in throat infection and whispering. The values of energy in samples disguised by these techniques showed significant variations from their control counterparts (at significance level of 0.05).

## Parameters useful for comparison and identification of voice samples disguised by different methods

**Disguise by constricting tract:** F0, F1, F2 and F5 were found to be essential parameters for identification of voice samples disguised by constricting tract, for both male and female voices. These parameters showed no significant variations from their respective control voice samples. F3, F4 and energy levels in such disguise condition showed more deviation from their control values.

**Disguise by lowering pitch:** F0, F1, F3, F4 and energy levels were found to be essential parameters for identification of voice samples disguised by lowering pitch, for male voices. These parameters showed no significant variations from their respective control voice samples. F2 and F5 in such disguise condition, in male samples, showed more deviation from their control values.

F2, F3, F4 and energy levels were found to be essential parameters for identification of voice samples disguised by lowering pitch, for female voices. These parameters showed no significant variations from their respective control voice samples. F0, F1 and F5 in such disguise condition, in female samples, showed more deviation from their control values.

**Disguise by pinching nostrils:** F3, F4, F5 and energy levels were found to be essential parameters for identification of voice samples disguised by pinching nostrils, for both male and female voices. These parameters showed no significant variations from their respective control voice samples. F0, F1 and F2 in such disguise condition showed more deviation from their control values.

**Disguise by pulling cheeks:** F0, F1, F3 and energy levels were found to be essential parameters for identification of voice samples disguised by pulling cheeks, for both male and female voices. These parameters showed no significant variations from their respective control voice samples. F2, F4 and F5 in such disguise condition showed more deviation from their control values.

**Disguise by raising pitch:** F1, F3 and energy levels were found to be essential parameters for identification of voice samples disguised by raising pitch, for both male and female voices. These parameters showed no significant variations from their respective control voice samples. F0, F2, F4 and F5 in such disguise condition showed more deviation from their control values.

**Disguise by changing accent/tone:** F0, F3, F4, F5 and energy levels were found to be essential parameters for identification of voice samples disguised by changing accent/tone. These parameters showed no significant variations from their respective control voice samples. F1 and F2 in such disguise condition showed more deviation from their control values.

**Disguise by covering mouth:** F0, F1, F3, and F4 were found to be essential parameters for identification of voice samples disguised by covering mouth, in both male and female voices. These parameters showed no significant variations from their respective control voice samples. F2, F5 and energy levels in such disguise condition showed more deviation from their control values.

**Disguise by simulating anger:** F0, F1, F4 and energy levels were found to be essential parameters for identification of voice samples disguised by simulating anger, in both male and female voices. These parameters showed no significant variations from their respective control voice samples. F2, F3 and F5 in such disguise condition showed more deviation from their control values.

**Disguise in state of cold:** F0, F1, F4 and F5 were found to be essential parameters for identification of voice samples disguised in state of cold, in both male and female voices. These parameters showed no significant variations from their respective control voice samples. F2, F3 and energy levels in such disguise condition showed more deviation from their control values.

**Disguise by mimicry:** F2 and F4 were found to be essential parameters for identification of voice samples disguised by mimicry, in both male and female voices. These parameters showed no significant variations from their respective control voice samples. F0, F1, F3, F5 and energy levels in such disguise condition showed more deviation from their control values.

**Disguise by obstacle in mouth:** F2 and F3 were found to be essential parameters for identification of voice samples disguised by obstacle in mouth, in both male and female voices. These parameters showed no significant variations from their respective control voice samples. F0, F1, F4, F5 and energy levels in such disguise condition showed more deviation from their control values.

**Disguise in state of throat infection:** F2, F4 and F5 were found to be essential parameters for identification of voice samples disguised in state of throat infection, in both male and female voices. These parameters showed no significant variations from their respective control voice samples. F0, F1, F3 and energy levels in such disguise condition showed more deviation from their control values.

**Disguise by whispering:** F2, F3 and F4 were found to be essential parameters for identification of voice samples disguised by whispering, in both male and female voices. These parameters showed no significant variations from their respective control voice samples. F0, F1, F5 and energy levels in such disguise condition showed more deviation from their control values.

## References

1. Benzeghiba M, De Mori R, Deroo O, Dupont S, Erbes T, et al. (2007) Automatic speech recognition and speech variability: A review. *Speech Communication* 49: 763-86.
2. Zhang C, van de Weijer J, Cui J (2006) Intra- and inter-speaker variations of formant pattern for lateral syllables in Standard Chinese. *Forensic Sic Int* 158: 117-24.
3. Fry D (1979) *The physics of speech*. Cambridge University Press, England.
4. Rogers H (2005) *The Sounds of Language: An Introduction to Phonetics* (1<sup>st</sup> edn) New York: Pearson Education Limited, Taylor & Francis Group.

5. Mathur S, Choudhary S, Vyas J (2014) Speaker Recognition System and its Forensic implications: A review. *Int J Latest Technol Eng Manag Appl Sci* 2: 1-6.
6. Van Lancker D (1986) Familiar voice recognition and unfamiliar voice discrimination are independent and unordered abilities. *J Acoust Soc Am* 79.
7. Endres W, Bambach W, Flösser G (1971) Voice Spectrograms as a Function of Age, Voice Disguise, and Voice Imitation. *J Acoust Soc Am* 49: doi.org/10.1121/1.1912589.
8. Clark J, Foulkes P (2006) Identification of familiar voices in disguised speech. Proceedings IAFPA. Department of Linguistics. Goteborg University.
9. Kunzel H, Rodriguez J and Garcia J (2004) Effect of Voice Disguise on the Performance of a Forensic Automatic Speaker Recognition System (1<sup>st</sup> edn), *Odyssey* 153-6.
10. Perrot P, Aversano G, Blouet R, Charbit M, Chollet G (2005) Voice forgery using ALISP: Indexation in a Client Memory. In Proc. ICASSP 17-20.
11. Perrot P and Chollet G (2008) The question of disguised voice. *J Acoust Soc Am* 123: 3878.
12. Amin T, German J, Marziliano P (2013) Detecting voice disguise from speech variability: Analysis of three glottal and vocal tract measures. *J Acoust Soc Am* 134: 4068.
13. Bansod N, Kawathekar S, Dabhade S (2012) Review of Different Techniques for Speaker Recognition System. *Adv Computational Res* 4: 57-60.
14. Beigi H (2011) Fundamentals of speaker recognition. Springer Science Business Media, LLC.
15. Benesty J, Sondhi M, Huang Y (2008) Springer handbook of speech processing Springer, Berlin
16. Chenafa M, Istrate D, Vrabie V, Herbin M (2008) Biometric system based on voice recognition using multiclassifiers. In *Biometrics and Identity Management* 5372: 206-15 Springer Berlin Heidelberg.
17. Mathur S, Choudhary S (2014) Voiceprint Evidence: Identifying anonymous callers. *Kavach; The Magazine of Raksha Shakti University* 3.
18. Mannell R (2009) Place of Articulation.
19. Markowitz J (2016) Glossary for Speaker Verification; Glossary for Speech Recognition; Glossary for Speech Analytics.
20. Masthoff H (1996) A report on a voice disguise experiment. *Int J Speech Lang Law* 3: 160-7.
21. Myers L (2016) An Exploration of Voice Biometrics.
22. Neuhauser S (2009) Voice Disguise Using a Foreign Accent: Phonetic and Linguistic Variation. *Int J Speech Lang Law* 15: 2.
23. Patra S (2007) Robust speaker identification system. Super Computer Education and Research Centre. Indian Institute of Science.
24. Paul T, George S (2013) Voice Recognition Based Secure Android Model for Inputting Smear Test Results. *Int J Eng Sci Emerg Technol* 6: 344-51.
25. Reynolds D (2008) Speaker and language recognition: a guided safari. In *Odyssey*, Stellenbosch, South Africa.
26. Rose P (2002) Forensic speaker identification. London: Taylor & Francis.
27. Shan Z and Yang Y (2009) Scores selection for emotional speaker recognition. Proceedings of Third Int Confer ICB Alghero, Italy, 494-502.
28. Stevens KN (1998) Acoustic phonetics. Cambridge, Mass, MIT Press, United States.
29. Strazny P (2005) Encyclopedia of linguistics. Fitzroy Dearborn, New York.
30. Stylianou Y, Faúndez Zanuy M, Esposito A (2007) Progress in nonlinear speech processing. Springer Berlin.
31. Zetterholm E, Elenius, D, Blomberg M (2005) A case study of impersonation from a security systems point of view (1st edn). Dept of Linguistics, Lund University, Sweden 239-55.

Submit your next manuscript to Annex Publishers and benefit from:

- ▶ Easy online submission process
- ▶ Rapid peer review process
- ▶ Online article availability soon after acceptance for Publication
- ▶ Open access: articles available free online
- ▶ More accessibility of the articles to the readers/researchers within the field
- ▶ Better discount on subsequent article submission

Submit your manuscript at

<http://www.annexpublishers.com/paper-submission.php>