XOURNALS

Academic Journal of Forensic Sciences

ISSN: 2581-4273 | Volume 03 | Issue 01 | April-2020

Speaker Recognition: On the Basis of their Habitual and Apprehensive Voice

Gurpreet Kaur[1], Ranjeet Kumar Singh[2]

Available online at: www.xournals.com

Received 07th March 2020 | Revised 26th March 2020 | Accepted 08th April 2020

Abstract:

Spoken language is the natural method of communication that contains the transfer of various information related to linguistics (accent etc.), information related to speakers (emotions, etc.), and also information related to the environment (background noise, etc.). The ability of humans to extract and decode spoken language automatically inspires various researchers to study the distinct prospects of spoken language, which includes recognition of accent or recognition of changed accent, recognition of emotions or gender, etc. "Voiceprint" is a collection of acoustic frequency spectrum which contains the significant features of a human speech that are used for the recognition of a speaker. The voiceprint of an individual has a distinct quality of uniqueness, durability, and strength. Every speaker has unique features of speaking besides those physiological dissimilarities such as the use of specific accent, intonation style, etc. An apprehensive speech is a disguised speech of the speaker recorded under the influence of any threat, nervousness, etc. used for various criminal purposes such as fraud or spam calls, etc. This paper focuses on the areas concerned with the information extraction of an individual's speech observable in speech signals such as emotional state, intentional accent change, belligerence, etc. will give better clues to the investigator for the differentiation. Some external factors (environmental noise or emotions etc.) impact the effectiveness of speaker identification. But, the basic components of their original voice remain unchanged such as formant frequency in the "Voiceprints" which helps in the recognition process even after using an apprehensive voice. The intonation pattern of formants of the speaker's original voices will almost be similar to the intonation pattern of formants of the speaker's deliberate apprehensive voice.

Keywords: *Apprehensive Voice, Disguised Voice, Speaker Recognition, Voiceprint, Intonation Pattern.*

Authors:

1. *Chandigarh University, Punjab, INDIA.*
2. *Sherlock Institute of Forensic Science, GTB Nagar, New Delhi, INDIA.*

Introduction

Ears have the unique ability to receive and decipher spoken language. Besides that, the ears also have various diverse functions, out of which one function of the ear is the identification of people by their voices **(Sharma and Bansal 2013)**. Forensic auditory analysis has been a subject of methodological and scientific discussion for a long time. It is a globally expanding tendency that the criminals are more willing to disguise their voices to hide their identity, particularly in cases of extortion, threatening calls, emergency calls to the police **(Zhang and Tan, 2007).** Every individual in this world has a unique voice. The voice of no two individuals is the same due to physiological dissimilarities. The individuality of the individual's voice can be employed to verify the person's identity. Every speaker has unique features of speaking besides those physiological dissimilarities such as the use of specific accent, intonation style, rhythm, suffering from any disorder that affects the speaking ability and causes tremors, etc. **(Zheng and Li., 2017)**. There are many possibilities available to the speaker to manipulate his/her voice to falsify an automatic recognition system or even the human ear **(Perrot, Aversano, and Chollet, 2007)**. Biometric access control systems, Automatic Speaker Recognition Systems, auditory analysis in forensics, etc. are several examples of speaker and voice recognition **(Lal and Nath N. J., 2015)**. Biometric accessed security systems are built based on the unique features of humans like voice, fingerprints, etc. These types of systems provide an additional barrier to stop the unauthorized approach for the protection of data by detecting the user's particular behavioral or physiological features. Biometric accessed security systems are more authentic than the standard traditional method. There are higher demands on speaker identification on modeling the vocal tract features of speakers such as an illness, to provide a more secure approach to financial or sensitive information. The verification of speakers gives more barriers to stop uncertified access to secure the data and also improves the protection provided by personal identification **(Li**, **Yang,** and **Dai, 2014)**.

Apprehensive voice is mainly that voice which reflects fear, anger, anxiety, nervousness, shivered voice because of illness which in final disguises the voice intentionally or unintentionally. An apprehensive speech is that disguised speech of the speaker which is recorded under the influence of any threat, fear, anger, and nervousness, etc. It also comes under the disguised voice category. This type of voice is used for various criminal and illicit purposes such as fraud or spam calls, threatening calls, and also during the sample collection process of suspects. In the forensic science field, finding a solution to differentiate between apprehensive voices from normal voices will give better indications to the investigator during the investigation process. Voice disguise is an intentional act of a speaker to alter, distort, deviate, or manipulate their normal voice to hide or falsify their identity **(Klevans & Rodman, 1998)**. In the field of acoustic analysis as well as in forensic science, the speaker recognition process or techniques are inexorable and they are used in the speaker identification. Speaker recognition based on Voiceprints is defined as the recognition of an individual or speaker's identity by using their voiceprints. Various researches proposed that the voiceprint of an individual has a distinct quality of uniqueness, durability, and strength, which always remain stable and become unchanged in adulthood except for some disorders. It is also suggested that the voice of no individuals is the same due to physiological dissimilarities. Every speaker has unique features of speaking besides those physiological dissimilarities such as the use of specific accent, intonation style, rhythm, suffering from any disorder that affects the speaking ability, and causes tremors, etc. Intonation patterns of voice are the patterns of variation generated by the rise and fall in the pitch of the voice. These intonation patterns are also helpful in speaker recognition in case of a disguised or apprehensive voice.

There are different techniques or methods available for the speaker identification, out of which some are based on the phonetic approaches while few others are based on the automatic algorithms. It is important to determine whether the voice is apprehensive or normal before proceeding to further voice analysis. An apprehensive voice is also a form of a disguised voice that means the intentional actions to hide the person's identity. But not all apprehensive voices are disguised,

as some of them are generated due to the reasons when some person is in stress, or when he/she is suffering from some disease or under the influence of someone's threat. It will also affect the performance of various speaker identification systems as they can either be affected by producing variations in the channels of communication or by producing variations in an individual's voice. There are two possible variations in the channel of communication: the first one is handset variations and the other is environmental variations. Both of these dissimilarities were deeply investigated and researched by several researchers after which they suggested various normalization methods to oppose these variations.

The cause of unintentional modifications is emotional conditions such as stress, fear, excitement, etc. or physical illness like cold, etc. and these unintentional modifications fall under the category of "Apprehensive voice". Voice disguising also has various good applications such as in radio and television interviews for information transmission without disclosing the speaker's identity. It is also used in speech coding, synthesis of speech, entertainment, etc. Various methods are available to alter their voice such as using a foreign accent, altering speaking rate, etc. There are three main ways to produce an apprehensive voice intentionally: the first one is by placing a hand over the mouth, and the others are trying to produce a high pitch or low pitch, and producing a strained nostril voice **(Künzel, 2004)**. To determine the apprehensive voice from the normal voice, firstly an analysis of spectrographic formants of the voice should be done and compared to the habitual voice. After that, an automatic classification is acquired and these acquired results and observations will, in turn, give fascinating clues of differentiation to the investigators.

There are two ways of variations in an individual's voice that is: **(Mireia Farrus, 2017)**

1. **Deliberate Variation –** A disguised voice that is deliberate mainly speaker-dependent. **Künzel (1)** described the dissimilarities in the approach between both the women and men. The deliberate disguise also categorized into:

- **Non-Electronic Disguise** – Imitation of voice is natural in human beings and observed in the communication of humans through the medium of acquisition of language, the transformation of voice, and impersonation. Impersonation is a type of imitation which aims to generate someone's voice/ speech **(Markham, 1997).**

- **Electronic Disguise** – In electronic disguise, a device is employed to alter someone's natural voice/speech. When used deliberately, it is frequently observed in the voice conversion form. The conversion of voice is the modification of the source speaker's voice into the target speaker's voice in a manner to mimic the target's voice.

2. **Non-deliberate Variation –** There are several changes due to the uncontrolled reasons occurring in the people's voice. The majority of those changes are caused naturally by modifications that impact the usual development of the body such as illness, age, etc. Other types of modifications can be initiated due to the use of electronic devices in the process of communication. They are known as "Non-deliberate Non-electronic Disguise" and "Non-deliberate Electronic Disguises".

- **Non-Electronic Disguise** – The best example of non-electric and non-deliberate disguise is a hoarse voice. Hoarseness in voice is a change in the quality of voice habitually manifested by rough or breathy voices and which is usually caused by illnesses like laryngitis etc. **(Sulica, 2011)**. The other example of this type of disguise is "Emotional changes" which contains the impact of emotions on the speech which have been studied widely. The other example is "Intoxication". Some speeches are also altered or manipulated under the influence of intoxication. "Ageing" is another example of a non-electronic disguise in which speech production is facing physiological as well as anatomical changes throughout life **(Schoetz, 2007)**.

- **Electronic Disguise** – Non-deliberate electronic disguise is defined as any distortion or deformity in speech because of the channel effects, for example, microphone use, etc. Aside from the population size employed apart from the size of the population used in the task of automatic speaker identification, the distortion produced by the noisy communication channels is evaluated as the huge factor impacting the performance of the system.

In the deliberate variation, the speaker/individual tries to mimic some other individual's voice to confuse the listener. In the non-deliberate, variations occurred due to the emotional condition or physical conditions like illness, sore throat, etc. Other two possible variations occurred in the voice of human which can be further categorized into:

1. Electronic Disguised/Apprehensive Voice
2. Non-electronic Disguised/Apprehensive Voice

In the former, the disguise in voice is produced electronically using various software tools such as Praat or Audacity. The latter voice is produced by changing the individual's voice mechanically such as by placing a hand over the mouth, by straining nostrils, etc. MFCC is the main feature that is employed widely for apprehensive voice identification and speaker identification. The MFC in the sound processing is a presentation of the short-term power sound spectrum depending on the linear cosine transform of a long-term power sound spectrum on the Mel scale of the frequency which is non-linear. MFC stands for Mel-frequency cepstrum which is a collection of the MFCC's **(Kurian and Kurup, 2016; George and George, 2015)**. Various distinct techniques have been employed in the past for identification of the author of an unidentified recording in several scenarios which ranges from entirely auditory to automatic. It was mentioned by the author **Künzel (1)** in the survey paper in 1975 that an average of 15 % of cases was provided to the Department of Speaker Recognition of the German Federal Police Office. The internal statistics of greater than 20 years disclosed that the disguise was found in most of the cases of some specific offense types such as kidnapping etc. Interestingly, the use of electronic equipment (voice changers) was exceptionally rare in Germany.

Review of Literature

Künzel, H. J. (2000), observed the effects of disguised voice on the fundamental frequency of speaking. In the longitudinal and simultaneous study, content (text) on five events were allotted to 100 peoples. They read the content within six months in which they initially used their habitual voice but later they started using the other two modes out of total three methods of voice manipulation. The total three modes of voice disguise are: (1) they increased their fundamental frequency; (2) they decreased their fundamental frequency, (3) denasalization by simply pinching their nostrils. This study mainly emphasizes the fundamental frequency. The results of this study show that most of the peoples who were employed in this investigation were capable of consistently altering their fundamental frequency as per selected disguise mode. However, in both sexes, there were some variations in the choice of the disguise modes as well as in the articulatory approach of individuals which they used to execute them. The results of the study corroborate with forensic casework experiences, which showed that the relation between the fundamental frequency of the individual's original voice and the type of disguise used in the inculpated voice call was constant and uniform. The speakers with the more than average fundamental frequency prefer to raise their fundamental frequency level for disguising their voice. But the speakers with less than average fundamental frequency prefer to decrease their fundamental frequency more which was ended up in persistently creaky voice. Generally, it was observed that the females were more reluctant than males to produce extreme changes to their patterns of fundamental frequencies. The speakers who tried to disguise their voice by lowering their fundamental frequency were much more observed in the males. Females are generally more reluctant to make drastic changes to their fundamental frequency patterns.

Künzel, H. J. *et al.* (2004), studied the effect/impact of disguised voice on the performance of FASR System. In their study, they examined the results of ongoing investigation or research on the impact of main types of disguised voices such as increased or decreased pitched voice, nose pinching during the speech, etc. on the techniques of forensic speaker recognition. Apprehensive (disguised) speech and the habitual (normal) speech of 100 German speakers which were recorded 5 times for over time duration of 7-9 months that were utilized in speaker identification experiments, by using the FASR system which was LR-based. The results of their study showed the three main selected types of disguised voice that impacts the system's performance slightly only if the referral population includes the voice data exhibiting a similar type of disguise. The impact was more critical if the referral population was collected only with the normal habitual speech and was also distinct for those three types of disguise which were under examination.

Perrot, Chollet, *et al.* (2007), studied the detection, identification, and recognition of the disguised voice. In their study, they evaluated the resolutions to differentiate between the normal and apprehensive (disguised) voice which gives important clues to the examiner that is useful in the forensic science field. Different techniques or methods were available, out of which some are based on the phonetic approaches while few others are based on the automatic algorithms. In this paper, they represented a differentiation study on the particular 4 voice disguise patterns. The main motive of their study was to determine whether the recorded voice is disguised and if yes, then what type of disguise pattern they used. The common types of disguises are: placing a hand over the mouth, producing a high or low pitch, and voice produced by pinching the nose. An analysis of formants is presented first on the four disguise patterns and then compared that with their normal (habitual) voice. After that, an automatic classification was then achieved. The achieved outcomes then give fascinating clues to the investigator for differentiation.

P. Perrot, G. Aversano *et al.* (2007), mainly emphasize on the voice disguise questions and its identification. As the voice disguise is an intentional

act, a wide range of options was available for the speaker to manipulate his/her voice and complicate the detection process. A manipulation of voice can be done by either exploiting intra-speaker variations or by scrambling electronically. The pitch modifications, articulator's position modification which in turn may impact the frequency of formants, were few examples of intra-speaker variation. In their study, they divided the work into three categories: out of which one is classifying various options exists for manipulating someone's voice, the second part consists of evaluation and review of various distinct techniques present in the literature and the third part explains the major indicators present in the literature to differentiate between the original and disguised voice and also proposes few aspects or outlooks that depend on the disturbed or diseased speech as well as emotive speech.

C. Zhang, T. Tan (2007), studied the disguised voice and its effects on the performance of automatic speaker recognition. A recently developed system which was named as "Forensic Automatic Speaker Recognition System (FASRS)" was established and also the impact of various (10 common types that were used in forensic) voice disguises patterns on the performance of "Forensic Automatic Speaker Recognition System" was examined. The 10 different types of disguised voices and 20 other normal voices from 20 male students of the college were studied in this study. 20 normal voice samples from male students were employed as "test samples". Each disguised sample was compared with every normal voice sample that was present in the database to produce speaker verification and recognition. In this study, speaker identification results were summarized and the impact of disguised voice on the system (FASRS) was analyzed.

Tiejun Tan (2010), proposed that the disguised voice impacts the Automatic Speaker Recognition. In their study, they analyzed the effect of 10 types of disguised voices on the functioning of the system named as "Forensic Automatic Speaker Recognition System". They employed 20 normal voices of 20 male students as "test samples" and 10 kinds of disguised speech samples. The comparison of all the disguised speech/voice with the original speech/voice in 2000 speaker's database was performed to identify and verify the speaker. In their paper, they presented the outcomes of speaker identification as well as they also evaluated the impact of disguised voice on the automatic system (FASRS).

Tim Polzehl *et al.* (2011), elaborated on the misuse of acoustic as well as linguistic feature modeling for the classification of anger. They produced statistics from the auditory audio descriptors regarding acoustic modeling such as loudness or characteristics of the spectrum. Regarding linguistic modeling, they applied entropy-based as well as probabilistic models of phrases and words, such as BOW, TF.IDF etc. compared the obtained results based on three distinct databases. Out of this, two databases were obtained from the IVR customer domain and the other one from WoZ data collection. They observed good scores for the IVR databases but the WoZ database exhibits overall lower results.

Pohjalainen J. *et al.* (2013), examined the automatic recognition of anger in the telephonic speech with the help of strong autoregressive modulation filtering. They proposed a new automatic recognition system of vexatious speech. They employed simulation of telephonic speech which was far-end-noise-corrupted and also largely utilized Berlin database of emotional speech, an autoregressive prophecy of the features over the speech frames was displayed to contribute remarkably to clear the speech performance as well as the system's durability. Also, in the feature segregation phase, the linear predictive spectrum analysis method performed better than the standard with regards to durability in the computation of the MFCC.

Sharma V, Bansal P. K (2013), proposed a review on the challenges and approaches in the speaker identification. In their paper, they gave a concise overview of areas of speaker identification, explained its system, several methods of modeling and extraction of features, its applications, techniques, and little information on the performance of the system. They also discussed various robustness and fragility of present technologies of speaker identification. They also summarized probable future trends in the field of research, application, and development.

Li D, Yang Y, Dai W (2014), focused on the cost-effective learning for the emotionally strong speaker recognition. This paper mainly covers the issue of error rate in speech due to several emotions which ultimately exasperate the speaker identification system's performance. They deal with this issue in their study by incorporation the cost-effective learning method to re-evaluate the possibility of test effective speech in a pitch envelope level, which in turn improves the robustness in emotion-based speaker identification respectively. A new design of the identification system and its components based on that technology was proposed in this study. The examination performed on the "Mandarin Affective Speech Corpus" exhibits improvement.

Abin Mathew George, Eva George (2015), investigated several disguising characteristics/factors that were used for the detection of the disguised voice.

In their paper, they studied and investigated those disguising factors that were used commonly in the rising criminal activities. Manipulating or altering their pitch to complicate the speaker identification was the most common factor used by the criminals in the disguised voice. The main emphasis of their study was on the enhancement of the identification of disguised voice which was manipulated using software such as Praat etc. first of all, they separated Mel frequency cepstral coefficient as the acoustic characteristic. After that, they used a classified named PNN (Probabilistic Neural Network) for the distinguishing between normal and disguised voice. They displayed the results of the PNN classifier for the distinguishing feature of -8 till now and as well as they also indicated the results of the Support Vector Machine classifier which was used for several disguising factors.

G. S. Didla and H. Hollien (2015), their study mainly emphasizes on the disguised voice and recognition of the speaker. Their research study was formulated to analyze the probability or chances of resistance of disguise effects by the formants of higher vowels which also indicates the identity of the speaker. According to the results obtained, it was observed that the values of higher formant frequency do not facilitate the speaker's identification based on voice examination that was carried out in this program.

Lini T Lal *et al.* (2015), worked on the identification of disguised voices by using the extraction and the classification of features. They emphasized mainly on the identification of impersonated suspect's voices as it is the most challenging task to determine whether the suspect uses the recordings of legitimate speakers. In this paper, they presented an algorithm for the identification of the manipulated voice. They employed the most commonly used technique for feature extraction was Mel Frequency Cepstral Coefficients (MFCC). Disguising voice may affect the frequency spectrum of a phonation signal by modifying it. They also used a few MFCC-based features to illustrate the properties of frequency spectra. The correlation coefficient and mean value of the system MFCC were used and also the regression coefficients were used as the auditory features. They also used a classifier named "Support Vector Machine" for classifying the disguised as well as normal voices that depend on the segregated features. The precise detection of disguised voice was obtained by using several methods and the algorithm performance was remarkable.

Kurian, Kurup (2016), recognized the disguised voice electronically by the methods of MFCC. In their paper, they tried to indicate the importance of recognition of disguised voice and normal habitual voice. In their study, they proposed an algorithm to identify electronic voice disguise by the method of MFCC. The statistical moments which include correlation coefficients and mean value of the MFCC are separated as acoustic characteristics. The calculations of the formants and pitch were estimated. At last, they used a support vector machine that was based on the algorithm for the extraction of disguised speech from the normal habitual speech. After all of that, they proposed the identification of disguised voices which was based on the separated characteristics, and HMM (Hidden Markov Model) classifiers were also approached.

P. Gangamohan *et al.* (2016), analyzed the speech that contains emotions. In this study, they showed a review of various analysis methods that were used for the analysis of emotional speech. Specifically, they focus on the problems of data collection, representation of features, and also establishment of automatic emotion identification systems. Mainly, they focused on the in-depth examination of the importance of the excitation originated components of speech production. The features of the obtained excitation source were exhibited to show the correlation between emotions.

Discussion

The voice of no two individuals is the same due to physiological dissimilarities. An apprehensive voice is a form of disguised voice and they are of two types: Deliberate and Non- deliberate. Deliberate apprehensive voice or disguised voices can confuse the human ears as well as Automatic Speaker Recognition Systems. It is not necessary that every apprehensive voice is disguised. A deliberate apprehensive voice is being used for illegal purposes. But a non-deliberate apprehensive voice is not necessarily used for illegal use. Deliberate apprehensive voices are most commonly observed during speech recording or when calling someone anonymously. The most common way of altering voices observed is changing their accent at the time of data collection (speech recording during court proceedings) or by changing their pitch. Using deliberate apprehensive voice, the speaker tries to create false voice/audio evidence which ultimately negatively impacts the evidence's authenticity. However, recognition of an apprehensive or disguised voice is unavoidable in the Audio forensic. The recognition of deliberate apprehensive or disguised voice is practiced as an initial step in the task of speaker recognition which helps to differentiate that the testing voice is deliberate apprehensive/disguised, non-deliberate apprehensive, or original. A review of distinct works and studies on some specific apprehensive as well as disguised voices are detailed in this paper. Have gone through the previous related

research such as "Gangamohan, P., Kadiri, S. R., & Yegnanarayana, B. (2016)"; "Zheng T.F & Li. L (Robustness-Related Issues In Speaker Identification)"; "Li D, Yang Y, Dai W (2104)"; "Jouni Pohjalainen and Paavo Alku (2013)"; "Tim Polzeh et al. (2011)" which enlighten a new idea of comparing the deliberate and non-deliberate apprehensive speech. After the analysis and review of various enlisted research papers, the present study represented an overview of the impact of deliberate and non-deliberate apprehensive voice on the speaker recognition. The basic components of their voice remain unchanged such as formant frequency in the "Voiceprints" which helps in the recognition process even after using apprehensive voice if the speaker tries to alter their accent deliberately to confuse the investigator or some parts of their voice gets changed due to illness (such as laryngitis etc.) or due to age. Intonation patterns are the patterns of variation generated by the rise and fall in the pitch of the voice. These intonation patterns are also helpful in speaker recognition in case of a disguised or apprehensive voice.

Conclusion

To conclude, this paper represents the issues of apprehensive voice under distinct aspects. The major contribution of this study in forensic audio or speaker recognition is to precisely differentiate between the deliberate disguised apprehensive speeches from non-deliberate apprehensive speech. This study mainly focuses on the probability of speaker recognition even after the use of the apprehensive voice or if some individual deliberately alters their accent to conceal their identity. This paper represents the possibility of speaker recognition even if the speaker is suffering from any anxiety, belligerence, indignation, fear, any pathological disorder, sad or happy at the time of speech recording (either call recorded or live sample collection). It is concluded that at some point during voice analysis, the deliberate and non-deliberate apprehensive voice (e.g. fear, emotions, etc.) affects some components or parts of voice but the basic components of the speaker's voice remain unchanged due to the individuality of the voice. This paper focuses on the areas which are concerned with the information extraction of an individual's speech such as emotional state, intentional accent change, threat, fear, belligerence, indignation, etc. that are observable in speech signals. There are various researches that show that some external factors such as environmental/background noise, or emotions impact the effectiveness of speaker identification. But, the basic components of their voice remains unchanged such as formant frequency in the "Voiceprints" which helps in the recognition process even after using apprehensive voice, if the speaker tries to alter their accent deliberately to confuse the investigator or some parts of their voice gets changed due to illness (such as laryngitis, etc.) or due to age. Various techniques are available to differentiate between the two and making a precise conclusion/recognition. These challenges or an issues motives future research direction to produce a benchmark result which involves: (1) further studies on the non-deliberate apprehensive voice, (2) employing more emotional and pathological disorder parameters, (3) emphasis on real-world robustness, (4) utilizing furthermore types of disguise and their impact on speaker identification, (5) further study on evaluation of the impact of apprehensive voice on the automatic speaker identification systems.

References:

C. Zhang, T. Tan. **"**Voice Disguise and Automatic Speaker Recognition"**,** *Forensic Science International*, Vol. 175, 2007, pp. 118–122.

G. S. Didla and H. Hollien. "Voice Disguise and Speaker Identification", *Acoustical Society of America, Proceedings of Meetings on Acoustics*, 02-06 November 2015, Vol. 25, doi: 10.1121/2.0000239.

Gangamohan, P., et al. "Analysis of Emotional Speech—A Review." *Toward Robotic Socially Believable Behaving Systems - Volume I Intelligent Systems Reference Library*, 2016, pp. 205–238., doi: 10.1007/978-3-319-31056-5_11.

George, A. M. et al. "Detection of Voice Disguise by Various Disguising Factors", *International Journal of Innovative Research in Computer and Communication Engineering*, Vol. 3, Issue 8, August 2015, doi: 10.15680/IJIRCCE.2015. 0308050.

Künzel, H. J. "Effects of Voice Disguise on Speaking Fundamental Frequency", *International Journal of Speech, Language and the Law*, vol. 7(2), 2000, pp.150-179, doi: https://doi.org/10.1558/sll.2000.7.2.149.

Künzel, Hermann J., et al. "Effect of Voice Disguise on the Performance of a Forensic Automatic Speaker Recognition System." *ODYSSEY04 -- The Speaker and Language Recognition Workshop Toledo, Spain*. 2004

Kurian, S. et.al. "Recognition of Electronic Disguised Voices by the Means of MFCC", *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, Vol. 5, Issue 6, June 2016, doi:10.15662/IJAREEIE.2016.0506014.

Li, Dongdong, et al. "Cost-Sensitive Learning for Emotion Robust Speaker Recognition." *The Scientific World Journal*, vol. 2014, 2014, pp. 1–9., doi:10.1155/2014/628516.

Lini T Lal et al. "Identification of Disguised Voices using Feature Extraction and Classification"**,** *International Journal of Engineering Research and General Science*, Volume 3, Issue 2, Part 2, March-April, 2015 ISSN 2091-2730.

Perrot, Chollet et al**. "**Detection and Recognition of Voice Disguise"**,** *Conference: IAFPA International Association for Forensic Phonetics and Acoustics*, 2007.

Perrot, Patrick, et al. "Voice Disguise and Automatic Detection: Review and Perspectives." *Lecture Notes in Computer Science Progress in Nonlinear Speech Processing*, 2007, pp. 101–117. doi: 10.1007/978-3-540-71505-4_7.

Pohjalainen, Jouni, and Paavo Alku. "Automatic Detection of Anger in Telephone Speech with Robust Autoregressive Modulation Filtering." *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, doi:10.1109/icassp.2013.6639128.

Sharma V, Bansal P.K, "A Review on Speaker Recognition Approaches and Challenges", *International Journal of Engineering Research & Technology (IJERT)*, Vol. 2, Issue 5, May – 2013.

Tiejun Tan. "The Effect of Voice Disguise on Automatic Speaker Recognition", *3rd International Congress on Image and Signal Processing (CISP2010)*, Volume 8, 16-18 October 2010, doi: 10.1109/CISP.2010.5647131.

Tim Polzeh et al., "Anger Recognition in Speech Using Acoustic and Linguistic Cues", *Speech Communication*, Volume 53, Issues 9–10, November–December 2011, Pages 1198-1209.