Acoustic Features Extraction of Non-Electronic Disguised Voice for Speaker Identification

Kumar, Sanjeev
CL Educate Ltd.

Kishore, D.
ECE Department, Aditya College of Engineering & Technology

Kiran, S. V.
ECE Department, Aditya College of Engineering & Technology

https://doi.org/10.5109/4843117

出版情報：Evergreen. 9 (3), pp.853-860, 2022-09. 九州大学グリーンテクノロジー研究教育センター
バージョン：

Acoustic Features Extraction of Non-Electronic Disguised Voice for Speaker Identification

Sanjeev Kumar[1*], D. Kishore[2] and S. V. Kiran[3]
[1]CL Educate Ltd., New Delhi, India
[2, 3]ECE Department, Aditya College of Engineering & Technology, Surampalem, A. P., India

E-mail: sanjeevjuet@gmail.com

(Received March 7, 2022; Revised July 5, 2022; accepted July 7, 2022).

Abstract: A lot of challenges remain despite the rapid advancement of speaker identification technology for non-electronic disguised or physically disguised voices. This is one of the most difficult components of voice disguised that has yet to be overcome by the experts. With a normal or ideal voice, the identification process is simpler and leads to a more conclusive result. A problem arises when a suspect is identified using speech samples that have been altered to hide their identity. There are both unintended and intentional disguises in this sample content. This way a person speaks can be temporarily altered by stress, anger, concern, anxiousness, or sorrow. It is common for people to disguise themselves when they get unidentified, payment or threat phone calls. The speaker changes their voice purposefully fear of being discovered. This manuscript aims to identify a speaker for multimedia applications using a non-electronically disguised voice. In speech signal processing applications requiring non-electronic disguised voice under physical speech fluctuation, it is challenging to identify the speaker. The primary goal of this study was to regulate whether or not it was conceivable to prompt precise sentiments under disguised speech circumstances. Auditory and spectrographic investigations were used to determine the impact of different disguises on personal identity and speaker recognition performances.

Keywords: Voice disguise, MFCC, Speaker identification, Feature extraction

1. Introduction

Forensic speaker identification is the application of science in criminal investigations to tackle issues regarding the identification of unfamiliar speakers. As the name suggests, voice is the characteristic sound produced by the vocal organs of living organisms. It's true that DNA cannot communicate, despite its importance as a tool in criminal investigations. It is possible to use a person's speech as a biometric characteristic since it is well-accepted by people and can be recorded rapidly with low-cost microphones and technologies[1, 2)]. Without the need to memorize a password or combination, or use keys or magnetic cards, it can be an unconventional and more secure way to get entry. A wide range of telephones, mobiles and recorders are readily available in today's society. This makes them a useful instrument for committing crimes such as kidnapping, extortion and threat of extortion[3,5,6)]. Criminals are increasingly exploiting numerous communication channels in the hope of remaining anonymous and unnoticed in modern society. The situation has fortunately changed. When a person's voice is heard, it can be used to identify them and pin the blame on them for the crime[4,7,8)].

To escape detection and confuse investigators, criminals are now able to deliberately alter their speech traits. So, a criminal drape a simple handkerchief over the speaker to alter his sound, for example. This is the biggest obstacle that voice experts around the country face[5,10,14)]. There are many challenges that might arise when a person's speech is disguised in various ways, and this study aims to help doctors examine these tough voice exhibits[6)].

A powerful means of communication that brings people from all over the world together in a number of ways, speech is a powerful mode of communication in its own right. Speech is used by the speaker to transmit his or her message to the audience as he or she formulates it[1,15,21)]. Currently, research in the science of speech communication is concentrated on significant topics including speech recognition, speech coding, speaker recognition, voice synthesis, and speech augmentation, amongst other things. Fig. 1. depicts how forensic departments can distinguish between the speech field and other areas of interest.

In speaker recognition, auditory or perceptual qualities are used to identify a speech sample's speaker. All of these aspects, including the speaker's characteristics, the said phrase, emotions, added noise, channel changes, and others, affect the information content of a spoken statement. Having the same voice mechanism and articulator coordination as the other person is highly rare[7].

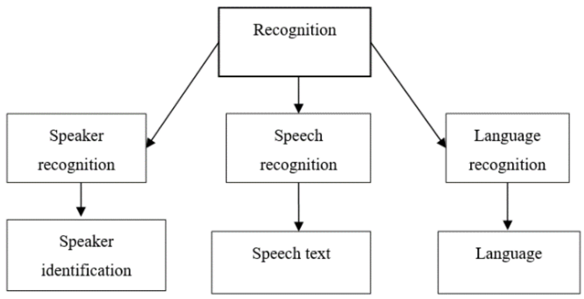

Fig.1: Block diagram speaker recognition system

However, the speech samples taken from the same person differ in certain respects. The reason for this is because a speaker cannot repeat the same phrase over and over again with perfect accuracy. So even a person's signature might vary from case to case. Identification, detection/verification and segmentation/clustering are three unique tasks within the general area of speaker recognition, depending on the application domain [21, 22,23]. To disguise, you can use several techniques, such as a cloth over your mouth, mimicking, adopting a different accent, using external items to alter vocal tract dynamics, moving your lips or tongue to alter formant frequencies, or using electronic gadgets [8,24,25].

The fundamental concept of exchange serves as the foundation for the science of criminal investigation. This principle states that when two objects come into contact with one another, material will be exchanged between them. The same is true for criminals, who are believed to leave all types of evidence no matter where they go or what they do, including DNA, fingerprints and footprints; hair; skin cells; blood; bodily fluids; fragments of clothing; and other fibers; and more, simply by coming into touch with stuff [9]. However, they will take something away from the circumstance as a result of their participation. To maintain their anonymity when there is no immediate crime scene, such as in situations involving blackmail, abduction, extortion and threatening phone calls, such as hoax and harassment calls, harassing calls, match fixing and other similar crimes, criminals make anonymous or ransom calls, as well as hoax and obscene phone calls. In these cases, the voice of an individual is the most important and dependable clue for identifying him or herself [10]. As defined by Merriam-Online Webster's Dictionary: "Any change, distortion, or divergence from the ordinary voice, regardless of the reason." Electronic and non-electronic disguises are divided into two groups by the author [11].

Most non-electronic disguises noticed in casework are deliberate non-electronic disguises, such as conscious adoption of other accents, voice quality alterations, and phonatory modifications such as whispering and falsetto. Non-deliberate disguise types include changes in a person's normal voice as a result of health factors, age factors, emotional variables, and distortions imposed by recording media and transmission channels as a result of phonetic and auditory anomalies, a person's voice can be misconstrued [12].

2. Related Works for Speaker identification System

A speaker's verification and identification are included in speaker recognition (also known as voice recognition). In general, it refers to the process of recognizing someone based purely on their voice [1, 13]. Understanding the language spoken, remembering the speaker's appearance, or using any other type of speaker recognition does not allow you to identify the speaker's identity. This statement is occasionally used when a person is confused whether a technique is one of verification or identification [14]. This task's goal is to determine which of a group of known speakers provided the input speech sample in order to perform speaker identification. There are two modes of operation for identifying voices: closed and open set modes. Assuming closed-set mode, the system believes the voice to be determined must come from a set of voices already known to the system. All other modes of operation are in open-set mode [20]. A multiclass classification problem can be used to describe the speaker identification problem in a closed-set setting.

In open-set mode, speakers who do not belong to the set of known voices are called impostors. Using voice evidence, this work can be used in forensic applications, such as identifying the criminal among several suspects [15].

A speaker identification exercise pitted the skills of phonetic experts against those of unskilled listeners. To perform a direct identification task, participants from both groups were asked which voice belonged to a certain target speaker. These experts performed substantially better than untrained listeners who had never been taught in phonetic speaker identification. According to the voice sample, speaker verification attempts to determine whether or not the individual in question is who they say they are. All of these words are used to describe the task of voice verification or authentication [16]. True or false, it's a binary decision dilemma. This challenge is often referred to as the open-set problem because it involves separating a speaker's voice that is known to the system from a potentially large collection of noises that are unknown to the system. The majority of speaker recognition apps today rely heavily on verification [17, 26, 27, 28].

The creation of a novel algorithm based on the physical variation of all speakers' speech by six methods and the extraction of acoustic features from text-dependent speech signals using the MFCC approach has been finished. Even when a speaker's voice has not been electronically altered, it is still possible to identify them in a multimedia environment. Applications for speech signal processing have a hard time recognizing the speaker in non-electronic disguised voice. The results show that using a disguised voice improved the speaker's simple intellectual recognition performance, which includes better recognition of well-known voices but less accurate matching with new speakers. As a result of this study, it was found that detecting disguised voices from known individuals was 79% accurate, whereas identifying disguised voices from unknown speakers was just 20.7 percent correct.

The speaker-specific information is extracted from the speech stream by the feature extraction module. Semantic, phonologic, phonetic, and acoustic changes occur at multiple stages of speech production to provide speaker-specific information. Speech signals undergo semantic alterations as a result of a speaker's communicative intent and dialogue interaction. Language choice and sentence phrasing, for example, can reveal a speaker's social class and/or educational background. The phonological level deals with the phonetic representation of communication intent [18)]. This information can be derived from the length and selection of phonemes, as well as the intonation of sentences. Vocal cord vibrations and motions of the vocal tract's articulators are considered part of the phonetic level (lips, jaw, tongue, and velum). For example, a speaker can use a distinct mix of articulator movements to create the same phoneme. Spectral aspects of the spoken signal are included in the acoustic level of analysis. The fundamental and resonant frequencies are influenced by factors such as the size of the vocal tract and the length and mass of the vocal folds [19)].

3. Acoustic Feature Analysis and Speaker Identification from Disguised Voice

The study included 200 participants, mostly from India, of varied sexes, faiths, and ages, who were given disguise and control samples. A high-resolution digital recorder was used to capture all of the voice samples. MFCC feature extraction and statistical analysis are used to identify speakers in this approach. In the case of a disguised speech identification system, the MFCC statistical coefficients are affected by a variety of disguised voice methods. The disguise voice samples of each speaker were rigorously recorded under a variety of conditions that caused considerable differences in the acoustic and perceptual features of recorded voice samples. In addition, each participant was given three control samples (normal voice samples) to measure the degree of difference between the disguised and genuine voices. The disguise criteria chosen by the various participants accounted for the following 5 approaches of disguising: Masking on mouth, Changing the accent, Pinching nostrils, Bite block pencil/pen, Variations in the vocal pitch etc.

Many obstacles remain in speech recognition technology despite its enormous advancements. This becomes problematic when disguised voices are met for the purpose of identification. These issues have severely affected the performance of speaker recognition (SR) systems. Someone who tries to mislead the listener by changing their voice poses a security concern. It is dependent on the obscuring element as to how much degradation takes place. A speaker's identity should be known when they make anonymous, ransom, or threatening calls on purpose. Therefore, before entering a voice into an ASR system, we must identify whether it is disguised. Due to this, the identification of disguised voices is the first and most crucial stage of a system of this kind The present voice disguise detection technologies, however, are poor according to several research. The method for speaker identification acoustic feature analysis and speaker identification from disguised voice given in Fig. 2. It was necessary to disguise the voice of a speaker multiple times in order to study the differences between speakers. A disguised speaker can be identified using an auditory characteristic, even while the mechanism of disguise is unclear. Each of the ten speakers was given a text to read. Each participant was obliged to recite 17 times. Three control recitations were also performed in their normal voices, as well as 14 disguised samples recited under the 14 disguise circumstances outlined in this thesis. To determine the degree of acoustic fluctuation within each speaker, they were compared to their respective controls.

Preparation of the files for analysis This means that each device's voice data are stored in its own format We therefore recommend that files in an incompatible format be changed to a suitable format: Sampling rate: 16KHz, File Format: .wav with the help of 'Audacity Software', Channel: Mono, Bit depth: 16 bits, Bit rate: 172 Kbps etc.

3.1 Feature Extraction: LPCC and MFCC

Linear predictive cepstral coefficient (LPCC) work well only in a noise-free environment, spectral analysis shows a huge spectral distortion in noisy conditions. The methods and computations required to calculate each of the attributes given above are performed by the module.

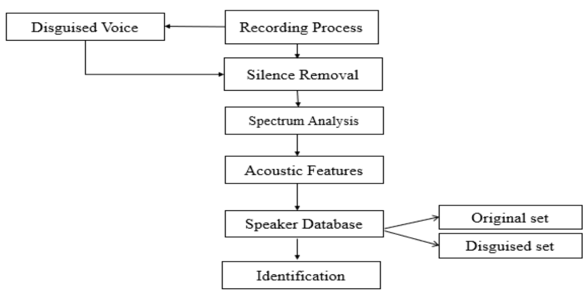

Fig.2: Speaker's acoustic features analysis method and identification from disguised voices

The features that are retrieved are either multidimensional or have a single dimension. A feature vector including all of these features is extracted for each frame. It improves the storage and processing efficiency of voice signals. There are a variety of feature extraction methods available, including LPCC shown in Fig. 3:

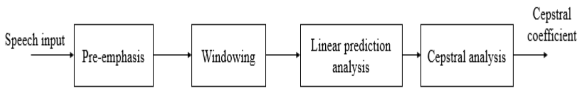

Fig.3: Feature extraction using LPCC

MFCCs are used to represent short-term speech properties. It's commonly used in the study of acoustic signals. The scheme of the MFCCs is well-known and well-accepted. It describes the link between the crucial bandwidth of the human ear and the mel frequency scale, which has linear filters below 1000 Hz and logarithmic filters above 1000 Hz.

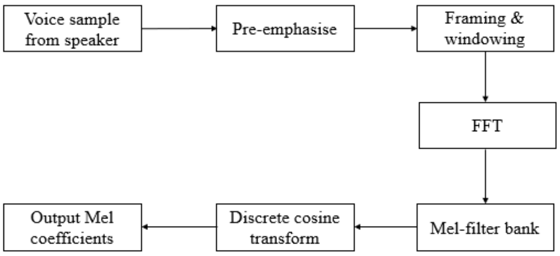

Fig.4: Block diagram of statistical analysis by MFCC algorithm

To remove gaps, the signal is first segmented into frames, and then each frame is multiplied by a hamming window. The discrete Fourier transform of each frame is then calculated, and the log of the amplitude spectrum is measured. Finally, the discrete cosine transform (DCT) is used to smooth the speech spectra, which facilitates in the production of cepstral feature vectors for each frame. The steps necessary for the creation of MFCCs are as follows. A block diagram representation of statistical analysis using MFCC feature extraction method shown in Fig. 4.

These filters are expressed in a linear manner using the Mel or Bark scale. The Bark scale is a vital band rate scale, whereas the Mel scale is a perceptual frequency scale, as indicated in the formulae below:

$$Bark = 13\, atan\left(\frac{0.76f}{1000}\right) + 3.5\, atan\left(\frac{f^2}{(7500)^2}\right) \quad (1)$$

$$melfreq = 2595\, log\, 10(1 + \frac{f}{700}) \quad (2)$$

The bandwidths chosen for the centre frequency should be equivalent to a crucial bandwidth. The following is a mathematical representation of critical bandwidth:

$$Critical\ BW = 25 + 75\left[1 + 1.4\frac{f}{(1000)^2}\right].^{0.69} \quad (3)$$

The power vector of each speech frame is first obtained, and then merged with other components to form a signal observation vector. The DCT of the log magnitude of the Mel scale filter bank outputs is then performed for each frame to obtain MFCC.

$$mfcc_{n^{\wedge}}[m] = \frac{1}{R}\sum_{r=1}^{R} log\left(MFn^{\wedge}[r]\right) cos\left[\frac{2\pi}{R}\left(r + \frac{1}{2}\right)m\right] \quad (4)$$

The problem of pattern recognition in speaker recognition has captivated people's interest for decades. Pattern recognition categorizes items of interest into one of several groups. In speaker recognition systems, each class represents a separate speaker. The MFCC approach, which converts a voice signal into a series of vectors, is thoroughly detailed in the preceding topic. Following this phase, you should build a speaker model that will be utilized for speaker categorization jobs. Support vector machine, decision tree, naive bayes, and linear discriminant analysis are some of the classifiers used to make conclusions regarding the relationship between the input test voice and the speaker(s) voice contained in the model.

4. Experimental Results for Acoustic Features and Identification

It's easy to hear someone's voice, just as it's easy to see their face. After listening to samples several times, specialists make a judgement about how similar or unlike the two speech events are. In a noisy setting, human ears are capable of distinguishing between different speakers. Other elements affecting the listener's performance include the signal to noise ratio and speech bandwidth, as well as coding and transmission systems. 20 terms were retrieved from both disguised and normal speech samples after an auditory comparison of each speaker's disguised and regular speech samples. Collection of speakers voice using 'Audacity software' in figure 5 and disguised voice in Fig. 6.

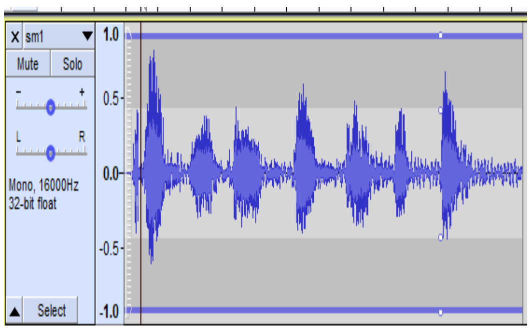

Fig.5: Voice sample collection of speakers using Audacity software (Snapshot of normal voice

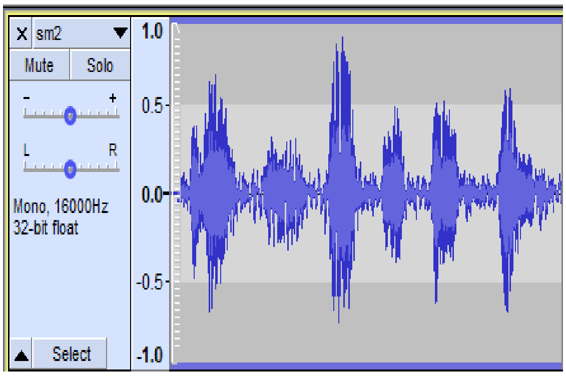

Fig.6: Voice sample collection of speakers using Audacity software (Snapshot of disguised voice).

For speaker recognition, a spectrographic analysis converts audio impulses into visual images. In order to turn sounds into images, an electromechanical acoustic spectrograph was created. During the course of a voice wave, it records the changing energy frequency distribution. Voiceprint identification uses spectrographic impressions of people's utterances to identify them, much like fingerprint identification. Determining questionable callers is another service they provide to law enforcement agencies A foolproof means of identification, it was thought. The "voiceprint" technique to voice identification has been in legal limbo. Fig. 7 shows spectrogram formant patterns in normal voice. Fig. 8 shows spectrogram formant patterns of disguise voice, pitch patterns, loudness, transitional features, and bandwidth, which can be used to determine the resemblance between two samples when an examiner is properly taught.

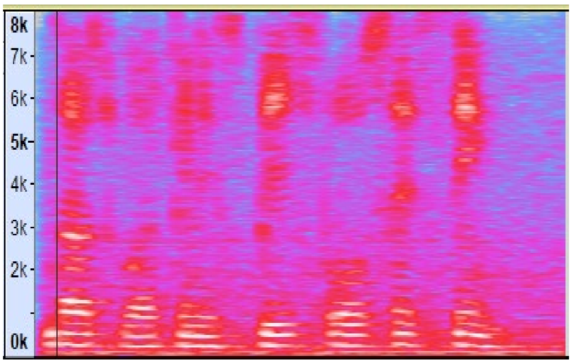

Fig.7: Spectrogram formant patterns of normal voice.

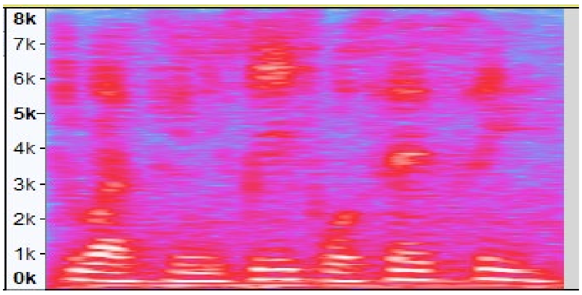

Fig.8: Spectrogram formant patterns of disguised voice.

Divide the screen into six windows that are arranged horizontally on the screen. To compare two words, select one from the disguised speech of a person on one side and the other from a control sample of that person on the other. Compare the spectrograms and format patterns of the two samples in terms of frequency, pitch, amplitude, and energy.

It was required that the participants modify their normal voice in some way in order to submit a voice sample. Due to the fact that there were 200 different subjects, the following disguise strategies were used: 5 approaches of disguising: Masking on mouth (32%), Changing the accent (16%), Pinching nostrils (22%), Bite block pencil/pen (12%), Variations in the vocal pitch (18%) shown in table 1.

Table 1. Speakers taken for variation in their voice

Disguised method	Spaeker's voice variation (%)
Masking on mouth	32
Changing the accent	16
Pinching nostrils	22
Bite block pencil/pen	12
Variations in the vocal pitch	18

To conceal oneself, the majority of respondents used handkerchiefs and hands to cover their mouths and then changed pitch levels by increasing or lowering normal values, and so on. The voice variation is shown in Fig. 8.

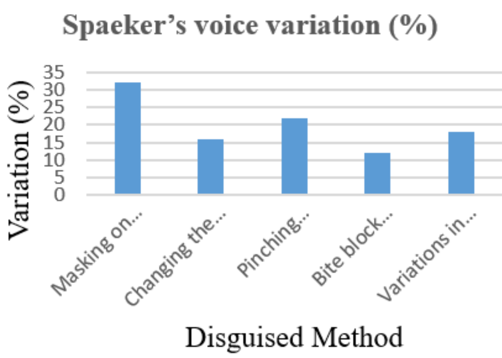

Fig.9: Variation in speakers voice variation.

The voice quality of an individual is determined by a variety of laryngeal and subpharyngeal features present throughout their speech. The majority (61%) of the total disguised speech samples obtained from 200 individuals, including males and females, had substandard or deteriorated speech quality, compared to just 12% of the control voice samples. This is because voice disguise disrupts the natural flow of discourse. The following pictures illustrate the MFCC feature vectors for each frame. Each column in this table corresponds to a feature

Table 2. Correlation coefficient in spectrographic parameters of disguised and normal voice samples of 10 individual speakers.

No. of Spea-kers	Normal Voice	Masking on mouth	Changing the accent	Pinching nostrils	Bite block pencil/pen	Variations in the vocal pitch
1	0.458	0.785	0.589	0.534	0568	0.123
2	0.253	0.236	0.689	0.458	0.254	0.853
3	0.257	0.568	0.453	0.254	0.652	0.569
4	0.568	0.784	0.324	0.698	0.368	0.365
5	0.236	0.236	0.754	0.236	0.125	0.145
6	0.145	0.451	0.127	0.784	0.256	0.254
7	0.125	0.364	0.698	0.457	0.365	0.563
8	0.546	0.245	0.745	0.659	0.458	0.892
9	0.258	0.354	0.235	0.521	0.698	0.258
10	0.458	0.365	0.698	0.890	0.780	0.658

vector, and the columns' components correspond to the MFCCs that follow. In this example, the top 12 DCT coefficients are chosen, resulting in a total of 12 entries in each column. The figures 10 and 11 illustrate the MFCC's two-dimensional plot, which displays the values of both the original and disguised speech of a single speaker inside the same frame. This method of calculation yields the mean speaker recognition rate. In the suggested method, the statistical coefficient mean and correlation coefficient are computed using MFFC methods. A classifier method is developed using the statistical mean and correlation coefficients of a normal voice characteristic and a specific kind of masked speech.

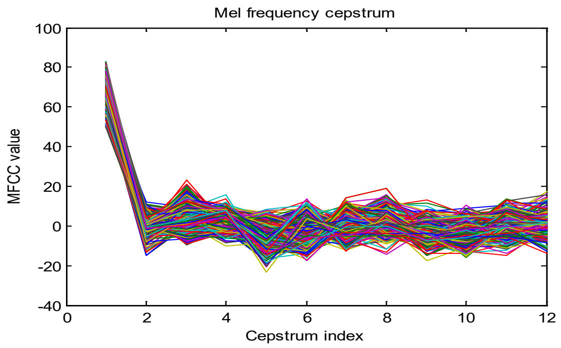

Fig.10: Plot of MFCC values of original signal

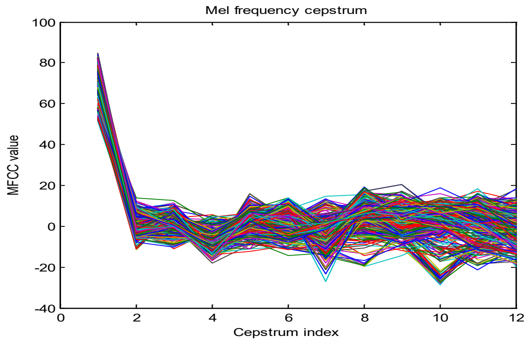

Fig.11: Plot of MFCC values of disguised voices

Speaker spectrographic parameters are in correlation with all disguised and control samples. Each of the 10 distinct speakers in the disguised sample exhibited a high connection with their control counterparts. F3 and F4 values of the disguised and control samples were highly correlated in just 4 speakers out of 10. Table 2 shows that the disguised voices of all 10 speakers showed a modest connection in F0, F1, F2, and F5. The degrees of nasality, dynamic loudness, and pauses in the disguised samples from each of the 10 speakers and their control counterparts were shown to be significantly related. Between the disguised and control samples of all 10 speakers, there was a moderate relationship (see table 2). For the identification of disguised speech samples, spectrographic characteristics proved to be more reliable than auditory criteria. A person's identity could be disguised successfully, but deceit was more likely to be detected. It is common for speakers to alter their speaking tempo, pitch and accent throughout the masking process. In most disguised scenarios, the impersonator's voice pitch rises significantly due to the increased stress level (unless when the speaker consciously reduces the pitch levels). Higher pitch values in disguised speech appear to be the result of greater muscular tension in the vocal organs. It's no secret that pitch and formant frequencies are interconnected. Larynx shrinkage and vocal tract lengthening are caused by low pitch, which results in a reduction in formant frequencies. Elevating the larynx lowers vocal tract size and enhances formant frequencies when the pitch is raised. Comparing masked voices to their control counterparts, the third formant and fourth formant measures showed reduced variances.

5. Conclusion

Uncertainty surrounds the method of disguise; nonetheless, the 10-participant experiment that evaluated intraspeaker changes shed light on the type of acoustic characteristic that may be useful for detecting disguised speakers. To assess the validity and reliability of various auditory cues for speaker identification, several techniques of disguising voice were examined. All ten speakers were properly recognized after accounting for nasality, pauses, and intensity levels. When used in a variety of situations, these qualities are 100 percent accurate. Depending on the disguise method, the validity of each aural signal varies slightly. Variations in auditory features observed under different hidden settings reveal inconsistency in a person's speech. The disguise is maintained by using several phoneme modifications. Based on variables such as nasality, pauses, and energy levels, a voice specialist can make the best judgement regarding any unknown disguise technique for a given speaker. The main assumption of this technique is that it is able to distinguish between genuine audio and voices that have been disguised using a non-electronic method. Regular speech databases and ten various kinds of non-electronic masking techniques are employed for both training and testing. The performance analysis of all the classifiers is still good despite the significantly higher voice identification rates of the disguise approach. The performance of non-electronic disguising methods for a variety of disguising approaches and their evaluation will be the subject of future research. This conclusion is supported by results. In addition to understanding the speaker's concealing approach, several features may be crucial. The divergence and convergence of all of these additional metrics have changed in different disguised scenarios.

References

1) Singh, Mahesh K., A. K. Singh, and Narendra Singh. "Multimedia analysis for disguised voice and classification efficiency." *Multimedia Tools and Applications,* **78** (20)29395-29411 (2019). doi: 10.1007/s11042-018-6718-6
2) Wu, Haojun, Yong Wang, and Jiwu Huang. "Identification of electronic disguised voices." *IEEE Transactions on Information Forensics and Security,* **9** (3) 489-500 (2014). doi: 10.1109/TIFS. 2014.2301912
3) Mahyudin, Alimin, et al. "Mechanical Properties and Biodegradability of Areca Nut Fiber-reinforced Polymer Blend Composites." *EVERGREEN Joint Journal of Novel Carbon Resource Sciences & Green Asia Strategy,* **7** (3) 366-372 (2020). doi: 2324/4068618
4) Weake, Nawaharsh, et al. "Optimising Parameters of Fused Filament Fabrication Process to Achieve Optimum Tensile Strength Using Artificial Neural Network." *EVERGREEN Joint Journal of Novel Carbon Resource Sciences & Green Asia Strategy,* **7** (3) 373-381 (2020). doi: 2324/4068614
5) Clark, Jessica, and Paul Foulkes. "Identification of voices in disguised speech." *International Journal of Speech, Language and the Law,* **14** 195-221 (2007). doi: 10.1558/ijsll.v14i2.195
6) Singh, Mahesh K., Narendra Singh, and A. K. Singh. "Speaker's voice characteristics and similarity measurement using Euclidean distances." In *2019 International Conference on Signal Processing and Communication (ICSC)*, 317-322 (2019). doi: 10.1109/ICSC45622.2019.8938366
7) Perrot, Patrick, and Gerard Chollet. "The question of disguised voice." *Journal of the Acoustical Society of America,* **123** (5) 3878-84 (2008). doi: 10.1121/1.2935782
8) Markowitz, Judith. "The many roles of speaker classification in speaker verification and identification." In *Speaker Classification I*, Springer, Berlin, Heidelberg, 218-225 (2007). doi: 10.1007/978-3-540-74200-5_12
9) Singh, Mahesh K., A. K. Singh, and Narendra Singh. "Multimedia utilization of non-computerized disguised voice and acoustic similarity measurement." *Multimedia Tools and Applications,* **79** (47) 35537-35552 (2020). doi: 10.1007/s11042-019-08329-y
10) Masthoff, Herbert. "A report on a voice disguise experiment." *The International Journal of Speech, Language and the Law,* **3** (1) 160-167 (1996). doi: 10.1558/ijsll.v3i1.160
11) Anushka, R. L., Jagadish, S., Satyanarayana, V., & Singh, M. K. (2021, October). Lens less Cameras for Face Detection and Verification. In 2021 6th International Conference on Signal Processing, Computing and Control (ISPCC) (pp. 242-246). IEEE. doi: 10.1109/ISPCC53510.2021.9609392
12) Nandini, A., Kumar, R. A., & Singh, M. K. (2021, October). Circuits Based on the Memristor for Fundamental Operations. In 2021 6th International Conference on Signal Processing, Computing and Control (ISPCC) (pp. 251-255). IEEE. doi: 10.1109/ISPCC53510.2021.9609439
13) Sawant, Swapnil, et al. "System Modelling of an Electric Two-Wheeled Vehicle for Energy Management Optimization Study." *EVERGREEN Joint Journal of Novel Carbon Resource Sciences & Green Asia Strategy,* 8 (3) 642-650 (2021). doi: 10.5109/4491656
14) Singh, Mahesh, Durgesh Nandan, and Sanjeev Kumar. "Statistical Analysis of Lower and Raised Pitch Voice Signal and Its Efficiency Calculation." *Traitement du Signal,* **36** (5) 455-461 (2019). doi: 10.18280/ts.360511
15) Zheng, Linlin, Jiakang Li, Meng Sun, Xiongwei Zhang, and Thomas Fang Zheng. "When Automatic Voice Disguise Meets Automatic Speaker Verification." *IEEE Transactions on Information Forensics and Security,* **16** 824-837 (2020). doi: 10.1109/TIFS.2020.3023818
16) Satya, P. M., Jagadish, S., Satyanarayana, V., & Singh, M. K. (2021, October). Stripe Noise Removal from Remote Sensing Images. In 2021 6th International Conference on Signal Processing, Computing and Control (ISPCC) (pp. 233-236). IEEE.

doi: 10.1109/ISPCC53510.2021.9609457

17) Nugraha, Gde Dharma, Budi Sudiarto, and R. Kalamullah. "Machine learning-based energy management system for prosumer." *Evergreen Joint Journal of Novel Carbon Resource Sciences & Green Asia Strategy,* **7** (2) 309-313 (2020). doi: 10.5109/4055238
18) Doherty, E. Thomas, and Harry Hollien. "Multiple-factor speaker identification of normal and distorted speech." *Journal of Phonetics,* **6** (1) 1-8 (1978). doi: 10.1016/s0095-4470(19)31081-2
19) Prasanna, GV Sai, K. Pavani, and Mahesh Kumar Singh. "Spliced images detection by using Viola-Jones algorithms method." *Materials Today: Proceedings* (2021). doi: 10.1016/ j.matpr.2021.06.300
20) Prasad, Swati, Zheng-Hua Tan, and Ramjee Prasad. "Multi-frame rate based multiple-model training for robust speaker identification of disguised voice." *16th international symposium on wireless personal multimedia communications (WPMC)*, 1-4 (2013).
21) Sebastian, Kesiya, and Leena Mary. "FASR: Effect of voice disguise." *International Conference on Emerging Technological Trends (ICETT)*, 1-4 (2016). doi: 10.1109/ICETT.2016.7873712
22) Matsuda, Mari J. "Voices of America: Accent, antidiscrimination law, and a jurisprudence for the last reconstruction." *Yale Law Journal*, 1329-1407 (1991). doi: 10.2307/796694
23) Tirumala, Sreenivas Sremath, Seyed Reza Shahamiri, Abhimanyu Singh Garhwal, and Ruili Wang. "Speaker identification features extraction methods: A systematic review." *Expert Systems with Applications,* **90** 250-271 (2017). doi: 10.1016/j.eswa.2017.08.015
24) Lu, Hong, AJ Bernheim Brush, Bodhi Priyantha, Amy K. Karlson, and Jie Liu. "Speakersense: Energy efficient unobtrusive speaker identification on mobile phones." In *International conference on pervasive computing*, Springer, Berlin, Heidelberg, 188-205 (2011). doi: 10.1007/978-3-642-21726-5_12.
25) Gupta, P., & Singh, B. (2021). Exploration of tool chatter in CNC turning using a new ensemble approach. *Materials Today: Proceedings*, *43*, 640-645. doi: https://doi.org/10.1016/j.matpr.2020.11.1005.
26) Gupta, P., & Singh, B. (2022). Ensembled local mean decomposition and genetic algorithm approach to investigate tool chatter features at higher metal removal rate. *Journal of Vibration and Control*, *28*(1-2), 30-44. https://doi.org/10.1177/1077546320971157
27) Gupta, P., & Singh, B. (2020). Local mean decomposition and artificial neural network approach to mitigate tool chatter and improve material removal rate in turning operation. *Applied Soft Computing*, *96*, 106714. doi: https://doi.org/10.1016/j.asoc.2020.106714.