International Journal of English Literature and Social Sciences, 5(3)
May-Jun 2020 |Available online: https://ijels.com/

A Review of Voice Disguise in a Forensic Phonetic Context

Dr. Grace Suneetha Didla

Assistant Professor, The English and Foreign Languages University, Hyderabad, India

Abstract— *Voice disguise entails altering one's voice deliberately with an intent to conceal one's identity. Voice disguise is employed predominantly for two reasons: 1) disguising for the purpose of entertainment and 2) disguising with a criminal intent of concealing one's identity. The thrust of this review is to better understand the second of these; i.e., its relationship to forensics. Voice disguise in the context of crime is usually employed by criminals in cases such as kidnapping, hoax calls, threatening calls etc. Owing to its formidable presence in the world of crime, it is imperative that numerous issues plaguing voice disguise must be addressed systematically. This paper presents an overview of what entails a disguise, its prevalence and the types of disguises usually employed by the criminals. It also reviews the research associated with the phonetic aspects that potentially contribute in the decoding of voice disguise.*

Keywords— ***Forensic Phonetics, Speaker Identification, Voice Disguise.***

I. INTRODUCTION

Forensic Phonetics is a fledgling discipline in the domain of Forensic sciences. It is the application of the knowledge of Phonetics to help solve crimes based on speech or voice. One of the key aspects of Forensic Phonetics is speaker identification. Recognizing unseen voices (probably over a telephone) of friends and family members is an everyday experience for many of us. Hollien relates this common experience to be a probable reason for the birth of the concept of 'speaker identification' [1].

The interest in speaker identification appears to be on the rise, primarily owing to its procedural applications in the forensic sciences. Forensic speaker identification (FSI) usually involves comparing the incriminating voice with one or more suspects' voices to determine if they are produced by the same speaker or not. While FSI can be carried out by lay listeners (ear-witnesses in a crime) or by expert listeners (people who are phonetically and acoustically trained), even under favorable circumstances, it can be quite demanding as many factors affect the identification process. A rather more daunting task faced by the forensic phoneticians is the identification of speech marked by voice disguise.

Definitions of voice disguise abound in the literature. While Nolan defines disguise as an "exploitation of the plasticity of the vocal tract for a very specific communicative effect" [2], Rodman describes voice disguise as "any alteration, distortion or deviation from the normal voice, irrespective of the cause" [3]. On the other hand, Künzel agrees that disguise is "a voluntary change of features of voice, speech and language, produced by a speaker in order to conceal his identity" [4].

There seem to be several reasons why people disguise their voices. Nevertheless, these can be narrowed down to primarily two: 1) disguising for entertainment purpose (as in mimicry) and 2) disguising with a criminal intent of concealing one's identity. The thrust of this review is to better understand the second of these; i.e., its relationship to forensics. Disguised speech is most often associated with crimes such as kidnapping, threats, extortion, hoaxes, and related. That is, such attempts typically occur when the criminal thinks that his or her 'identity' must be protected [5] or especially if he knows he is being recorded. Therefore, "keeping his identity covered is to the advantage of the speaker in question." [6].

Relevant studies carried out in prior years [3], [4] & [6] confirm that there are several ways by which the voice may be disguised. Among the more frequent are the changes or modifications involved in:

1.1. **Voice source**: falsetto, raised pitch, lowered pitch, vocal fry, and whispered speech,

1.2. **Resonance**: Hypo- or hyper-nasality, foreign objects placed in mouth, clenched in the teeth or other modifications of the vocal tract,

1.3. **Language**: varied dialect, foreign accent, and

1.4. **Manner of speaking**: variation of tempo, stress patterns, monotonous voice production.

If speakers are competent in disguising their voices, the effort can be markedly detrimental to effective speaker identification [1]. That is, the examiner's efforts can be frustrated by distortion of the talker's speaker-dependent features.

Kunzel reports that "over the last two decades, between 15 and 25 percent of the annual cases dealt with at the German Federal Police Office speaker identification section, exhibited at least one kind of disguise: falsetto, pertinent creaky voice, whispering, faking a foreign accent and pinching one's nose being the perpetrators' favorites" [4]. At JP French Associates (a leading forensic speech and acoustics laboratory in the UK), "it is estimated that the one in forty cases of speaker identification involves some form of disguise" [7].

Voice disguise, undoubtedly, has a considerable detrimental effect on speaker identification. Given the frequency of adoption of voice disguise in the world of crime and its crippling effects on speaker identification, this paper aims to give an overview of the relevant studies which have contributed towards better understanding of the issue of voice disguise.

II. REVIEW OF LITERATURE

Voice disguise is a relatively new domain in the forensic phonetic milieu. So far, a few researchers have reported on the different modes of voice disguise employed, their effects and also their influence on FSI. In forensic speaker identification context, human voices can be disguised by means of human impersonation and electronic voice conversion. Going by Rodman's [3] classification, these may be termed as 'deliberate non-electronic' and 'deliberate electronic' voice disguises respectively. Over the years, different researchers have adopted different methods in the analysis of voice disguise. These methods can be identified under four headings: 1) spectrographic speaker identification, 2) aural-perceptual speaker identification and 3) acoustic speaker identification and 4) automatic speaker identification.

2.1. **Spectrographic Speaker Identification**: Spectrographic speaker identification popularly referred to as 'Voice print analysis' refers to the visual examination of spectrograms of the questioned sample and the suspect's sample to observe similarities/dissimilarities in patterns. Voice print analysis had its genesis during the world war 11. In the aftermath of war, there was a period of silence/no progress in the said domain. The interest in voice prints yet again resumed in the early 1960s and continued its presence in the following two decades. Its proponents made their way into the courts justifying the validity of this approach. For a long time, this method prevailed for want of any opposition and also on the insistence of the proponents that it is a scientific and valid approach with a negligible error rate.

In one of the earliest publications, Kersta [8] argued that voiceprint is as unique as fingerprints and that speakers can be identified safely through voice print analysis. In his study he compared disguised speech samples with undisguised and claimed that the speech spectrograms were unaffected by attempts of voice disguise. Unfortunately, no details were provided as to how he carried out his analysis. In any case, he claimed that the process was a reliable one with very low error rates. This caught the interest of the law enforcement authorities and 'voice print analysis' became a new buzz word to capture the perpetrators. Not long after, this claim was refuted by Endres, Bambach, and Flosser [9]. Through their investigation they established that speech spectrograms of utterances spoken in normal and disguised voice (changes in F0, rate of articulation pronunciation and dialect) reveal strong variations in formant structure.

Subsequently, Reich, Moll, and Curtis [10] have investigated the effects of selected disguises upon spectrographic speaker identification. In a matching task that was carried out, they found that identification of disguised speech samples posed a greater challenge and had a significant effect on the types of errors made by the examiners.

In the early 1980s, however, the reliability of spectrographic SI (commonly referred to as 'voice print analysis') was severely questioned by the then scientific community and it did not stand the test of time and eventually gave way to other approaches.

2.2. **Aural-Perceptual Speaker Identification**: Several experiments in the following decades have been dedicated to the aural-perceptual identification of disguised speech by listeners. This method predominantly focused on how a listener (whether naïve or expert) perceives speech. Naïve speaker identification gains significance in the real world of crime, because, more often than not, it is a naïve person who is either a witness or a victim of a crime. Therefore, it is very important to understand how speech is perceived by a lay listener. Subsequent research focused on recognizing the challenges involved in identifying a voice disguise, the hierarchy in the types of disguise in terms of their difficulty in recognition and the nature of specific types of disguise.

Indeed a review of relevant studies suggests that disguised voices may be much more difficult to identify than if they are not disguised. In this regard, Reich and Duke [11] have employed aural-perceptual techniques to investigate the effects upon SI of selected voice disguise (hyper-nasal, slow-rate, hoarse voice, speaking like an elderly individual and free disguise). They further investigated to see if certain disguises had markedly interfered with SI than the others. Their experiment consisted of 360 discriminations of paired samples presented in a fixed sequence mode. Two listener groups (‘naïve’ undergraduate students and ‘sophisticated’ doctoral students and professors) were trained for the task. The listeners were asked to decide whether the paired sentences were uttered by the same speaker or two different speakers. The results obtained from both the groups indicated that speaker recognition rates fell from 92% correct identification for undisguised voices to 59-81% (depending upon the disguise) for those that were disguised.

Another useful study which underlined the difficulty in recognizing disguised voices was carried out by Hollien, Majewski, and Doherty [12]. They have worked on the identification of voices perceptually under three speaking conditions: normal, stress and disguise. This experiment investigated to estimate the listeners’ capabilities in identifying the voices and assess how familiarity of the talker’s voice impacted the auditors. The experiment included three groups of listeners: a) who were familiar with the talkers, b) who were unfamiliar with the talkers but were trained to identify them and c) who were unfamiliar with both the talkers and the language. It was reported here that the disguised condition resulted in markedly lower identification rates for all the three groups compared with the undisguised condition. (Group A: 98 to 79%, Group B: 40 to 21%, Group C: 27 to 18%).

In addition to identifying the most common disguises employed, the question of whether the type of disguise has any influence on speaker identification is also of significance. In this regard, a significant study [11] addressed this issue and confirmed that certain disguises (hyper-nasal and free disguise) were more effective than others (slow-rate, hoarse voice and speaking like an elderly individual).

Experiments focusing on the effect of a particular voice disguise have also been carried out [13], [14]. While the former study assessed the nature of creak (vocal fry) and its effectiveness as a voice disguise, the latter focused on falsetto as a form of phonation. In the experiment on creaky voice, results have shown that phonetically trained listeners were able to match speakers with 90% accuracy for the undisguised condition as compared with 65% for the disguised voices. On the other hand, an SI experiment carried out on identification of falsetto disguise (by familiar listeners) showed significantly poor results (4% match) compared to the normal voice (97% match). These results clearly show that falsetto can be an effective disguise. Nevertheless, this result should be validated by other research.

2.3. **Acoustic speaker Identification:** While it is important to understand how the listeners (both naïve and expert) perceive disguise, it is also equally important to gain knowledge on the acoustic characteristics of a disguise. In this regard, Neuhauser [15] examined how well native German speakers could produce a foreign accent (French) and described the accent’s main and consistent features. The results of auditory, acoustic and linguistic (non-phonetic) analyses have shown that speakers were able to use several forms of variations (articulatory and pitch) during voice disguise by using a foreign accent. The varied features partially matched with those which would be expected from French natives speaking German, but speakers were generally unable to perform consistently.

Also of relevance to this review is research on the most common types of disguises employed by speakers and their acoustic features. Masthoff [6] carried out a study which provided insight relative to this issue. The goals of the study were to identify the disguise preferred by the speakers and to observe the similarities/ differences between the modal voice and the disguised voice chosen. His experiment employed 20 students disguising their voices with no restrictions. They also were allowed to use multiple disguises. The resulting data showed that the plurality of disguises involved an alteration of phonation (35%). Furthermore, single disguises (55%) outnumbered the multiple disguises (45%). It is interesting to note that raised pitch was used only by males and lowered pitch only by females.

In yet another experiment, Kunzel [4] investigated the effects of voice disguise on speaking fundamental frequency. The results indicate that the speakers were adept at consistently changing their F0 in accordance with the selected disguise. Results corroborated that there is an underlying relation between the F0 of a speaker’s natural voice and the choice of disguised voice one would employ in an incriminating phone call.

It is established in the literature that the higher formant frequencies provide speaker-specific cues. Earlier research by Stevens wherein he explores the sources of inter- and intra- speaker variability in the acoustic properties of speech sounds, states that mean F3 is a good indicator of a speaker's vocal tract length [16]. The same view has been echoed by Baldwin and French that "most of the significant information about voice quality is carried by the third and fourth formants" [17] These data imply that the higher formants provide naturalness to voice quality and suggest that they assist in identifying speakers. Didla and Hollien [18] carried out an experiment to test if the higher formants frequencies (of certain vowels) are affected by the use of voice disguise and thereby if they might be useful in identifying speakers. Four sets of speech samples were obtained from each speaker (normal voice, low pitch, falsetto and disguise sample created by clenching a pencil between the subject's teeth while pinching the nose). They concluded that the higher formant frequency values are not reliable measures in the speaker identification process.

2.4. **Automatic Speaker Identification:** In the recent years, owing to the increased use of technology, there have been attempts to understand the effectiveness of the automatic speaker recognition systems in identifying the disguised voices.

Zhang and Tan [19] introduced a newly developed Forensic automatic speaker recognition system (FASRS). To study the effectiveness of this system in identifying disguised voices, an experiment was set up with 10 types of disguises. A speaker recognition task was carried out which involved comparison of each disguised voice with all the normal voices in the data base. The result of speaker recognition was summarized and the influence of voice disguises on the FASRS was evaluated.

In a more recent experiment, Farrus et al. [20] analyzed a) the prosodic features employed by professional impersonators when mimicking a voice and 2) intra- and cross-gendered converted voices in a spectral-based speaker recognition system. The results indicated that when imitated and converted voices were used, the identification error rate increased, especially the cross-gender conversions.

Very few studies have been carried out in the direction of identifying the influence of the use of a particular language by a bilingual speaker on the identification of voice disguise. A very recent study by Kunzel [21] tried to explore the efficacy of automatic speaker recognition with cross-language speech material. In the same article he stated "For obvious reasons, the impact of the cross-language problem on these systems remains undisclosed, but neither has it received much attention in published research on auditory or automatic speaker recognition. This is all the more surprising since probably the majority of countries has become, or has always been, multi-ethnic and/or multi-lingual."

Research on voice disguise has met with little success in the area of Forensic speaker identification (FSI) owing to the infinite ways humans can disguise their voices. The studies carried out thus far are limited in their scope as 'voice disguise' is influenced by a number of variables such as type of disguise, familiarity with the speaker, language, dialect etc. Given its formidable presence in the world of crime, the numerous issues plaguing voice disguise must be addressed systematically to achieve the desired results.

REFERENCES

[1] Hollien, Harry. (2002). Forensic Voice Identification. London: Academic press.

[2] Nolan, Francis. (1983). The Phonetic Bases of Speaker Recognition. Cambridge: Cambridge University press.

[3] Rodman, Robert D. (1998). "Speaker recognition of disguised voices: a program for research." Proceedings of the Consortium of Speech Technology Conference on Speaker recognition by Man and Machine: directions for forensic application: 9-22.

[4] Künzel, Hermann J. (2000). "Effects of Voice Disguise on Speaking Fundamental Frequency." Forensic Linguistics 7.2:149-79.

[5] Hollien, Harry. (2014). "Forensic Phonetics: an introduction." Forensic Linguistics. 3 ed. John Olsson, John and June Luchjenbroers. London, New Delhi, New York, Sydney: Bloomsbury: 83-136.

[6] Masthoff, Herbert. (1996). "A report on a voice disguise experiment." Forensic Linguistics 3: 160-67. Print.

[7] Clark, Jessica and Paul Foulkes. (2007). "Identification of voices in electronically disguised speech." IJSLL 14.2: 195-221.

[8] Kersta, Lawrence G. (1962). "Voiceprint Identification." Nature 196: 1253-1257.

[9] Endress, W., W. Bambach, and G. Flosser. (1971). "Voice spectrograms as a function of age, Voice Disguise and Voice Imitation." JASA 49.6: 184-1848.

[10] Reich Alan, et al. (1976). "Effects of selected vocal disguises upon spectrographic speaker identification." The Journal of the Acoustical Society of America 6.4: 919–925.

[11] Reich, Alan R. and James E. Duke. (1979). "Effects of selected vocal disguises upon speaker identification by listening." The Journal of the Acoustical Society of America 66.4: 1023–1027.

[12] Hollien Harry, Wojciech Majewski and Thomas E. Doherty. (1982). "Perceptual identification of voices under normal, stress and disguise speaking conditions." Journal of Phonetics 10: 139–148.

[13] Hirson, Allen and Martin Duckworth. (1993). "Glottal Fry and Voice Disguise: a Case Study in Forensic Phonetics." Journal of Biomedical Engineering 15:193-200.

[14] Wagner, Isolde and Olaf Köster. (1999). "Perceptual recognition of familiar voices using falsetto as a type of

voice disguise.” Proceedings of the 14th International Congress of Phonetic Sciences 2: 1381-1384.

[15] Neuhauser, Sara. (2008). “Voice Disguise using a foreign accent.” IJSLL 15.2: 131–35.

[16] Stevens, K.N. (1971). “Sources of inter- and intra-speaker variability in the acoustic properties of speech sounds,” Proc. 7th International Cong. Phonetic Sci., Montreal 206-32.

[17] Baldwin, John and Peter French. (1990). *Forensic Phonetics.* London: Pinter.

[18] Didla, Grace S. and Harry Hollien. (2016). “Voice Disguise and Speaker Identification.” Proceedings of Meetings on Acoustics 25:060006: 1-8.

[19] Zhang, Cuiling and Tiejun Tan. (2008). “Voice disguise and Automatic speaker recognition.” Journal of Forensic Science International 175: 118-122.

[20] Farrus Mireia, Michael Wagner, Daniel Erro and Javier Hernando. (2010). “Automatic Speaker Recognition as a measurement of voice imitation and conversion.” IJSLL 17.1:119-142.

[21] Künzel, Hermann J. (2013). “Automatic speaker recognition with cross-language speech material.” IJSLL 20.1: 21-44.