

Literature Review of Voice Disguise Identification

Chao Zhang

Abstract—Voice disguise has raised people’s attention recently for been widely used in illegal applications, such as wire fraud and kidnapping which has a huge negative impact on audio authenticity. Identification of disguised voices seems extremely important especially in some fields where audio forensics are required such as law enforcement. This study will review types of disguised voices, collection of disguised voices, disguise voice application area, related identification softwares and analysis of experimental results.

Index Terms—voice disguise, audio authenticity, audio forensics, identification.

I. INTRODUCTION

Nowadays, our society, compared to one century ago, seems to contain more and more fake information due to highly development of technology. It is a worldwide tendency that voice disguise has been used frequently for identities concealing. Since disguised voices has been used for threatening calls, kidnapping, even emergency police calls and human can not directly identify the disguised voices without any tools or techniques, this literature review gives illustrations to basic concepts of disguised voices and detailed implemetations in non-electronic deliberate disguised voices.

II. REVIEW OF LITERATURE

A. Types of Disguised Voices

The voice disguise can be classified into two independent dimensions: non-electronic versus electronic and deliberate versus non-deliberate as shown in Table1. For deliberate non-electronic voice disguise, speakers have many ways to alter their voices by using foreign accent, pinched nostril, distortion mouth, speed rate acceleration and so on. For deliberate electronic voice disguise, some softwares such as voice changer, Audacity, Cool Edit are used to change voice for confusing recognition. In non-deliberate situations, disguise voice is mainly generated by external factors. Illness or emotional feelings may cause nonelectronic voice disguise and voice channels with noise or bandwidth limitations generate electronic voice disguise.

TABLE I: Four types of voice disguise

Comparison of different types of disguised voices		
Taxonomy	Eletronic	Non-electronic
Delibrate	Electronic scrambling	Speaking in a falsetto
Non-delibrate	Limitation of bandwidth	Speaking with emotions

B. Data Collection for Disguised Voices

Due to electronic deliberate voice disguise is relatively uncommon, occurring in only one to ten percent of voice disguise situations(Rodman, 2000)[1]. Non-deliberate voice disguise is

a poorly researched area expect for some special researchers such as medical personnel for non-deliberate non-electronic voice disguise or electrician for non-deliberate electronic voice disguise. So, non-electronic deliberate voice disguise is focused in futher research and review of collection of this kind of disguised voices is necessary for implementation. According to Rodman and Powell(2000), there are not any standardized databases of voice disguises[1]. Collection of voice data and integrating them into database is the begining of study of voice disguise. The data collection should follow the specifications and standards which were formated by Boves and Carlo with their teammates in 1994[4][5]. The recomendation requirement is: 1. Voice source is from 30 to 40 speakers with multiple sessions. 2. The recordings should be digital, sampled by high quality recording devices at 22kHz, 16 bit quantization in a low noise environment. 3. Data should be permanently stored in superior media for future use.

C. Voice Disguise Identification Application Areas

1) *Law Enforcement*: It is important for law enforcement agencies to make sure the audio authenticity when matching voice of a suspect with different recorded voices. Voice disguise identification is applied to avoid suspects using methods deliberately change their voices for getting rid of crimes.

2) *Speaker Verification*: Speaker recognition of intentional disguised voices is widely used in the forensic field, however, it is also necessary to establish methodologies for research into unintentional disguised voice speaker recognition. When someone has illness problem casuing non-deliberate voice disguise, the speaker verification for access to sites may be a problem. So the speaker verification system should be trained with a large number of different disguised voices to tell the difference between deliberate and non-deliberate disguised voices.

D. Voice Disguise Identification Implementation

1) *Forensic Automatic Speaker Recognition System*: In implementation, Forensic Automatic Speaker Recognition System(FASRS) developed by the department of computer science in Tsinghua University is used. It can both identify and verify speakers with a database collecting more than 2000 speakers’ voices in different channels and automatic speaker recognition which is independent of speech text and language making it better than traditional spectrogram comparison[2].

2) *Method*: Forensic Automatic Speaker Recognition System can indentify voice disguise automatically and output the disguise rate by two input voices which are normal voices and disguise voices with a threshold. The system compares each testing voice with all voices in database to determine if the voice is disguised.

Input:

- Sample A: Normal voices modeling known speakers.
- Sample B: Disguised voices modeling unknown speakers.
- The setting threshold for similarity comparison.

Output:

- Score of similarity rate.
- The result of verification according to the obtained score.

3) Results Summary:

- Speaker recognition by normal voice: The system performs well for normal voices recognition which shows it can correctly recognize almost all input voices of speakers. From the similarity between one speaker's two different ways of speaking, it displays intra-speaker variation differs among speakers.
- Speaker recognition by disguised voices: The speaker recognition performs badly because the voice has been disguised. Except for foreign accent disguising, disguised voice recognition rates are significantly decreased. According to different disguising types, the effects of disguising are different. One of types, which is pinched nostril, has a great effect because of large changes of vocal tract configuration and voice quality.
- Inter-speaker variation on disguising skill: Both for normal voices and disguised voices, speaker recognition has inter-speaker variation. Different speakers have different disguise skills. If a speaker is good at disguising his/her voice, the disguised voices then change a lot and the results show lower speaker recognition rate and correct recognition rate.

E. Electronic Voice Disguise

Above reviews are mainly based on non-electronic voice disguise because electronic voice disguise is very uncommon. However, in some television and radio interviews, voices are often disguised electronically for privacy protection. There are few researches related to electronic disguised voices but Wu and Huang(2014) found an algorithm for electronic disguised voices identification according to two frequency spectral properties which are modification of the frequency spectrum of speech signals, and mel-frequency cepstrum coefficients (MFCCs)[3]. They extracted MFCC statistical moments including mean values and correlation coefficients as acoustic features and then separate disguised voices from original signals. Their algorithm is remarkable which can detect more than 90% voices from various databases and disguised by different methods.

III. CONCLUSION

This article reviews a general understanding of voice disguise from disguise types, data collection, application areas and implementation of identification of disguised voices. It focuses on non-electronic deliberate voice disguise which is a rich research area related to our daily life. For the implementation of non-electronic deliberate voice disguise, Forensic Automatic Speaker Recognition System has been used and implementation details are covered. It forms a framework for further research in voice disguise especially in non-electronic deliberate voice disguise.

REFERENCES

- [1] Rodman, R., & Powell, M. (2000, October). Computer recognition of speakers who disguise their voice. In The international conference on signal processing applications and technology ICSPAT2000.
- [2] Zhang, C., & Tan, T. (2008). Voice disguise and automatic speaker recognition. *Forensic science international*, 175(2-3), 118-122.
- [3] Wu, H., Wang, Y., & Huang, J. (2014). Identification of electronic disguised voices. *IEEE Transactions on Information Forensics and Security*, 9(3), 489-500.
- [4] Boves, L., Bogaart, T., & Bos, L. (1994). Design and Recording of large data bases for use in speaker verification and identification. In *Automatic Speaker Recognition, Identification and Verification*.
- [5] A. di Carlo, M Falcone, A. Paoloni. Corpus design for speaker recognition assessment. *Proceedings of the ESCA Workshop on Automatic Speaker Recognition, Identification and Verification*. Pp. 47-50. Martigny, Switzerland. April 1994.