PROCEEDINGS OF SCIENCE

Recording Device Identification Based on Cepstral Mixed Features

Wei Zhong[12]
School of Information and Communication Engineering,Dalian University of Technology,Dalian, 116024, Liaoning, P.R. China
*E-mail:*zww110221@163.com

Xiangwei Kong, Xingang You
School of Information and Communication Engineering,Dalian University of Technology,Dalian, 116024, Liaoning, P.R. China
*E-mail:*kongxw@dlut.edu.cn , youxg@dlut.edu.cn

Bo Wang[3]
School of Information and Communication Engineering,Dalian University of Technology,Dalian,116024 Liaoning, P.R. China
E-mail: bowang@dlut.edu.cn

The authenticity of the recording evidence is the foundation of legitimacy and relevance, which is the primary condition of recording evidence. With the springing up of private recording evidence, there is an urgent need for authenticity identification of recordings. That the evidence shall be from an accurate and legitimate source is a prerequisite for three elements. Recording equipment identification is the core content of sources of evidence. This article studies the characteristics of the recording device parameters, proposing three characteristic parameters of recording equipment such as the proportion of time-domain low roughness, etc. And combined with improved Mel Frequency Cepstrum Coefficient (MFCC) feature parameters characteristic parameters constitute a hybrid 92-dimensional. According to experimental analysis, with 10 different brands and models of recording device (including five different brands and models commonly used in voice recorder and five kinds of commonly used different brands and models of mobile phones), 60 young men and women, each of 10 different voice, the same type of equipment to record each 2, shows that mixed characteristic parameters can effectively characterize the characteristics of the recording equipment. Recognition rate increases by more than 6% compared with ordinary cepstrum.

CENet2015
12-13 September 2015
Shanghai, China

[1]Speaker

[2]This work is supported by the Research Fund for the Doctoral Program of Liaoning Province (Grant No. 20131014), the Open Fund of Artificial Intelligence Key Laboratory of Sichuan Province (Grant No. 2012RZJ01), and also the Fundamental Research Funds for the Central Universities (Grant No. DUT13RC201).

[3]Corresponding Author

PoS(CENet2015)037

1. Introduction

The recording equipment classification is the latest audio forensics research hotspot [1]. In the course of the audio evidence provided, somebody claims that he used a device to record audio evidence, but there is no effective way to verify it, hence, people carry out researches in this area [2-3]. In 2006, Lukas [4] studied on the effects of the sensor output noise on VCR recognition. Since 2007, Dirik [5-6], who studied the impact of dust characteristics of the sensor to VCR recognition, achieved valuable results. Tsai and Li *et al.* had a in-depth study cellular phone recognition[7-8] . Cemal *et al.* extracted cell phone's characteristics from the cell phone recording signal[9], and using the MFCC parameters as feature parameters and SVM as a recognition model, a high recognition rate of 96% is achieved for 14 different phones.

Cemal had studied and analyzed the characteristic parameters and recognition model of recording equipment, its characteristic parameters and recognition model are based on the existing speaker recognition features and models, either Fourier Transformation parameters or MFCC is not for a special recording device identification parameters [10]. Characteristic parameters that specifically for recording equipment are still very few. In terms of the MFCC, low-dimensional parameters generally reflect the speaker's semantic features, and high-dimensional parameters generally reflect the speaker's personality traits. The MFCC will definitely affect recognition of recording devices accuracy rate, when it is used as a characteristic parameter of recording equipment. Therefore, we must find or construct characteristic parameters consistent with the characteristics of the recording device.

From the recording equipment itself, taking into account copyright and other reasons, there may be a difference in terms of recording circuit and chip, sampling rate, the number of quantization bits and the compression algorithm, where we can find the recording equipment personality characteristics. Also, recording equipment parameters are not only mixed in semantic features bands but also mixed in speaker feature parameters.

Hence, considering the lack of special characteristic parameters of recording equipment, we study and propose a number of characteristic parameters characterizing feature of recording equipment firstly, and then combining with existing audio feature consist of mixed feature of recording equipment.

2. Propose of two Time-frequency Domain Characteristic Parameters

Currently on the market a lot of recording equipment or phone recording material have adopted the compression. Different compression algorithms and filtering algorithm makes audio signal present different time-frequency domain features, and at present there is no research on this aspect. Therefore, it is necessary to analyze the new characteristics parameters of the recording device according to this situation.

2.1 Amplitude Proportion

For the recording device, considering patents and other reasons, recording equipment differ from each other in circuit and personality characteristics, which constitute the personality characteristics of the recording equipment. Minimum amplitude proportion is a parameter reflecting quantization bit number of the device. In the recording signal, the amplitude of the smaller sampling points occupies a certain proportion. After normalized quantifying the signal, the minimum amplitude and the number of quantization bits show the following relationship:

$$x_{\min K} = K2^{-M} \tag{2.1}$$

in which M is quantization bit number, $x_{\min K}$ is the K-th minimum amplitude. Any amplitude is an integer multiple of the minimum quantization value.
Amplitude proportion is:

$$Aratio_K = num_{\min K} / num_{total} \tag{2.2}$$

As the statistical properties of the speech signal satisfies Laplace distribution, amplitude distribution of the speech signal satisfies the following equation:

$$p_L(x) = 0.5ae^{-a|x|}, \ a = \sqrt{2}/\sigma_x \tag{2.3}$$

2.2 Time-domain Low Proportion Roughness

In the speech signal processing, in order to improve the quality of hearing or speaker recognition rate, people pay more attention to vowel, larger amplitude of a signal, and optimize in the spectrum. They often overlook the processing of the auditory insensitive low amplitude sampling points, which often carry characteristics of the amplifier circuit's non-linear area and compression algorithms personality characteristics such as information of quantify bits which reflects the characteristics of the recording equipment.

According to the probability distribution of the voice, the voice in the amplitude of the lower case, were evenly distributed. However, the proportion of low amplitude is not uniformly distributed. Each device presents a unique personality trait in the low-amplitude. The proportion of time-domain low roughness's definition process is given as following.

It can be defined by:

$$x_i = i2^{-M} \tag{2.4}$$

The proportion a_i in each frame is defined as follows:

$$a_i = \frac{count(x_i)}{count\_total} \tag{2.5}$$

$count(x_i)$ denotes the number of the data whose amplitude is x_i in the frame, $count\_total$ is frame length. Let:

$$b_i = | a_i - a_{i-1} | \tag{2.6}$$

when

$$i = 1, b_1 = 0 \tag{2.7}$$

It can be defined as follows：

$$b_{i\sim j} = \{b_i, b_{i+1}, ..., b_{i+j}\} \tag{2.8}$$

Then we can make the following definition:

$$c_{ij} = \frac{b_{i\sim j} b_{i\sim j}^H}{j} \tag{2.9}$$

c_{ij} gives roughness of a total of j points starting of the i-th minimum amplitude. If the low amplitude were evenly distributed, and a_i satisfy： $a_i = a, \forall i = 1, 2, 3, \cdots$, where a is a constant.

Equation (2.6) may be represented as:

If $b_i = 0, \forall i = 1, 2, 3, \cdots$, then:

$$c_{ij} = \frac{b_{i\sim j} b_{i\sim j}^H}{j} = 0 \tag{2.10}$$

2.3 Characteristic Mixing Parameters of Recording Device

According to the above analysis, this chapter intends to adopt the following mixing 92-dimensional feature mixing parameters as the characteristic parameter of recording equipment. Table 1 shows the details.

The MFCC and DCT minimum amplitude proportion features based on frequency domain have been proved to be effective in prior works. In last two subsection, we have demonstrated that the effect of quantization step of difference devices. The feature vectors in spacial domain are sensitive to the effect. A reasonable approach can be obtained to combine the time-domain and frequency-domain features to construct a better classifier. Base on this, we mix 44-dimensional MFCC features, 10-dimentional DCT minimum amplitude proportion features, 20-dimensional time domain minimum amplitude proportion features and 20-dimensional time-domain low proportion roughness features for the feature vector.

Mixed characteristic parameters	Description
MFCC1-10，33-64	Using 64-dimensional MFCC parameters' low-dimensional and high-dimensional parts
10-dimensional DCT minimum amplitude proportion	Frequency domain features, after DCT transform, calculate the number of the minimum value of 10 points in the proportion of all point values.
20-dimensional minimum amplitudeproportion	Time-domain characteristics, calculate the number of the minimum value of 20 points in the proportion of all point values. Definition is shown in Equation (2.3).
20-dimensional time-domain low proportion roughness	Definition is shown in Equation (2.6).

Table 1: Time-frequency mixing characteristic parameters of recording equipment

3. Experimental Results and Analysis

The recording device used in the experiment are five recording device (each type of equipment is two). Recording subjects were 60 persons consist of 30 young men and 30 women. Everyone speaks 10 different Mandarin, and every word is about 10 seconds, generating 6000 wav audio data. The sampling frequency is 44.1KHz, quantization bits are all 16-bit, frame length is 2048 points, a frame shift of 50%.Take a word each person and each device as training audio, the other as a test audio.

The basic situation of these five voice recorder are as follows:

(1) Sony PCM-M10: Recordable: MP3 format, sampling frequency is 44.1KHz (bit rate is 64Kbps, 128Kbps, 320Kbps); PCM format, sampling frequency selectable from 22.05KHz, 44.1KHz, 48KHz, 96KHz, respectively, can be quantified into a 16bit / 24bit; hereinafter referred to by Sony.

(2) Tong Fang TF-A20: MP3 (sampling frequency is 32KHz, 192Kbps), hereinafter referred to by Tong Fang;

(3) Jing Hua DVR-818: MP3 (sampling frequency of 32KHz, 128Kbps), hereinafter referred to by Jing Hua;

(4) Modern HYM-3698: MP3 (sampling frequency of 44.1KHz, 128Kbps), hereinafter referred to as the Modern;

(5) Sanyo ICR-PS004M: MP3 (sampling frequency of 44.1KHz, bit rate of 192Kbps), hereinafter referred to by Sanyo.

Baseline system uses 12-dimensional MFCC parameters that Cemal proposed in 2012 as a baseline characteristic parameters. Actually voice signal characteristic parameters including the speaker characteristic parameters have the best noise immunity. MFCC parameters characterize personality traits of the most effective.

Recognition model uses SVM classifier. The proposed method with hybrid characteristic parameters is compared with a baseline proposed in a paper [9]. Experimental results are listed in Table 2.

	Sony	Sanyo	Modern	Tong Fang	Jing Hua	AVGERAGE
Baseline proposed	82.2%	74.6%	76.9%	65.1%	68.4%	73.4%
Proposed method	91.7%	78.5%	81.4%	73.0%	75.5%	80.0%

Table 2: Identify performance comparison of no projection of hybrid feature parameters

Table 2 gives a comparison of recognition rate between hybrid characteristic parameters and the baseline system. Recognition mode uses the text-independent manner. From the table, recognition rate of hybrid characteristic parameters increases by more than 6% compared with baseline system. The most obvious improvement is Sony, which improve by 9.5%. For a variety of devices, recognition rate of Sony is highest, Sanyo and modern secondly, between 75% to 83%. Tong Fang and Jing Hua are poor, around 70%. An average accuracy of 80.0% is achieved, compared with that of 73.4% obtained by the baseline. The results shows that combination of the proportion of low time-domain roughness and MFCC can improve the performance of the device identification

Table 3 shows the result of picking up characters from characteristic parameters of base line and mixing characteristic parameters through the way of orthogonal projection operator. From the table, it is obvious that adopting the orthogonal projection operator improves the recognition rate of system. For example, equipments like Sony, Sanyo and Modern get a significant improvement of 3% to 5% approximately. However, the improvement of property seems not very obvious for Tong Fang and Jing Hua, whose improvements are approximately below 1%.

	Sony	Sanyo	Modern	Tong Fang	Jing Hua	AVGERAGE
with orthogonal projection operator	86.3%	77.9%	80.6%	66.7%	69.2%	76.1%
Proposed method with orthogonal projection operator	93.1%	83.2%	84.0%	74.4%	75.9%	82.1%

Table 3: Comparison of Identifying performance by orthogonal projection of mixing characteristic parameters

4. Conclusion

The original-evidence research mainly consists of obtaining evidence with the recording equipment, recognizing the time and place of recording and so on. The progress of recognizing recording time and place achieve less among home and abroad. The judge mainly depends on the relevance of other evidence during the actual operation. But research of obtaining evidence of recording equipment is still the hot issue among domestic and overseas in terms of speech single processing , which remains in the technology trigger and has not raised or analyzed the special characteristic parameter of recording evidence. The article goes deep into the characteristic parameter of recording evidence, raises the time-domain low proportion roughness and other two characteristic parameters of recording evidence, which constitutes 92-dimensional feature mixing parameters combined with the modified MFCC characteristic parameters . The experiment demonstrates that the mixed characteristic parameters are able to represent the feature of recording evidence effectively, by collecting sixty youth that ten different speech each of them and two speech of the same model with five different brand of recording evidence, whose recognition rate raises up by 10.4 percent comparing with the ordinary parameters of cepstrum.

References

[1] Y. Panagakis, C. Kotropoulos. *Automatic telephone handset identification by sparse representation of random spectral features*[C]. MM and Sec'12 - Proceedings of the 14th ACM Multimedia and Security Workshop, ACM, USA. pp, 91-95(2012).

[2] O. Farooq, S. Datta, J. Blackledge. *Blind tamper detection in audio using chirp based robust watermarking*[J]. WSEAS Transactions on Signal Processing, 4(4): 190-200(2008).

[3] M. Unoki, R. Miyauchi, *Detection of tampering in speech signals with inaudible watermarking technique*[C]. Proceedings of the 2012 8th International Conference on Intelligent Information Hiding and Multimedia Signal Processing(IIH-MSP), IEEE, USA. pp, 118-121(2012).

[4] J. Lukas, J. Fridrich , M. Goljan. *Digital camera identification from sensor pattern noise*[J]. IEEE Transaction on Information Forensics and Security, 1(2): 205–214(2006).

[5] E. Dirik, H. T. Sencar, N. Memon. *Source camera identification based on sensor dust characteristics*[C]. Proceedings IEEE Workshop Signal Processing Applications Public Security Forensics, IEEE, USA. pp,1-6(2007).

[6] A. E. Dirik, H. T. Sencar, N. Memon. *Digital single lens reflex camera identification from traces of sensor dust*[J]. IEEE Transaction on Information Forensics and Security, 3(3): 539–552(2008).

[7] M. J. Tsai, C. L. Lai, J. Liu. *Camera/mobile phone source identification for digital forensics*[C]. Proceeding of IEEE International Conference on Acoustics, Speech Signal Processing, IEEE, USA. pp, II-221 - II-224(2007).

[8] O. Celiktutan, B. Sankur, I. Avcibas. *Blind identification of source cell phone model*[J]. IEEE Transaction on Information Forensics and Security, 3(3): 553–566(2008).

[9] C. Hanilci, F. Ertas. *Recognition of Brand and Models of Cell-Phones From Recorded Speech Signals*[J]. IEEE Transaction on Information Forensics and Security, 7(2): 625-634(,2012).

[10] S. Gupta, S. Cho, C.-C.J. Kuo. *Current Developments and Future Trends in Audio Authentication* [J]. IEEE MultiMedia, 19(1): 50-59(2012).