Forensic Science International 266 (2016) 427–439

Contents lists available at ScienceDirect

Forensic Science International

journal homepage: www.elsevier.com/locate/forsciint

High-performance combination method of electric network frequency and phase for audio forgery detection in battery-powered devices

Maryam Savari *, Ainuddin Wahid Abdul Wahab *, Nor Badrul Anuar

Faculty of Computer Science and Information Technology, University of Malaya, 50603 Kuala Lumpur, Malaysia

ARTICLE INFO

Article history:
Received 10 January 2016
Received in revised form 27 June 2016
Accepted 4 July 2016
Available online 11 July 2016

Keywords:
ENF
Phase
Audio
Forgery
Battery-powered devices
Accuracy

ABSTRACT

Audio forgery is any act of tampering, illegal copy and fake quality in the audio in a criminal way. In the last decade, there has been increasing attention to the audio forgery detection due to a significant increase in the number of forge in different type of audio. There are a number of methods for forgery detection, which electric network frequency (ENF) is one of the powerful methods in this area for forgery detection in terms of accuracy. In spite of suitable accuracy of ENF in a majority of plug-in powered devices, the weak accuracy of ENF in audio forgery detection for battery-powered devices, especially in laptop and mobile phone, can be consider as one of the main obstacles of the ENF. To solve the ENF problem in terms of accuracy in battery-powered devices, a combination method of ENF and phase feature is proposed. From experiment conducted, ENF alone give 50% and 60% accuracy for forgery detection in mobile phone and laptop respectively, while the proposed method shows 88% and 92% accuracy respectively, for forgery detection in battery-powered devices. The results lead to higher accuracy for forgery detection with the combination of ENF and phase feature.

1. Introduction

Audio forensics involves the analysis, identification, and assessment of digital audio as evidence in the court of law and other legal venues [1]. Primarily, the aspects of audio forensics consists of (i) enhancement, to improve the quality and clarity of the audio; (ii) interpretation, to interpret and document the audio evidence such as speech or speaker identification; and (iii) authenticity, to authenticate sonic documents for civil or law enforcement investigation [1,2]. In order to investigate the validity of an audio clip in any official venue such as the court of law, audio authentication can play an essential role in detecting possible forgery in the submitted evidence. Therefore, researchers have been proposing and developing various techniques for authenticating audio since the 1970s. Among them, ENF stands as a promising audio authentication technique and can be detected in battery and plug-in powered devices. Moreover, ENF has registered a high level of accuracy, especially in plug-in powered devices [3–6]. However, ENF is highly dependent on databases [7] and weak against high levels of noise [3]. The performance of ENF in terms of accuracy in battery-powered devices is unstable due to the source of ENF in these devices. The source of ENF in battery-powered devices is audible hum, which is created by electrical devices placed around the battery-powered recorder. The created ENF by the audible hum in battery-powered devices is a weak feature for audio authentication [8], which is considered as the focal problem in this study. In this study, the influence of the ENF feature on battery-powered devices is investigated.

This study is conducted to focus on the authenticity for solving a particular problem in this field, which is considered a weakness of ENF in battery-powered devices. A combination method of ENF and phase is proposed and employed as a novel and accurate method. The accuracy of the combination method of ENF and phase for forgery detection in battery-powered devices is investigated experimentally.

2. Literature review

A brief review of audio authentication fields, which include evaluation, vision and goals is explained in this part of study. Audio authentication has emerged as the most important field in audio forensics. It can be described in a series of evolution, vision, and goals. Audio authentication began as early as the 1950s concurrently with the use of audio outside the recording studio [2]. In the United States, the use of recorded conversations as evidence in legal cases was allowed for the first time in the

* Corresponding authors.
E-mail addresses: maryamsavari@gmail.com, maryamsavari@siswa.um.edu.my (M. Savari), ainuddin@um.edu.my (A.W. Abdul Wahab), badrul@um.edu.my (N.B. Anuar).

http://dx.doi.org/10.1016/j.forsciint.2016.07.001
0379-0738/

1958 McKeever case. Another significant stage in audio authenticity is the 1973 Watergate case. This case has influenced the advisory panel to establish an important methodology to determine the authentication of recorded audio in the White House [2]. The methodology is called "panel methodology". Over time, the essential requirement for audio authentication has enhanced, and new solutions have been proposed. The authentication techniques are commonly proposed based on the features of the audio.

Also, the field of audio authentication has been shaped by several high-profile cases such as the 1973 Watergate scandal. This political scandal involved an audio recording system installed in the White House and Democratic National Committee headquarters at the Watergate office complex in Washington, DC, by president of the United States, Richard M. Nixon. During the transcription of the recorded tape, there was unexplained 18½ min gap of conversation between President Nixon and his chief of staff. The scandal has led to the investigation by the advisory panel to identify if the recorded audio had been tempered with intentionally [2]. Other high-profile cases that involved audio forensic are assassination of President John F. Kennedy in 1961 and presidential candidate Senator Robert Kennedy in 1968 [2]. These investigations have highlighted the importance of authentication in the audio forensics field.

In addition, audio authentication has different goals between technical authentication and the legal system [9]. In recent years, audio authentication through establishing a chronology of recorded events, forgery detection, and matched recording devices have become more important, especially in the court of law [10]. Various techniques have been proposed for authentication based on the features in the audio, which are created from the condition of the recording, environment, and devices. These techniques were established in the 1990s and have since rapidly enhanced. Generally, the available authentication techniques lack the following aspects.

(i) Lack of accurate technique in detecting all type of forgeries is considered as one of the main problems in this field; thus, technical improvement of the available techniques is required [11].
(ii) ENF is an audio forgery detection feature, which is eligible for its high level of accuracy in plug-in devices [3,5,12]. ENF features can be detected in battery and plug-in devices. To the best of our knowledge, an accurate solution to achieve high-level validity of ENF features in battery-powered devices has not yet been accomplished. Therefore, a combination method of ENF and phase is proposed in this study to improve the accuracy of ENF in battery-powered devices.

Grigoras [3] mentioned that, due to the absence of an ideal voltage regulator, digital equipment captures 50/60 Hz electric network frequency in speech sound during the recording process. Therefore, he proposed a method to detect forgery using electric network frequency. In this method, ENF of recorded audio is compared with a database retrieved from an electric company. Based on this method, a digital audio recorder (DAR) can be divided into two groups that record digital audio evidence with or without the 50/60 Hz frequency. The author analyzed ENF variation over six months and found that the 50 Hz frequency is constantly changing up to ±0.6 Hz every 10 s. Thus, the ENF can be defined as follows:

$$f = [50 \pm \Delta f]\ \mathrm{Hz}, \tag{1}$$

where Δf is the deviation between the instantaneous frequency and the set point frequency. In the recorded evidence, the retrieved ENF from the electric company must be the same time and date of the recorded evidence. The mismatch between the ENF of recorded audio and ENF of the electric company will lead to suspicion of tampering with the submitted audio. It is recommended to use ENF in high-quality sampling with low level of noise for forgery detection [3,13].

The ENF has weak visibility in an audio signal; therefore, it is highly dependent on the capability of extraction technique [14,15]. There are three methods for extracting ENF, which are described as follows [5,12]:

(i) "time/frequency domain" spectrograms method, which comprises computing spectrograms for comparing questioned signal versus referenced value of ENF retrieve from an electric company visually;
(ii) "frequency domain" method, which consists of computing FFT over a short period of time and followed by extracting the maximum magnitude value and comparing questioned signal versus referenced value of ENF retrieve from the electric company, which is used in the current study for ENF extraction;
(iii) "time domain" analysis, which comprises zero-cross measurement and comparing questioned signal versus referenced value of ENF retrieved from the electric company [5,12].

Conventionally, ENF is dependent on the audio recording time. In the case that the audio recording time is unclear, forgery detection will be impossible using this method. Also, this method needs an ENF database, which must always be updated. The process of completing and updating the database requires a significant time period.

Rodríguez et al. [7] worked on the database problem discussed above; they proposed a method, which is independent from the database and is based on the discontinuity of the power grid signal. A power grid signal is the single sinusoidal waveform with a fixed frequency, which is called ENF. Most electricity power in the city, in particular in the developed countries, is supplied from turbines and the rotation velocity of these turbines generates the ENF signal, which is 50 or 60 Hz. In this method, phase of the power grid signal, which is accommodated in the audio, can help in detecting forgery when there is a lack of reference value. The power grid signals are considered as a single tone; thus, the frequency and phase of a single tone are approximated. They used the discrete Fourier transform (DFT) to estimate the signal phase. The phase tone is calculated by the argument (angle) of the maximum value [7]:

$$x(n) = s_{\mathrm{tone}}(n) w(n), \tag{2}$$

where $s_{\mathrm{tone}}(n)$ is an M-sample single tone sequence, whose frequency and phase are to be estimated and $w(n)$ is the smoothing window. Also, $X(k)$ is the N_{DFT}-point DFT of $x(n)$ when $N_{\mathrm{DFT}} \geq \mathrm{M}$. Let k_{peak} be the integer index associated with the maximum value of $|X(k)|$ and f_s is the sampling frequency of $s_{\mathrm{tone}}(n)$. Then, the estimated value of the tone frequency is as follows [7]:

$$f_{\mathrm{DFT}} = k_{\mathrm{peak}} \frac{f_s}{N_{\mathrm{DFT}}}. \tag{3}$$

In Eq. (3), the resolution of f_{DFT}, which can only be assume as discrete values, is f_s/N_{DFT}. This means that increasing the value of N_{DFT} will cause better accuracy of f_{DFT}. The tone phase is the argument (or angle) of $X(k_{\mathrm{peak}})$ as follows [7]:

$$\phi_{\mathrm{DFT}} = \arg[X(k_{\mathrm{peak}})]. \tag{4}$$

Their method is divided into two sections of visual and automatic methods. The automatic method is utilized based on the detection of abrupt phase change in the power grid signal of the original and the forged audio. Although the advantage of this method is independent from the reference value, it is, however, dependent on the power grid signal, which is only supplied in some developed countries [7].