2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing

Electrical Network Frequency as a Tool for Audio Concealment Process

Feng-Cheng Chang
Dept. of Innovative Information and Technology
Tamkang University
Taipei, Taiwan
135170@mail.tku.edu.tw

Hsiang-Cheh Huang
Dept. of Electrical Engineering
National University of Kaohsiung
Kaohsiung, Taiwan
hch.nuk@gmail.com

***Abstract*—We live in a digital era. Digital contents may be produced by digital equipments or by converting old analog recordings. With the rapid growth of digital contents, digital archiving technology is demanded. Different types of contents require different processing techniques. In this paper, we focus on digital audio contents. The related techniques, such as forensics, authentication, and error concealment, were studied. When converting audio tapes to digital files, sometimes a certain automatic error detection and concealment is needed. However, traditional audio tapes were recorded without any error recovery information. Based on the restriction, we proposed a scheme that incorporates the electrical network frequency (ENF) as a tool for detecting damaged audio segments. The goal is to help people identifying candidate concealment segments. When using in an archiving application, it reduces the manpower as well as increases the accuracy of the generated meta-data.**

Keywords-audio; forensics; ENF; electrical network frequency; concealment; digital archive

I. INTRODUCTION

With the maturity of digital technology, not only the digital contents are easily produced but also the digitization of traditional analog contents is demanded. In a typical digital archiving application, one of the important tasks is to digitize the old analog data. Sometimes it also requires that a given digital object is authenticated when archiving. In this paper, we focus on archiving audio data.

Given an analog audio tape, there are several forensic methods to authenticate the data[1]. Some of the techniques are based on the signal characteristics[2], but they are not reliable. Some of the techniques are based on the signature of the recorder. However, recorder signatures are not available in digital recorders because there are no electromagnetic effects among the circuits and the recording media. In order to authenticate digitally recorded audio data, we have to use other signature sources, such as the microphones[3][4] and the electrical network frequency (ENF)[5][6].

Digitizing audio tapes is not difficult. In terms of general archiving concepts, any defects in the analog version should be digitally preserved as well because they are part of the history of the archived object. The procedure becomes complicated when we consider in terms of the database. A digital archive is different from analog archives in that we can extract the features (the so-called meta-data) from the contents and store them for content-based searching applications. If we allow defects in the digital object, they may interfere with the meta-data extraction and produce imprecise description of the content. This implies that we need two digitized versions of an analog object: one for archive and the other for feature-extraction. The version for feature extraction should be processed so that most of the errors are concealed. This may involve a lot of human interactions because we have to make sure that the processed version conforms to the original semantics. An alternative approach is to exclude the defected segments out of the feature extraction process. To identify the defected segments, human interactions and/or audio specific analysis are involved. In short, to generate sensible meta-data, we need audio concealment[7][8][9] and/or damaged audio segment detection techniques. Because we focus on analyzing abnormal audio segments, audio forensic tools and concealment techniques are useful for this purpose.

This paper is organized as follows. In Sec. II, we describe the general audio forensics methods and the electrical network frequency criterion. In Sec. III, we describe the concepts and approaches of audio concealment. In Sec. IV, we propose a scheme which incorporates the ENF as a tool for detecting erroneous audio segments. Then, we conclude our design and the future work in Sec. V.

II. AUDIO FORENSICS

In this section, we briefly introduce the concepts of audio authentication in Sec. II.A. Then, one of the useful tools to authenticate recorded audio data, called electrical network frequency (ENF) criterion, is described in Sec. II.B.

A. Analysis of Audio Recordings

Given a recorded audio tape for forensics, we may want to authenticate the people, the time, the location, etc in the recording scene. The analysis is based on that the audio tape is original or a trusted copy. To determine the originality of an audio tape, we could use the idiosyncrasies and characteristics of a recorder. An audio tape recorder leaves its signatures on the tape when a button (start, stop, or pause) is pressed. Many research works showed the effectiveness of the signature by analyzing the waveform spectral component, spectrograms, and magnetic patterns.

The signature of a recorder is emitted from the recording and erasing heads as magnetic signals. The heads are physically wired (by electric circuits) to the power supply and the switches. When a button is switched, the electric

978-0-7695-4222-5/10 $26.00 © 2010 IEEE
DOI 10.1109/IIHMSP.2010.51

current in the circuits is changed abruptly. Because a recorder is a transformer between electric and magnetic signals, the abrupt change in electric current generates magnetic field change near the heads. Two recorders of the same design are expected to behave the same in the electromagnetic transitions. In addition, it is not possible to produce two exactly same components due to the manufacturing process. The errors in the electric components and the recording heads make it unlikely to produce two same recorders. Therefore, we can assume that each recorder generates its unique magnetic signatures. The tape is very close to the head by design. Therefore, it is inevitable to leave a trace of magnetic field change on the tape when we operate the recorder. In short, each recorder leaves its signatures on the tape.

The signatures of a recorder are a substantial property of an analog audio recording device because we cannot suppress the electromagnetic effects. However, this property does not exist when we use a digital audio recorder. Without the assistance of recorder signatures, two recording devices are not distinguishable. This implies that recorders are unreliable for forensics, and we have to use another representative property as the "signature". Moreover, digitized audio data are usually compressed. Most of the popular audio compression algorithms are based on psychoacoustic models and perform the lossy (quantization) computations. If the forensic criterion is not salient enough, the compression would diminish the property to a useless level.

There are several alternative forensic criteria, such as the microphone characteristics and the electrical network frequency (ENF). In the former approach, we shift the signature from the electromagnetic components to the microphone; and in the latter one, we shift the signature to outside of the recording device. In the following section, we will introduce the ENF criterion.

B. The ENF and Digital audio Recordings

A power plant generates electricity with the designated utility frequency. Different regions may use different utility frequencies. For example, the utility frequency is 60Hz in Taiwan and all the AC-powered electric equipments should be compliant to this specification. The transmission of electricity from the power plants to the end-users is typically organized as a number of grids. In a grid, the power cables are directly connected. This implies that any variation in the utility frequency in a grid is almost the same.

The electricity is generated by mechanical devices in a power plant. The variations in heat source, pressure, and the rotation speed of the coils affect the generated frequency and current. To provide almost-constant electric frequency to the users, a power plant has to monitor and regulate the generated frequency. For example, we have to increase the heat when the frequency is decreasing, and vice versa. Since all the control processes are based on mechanical operations, the adaptation delay is inevitable. This induces the time-varying, slowly and slightly deviation around the specified utility frequency. This variation is then distributed by the electrical grids. Sometimes we call the distributed time-varying electrical frequency the electrical network frequency (ENF).

When electronic equipment is used to record audio data, it captures not only the sound but also the implicit electric characteristics in the device. For an AC-powered recording device (assume that it is not professional-grade), the ENF is also recorded due to the coupling of the powering circuit and the capturing circuit. In some audio forensic researches, the ENF is used as a criterion. There are several reasons that the ENF is suitable for forensics:

- The ENF is distributed only in the connected grids. This property can be used to localize the location of the recording device.
- The ENF is time-varying. This property can be used to localize the time duration of the recorded audio clip.
- The ENF is generated by the power plants. They can be the trusted sources to provide the ENF histories. This makes matching the ENF in a given audio clip meaningful.
- Even without the trusted ENF history for matching, the phase continuity of the ENF in the given audio clip can be used to detect whether the clip is manipulated.

According to the above characteristics, the typical usage is to match the extracted ENF components with the trusted ENF components. If they are not matched, either the time/location is wrong or the given audio clip is tampered. The trusted ENF source could be from the power plant or from an authenticated audio clip. The restriction to apply the ENF techniques is that the recording device should be AC-powered. Otherwise, there is no ENF in the recorded data.

III. AUDIO CONCEALMENT

Converting traditional analog audio tapes to digital form is one of the important steps in a digital archiving application. Strictly speaking, any damage is part of the archived object and should also be preserved. However, there are cases that we want to conceal the damage from the original version. For instance, we may want to detect the damaged (or tapped) duration in an audio clip and exclude it from the meta-data extraction process as described in Sec. I. We may further conceal the damaged data so that nonprofessional people do not experience the noisy or weird sound. In this section, we will briefly describe the approaches of audio concealment.

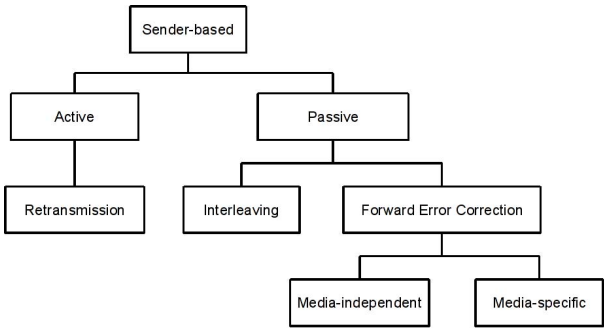

Figure 1. Sender-based audio concealment[7]

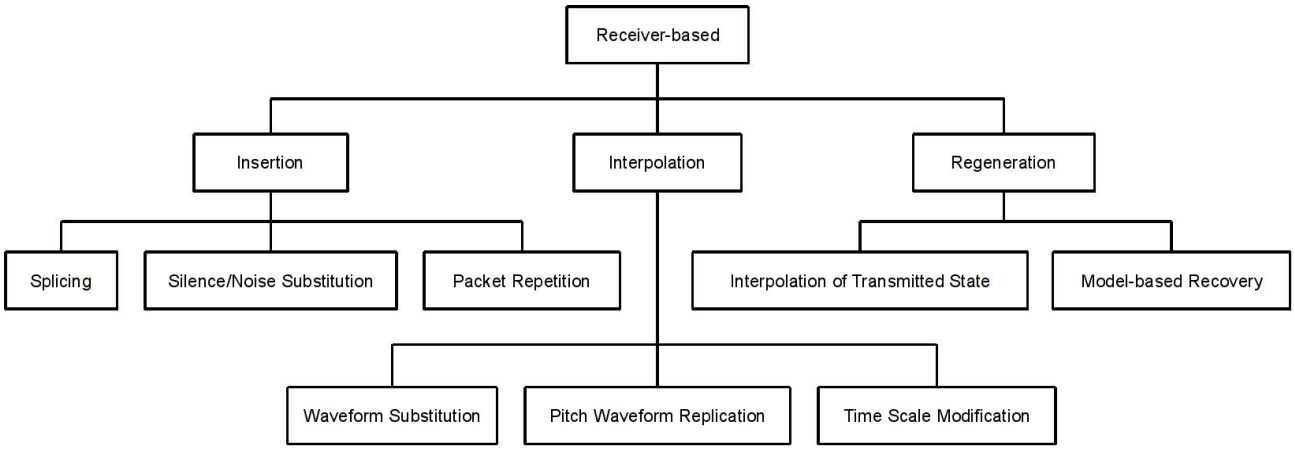

Figure 2. Receiver-based audio concealment[7]

A. *Sender-based Approach*

Sender-based approach requires the involvement of the audio sources. There are two sub-categories: active and passive approaches (Fig. 1). An active method is used in a networked delivery environment. Damages or losses can be detected by the network management or the receiver. Once the sender is notified, it retransmits the data for concealment. This approach is able to precisely conceal the errors with the cost of retransmission delay. The delays are affordable because most of the players are designed to buffer a certain amount of data before decoding. Therefore it is a practical solution for audio streaming applications.

A passive method does not require a real-time networking infrastructure. Instead of retransmitting concealment data when necessary, the sender always transmits additional information along with the media data. Once the receiver detects an error, it conceals the error by the additional information. One of the examples is to insert the forward error correction (FEC) code to the audio stream. The receiver can recover the erroneous data to the original version, as long as the number of errors is within the designed tolerance.

If the audio data is treated as an ordinary bit stream, conventional channel coding techniques can be adopted as the FEC code. This is called the media-independent approach. If we take the media coding scheme into consideration, we may design a specific coding structure so that multiple correlated streams are transmitted. The more the correct streams are obtained by the receiver, the more precise the reconstructed data is produced. This is called the media-specific approach. Some of the techniques are general, such as the interleaving method. It is a general technique in communication systems to reduce the impact of burst errors. When transmitting, the data packets are interleaved to avoid burst errors in the received stream. The costs are the enlarged receiver buffer and the increased decoding delay.

B. *Receiver-based Approach*

Without the assistance of the sender, the receiver is not able to recover the data precisely. The receiver can only try it best to conceal the errors, in a sense of human perceptions. These methods only require the receiver to participate in the error concealment process and thus called receiver-based approaches. Of course, it is very difficult to conceal the errors and retain the semantic correctness without ambiguity. Therefore, most of the concealment methods are designed to work with sound quality (e.g., psychoacoustic models). The semantic correctness of the recovered data is assumed to be checked manually.

As shown in Fig. 2, receiver-based methods are based on insertion, interpolation, and regeneration of data. Researches show that human psychoacoustic model is quite different from the human visual model. For audio signals, the masking effect, the noise tolerance, and the automatic concealment are important to the computer-based concealment algorithms. Because human can automatically conceal short-duration noise and discontinuity (i.e., human are not sensitive to short-duration errors), the computer-based concealment does not have to be perfect. The insertion-based methods are based on simple splicing, noise substitution, or even repetition of adjacent audio data.

Interpolation-based concealment methods are based on the assumption that the waveforms can be predicted. It is also known that audio signals are somewhat self-similar. Based on the two properties, concealment algorithms detect the pitches and the waveforms, and then try to interpolate the existent (correct) data to replace the damaged segments. In general, the interpolation-based approach is better than the insertion-based. It not only relies on human neurons to compensate the recover error but also predicts the probable waveform according to the nature of audio signals.

The most complicated among the three is the regeneration-based approach. It requires sophisticated analysis so that the coding states and/or the audio models are well estimated. After having a good model for the audio segment, we can regenerate data to replace the damaged part. The regeneration rule is to make the generated data conform to all the properties of the estimated audio segment. This approach usually requires a lot of computation to estimate the model. The advantage is the higher quality of the regenerated audio data. On the contrary, if the estimation

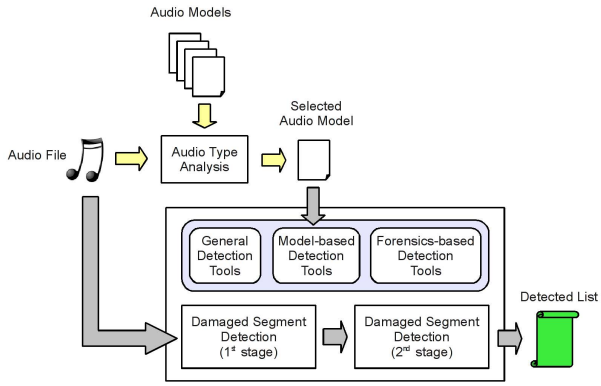

Figure 3. ENF-enabled damaged audio detection

procedure somehow fails, the quality of the regeneration would be worse than the former two approaches.

IV. Detecting Damaged Audio Clips

In the above audio concealment approaches, sender-based methods are not suitable for digital audio archiving applications. The reason is that traditional analog recording devices are not equipped with error correction mechanisms. Since the senders (original audio sources) are not available today, the receiver-based techniques are suitable for digital archiving applications.

Many receiver-based concealment techniques can also be used for analyzing audio properties. However, there are not many error detection methods for raw audio data (due to the semantic issues). It thus costs a lot of manpower to identify the damaged audio segments in a large collection of tapes. Although automatically detection is not likely to work perfectly, a program that analyzes the data and suggests the possible erroneous segments is useful for decreasing costs.

Fig. 3 shows our proposed scheme for detecting damaged audio data. For an input audio stream, we analyze it and estimate the parameters based on the audio models. During this process, some generic signal processing tools may be used. The feature of this scheme is that we incorporate the forensics-based techniques as the detection tools. Currently we concentrate on the use of the ENF criterion because a large portion of the traditional recordings are AC-powered. The detection accuracy could be increased by taking the ENF properties into consideration:

- Divide the raw audio stream into a sequence of packets (may be overlapped).
- Use a band-pass filter to extract the spectrum around the utility frequency.
- We assume that the ENF is a slow-varying function of time. If an abrupt frequency change is detected, it is possible that the packet contains a discontinuity of the audio data. This implies that it is an intended recording operation, a manipulation, or a damaged duration.
- The marked packets represent the set of candidate inconsistent durations. They can be part of the information for deciding errors.

V. Conclusions and Future Work

In this paper, we discussed the analysis of audio recordings and one of the audio forensic methods called the electrical network frequency (ENF) criterion. For digital audio recordings, the electromagnetic signatures of a recorder are not available. If the recording device is AC-powered, no matter it is an analog device or a digital device, the ENF is usually coupled into the recorded data via the circuits. Therefore, matching the ENF history and localizing the location/time of the recording is feasible.

We also discussed a few audio concealment techniques, and related them to the digital archiving applications. Our conclusion is that receiver-based methods are suitable for this purpose. To reduce the manpower in identifying the damaged audio segments, we proposed that incorporating the ENF as a detection tool in the process. Instead of being used as a matching criterion, the ENF is used to detect discontinuities of a given audio clip. It is based on the assumption that the ENF is a slow-varying function of time.

Currently we have not found a similar idea in the literature. In the future, we would like to verify the effectiveness of the ENF-based tool. After the verification, we will refine the scheme so that the ENF-based detection results are formulated to increase the detection accuracy.

Acknowledgment

This work was partially supported by the NSC, Taiwan, under Grants NSC 98-2218-E-032-008.

References

[1] R. Maher, “Audio forensic examination,” IEEE Signal Processing Magazine, vol. 26, no. 2, Mar. 2009, pp. 84-94.

[2] Y. Wang and M. Vilermo, “A compressed domain beat detector using MP3 audio bitstreams, ” MULTIMEDIA '01: Proceedings of the ninth ACM international conference on Multimedia, Ottawa, Canada, 2001, pp. 194-202.

[3] C. Kraetzer et al., “Digital audio forensics: a first practical evaluation on microphone and environment classification,” MM&Sec '07: Proceedings of the 9th workshop on Multimedia & security, Dallas, Texas, USA, 2007, pp. 63-74.

[4] A. Oermann, “Verifier-tuple for audio-forensic to determine speaker environment,” MM&Sec '05: Proceedings of the 7th workshop on Multimedia and security, New York, NY, USA, 2005, pp. 57-62.

[5] D.P. Nicolalde and J.A. Apolinario, “Evaluating digital audio authenticity with spectral distances and ENF phase change,” IEEE International Conference on Acoustics, Speech and Signal Processing, 2009, Apr. 2009, pp. 1417-1420.

[6] M. Kajstura et al., “Application of the Electrical Network Frequency (ENF) Criterion: A case of a digital recording,” Forensic Science International, vol. 155, no. 2-3, 2005, pp. 165-171.

[7] C. Perkins, O. Hodson, and V. Hardman, “A survey of packet loss recovery techniques for streaming audio,” Network, IEEE, vol. 12, no. 5, pp. 40-48, Sep/Oct 1998.

[8] B. W. Wah, X. Su, and D. Lin, “A survey of error-concealment schemes for real-time audio and video transmissions over the internet*,” in MSE '00: Proceedings of the 2000 International Conference on Microelectronic Systems Education. Washington, DC, USA: IEEE Computer Society, pp.17. 2000.

[9] E. Gunduzhan and K. Momtahan, “Linear prediction based packet loss concealment algorithm for pcm coded speech,” IEEE Transactions on Speech and Audio Processing, vol. 9, no. 8, pp. 778-785, Nov 2001.