Submitted to Workshop on Multimedia Technologies

Content-based Digital Watermarking for All Audio Formats

Changsheng Xu and Qi Tian
Laboratories for Information Technology
21 Heng Mui Keng Terrace, Singapore 119613
Tel: 65-68748248, Fax: 65-67744998, Email: xucs@lit.a-star.edu.sg

ABSTRACT

This paper proposes a set of digital watermarking schemes for PCM audio, WAV-table synthesis audio and compressed audio. The watermark embedding scheme is high related to audio content and based on human auditory system. The proposed methods are also very useful and effective for copyright protection, trace of illegal distributions and other applications.

Keywords: Digital Watermarking, PCM Audio, Synthesis Audio, Compressed Audio

1. INTRODUCTION

The rapid development of computer networks and the increased use of multimedia data via the Internet have resulted in the faster and more convenient exchange of digital information. However, the open environment of the Internet creates consequential problems regarding copyright of artistic work, and, in particular, the illegal distribution of digital multimedia work without owner's authorization. In 1998, the global pirate music market topped 2 billion units, worth an estimated US$4.5 billion. Sales of pirate music CDs rose to 400 million units, nearly 20% up on the previous year. To dissuade and perhaps eliminate illegal copying, a need exists for strengthening and assisting the enforcement of copyright protection of such work.

Digital multimedia contents include digital audio, video, image, and document. The ubiquity of digital multimedia content in Internet and digital library applications has called for new methods in digital copyright protection and new measures in data security. Digital watermarking techniques have been developed to meet the needs for these growing concerns and have become active areas of research.

Digital watermark is an invisible structure to be embedded into the host media. To be effective, a watermark must be imperceptible within its host, discrete to prevent unauthorized removal, easily extracted by the owner, and robust to incidental and intentional distortions. Many watermarking techniques in images and video are proposed, mainly focusing on the invisibility of the watermark and its robustness against various signal manipulations and hostile attacks. Most of recent work can be grouped into two categories: spatial domain methods[1-2], and frequency domain methods[3-4]. There is current trend towards approaches that make use of information about the human visual system (HVS) to produce a more robust watermark. Such techniques use explicit information about the HVS to exploit the limited dynamic range of the human eye.

Compared with the development of digital video and image watermarking, digital audio watermarking provides a special challenge because the human auditory system (HAS) is extremely more sensitive than HVS. The HAS is sensitive to a dynamic range of amplitude of one billion to one and of frequency of one thousand to one. Sensitivity to additive random noise is also acute. There is always a conflict between inaudibility and robustness in current audio watermarking methods. How to get a satisfactory trade-off between these two aspects becomes an important index to evaluate digital audio watermarking techniques.

Digital audio can be classified into three categories: PCM audio, synthesis audio and compressed audio. Current digital audio watermarking methods focus on PCM audio. The popular methods include spread spectrum coding [5], phase coding [6] and echo hiding [7], but none of these methods can achieve a good performance in audio quality and robustness. There are fewer watermarking methods related to compressed audio [8], but they are very weak in robustness. So far there is no any watermarking method for WAV-table synthesis audio.

In this paper, we will present novel content-adaptive watermarking schemes for PCM audio, compressed audio and a WAV-table synthesis audio. The embedding is highly related to the audio content. In view of common signal manipulations, watermark detection does not need the original audio signal as a reference.

2. DIGITAL WATERMARKING FOR PCM AUDIO

In this section, a novel content-adaptive watermark embedding scheme is described. The embedding design is based on audio content and Human Auditory System (HAS). With content-adaptive embedding scheme, the embedding parameter for setting up the embedding process will vary with the content of the audio signal. For example, because the content of a frame of digital violin music is very different from that of a recording of a large symphony orchestra in terms of spectral details, these two respective music frames are treated differently. By doing so, the embedded watermark signal will better match the host audio signal so that the embedded signal is perceptually negligible. The content-adaptive method couples audio content with embedded watermark signal. Consequently, it is difficult to remove the embedded signal without destroying the host audio signal. Since the embedding parameters depend on the host audio signal, the tamper-resistance of this watermark embedding technique is also increased.

In broad terms, this technique involves segmenting an audio signal into frames in the time domain, classifying the frames as belonging to one of several known classes, and then encoding each frame with an appropriate embedding scheme. The particular scheme chosen is tailored to the relevant class of audio signal according to its properties in the frequency domain. To implement the content-adaptive embedding, two techniques are disclosed. They are audio frame classification and embedding scheme design.

Fig.1 illustrates the watermark embedding scheme. The input original signal is divided into frames by audio segmentation. Feature measures are extracted from each frame to represent the characteristics of the audio signal of that frame. Based on the feature measures, the audio frame is classified into one of the pre-defined classes and an embedding scheme is selected accordingly, which is tailored to the class. Using the selected embedding scheme, watermark is embedded into the audio frame using multiple-bit hopping and hiding method. In this scheme, the feature extraction method is exactly the same as the one used in the training processing. The parameters of the classifier, and the embedding schemes are generated in the training process.

Fig.2 depicts the training process for an adaptive embedding model. Adaptive embedding, or content-sensitive embedding, embeds watermark differently for different types of audio signals. In order to do so, a training process is run for each category of audio signal to define embedding schemes that are well suited to the particular category of audio signal. The training process analyses an audio signal to find an optimal way to classify audio frames into classes and then design embedding schemes for each of those classes. To achieve this objective, the training data should be sufficient to be statistically significant. Audio signal frames are clustered into data clusters and each of them forms a partition in the feature vector space and has a centroid as its representation. Since the audio frames in a cluster are similar, embedding schemes can be designed according to the centroid of the cluster and the human audio system model. The design of embedding schemes may need a lot of testing to ensure the inaudibility and robustness. Consequently, an embedding scheme is designed for each class/cluster of signal, which is best suited to the host signal. In the process, inaudibility or the sensitivity of human auditory system and resistance to attackers must be taken into considerations.

The training process needs to be performed only once for a category of audio signals. The derived classification parameters and the embedding schemes are used to embed watermarks in all audio signals in that category.

As shown in Fig.1 in the audio classification and embedding scheme selection, similar pre-processing will be conducted to convert the incoming audio signal into feature frame sequences. Each frame is classified into one of the predefined classes. An embedding scheme for a frame is chosen, which is referred to as content-adaptive embedding scheme. In this way, the watermark code is embedded frame by frame into the host audio signal.

Fig.3 illustrates the scheme of watermark extraction. The input signal is converted into a sequence of frames by feature extraction. For the watermarked audio signal, it will be segmented into frames using the same segmentation method as in embedding process. Then the bit detection is conducted to extract bit delays on a frame-by-frame basis. Because a single bit of the watermark is hopped into multiple bits through bit hopping in the embedding process, multiple delays are detected in each frame. This method is more robust against attackers compared with the single bit hiding technique. Firstly, one frame is encoded with multiple bits, and any attackers do not know the coding parameters. Secondly, the embedded signal is weaker and well hidden as a consequence of using multiple bits.

3. DIGITAL WATERMARKING FOR SYNTHESIS AUDIO

Typically, watermarking is applied directly to data samples themselves, whether this be still image data, video frames or audio segments. However, such systems fail to address the issue of audio coding systems, where digital audio data is not available, but a form of representing the audio data for later reproduction according to a protocol is. It is well known that tracks of digital audio data can require large amounts of storage and high data transfer rates, whereas synthesis architecture coding protocols such as the Musical Instrument Digital Interface (MIDI) have corresponding requirements that are several orders of magnitude lower for the same audio data. MIDI audio files are not files made entirely of sampled audio data (i.e., actual audio sounds), but instead contain synthesizer instructions, or MIDI message, to reproduce the audio data. The synthesizer instructions contain much smaller amounts of sampled audio data. That is, a synthesizer generates actual sounds from the instructions in a MIDI audio file.

Expanding upon MIDI, Downloadable Sounds (DLS) is a synthesizer architecture specification that requires a hardware or software synthesizer to support all of its components. DLS permits additional instruments to be defined and downloaded to a synthesizer besides the standard 128 instruments provided by MIDI system. The DLS file format stores both samples of digital sound data and articulation parameters to create at least one sound instrument. An instrument contains “regions” which point to WAVE “files” also embedded in the DLS file. Each region specifies a MIDI note and velocity range that will trigger the corresponding sound and also contains articulation information such as envelopes and loop points. Articulation information can be specified for each individual region or for the entire instrument.

DLS is expected to become a new standard in musical industry, because of its specific advantages. On the one hand, when compared with MIDI, DLS provides a common playback experience and an unlimited sound palette for both instruments and sound effects. On the other hand, when compared with sampled digital audio, it has true audio interactivity and smaller storage requirement. One of the objectives of DLS design is that the specification must be open and non-proprietary. Therefore, how to effectively protect its copyright is important.

So far there is no watermarking method for WAV-table synthesis audio copyright protection. Compared with the previously mentioned PCM audio, it has its own special format. In view of this format, the watermarking scheme for WT synthesis audio is totally different from that for PCM audio. Fig.4 illustrates the watermark embedding scheme. Generally, a WT synthesis audio file contains two parts: articulation parameters and sample data such as DLS, or only contains articulation parameters such as MIDI. Unlike traditional PCM audio, the sample data in WT synthesis audio are not the prevalent components. On the contrary, it is the articulation parameters in WT audio that control how to play the sounds. Therefore, in our embedding scheme we not only embed watermarks into sample data (if they are included in the WT synthesis audio) but also into articulation parameters. Firstly, original WT synthesis audio is divided into sample data and articulation parameters. Then, we use two different embedding schemes to process them respectively and form the relevant watermarked outputs. Finally, the watermarked WT synthesis audio is generated by integrating the watermarked sample data and articulation parameters. In order to guarantee inaudibility and robustness, the watermark is embedded in both sample data and articulation parameters of a WT synthesis audio. A finite automaton is used to implement adaptive coding of the sample data. Virtual parameters are generated to hide watermark information in the articulation parameters. Furthermore, there is no need to use the original file when conducting watermark detection. In this method, inaudibility and robustness of watermarked WT synthesis audio are fully taken into consideration.

Fig.5 shows the scheme of watermark extraction. In the extracting process, the original WT audio is not needed. For a watermarked WT audio, it is also divided into sample data and articulation parameters at first. Then the watermark sequence in the coding bits of the sample data and the encrypted watermark information in the articulation parameters are detected. If the watermark sequence in sample data is obtained, it will be compared with the watermark in articulator parameters to make the verification. If the sample data suffered from distortions and the watermark sequence can not be detected, the watermarked bit sequence in the articulation parameters will be used to restore the watermarked bit information in the sample data and make the detection in the restored data. Similarly, the detected watermark will be verified by comparing with that embedded in articulation parameters.

4. DIGITAL WATERMARKING FOR COMPRESSED AUDIO

Compression algorithms for digital audio can preserve audio quality as well as reduce bit rate dramatically, increase network bandwidth, and save density storage of audio content. Among various kinds of compressed digital audio currently used, MP3 is the most popular one and gets more and more welcomed by music users. MP3 audio compression is based on psycho-acoustic models of human auditory system (HAS). It is an ideal format for distributing high-quality sound files online because it can offer near-CD quality at the compression ratio of 11 to 1 (128kb/s). One possible method to protect compressed audio is to decompress it first, then embed watermark into decompressed audio, and finally recompress the watermarked decompressed audio. This can probably ensure the robustness of the watermark, but it is very time-consuming because the compression process will take a long time. For example, it will take more than 30 minutes to compress a five to six-minute audio of WAV format to the MP3 format with the bit rate of 128k/sec. Therefore it is not suitable for on-line transaction and distribution. In order to improve the embedding speed as well as maintain the robustness of watermark, fast and robust embedding schemes for compressed audio must be taken into consideration.

In order to improve the robustness of the watermark embedded into the compressed audio as well as ensure the embedding speed, a content-based watermark embedding scheme is proposed in this section. According to this scheme the watermark will be embedded into partially uncompressed domain and the embedding scheme is high related to audio content. Fig.6 illustrates the block diagram of the content-based watermark embedding scheme in partially uncompressed domain.

The incoming compressed audio is first segmented into frames according to the coding algorithm. All the frames are decoded from compressed domain to uncompressed domain. Then the feature extraction model and the psychoacoustic model [9] are applied to each decoded frame to calculate the features of the audio and masking threshold in each frame. According to the features and masking threshold, a pre-designed filter bank [10] is used to select the candidate frames suitable for embedding watermark. The watermark will be embedded into these selected frames using an adaptive multiple bit hopping and hiding scheme depicted in Fig.7. The embedded frames will be re-encoded to generate the coded frames using the coding algorithm. Finally, The re-encoded frames and the non-embedded frames will be reconstructed to generate the watermarked compressed audio. Compared with the embedded scheme in wholly uncompressed domain, this scheme can not only get the same performance in audibility and robustness but also embed the watermark much faster. It is suitable for on-line embedding and distribution.

Fig.7 illustrates the block diagram of detailed watermark embedding scheme for decoded frames from the compressed audio. Since audio coding is a lossy processing, the embedded watermark must exist after audio compression. Furthermore, the embedded watermark must not affect the audio quality perceptually. In order to satisfy these requirements, the embedding scheme fully considers the human auditory system and the features of audio content. For the decoded frames from the original compressed audio which will be selected to embed watermark, feature parameters are extracted from each selected frame to represent the characteristics of the audio content in that frame. In the meantime, each selected frame will pass through a psychoacoustic model to determine the ratio of the signal energy to the masking threshold. Based on the feature parameters and masking threshold, the embedding scheme for each selected frame is designed. The watermark is embedded into these frames using a multiple-bit hopping and hiding method. The watermarked audio frame will be compressed to generate the compressed audio frame.

In order to correctly detected the watermark from a compressed audio, the frames embedded watermark must be extracted at first. Fig.8 illustrates how to extracted the frames including watermark from a compressed audio. This process is similar to the watermark embedding scheme to select candidate frames to embed watermark. The watermarked compressed audio is first segmented into frames according to the coding algorithm. These frames are decoded and each decoded frame is analyzed by the feature extraction model and the psychoacoustic model. According to the calculated feature parameters and masking threshold, a filter bank is applied to select the frames including watermark information. The watermark will be detected from these frames using the extraction scheme depicted as Fig.9.

Fig.9 illustrate the block diagram of watermark extraction from the selected frames. For each incoming frame, we examined the magnitude (at relevant locations in each audio frame) of the autocorrelation of the embedded signal's cepstrum. From the diagram of autocorrelation of the cepstrum, the bits of a watermark in each frame can be found according to a "power spike" at each delay of the embedded bits. Since we use multiple-bit hopping method to embed the bits into the frames, for detected bits in each frame, they will pass through a matched filter bank that can map the bits into the actual code (1 or 0). Finally, the watermark is recovered by correlate the detected codes with the original watermark.

5. CONCLUSION

Compared with digital image and video watermarking technologies, digital audio watermarking technology provides a special challenge because the human auditory system is extremely more sensitive than human visual system. In this paper, we propose a set of novel content-based digital watermarking methods for PCM audio, synthesis audio and compressed audio. The watermark embedding is highly related to audio content and based on human auditory system. In order to improve the robustness and security of the embedded watermark in PCM audio and compressed audio, a multiple bit hopping technique is applied in watermarking embedding and extraction. The proposed watermarking methods can attain an optimal balance between the audibility and robustness in the embedded audio.

6. REFERENCES

[1] R.B.Wolfgang and E.J.Delp, A Watermark for Digital Images, *Proc. of IEEE Int. Conf. On Image Processing*, Vol.3, pp.219-222, 1996.

[2] I.J.Cox and M.L.Miller, A Review of Watermarking and the Importance of Perceptual Modelling, *Proceedings of SPIE Human Vision and Electronic Imaging*, Vol.3016, pp. 92-99, 1997.

[3] I.J.Cox, J.Kilian and T.Leighton, Secure Spread Spectrum Watermarking for Multimedia, *IEEE Trans. on Image Processing*, 6(12):1673-1687, 1997.

[4] D. Kundur and D. Hatzinakos, A Robust Digital Image Watermarking Method Using Wavelet-Based Fusion, *Proc. of IEEE Int. Conf. On Image Processing*, Vol.1, pp.544-547, 1997.

[5] M.D.Swanson, B.Zhu, A.H.Tewfik and L.Boney, Robust Audio Watermarking Using Perceptual Masking, *Signal Processing*, Vol.66, pp.337-355, 1998.

[6] Y.Yardimci, A.E.Cetin and R.Ansari, Data hiding in speech using phase coding, *ESCA*, *Eurospeech97*, Greece, pp.1679-1682, 1997.

[7] D.Gruhl, A.Lu and W.Bender, Echo hiding, *Proc. of information Hiding Workshop*, Univ. of Cambridge, pp. 295-315, 1996.

[8] S.Sandford et.al., Compression embedding, US Patent 5,778,102, 1997.

[9] B.J.C.Moore, An introduction to the psychology of hearing, Academic Press, Fourth edition, 1997.

[10] M.Kahrs, K.Branderburg, Applications of digital signal processing to audio and acoustics, Kluwer Academic Publishers, 1998.

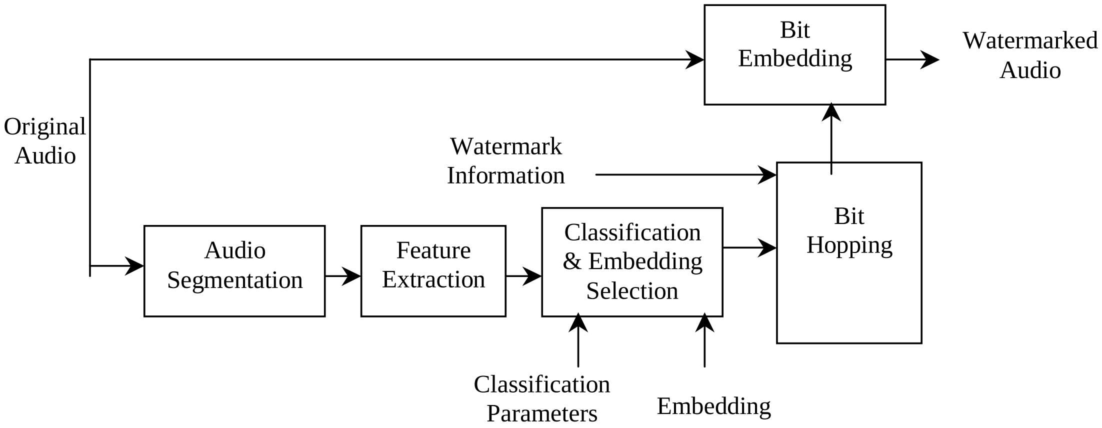

Fig.1 Watermark embedding scheme for PCM audio

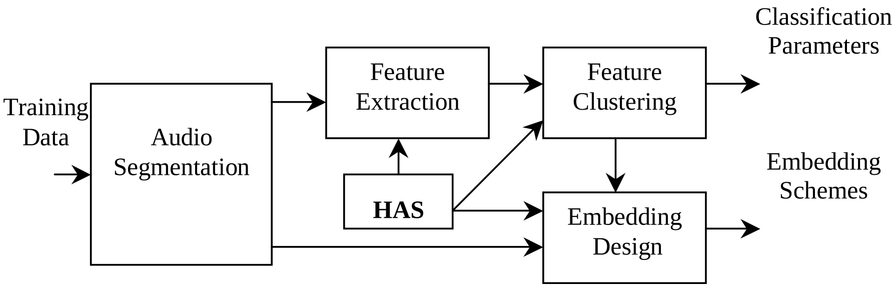

Fig.2 Training and embedding scheme design

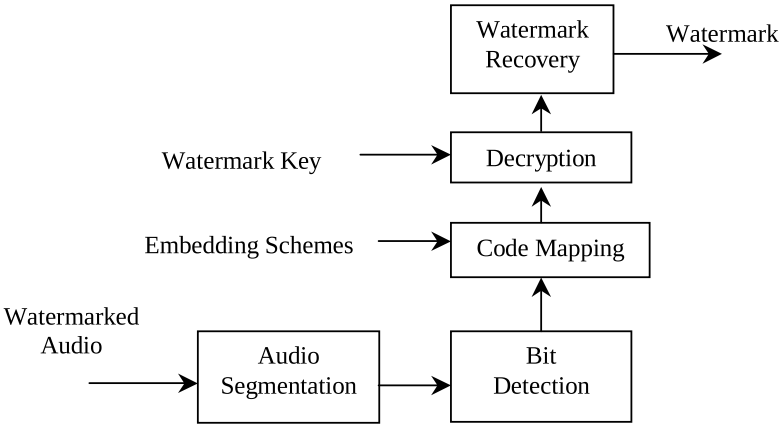

Fig.3 Watermark extracting scheme for PCM audio

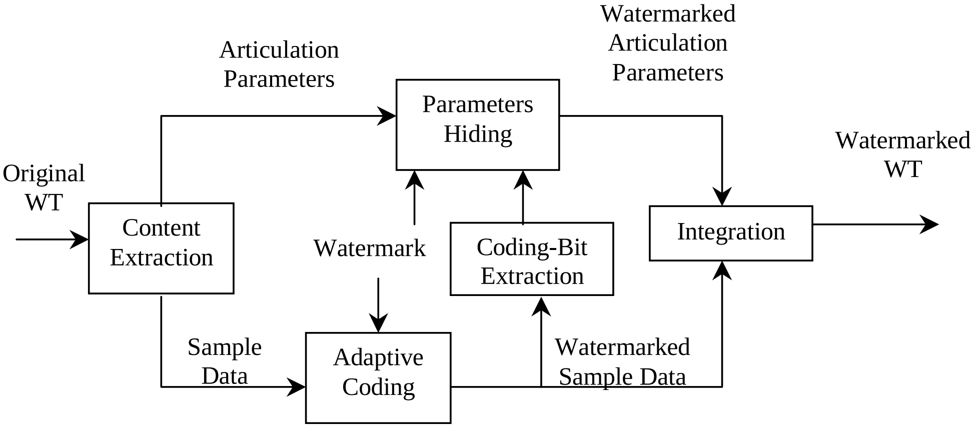

Fig.4 Watermark embedding scheme for synthesis audio

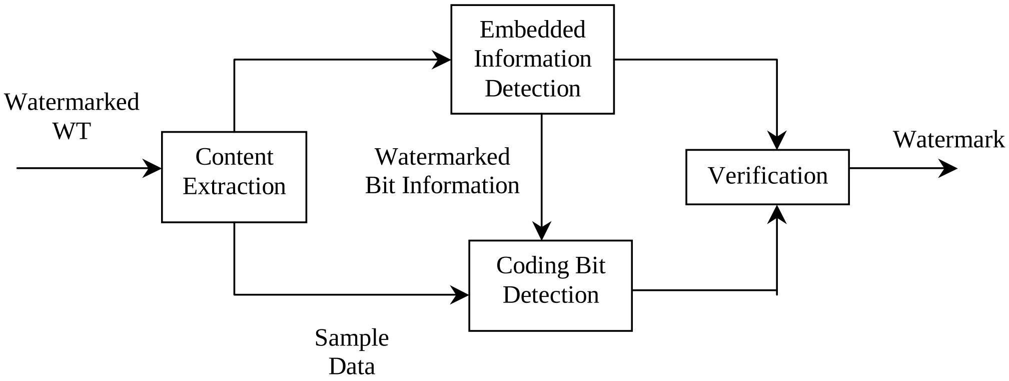

Fig.5 Watermark extracting scheme for synthesis audio

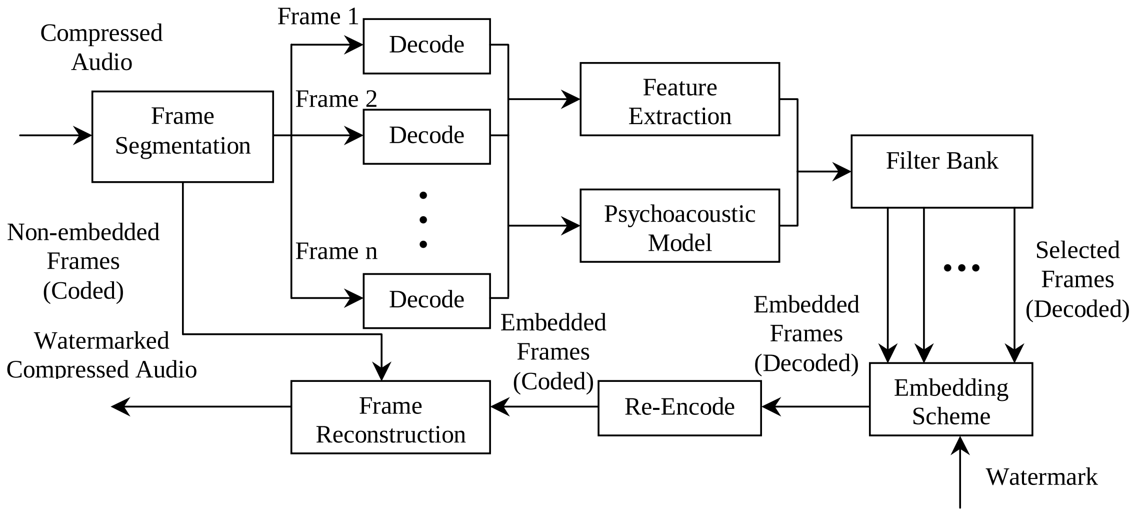

Fig.6 Digital watermark embedding scheme for compressed audio

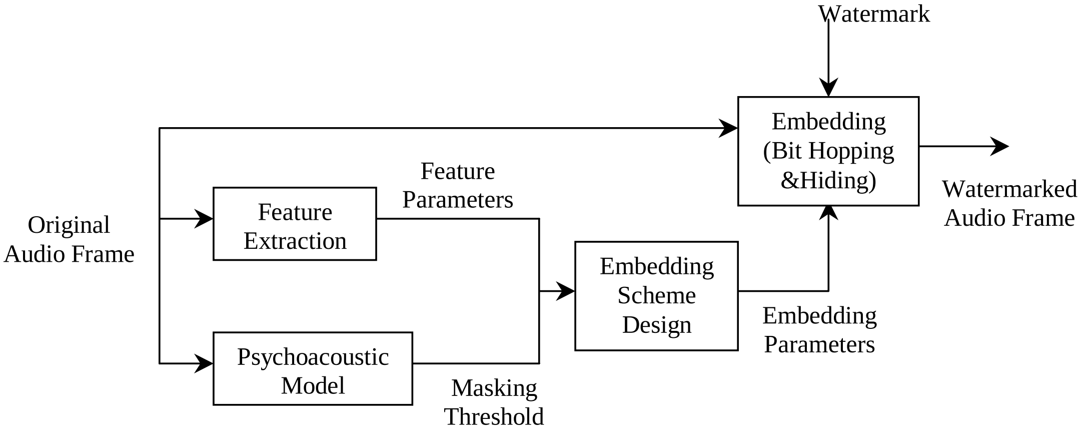

Fig.7 Watermark embedding scheme for single frame

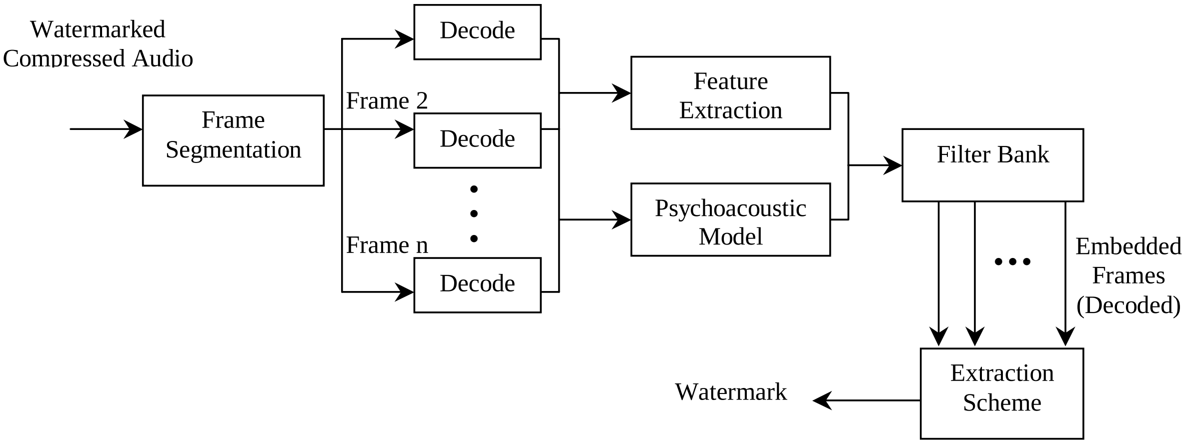

Fig.8 Frames and watermark extraction scheme for compressed audio

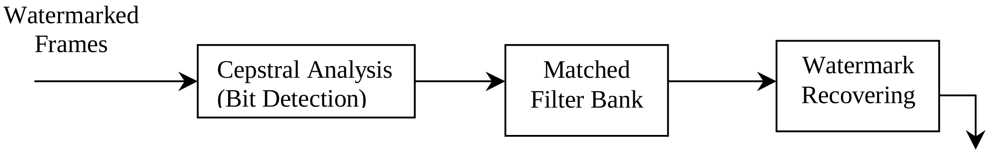

Fig.9 Watermark extraction in uncompressed domain