AN ADAPTIVE SCALABLE WATERMARK SCHEME FOR HIGH-QUALITY AUDIO ARCHIVING AND STREAMING APPLICATIONS

Zhi Li[1, 2], *Qibin Sun*[1] *and Yong Lian*[2]
[1] Institute for Infocomm Research (I^2R), A*STAR, Singapore 119613
[2] Dept. of ECE, National University of Singapore, Singapore 119260
Email: {stuzl, qibin}@i2r.a-star.edu.sg; eleliany@nus.edu.sg

ABSTRACT

In this paper, we present a scalable (i.e. lossy-to-lossless) watermark scheme based on a recently standardized scalable audio coder – AAZ [4]. The proposed framework enables the recovery of the original lossless audio after watermark embedding, and in the meanwhile, is able to make the watermark adaptive such that the watermark distortion to the lossy host audio is minimized. An encryption mechanism is further employed for restricting unauthorized access to lossless audio and watermark removal. Based on this framework, we elaborate its possible applications on high-quality audio archiving and streaming. Experimental results demonstrate the validity of our proposal.

1. INTRODUCTION

Most of the traditional watermark methods protect the media content in a *lossy* way, i.e. once the watermark is embedded into the media content, the distortion introduced is non-invertible. While this distortion is relatively small and inaudible for most of the average users, it is not suitable for applications with lossless quality requirement such as audio archiving, studio, high-quality streaming and high-end consumer electronics applications. Such applications are usually based on lossless compression; therefore, further applying a lossy watermark scheme would render the lossless compression meaningless. These applications motivate us to develop *lossless* watermark in which the original content is still recoverable after watermark embedding.

Some lossless watermark methods have been proposed for images [1, 2, 3]. In [1], Fridrich *et al.* presented a method which is based on losslessly compressing some selected bit-plane to "make some space" for the embedded bits. In [2], Vleeschouwer *et al.* proposed to map the image gray levels to the points on the circle; watermark embedding and image reversion correspond to rotation of the vector that points to the gray level mass center. In [3], Ni's embedding algorithm rests on the gray level value shifting and error correcting coding (ECC). However, all these methods have a common shortness – the watermark strength has to be independent of the local gray level values in order to make it invertible. While this is an acceptable requirement for images, it may not be applicable for audios, because the human auditory system (HAS) is much more sensitive than the human visual system (HVS). Therefore, in order to be more imperceptible and robust, the embedded watermark has to be more adaptive to the host signal.

To circumvent this problem, in this paper, we present an alternative approach – *scalable* watermark which builds a bridge between lossy and lossless watermark. The proposed framework is based on the Advanced Audio Zip (AAZ) coder [4], which has been adopted in the Commission Draft for the on-going scalable audio coding standard under MPEG4 [5]. We therefore call our system "AAZ-WM". The AAZ coder is a two-layer scalable audio coder, in which the core layer is backward compatible with the well known AAC coder [6] whereas the enhancement layer (LLE) is an embedded entropy coder, named Bit-Plane Golomb Code (BPGC) [7], serving the transcoding purpose. The AAZ-WM is fully incorporated into the AAZ coder by watermarking both layers. The watermarks in the two layers are designed in such a way that the watermark in the LLE layer compensates that in the core layer. When merging the two layers at the decoder, the watermarks cancel out and the original lossless audio can be recovered.

Under this framework, the embedded watermark can be made well adaptive to the local host audio. We present three techniques – i) *perceptual shaping* by using a HAS model ii) *host signal compensation* and iii) *adaptive watermark allocation* – to "tailor" the watermark to the local strength such that the watermark impact on the audio quality is minimized.

Note that once the audio is losslessly recovered, the watermark is fully removed. To prevent illegal watermark removal and lossless audio access, we encrypt the LLE layer data. The feasibility of this approach will be further elaborated in Section 3 and experimentally examined in Section 5.

AAZ-WM's another great feature is that the watermark is "scalable" in the sense that the watermark strength is a continuous function of the transcoding rate – when the LLE layer is fully transcoded, the watermark strength in the final decoded audio signal is the strongest; when there is no transcoding and the LLE layer fully compensates the core layer, the watermark is gone. This feature motivates an innovative application – since the watermark strength is a deterministic function of the transcoding rate, we could use the watermark to "blindly" (i.e. in the absence of the original audio signal) assess the audio quality. This application can be employed for the purpose of *quality control* in one of the newly proposed P2P-based music sharing architecture [8].

In brief, the watermark functions as follows. In the compressed domain, it serves the purpose of content annotation, information hiding and etc.; after decoding, it can be used as a "blind" indication of the audio quality. The AAZ-WM framework is also well aligned with the newly proposed concept called Quality of Protection [9] – the scalable watermark could

0-7803-9332-5/05/$20.00 ©2005 IEEE

flexibly protect the audio based on the received quality.

The paper is organized as follows. Section 2 briefly describes the AAZ coder. In Section 3, we present the structure of the AAZ-WM framework. Potential applications of the proposed framework are elaborated in Section 4. Experiment results demonstrating the properties of this system are given in Section 5. Section 6 concludes this paper.

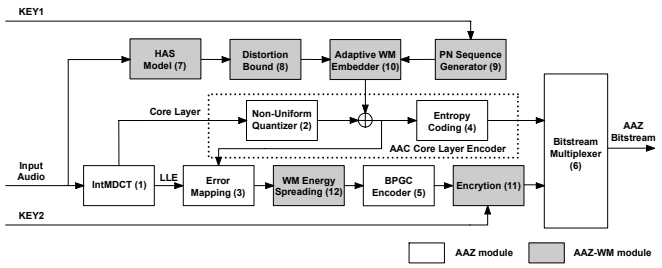

Fig. 1 System diagram of the AAZ encoder (the unshaded blocks) and the incorporated AAZ-WM embedding framework (the shaded blocks)

2. STRUCTURE OF THE AAZ CODER

The AAZ encoder modules (i.e. the unshaded blocks) are illustrated in Fig. 1. For the full reference, please see [5].

The AAZ coder consists of two layers – the core layer which is essentially an AAC encoder [6] and the enhancement layer (LLE). First of all, the time-domain audio signal is losslessly transformed into frequency-domain coefficients *c(k)* by using integer Modified Discrete Cosine Transform (intMDCT) (Module 1). *c(k)* is then passed to the AAC core layer encoder, quantized by an non-uniform quantizer (Module 2), and further Huffman-coded (Module 4) to produce the core layer bitstream. In the LLE layer, the output of the non-uniform quantizer *i(k)* is fed to the error-mapping process (Module 3), in which *i(k)* is reconstructed to be $\tilde{c}(k)$ and then subtracted from *c(k)*, to produce the residue *e(k)*. *e(k)* is further bit-plane coded in the BPGC encoder (Module 5) to generate the LLE bitstream. In order to facilitate transcoding, *e(k)* is bit-plane coded progressively from the MSB plane to the LSB plane. In the final stage, the core layer and LLE layer bitstreams are multiplexed to generate the final AAZ bitstream (Module 6).

3. STRUCTURE OF THE AAZ-WM FRAMEWORK

3.1. Watermark Embedding

Our main design requirement is to embed robust watermark to both layers in a way that when merging the two layers together in the decoder, the lossless audio signal can be recovered. The underlying watermarking algorithm is a spread- spectrum (SS) based method (i.e. spreading one bit energy to many coefficients during embedding and collecting the bit energy by computing correlation during extraction). Refer to Fig. 1, the watermark is embedded to the quantized intMDCT coefficients *i(k) after* the non-uniform quantization in the core layer, such that the embedded watermark survives the quantization process "by nature". Besides, the watermark embedding takes place *before* *i(k)* is fed to the error-mapping process. Consequently, the reconstructed intMDCT coefficient is modified by the watermark, and therefore the residue *e(k)* is also modified. In other words, the LLE layer bitstream is "automatically" watermarked. In fact, the watermark added to the LLE layer is a negative component of that in the core layer, and thus a similar extraction rule (i.e. by computing the correlation between the spreading sequence and the bitstream) as in the core layer can be used to extract the watermark in the LLE layer. More importantly, this negative component cancels out the positive component in the core layer when the two bitstreams are merged in the decoding process, therefore the lossless audio can be recovered. Also note that the watermark is only embedded to the near-DC perceptually significant coefficients.

To ensure the watermark inaudibility, before embedding it to the core layer bitstream, it must be pre-processed to adapt to the local strength. Three techniques are used for this purpose: i) perceptual shaping by using a HAS model, ii) host signal compensation and iii) adaptive watermark allocation. Firstly the HAS model (Module 7) is used to find the bound of maximally allowed distortion imperceptible by human ears. The HAS model output is in the form of signal-to-masking ratio (SMR) for each transform coefficient. Then the lower bound $c_L(k)$ and upper bound $c_U(k)$ of the allowed alteration is computed as:

$$\begin{aligned} c_L(k) &= \tilde{c}(k)[1-\varepsilon/\sqrt{SMR(k)}] \\ c_U(k) &= \tilde{c}(k)[1+\varepsilon/\sqrt{SMR(k)}] \end{aligned} \quad (1)$$

where ε is the global strength bound, usually set to 1. Next, by passing $c_L(k)$ and $c_U(k)$ through the non-uniform quantizer, the lower and upper bound of allowed alteration for the quantized intMDCT coefficients, $i_L(k)$ and $i_U(k)$, can be computed respectively (Module 8).

In the adaptive watermark embedder (Module 10), the amount of watermark to be embedded for each watermark message bit is firstly computed. Since for SS-based watermark method, the host signal influence is the main source that affects the watermark extraction correctness, the embedded amount of watermark must be able to compensate the host signal influence. Based on the idea in [10], we propose the following *host signal compensation* strategy – as in the embedding process, we have the complete knowledge of the host signal, we can effectively calculate the correlation between the host signal and the spreading sequence, estimate its impact on the watermark extraction, and thereby determine the amount of watermark needed in order to maintain a pre-determined level of extraction correlation value. In other words, the watermark embedding works on a "supply-on-demand" basis, thus its impact on the audio quality is minimized.

Next, a spreading sequence is generated according to KEY1, which the secrecy of the hidden information fully depends on (Module 9). The computed watermark is to be adaptively allocated with reference to the spreading sequence, which determines the polarity (i.e. positive or negative) of the watermark. Let us define *watermark capacitance* for coefficient

i(k) as the maximum amount of watermark allowed for *i(k)*, as bounded by $i_L(k)$ or $i_U(k)$. (If the spreading sequence bit is -1, $i_L(k)$ is the bound; otherwise the bound is $i_U(k)$.) The *adaptive watermark allocation* strategy is based on the following rule: the watermark embedded in the coefficient with higher watermark capacitance must be less audible than others, thus have higher priority for watermark allocation. Based on this rule, the watermark is allocated iteratively, until either the needed amount of watermark is fully allocated, or there is no more capacitance for allocation.

We notice that the watermarked residue *e(k)* in the LLE layer displays significant magnitude increase. This non-adaptiveness will cause inefficiency in the subsequent BPGC encoding. As a solution, the excessive bit-planes of the watermarked residue are spread to residues in a wider frequency range in each intMDCT block (Module 12). Fig. 2 demonstrates the spreading effect.

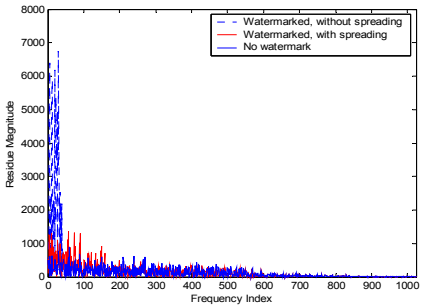

Fig. 2 Watermarked residue energy spreading. After spreading the watermark energy from coefficients indexed from 0 - 39 to 0 - 159, the watermark becomes much less visible. Further improvement can be achieved by spreading the energy to even wider range in the spectrum.

3.2. Watermark Extraction

The watermark extraction can be performed in either the core layer or the LLE layer of the compressed AAZ bitstream. Alternatively, the watermark extraction can also be performed on the decoded raw audio. Due to page limits, we omit the details here. Interested readers can refer to [11] for more information.

3.3. LLE Layer Access Control

To restrict unauthorized watermark removal and access to the lossless audio, the LLE layer bitstream is encrypted using KEY2 before it is multiplexed with the core layer bitstream (Module 11). Consider an authorized user who possesses KEY2 and has access to the watermarked and encrypted AAZ bitstream. By using a modified AAZ decoder (i.e. with the LLE decryption mechanism) and the key, he can conveniently recover the lossless audio and remove the watermark. For another unauthorized user who does not have the key, if he uses a standard AAZ decoder, the encrypted LLE bitstream will be decoded into some random noise, so the watermark is not compensated, and therefore still preserved in the audio. Note that since the LLE layer bitstream is perceptually less important for the audio quality, the random noise does not introduce severe distortion, therefore the unauthorized user can still decode an audio with acceptable quality. Experiments in Section 5 will demonstrate these properties.

The encryption algorithm must be carefully designed such that the transcoding is applicable to the encrypted LLE bistream. Some related work on JPEG2000 secure transcoding has been made in [12]. In this paper, a simple stream cipher algorithem is used for illustration purpose: we generate a PN-sequence according to KEY2. The BPGC-encoded bitstream is XORed with this PN-sequence to produce the encrypted bitstream.

4. APPLICATIONS

The proposed watermark framework meets the requirement of many high-end applications. In this section, we elaborate the following potential applications, in which the AAZ-WM can be employed – i) lossless audio archiving and ii) quality-differentiated online audio streaming. The schematic diagram of these applications is shown in Fig. 3.

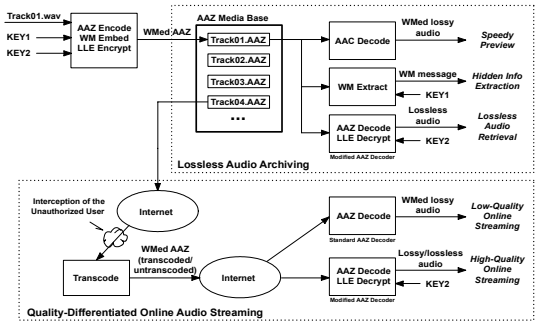

Fig. 3 AAZ-WM applications of lossless audio archiving and quality-differentiated online audio streaming

Lossless Audio Archiving – In Records Company or studio, thousands of tracks needs to be stored in lossless format. In the meanwhile, watermark is preferred to embed information such as unique identification numbers, since it provides resilience to malicious information alteration. In this scenario, AAZ-WM provides a good solution. The raw audios are encoded into AAZ format (both watermarked and encrypted) and then stored in the AAZ media base. When browsing for a particular track, low-quality audio for speedy preview is generated by decoding the core layer bitstream with an AAC decoder. For authorized users with KEY1, the hidden information can be extracted. For lossless audio retrieval, the authorized user can use KEY2 to decrypt the LLE layer and thereby remove the watermark and recover the lossless audio.

Quality-Differentiated Online Audio Streaming – In this application, the watermarked / encrypted AAZ bitstream is generated on the server side. During audio streaming, depending on the traffic conditions, a remote transcoding server may or may not transcode the LLE bitstream. On the client side, an unregistered user could use a standard AAZ decoder to produce low-quality and watermarked audio. However, if the user is not satisfied with the quality and wants even better service, the user can pay and register as a VIP user to obtain a key. With a modified AAZ decoder and the key, the user can enjoy high-quality audio services. Note that "watermark + encryption"

solution prevents illegal interception of the lossless audio. If an unauthorized user intercepts the bitstream somewhere before the transcoding server and attempt to recover the lossless audio, he will fail since he can only generate the watermarked low-quality audio. In addition, when the watermark embeds anti-collusion fingerprint [13], the preserved watermark will help to trace the source of the interception.

5. EXPERIMENT RESULTS

Some useful experiment results demonstrating the properties of the AAZ-WM framework are presented in this section. The test sequence is single-channel and sampled at 48 kHz.

The rate-distortion characteristic is firstly examined. Fig. 4 presents the curves of decoding an unwatermarked AAZ bitstream using the standard AAZ decoder, and decoding a watermarked AAZ bitstream (with the LLE layer encrypted) using the standard AAZ decoder and modified AAZ decoder, respectively. The curves show that the watermarked audio decoded with the modified AAZ decoder has similar rate-distortion characteristic as the unwatermarked audio, except at low bitrate, there is a degradation of about 5 dB. However, after conducting subjective quality measurement, we found that since we have used a HAS model, the 5-dB degradation is merely due to the changes in perceptually insignificant components, and thus does not harm the perceptual quality. The SNR increases as the streaming bitrate increases, and eventually the audio becomes lossless when no transcoding occurs. Also note that the watermarked and encrypted audio decoded with the standard AAZ decoder still has SNR of around 25dB, which is perceptually acceptable.

The watermark scalability is demonstrated in terms of the correlation value (i.e. watermark strength) as a function of the streaming bitrate. In Fig. 5, when the extraction is performed on an audio decoded with the modified AAZ decoder (with LLE decryption mechanism), the correlation value decreases as the streaming bitrate increases, which agrees with what we have expected. When the extraction is on an audio decoded with the standard AAZ decoder, the correlation value remains high, suggesting that the watermark is preserved.

For more experiment results regarding the watermark payload, robustness, data size expansion property etc., please refer to [11] for more details.

6. CONCLUSION

We have presented a novel scalable watermark scheme for high-quality audio archiving and streaming applications. The favorable features of this scheme – recovery of the original lossless audio, watermark adaptiveness and watermark scalability – are elaborated and experimentally demonstrated.

REFERENCE

[1] J. Fridrich, M. Goljan and R. Du, "Invertible Authentication", in *Proc. SPIE Security and Watermarking of Multimedia Contents*, San Jose CA, Jan. 23-26, 2001.

[2] C. De Vleeschouwer, J. F. Delaigle and B. Macq, "Circular Interpretation of Bijective Transformations in Lossless Watermarking for Media Asset Management", *IEEE Tran. Multimedia*, vol. 5, pp. 97-105, March 2003.

[3] Z. Ni, Y. Q. Shi, N. Ansari, W. Su, Q.B. Sun and X. Lin, "Robust Lossless Image Data Hiding", *Proc.IEEE Intl. Conf. on Multimedia and EXPO*, 2004.

[4] R.S. Yu, X. Lin, S. Rahardja and C.C. Ko, "A Scalable Lossy to Lossless Audio Coder for MPEG-4 Lossless Audio Coding", *Proc. IEEE Intl. Conf. on Acoustics, Speech, and Signal Processing*, 2004.

[5] "Scalable Lossless Coding (SLS)", SC29/WG11/N6673, Text of 14496-3:2001/PDAM 5, Seattle, USA, July, 2004.

[6] M. Bosi, et al, "ISO/IEC MPEG-2 Advanced Audio Coding", J. *Audio Eng. Soc.*, Vol. 45, No.10, pp. 789-814, 1997.

[7] R. Yu, C.C. Ko, S. Rahardja and X. Lin, "Bit-plane Golomb code for sources with Laplacian distributions," *Proceeding of ICASSP* 2003.

[8] T. Kalker, D.H.J. Epema, P.H. Hartel, R. L. Lagendijk, and M. V. Steen, "Music2Share – Copyright-Compliant Music Sharing in P2P Systems", *Proceedings of the IEEE*, June 2004.

[9] C. S. Ong, K. Nahrstedt and W. Yuan, "Quality of Protection for Mobile Multimedia Applications", *Proc.IEEE Intl. Conf. on Multimedia and EXPO* 2003.

[10] H. S. Malvar, D. A. F. Florencio, "Improved Spread Spectrum: A New Modulation Technique for Robust Watermarking", *IEEE Trans. Signal Processing*, Vol. 51, No. 4, April 2003.

[11] Z. Li, Q.B. Sun, Y. Lian and R.S. Yu, "A Scalable Watermark Scheme for the Scalable Audio Coder", to appear in *IEEE Intl. Conf. on Communications*, 2005.

[12] S. Wee and J. Apostolopoulos, "Secure Transcoding with JPSEC Confidentiality and Authentication", *Proc. IEEE Intl. Conf. on Image Processing*, 2004.

[13] W. Trappe, M. Wu, Z.J. Wang and K.J.R. Liu, "Anti-collusion Fingerprinting for Multimedia", *IEEE Trans. Signal Processing*, Vol. 51, No. 4, April 2003.

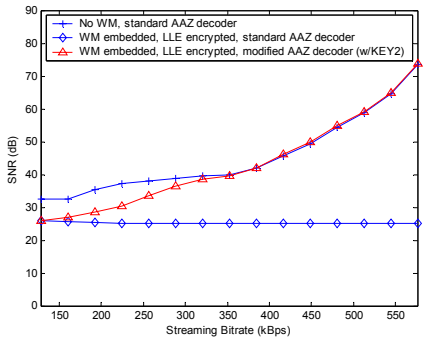

Fig. 4 Comparison of rate-distortion characteristics. The SNR of the decoded audio is calculated w.r.t. to the original audio. The streaming bitrate is the sum of the core layer bitrate (fixed at 128kbps) and LLE layer bitrate (from 0kbps onwards).

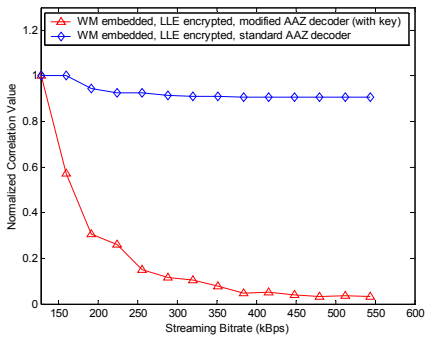

Fig. 5 Correlation value (i.e. watermark strength) vs. streaming bitrate. The watermark correlation value is extracted from decoded raw audio.