

ROBUST DATA-HIDING IN AUDIO

Hafiz Malik, Ashfaq Khokhar, Rashid Ansari

Dept. of Electrical and Computer Engineering University of Illinois at Chicago, Illinois, USA

ABSTRACT

A novel high capacity data hiding technique for digital audio is proposed. Imperceptibility of the embedded data is ensured based on the masking property of the human auditory system (HAS). Audio signal is decomposed into subband signals, some of which are selected for embedding data using finite-length impulse response approximations to allpass digital filters. Data detection is based on finding the filter pole-zero locations, which is achieved by power spectrum estimation of the data embedded audio signal. Performance of the proposed scheme is evaluated for different data encoding strategies. The proposed method is robust to desynchronization attacks as well as other standard data manipulation attacks.

1. INTRODUCTION

Digital media production and distribution has witnessed dramatic growth in recent years. Widespread access to multimedia data on the Internet, connectivity via high-speed networks, and availability of low-cost and reliable storage devices has made it possible to replicate and distribute digital information easily. This has created a need for protection and enforcement of intellectual property rights for digital media to prevent its illegal copying/reproduction. The need has spurred research in data hiding in digital media due to its applications in watermarking, annotation, and steganography [2].

The development of data-hiding methods requires many design and quality tradeoffs. An important requirement is that embedded data should be imperceptible. In addition data embedding should be robust to standard digital data manipulations such as lossy compression, noise addition, and sampling rate conversion. Moreover, embedded data should be tamperproof against adversary attacks.

Perception-based data hiding schemes for audio are influenced by the masking property of HAS. In the past, different perception-based algorithms were proposed for data hiding/watermarking for audio data [4 - 9]. These algorithms can be broadly classified according to the underlying technique of data embedding: perceptual masking [4, 5, 9], direct sequence spread spectrum (DSSS) [4], and phase coding [4,7,8]. Algorithms based on phase coding work well as far as imperceptibility of the embedded data is concerned, but suffer from some limitations e.g. low robustness to standard data

manipulations and limited capacity for payloads, i.e. the amount of information embedded. The phase coding technique proposed in [4] embeds only 16-32 bits of data in audio samples of one-second duration, and detection performance deteriorates rapidly in the presence of random noise. The algorithm based on echo-based coding [6] can embed about 40-50 bits of data in one-second duration of an audio signal.

In this paper, a novel perception-based data hiding method is proposed. The following properties of HAS are exploited in this method: the magnitude distortion at a specific frequency in an audio signal is inaudible if it is below masking threshold; and human perception is less sensitive to absolute phase changes in a certain frequency range [1]. Not all of the customary full range of audible frequencies, i.e. 20 Hz ~ 20 kHz, is suitable for data embedding. In the higher frequency range ($\approx f > 10$ kHz) detection of small magnitude changes is unreliable due to insignificant signal energy. On the other hand human perception is more sensitive to phase distortion in the lower frequency range ($\approx f < 4.0$ kHz). The frequency range (i.e. $4.0 < f < 10.0$ kHz) is therefore suitable for making embedded data imperceptible and robust to the standard manipulations.

The signal content in the above range is partitioned into subband signals using discrete wavelet packet analysis filter bank (DWPA-FB). Data is embedded in selected subband signals by introducing inaudible magnitude and phase distortion using finite-length impulse response (FIR) approximations of allpass filters (APFs). Data is detected by estimating the parameters (pole-zero) of the APF by estimating the power spectrum of the audio. In our method the power spectrum is estimated using parametric signal models, i.e. moving average and autoregressive models. The performance of this method is evaluated for detecting data that is embedded in an audio clip using binary and 4-ary encoding and is subjected to signal manipulations such as addition of noise, lossy compression, resampling, and random chopping. Compared with existing methods [5-9] the proposed technique is shown to embed 5-8 times more data for binary encoding and twice for 4-ary encoding.

2. FIR APPROXIMATION OF APF

An APF is suitable for data embedding because the phase distortion it introduces in the chosen frequency range is largely inaudible. Let the frequency response of the APF

be $H(e^{j\omega}) = ke^{j\phi(\omega)}$. Estimation of APF parameters from the processed audio consists of finding local extrema in the magnitude spectrum of the processed audio along radial lines of pole-zero locations of APF.

The transfer function $H_{AP}(z)$ of a stable and causal first-order allpass filter is

$$H_{AP}(z) = \frac{z^{-1} - \alpha^*}{1 - \alpha z^{-1}} \quad (1)$$

where $\alpha \in \mathbb{C}$ and $|\alpha| < 1$ and the region of convergence is $|\alpha| < |z|$. The transfer function of a higher order APF can be expressed as a product of first-order allpass sections specified in Eq.(1). An allpass filter has an infinite-duration impulse response (IIR). Data is embedded by introducing controlled phase distortion using a fixed set of pole-zero locations. We use an FIR approximation of length L of an n^{th} order APF. This introduces both magnitude and phase distortion. The magnitude distortion tends to zero as $L \rightarrow \infty$. This is shown below.

Consider a stable, causal first order APF defined in Eq (1)

$$\begin{aligned} H_{AP}(z) &= \frac{z^{-1} - \alpha^*}{1 - \alpha z^{-1}} = (z^{-1} - \alpha^*) \times \frac{1}{1 - \alpha z^{-1}} \\ &= (z^{-1} - \alpha^*) \left[\sum_{k=0}^{\infty} (\alpha z^{-1})^k \right] \\ &= (z^{-1} - \alpha^*) \left[\sum_{k=0}^L (\alpha z^{-1})^k + \sum_{k=L+1}^{\infty} (\alpha z^{-1})^k \right] \\ &= (z^{-1} - \alpha^*) \left[\sum_{k=0}^L (\alpha z^{-1})^k \right] + (z^{-1} - \alpha^*) \left[\sum_{k=L+1}^{\infty} (\alpha z^{-1})^k \right] \quad (2) \\ &= \left(\frac{z^{-1} - \alpha^*}{1 - \alpha z^{-1}} \right) (1 - (\alpha z^{-1})^{L+1}) + (\alpha z^{-1})^{L+1} \left(\frac{z^{-1} - \alpha^*}{1 - \alpha z^{-1}} \right) \quad (3) \end{aligned}$$

First term on right hand side in Eq (3) is FIR approximation of an APF referred as $H_{FIR-AP}(z)$. $H_{FIR-AP}(z)$ can be expressed as,

$$H_{FIR-AP}(z) = H_{AP}(z) (1 - (\alpha z^{-1})^{L+1}) \quad (4)$$

The factor $(1 - (\alpha z^{-1})^{L+1})$, introduces $L + 1$ zeros at $z_i = \alpha e^{j2\pi i/(L+1)}$ $i = 0, 1, \dots, L$. These $L + 1$ zeros are uniformly distributed on the circle $|z| = \alpha$. The zero for $i=0$, i.e. $z_0 = \alpha$, cancels the pole at the same location, therefore $H_{FIR-AP}(z)$ has $L + 1$ zeros altogether, where L zeros are located at $z_i = \alpha e^{j2\pi i/(L+1)}$ $i = 1, 2, \dots, L$; and remaining one zero at $|z| = 1/\alpha$. The transfer function $H_{AP}(z)$ of a single pole-zero pair is obtained from $H_{FIR-AP}(z)$ as L goes to infinity. The nature of distortion for different L is illustrated in Figure 1.

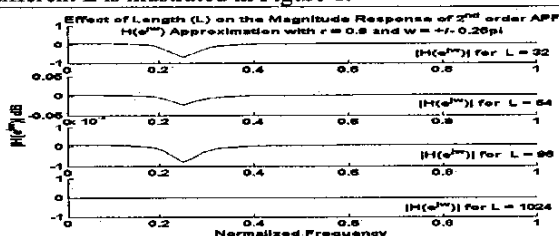


Figure1: Magnitude Response of APF $H(e^{j\omega})$ Approximation for Different values of Length (L).

An n^{th} order APF is used in our method for embedding data. The n^{th} order APF is realized with a cascade of $n/2$ second order allpass filters. The use of

cascaded form realization reduces the effect of quantization of APF coefficients. Parameter α_i of the transfer function $H_{APi}(z)$ used for data embedding is defined as: $\alpha_i = r e^{j\omega_i}$, where $0 < r < 1$, $0 < \omega_i < \pi$, where $i = 0, 1$ in binary encoding, and $i = 0, 1, 2, 3$ in 4-ary encoding. The transfer function $H_{APi}(z)$ of the APF used for data embedding is expressed as:

$$H_{APi}(z) = \left(\frac{(z^{-1} - \alpha_i^*)(z^{-1} - \alpha_i)}{(1 - z^{-1}\alpha_i^*)(1 - z^{-1}\alpha_i)} \right)^{n/2} \quad n = 2, 4, \dots \quad (5)$$

Note that $H_{APi}(z)$ has $n/2$ poles at each α_i and α_i^* , and $n/2$ zeros at each $1/\alpha_i$ and $1/\alpha_i^*$ locations. The parameter α_i of the transfer function $H_{APi}(z)$ for binary and 4-ary encoding used in data embedding are given in Table 1.

Table1: APF parameters for binary and 4-ary schemes

Binary Encoding Scheme	4-ary Encoding Scheme				
	r	ω	R	Ω	
α_0	0.9	0.25π	α_0	0.95	0.2π
			α_1	0.95	0.4π
α_1	0.9	0.75π	α_2	0.95	0.6π
			α_3	0.95	0.8π

The pole-zero layouts of the 2nd order allpass filters used in binary encoding and 4-ary encoding are illustrated in Figure 2 and 3 respectively.

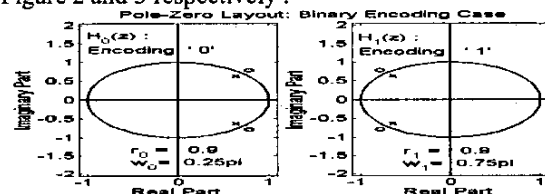


Figure2: Pole-Zero Layout of $H_{AP}(z)$ for Binary Encoding

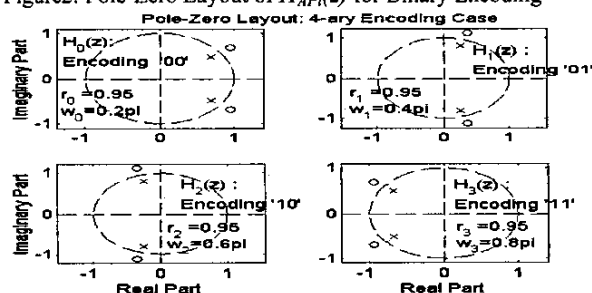


Figure3: Pole-Zero Layout of $H_{AP}(z)$ for 4-ary Encoding

3. DATA EMBEDDING

The data embedding process begins by dividing the input audio signal into non-overlapping blocks of N samples. Each block is then decomposed into 2^l subbands using l -level DWPA-FB, where l and N are positive integers. According to the human auditory perceptual model, subbands corresponding to 4 to 10 kHz range are relatively less sensitive to the phase distortion and robust to data manipulations. These subbands are selected for data embedding. To this end ten subbands (from subband # 6 to subband # 15 at 44.1 k Hz sampling rate with $l=5$)

are selected for data embedding for each block. One bit of data is embedded in each subband for binary encoding scheme and two bits of data for 4-ary scheme. For example, in binary encoding, bit 'm', $m=0,1$, is embedded by passing a selected subband through an APF with transfer functions $H_m(z)$.

In order to cope with the desynchronization attacks, such as signal chopping, synchronization locations called salient points are identified for inserting data. Salient points are attack-sensitive locations in the input audio that can be used for synchronization. The salient points correspond to audio features to which HAS is sensitive such as fast energy climbing points. If an adversary alters these features, audible distortion is introduced [2]. In our method we adopt the salient point extraction method described by C.-P. Wu et al. [2]. In our implementation thresholds Th_1 , Th_2 and Th_3 of [2] are suitably set in order to ensure 1- 2 salient points per second. We set $r = 4000$ samples, $Th_1 = 2$, $Th_2 =$ mean energy of one second duration audio window around n , and $Th_3 = 200$ samples. The following steps outline the data-embedding scheme:

- A list of salient points is extracted for a given audio signal using the method described in [2].
- Starting with the first salient point, the audio signal is segmented into non-overlapping frames of N-samples.
- Each frame is decomposed using a 5-level DWPA-FB and ten subbands (from subband 6 to 15) are selected for data embedding.
- One bit of channel-encoded data is embedded in each selected subband in the case of binary encoding, and two bits for 4-ary encoding.
- All frames that contain a salient point are embedded with synchronization code (using a suitable bit sequence).
- Finally each frame is re-synthesized using discrete wavelet packet synthesis filter bank (DWPS-FB).

A detailed block diagram of the data embedding process is given in Figure 4.

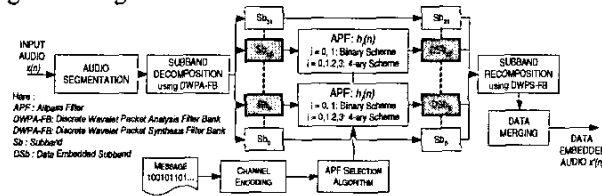


Figure4: Block Diagram of the Data Embedding Scheme

4. DATA DETECTION

The detector first analyzes the data-embedded input audio to extract the list of salient points using the method described in [2]. Then, starting from the first salient point, the input audio signal is segmented into non-overlapping frames of N-samples. Each frame is decomposed into subband signals using a 5-level DWPA-FB (as in the case of data embedding) after which ten-subbands (from # 6 to

15) are selected for data recovery. Data recovery consists of power spectrum estimation of the selected subband signals using a priori knowledge of the signal model followed by APF parameter estimation to recover the embedded information.

4.1 SPECTRUM ESTIMATION

Our parametric spectrum estimation approach assumes an appropriate model of the process based on a priori knowledge of the signal. We know that the subband signals have been processed by an FIR approximation of an APF. Therefore, both autoregressive (AR) and moving average (MA) signal models of sufficient order can be used [3] for spectrum estimation.

An autoregressive process, $x(n)$, can be represented as the output of an all-pole filter excited by unit variance white noise. The estimated power spectrum of a p^{th} order AR process is

$$\hat{P}_{AR}(e^{j\omega}) = \frac{|\hat{b}(0)|^2}{|1 + \sum_{k=1}^p \hat{a}_p(k) e^{-j\omega k}|^2} \quad (6)$$

where $\hat{a}_p(k)$ and $\hat{b}(0)$ are the estimates of the process model parameters. These $p+1$ estimated can be obtained from the data methods such as autocorrelation method, covariance method, modified covariance method, Burg algorithm etc [3]. We use Burg algorithm for p^{th} order AR model parameter estimation.

A moving average process, $x(n)$, can be generated by exciting a q^{th} order FIR filter by unit variance white noise. The estimated power spectrum of q^{th} order MA process is,

$$\hat{P}_{MA}(e^{j\omega}) = |\sum_{k=0}^{q-1} \hat{b}_q(k) e^{-j\omega k}|^2 \quad (7)$$

where $\hat{b}_q(k)$ are the estimates of the process model parameters. We use Durbin's method [3] for q^{th} order MA model parameter estimation.

The next step for data recovery is to estimate APF parameter $\hat{\alpha}(r, \omega)$ from the estimated spectrum $\hat{P}(e^{j\omega})$.

4.2 ALLPASS FILTER PARAMETER ESTIMATION

For APF parameter $\hat{\alpha}(r, \omega)$ estimation we need to estimate \hat{r} and $\hat{\omega}$ from the estimated spectrum $\hat{P}(e^{j\omega})$ as α is function of r and ω . In our method r is fixed for all APF parameters and only the frequency ω is varied for the information encoding schemes given in Table 1. Therefore, we need to estimate frequency ω for $\hat{\alpha}(r, \omega)$ estimation. Frequency ω is estimated from the estimated spectrum $\hat{P}(e^{j\omega})$ based on the results of FIR approximation of APF (discussed in Section 2).

We know, from Section 2, that the FIR approximation introduces magnitude distortion that manifests as a local extremum at $z = e^{j\omega}$. The extremum becomes more pronounced as we traverse from $re^{j\omega} \rightarrow (1/r)e^{j\omega}$. Moreover, this extremum is stronger and evident for small value of the duration L of the FIR approximation to the APF (as illustrated in Figure1). Therefore, to estimate frequency $\hat{\omega}$ we need to estimate consistent local extrema from the estimated spectrum $\hat{P}(e^{j\omega})$. For more accurate

estimate of $\hat{\omega}$ spectra based on both AR signal model as well as MA signal model are used. Finally nearest neighborhood hypothesis testing is applied to decode embedded information.

Hypothesis Testing: binary decoding scheme,

$$\begin{aligned} H_0 &: |\hat{\omega} - \omega_0| < Th = 0 & Th = 0.15\pi \\ H_1 &: |\hat{\omega} - \omega_1| < Th = 1 \end{aligned} \quad (8)$$

Hypothesis Testing: 4-ary decoding scheme

$$\begin{aligned} H_0 &: |\hat{\omega} - \omega_0| < Th = 0.0 & Th = 0.05\pi \\ H_1 &: |\hat{\omega} - \omega_1| < Th = 0.1 \\ H_2 &: |\hat{\omega} - \omega_2| < Th = 1.0 \\ H_3 &: |\hat{\omega} - \omega_3| < Th = 1.1 \end{aligned} \quad (9)$$

After estimating the coded bit sequence, channel decoding is applied to recover the original message. The block diagram in figure 5 illustrates the data detection in detail.

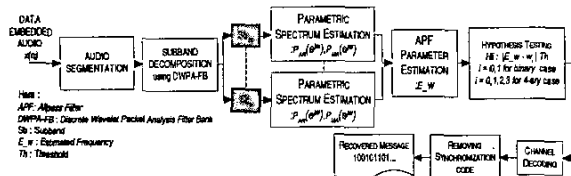


Figure 5: Block Diagram of the Data Detection Process

5. SIMULATION RESULTS

Imperceptibility and robustness are the two benchmarks used for performance evaluation of the proposed data hiding scheme. Robustness is measured based on the probability of error in the received data under different degradations: a) noise addition, b) lossy compression, c) random chopping and d) resampling; for both encoding schemes. Probability of error P_e is defined as

$$P_e = \left(1 - \frac{\text{Number of Bits Correctly Detected}}{\text{Number of Bits Embedded}} \right) \times 100 \quad (10)$$

For robustness test we have following observations:

- White gaussian noise with 0% to 100% of the audio power is added to the data-embedded audio signal. The probability of error P_e of the recovered data for different values of signal to noise ratio (SNR in dB) and for both encoding schemes is given in Figure 6.

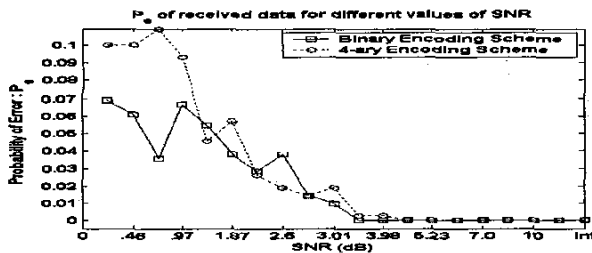


Figure 6: P_e vs. SNR Plot for Both Encoding Schemes

- Data embedded audio signal is compressed using MPEG layer III coder [10]. Despite the lossy compression the P_e value of the recovered data was below 1% for binary encoding scheme. The P_e value for 4-ary encoding was higher.
- To test the robustness against desynchronization attacks, 2 – 4 samples out of every 100 samples of the

data-embedded audio are dropped randomly, probability of error P_e for the resulting audio was error for both encoding schemes.

- In this test data-embedded audio is first down-sample to 22.05 kHz and then interpolated to 44.1 kHz. Probability of error P_e of the recovered data after resampling was 0 % for binary encoding case, and 0.5% for 4-ary encoding case.

6. CONCLUSION

We have proposed a novel method of high-capacity data hiding based on the controlled inaudible frequency response distortion introduced in selected subbands of an audio signal using FIR approximations of an n^{th} order APF. The proposed technique is robust to standard data manipulations yielding low error probability. The error probability performance can be improved by using channel coding with higher error correction capability. Performance was evaluated with informal listening tests. We are seeking approval for subjective test for evaluations using formal listening tests. We are currently investigating the extension of the proposed scheme for its potential for copy-control and digital watermarking applications for audio as well as for other multimedia data types such as images and videos.

7. REFERENCES

- [1] E. Zwicker, and H. Fastl, "Psychoacoustics: Facts and Models," Springer-Verlag, Berlin, 1999.
- [2] C.-P. Wu, P.-C. Su, and C.-C. J. Kuo, "Robust Audio Watermarking for Copyright Protection," *SPIE's 44th Ann. Meet. Adv. Sig. Proc. Alg. Arch. Impl. IX (SD39)*, Denver, Colorado, July 18-23, 1999.
- [3] M. H. Hayes, "Statistical Digital Signal Processing and Modeling," John Wiley & Sons, Inc., NY, 1996.
- [4] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding," *IBM Systems Journal*, vol.35, nr. ¾, 1996.
- [5] P. Bassia and I. Pitas, "Robust audio watermarking in the time domain," *Proc. Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP99)*, 1999.
- [6] M. F. Mansour, and A. H. Tewfik, "Time-scale invariant audio data embedding," *Proc. IEEE Int. Conf. on Multimedia & Expo (ICME'01)*, Japan, August 2001.
- [7] Y. Yardimci, A. E. Cetin, and R. Ansari, "Data hiding in speech using phase coding," *Proc. Eurospeech Conference, 1997*.
- [8] D. Kirovski, and H. S. Malvar, "Spread Spectrum watermarking of Audio Signals," *IEEE Trans. Signal Proc.* Vol. 51, no. 4, pp. 1020-1033, April, 2003.
- [9] C. I. Podilchuk and E. J. Delp, "Digital watermarking algorithms and applications," *IEEE Signal Processing Magazine*, pp. 33-45 July, 2001.
- [10] D. Pan, "A Tutorial on MPEG/ Audio Compression," *IEEE Multimedia Magazine*, 2(2):60-74, 1995.