International Journal of Electrical, Electronic and Communication Sciences
ISSN: 2517-9438
Vol:1, No:2, 2007

Watermark Bit Rate in Diverse Signal Domains

Nedeljko Cvejic, Tapio Seppänen

***Abstract*—**A study of the obtainable watermark data rate for information hiding algorithms is presented in this paper. As the perceptual entropy for wideband monophonic audio signals is in the range of four to five bits per sample, a significant amount of additional information can be inserted into signal without causing any perceptual distortion. Experimental results showed that transform domain watermark embedding outperforms considerably watermark embedding in time domain and that signal decompositions with a high gain of transform coding, like the wavelet transform, are the most suitable for high data rate information hiding.

***Keywords*—**Digital watermarking, information hiding, audio watermarking, watermark data rate.

I. Introduction

Information hiding techniques have developed a strong basis an area with a growing number of applications like digital rights management, covert communications, annotations, etc. In all applications given above, data hiding techniques have to satisfy two basic requirements. The first requirement is perceptual transparency, i.e. cover object (object not containing any additional data) and stego object (object containing secret message) must be perceptually indiscernible. The second constraint is high data rate of the embedded data.

The simplest visualization of the requirements of information hiding in digital audio is so called magic triangle, given in Figure 1. This model is convenient for a visual representation of the required trade-offs between the capacity of the watermark data and the robustness to certain watermark attacks, while keeping the perceptual quality of the watermarked audio at an acceptable level. It is not possible to attain high robustness to signal modifications and high data rate of the embedded watermark at the same time. Therefore, if a high robustness is required from the watermarking algorithm, the bit rate of the embedded watermark will be low and vice versa, high bit rate watermarks are usually very fragile in the presence of signal modifications. However, there are some applications that do not require that the embedded watermark has a high robustness against signal modifications. In these applications, the embedded data is expected to have a high data rate and to be detected and decoded using a blind detection algorithm. While the robustness against intentional attacks is usually not required, signal processing modifications, like noise addition, should not affect the covert communications [1].

An interesting application of the high capacity covert communications is public watermark embedded into the host multimedia, used as the link to external databases that contain certain additional information about the multimedia file itself, e.g. copyright information and licensing conditions [2,3,4].

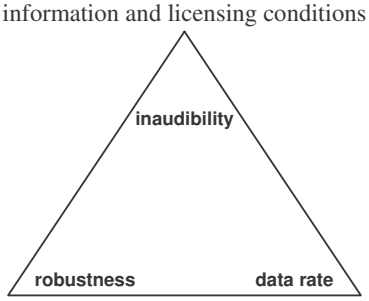

Figure 1. Magic triangle for data hiding

II. Perceptual Entropy of Audio Signals

Experimental results, obtained during decades of audio compression research, showed that only a few bits per sample are needed to represent compact disk quality music [5,6]. When performing a bit rate reduction of audio or speech signals that will be presented to the human auditory system (HAS), the objective is to introduce either imperceptible or inoffensive distortion during the compression process. This implies that for uncompressed music, noise can be injected into the host audio signal without being audible to the end user [5]. In audio data hiding, this is not used for compression, but for embedding additional data. An estimate of the perceptual entropy of audio signals is created from a combination of several noise masking measures. The results of tone-masking-noise and noise-masking-tone, as well as research on critical bands and spreading functions are combined to estimate the short term masking templates for audio signals [6].

The perceptual entropy of each short-term section of the audio signal is estimated as the number of bits required to encode the short-term spectrum of the signal to the resolution required to inject noise below the masking template level. This model is attractive, because it takes into account all of the artifacts and redundancies in the audio signal in the same manner as the HAS does (pitch, short term spectral model, etc.). There are three main parts of the perceptual entropy calculation algorithm [6], given in Figure 2:

Manuscript received November 5, 2004. This work is part of the Stego project, which is supported by the Finnish National Technology Agency (TEKES), Nokia and Yomi Solutions. The research topic has been supported by Nokia Oyj Foundation and Tauno Tönning Research Scholarship.

N. Cvejic and T. Seppänen are with the MediaTeam Oulu, Information Processing Laboratory, University of Oulu, Finland. (Contact email: {cvejic, tapio}@ee.oulu.fi)

1. Windowing of audio signal and transformation to Fourier domain
2. Calculation of the masking threshold
3. Calculation of the number of bits required to quantize spectrum of the signal

The windowing of the signal is performed using a Hanning window and frequency transformation by FFT of length 2048. The first 1024 complex lines are kept (including the DC and lines counted as one line). The steps involved in calculating the masking threshold are critical band analysis, applying the spreading function to critical bands, calculating the spread masking threshold, accounting for absolute thresholds and, finally, relating the spread masking threshold to the critical band masking threshold.

As noted above, the perceptual entropy is calculated by measuring the actual number of quantizer levels to follow the signal in the frequency domain, given a step size in the quantizer that will result in noise energy equal to the audibility threshold [5]. Audibility threshold T_i is usually defined in the power domain and quantization energy is spread across k spectral lines in each critical band. It is also assumed that the quantization noise is spread uniformly across the critical band. The distribution of the quantization error is uniform in the amplitude domain; it gives noise variance equal to $\sigma^2/2=12$.

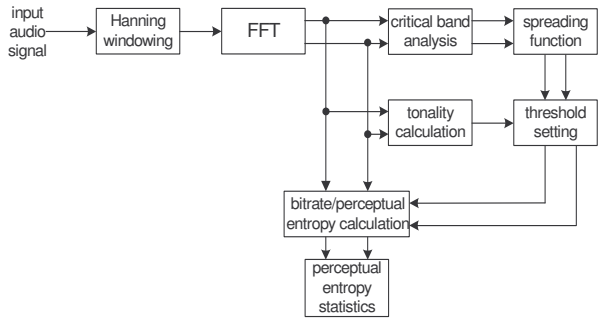

Figure 2. Perceptual entropy calculation algorithm

The step size S_i is calculated as follows. First, the energy is spread across the entire band, i.e. the energy at each spectral frequency is equal to T_i/k_i. Since the real and imaginary parts of the spectrum are quantized independently, the energy at each frequency must be divided in half, specifically the energy at each spectral component is $T_i/2k_i$. The noise energy, due to quantization is $\sigma^2/2=12$, therefore $\sigma^2/2=12=T_i/2k_i$ and since $\sigma=S_i$ we obtain $S_i = \sqrt{6T_i / k_i}$, where S_i is the quantizer step size. This is done in each of the n critical bands:

$$N_{\mathrm{Re}}(\omega) = abs\left(n\mathrm{int}\left(\frac{\mathrm{Re}(\omega)}{S_i} \right) \right), \qquad N_{\mathrm{Im}}(\omega) = abs\left(n\mathrm{int}\left(\frac{\mathrm{Im}(\omega)}{S_i} \right) \right)$$

for each σ within the critical band i. The function abs(·) represents the scalar absolute value function and nint(·) a function that returns the nearest integer to its argument. $N_{Re,Im}(\omega)$ represents the integer quantized value of the each spectral line. Then, for each ω, and individually for real and imaginary parts, $N_{Re,Im}(\omega)$ is altered as follows:

if $N_{Re,Im}(\omega)=0$, then $N'_{Re,Im}(\omega)=0$

if $N_{Re,Im}(\omega)\neq 0$, then $N'_{Re,Im}(\omega)=\log_2(2N_{Re,Im}(\omega)+1)$.

This operation assigns a bit rate of zero bits to any signal with an amplitude that does not need to be quantized, and assigns a bit rate of $\log_2$(number of levels) to those that must be quantized. If, for example, the integer number is 1, three levels (-1, 0, +1) are required to quantize the particular line. As the signs of different spectral lines are random, the sign information must be included. When no levels are necessary, the transmission of the sign bit is unnecessary as well, and a 0 is assigned to that line. The total bit rate is then calculated as:

$$\text{Total Rate}=\sum_{\omega=0}^{\pi}\left(\mathrm{N'}_{\mathrm{Re}}(\omega)+\mathrm{N'}_{\mathrm{Im}}(\omega)\right)$$

and the rate per sample (perceptual entropy) of the audio sequence is given by:

$$\text{Perceptual Entropy} = \text{Total Rate}/2048$$

The term perceptual entropy, used throughout this section, therefore indicates the 2048 sample perceptual entropy, regardless of the sampling rate or bandwidth of the signal. The block-to-block changes in perceptual entropy values increase as the window length decreases, but the mean and extreme values do not change radically [6].

Reported perceptual entropy for wideband monophonic audio signals is in the range of 4-5 bits per sample, taking into account all the spectral complexity, spectrum range and dynamic range requirements. This implies that for an uncompressed audio signal, a significant amount of additional information can be inserted into signal without causing a perceptual distortion. There is obviously a considerable gap between the currently available data rates for high capacity covert communications and theoretically obtainable data rates [2,3,7]. Therefore, a theoretic analysis of the capacity of information hiding channel is necessary in order to design a scheme that can offer higher data rates.

III. Capacity of the Data Hiding Channel

First we consider a simple data-hiding channel shown in Figure 3 [8,9]. Here, $\mathbf{X}\sim(f_X(x),\sigma_x^2)$ is the message to be embedded, $\mathbf{Z}\sim(f_Z(x),\sigma_z^2)$ is the additive noise channel and $\mathbf{Y}\sim(f_Y(x),\sigma_y^2)$ is the received signal at the output of the channel. We also assume **X** and **Z** are independent, implying that $\sigma_y^2=\sigma_z^2+\sigma_x^2$. The channel capacity is given by:

$$C=\max_{f_X(x)} I(X,Y)=\max_{f_X(x)} H(Y)-H(Y\mid X)=\max_{f_X(x)} H(Y)-H(Z)$$

I(**X**,**Y**) is the mutual information between **X** and **Y**. For a given statistics $f_Z(z)$ and σ_z^2, the entropy of **Y** should be maximized, $H(Y)=-\int f_Y(y)\log_2(f_Y(y))dy[bits]$, using a suitable distribution $f_X(x)$ of the message **X**. For a given σ_y^2 the maximum value of $H(Y)=\frac{1}{2}\log_2(2\pi e\sigma_y^2)$ bits is achieved when **Y** has a normal distribution. For instance,

the maximum value of H(**Y**) is achievable if both $f_Z(z)$ and $f_X(x)$ are normally distributed. However, for an arbitrary distribution $f_Z(z)$ and a fixed σ_x^2, the maximum achievable value of h(**Y**) is not immediately obvious. This is because **Z** is usually altered in such a manner that the amount of information in **Z** is not altered, but the statistics of **Z** is changed to Gaussian distributed $\mathbf{Z_g}$. Therefore, for the purpose of calculating the channel capacity, we can replace $f_Z(z)$ by $N\left(0,\sigma_{zg}^2\right)$ and $H(Z)=H(Z_g)=\frac{1}{2}\log_2\left(2\pi e\sigma_{zg}^2\right)$ and we get:

$$C=\max_{f_X(x)} H(Y)-H(Z)[bits]=\frac{1}{2}\log_2\left(1+\frac{\sigma_x^2}{\sigma_{zg}^2}\right)[bits].$$

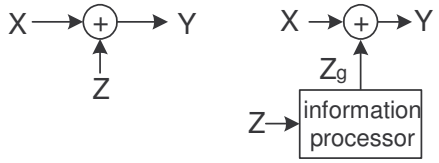

Figure 3. Data-hiding channel modeling

The general data-hiding channel is decomposed into multiple channels, as hiding process is performed in a transform domain [8]. The decomposition is performed by the forward and inverse transform (Figure 4). Signal decomposition into L bands results in L parallel channels with two noise sources in each channel. Let σ_{ij}^2, j=1,…,L be the variances of the coefficients of each band of the decomposition.

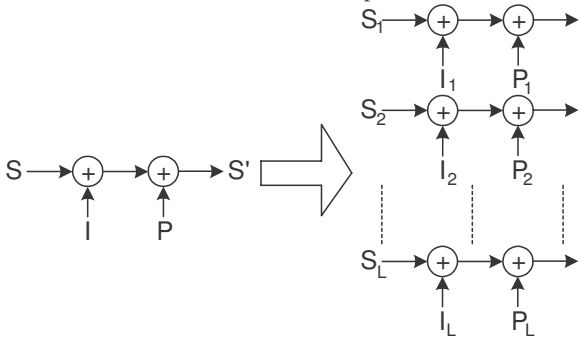

Figure 4. Data-hiding channel decomposition into multiple channels

Let the corresponding Gaussian variances be σ_{igj}^2. If σ_{pj}^2 is the variance of the processing noise in the jth channel, the total capacity of the *L* parallel channels is given by:

$$C_h=\frac{N^2}{2L}\sum_{j=1}^{L}\log_2\left(1+\frac{T_j^2}{\sigma_{igj}^2+\sigma_{pj}^2}\right)[bits]$$

for a sequence of N samples. In the equation above, T_j is the masking threshold of band j, in other words, the maximum power of the embedded message permitted in band j. In the case of no-processing noise (or if the processing noise is negligible), and we assume that all the channels have the same probability distribution function (such that $K\sigma_{ij}=K\sigma_{igj}$), the channel capacity is given by:

$$C_h=\frac{N^2}{2L}\sum_{j=1}^{L}\log_2\left(1+\frac{K}{\sigma_{ij}^2}\right)\approx\frac{N^2}{2L}\log_2\left(1+\sum_{j=1}^{L}\frac{K}{\sigma_{ij}^2}\right)[bits]$$

It is clear that the minimum channel capacity is obtained when $\sigma_{ij}=\sigma$, $\forall$j or when no decomposition is employed [9]. A transform with a good energy compaction or high gain of transform coding (GTC) [9] would result in more imbalance of the coefficient variances, resulting in an increased channel capacity. Therefore, discrete wavelet transform (DWT) or discrete cosine transform (DCT) are good decompositions for low processing noise scenarios. The term processing noise here refers to equivalent additive noise which accounts for the reduction in correlation between the transform coefficients of the original signal and the transform coefficients of the audio signal obtained after MPEG compression, noise addition, low pass filtering, etc. On the other hand, the reduction in capacity with an increase of processing noise tends to be lower for transforms which are not used in compression methods, like DFT. While severe MPEG compression is certain to remove almost all high frequency components of DCT coefficients, it will not affect the high frequency DFT at the same extent. Signal decomposition with a low GTC is generally more immune to processing noise than decomposition with a high GTC and should predominantly be used in applications demanding robust watermarks. Therefore, signal decompositions with a high GTC, like the DWT or DCT, are more suitable for high data rate steganography applications, where processing noise variance is low, because no intentional attacks are expected.

IV. Information Hiding Using LSB Coding

The information hiding algorithm that fulfils the requirements of high data rate and low robustness against signal modifications is the algorithm that uses LSB coding. It is one of the earliest and simplest information hiding techniques and, as in cases of other known algorithms; it has first been developed for watermarking of images [10] and video sequences [11]. The watermark encoder uses a subset of all available host audio signal samples chosen by a secret key. The substitution operation on the LSBs is performed on this subset. The extraction process simply retrieves the embedded data by reading the value of these bits.

The main advantage of the method is a very high watermark channel capacity; the use of only one LSB of the host audio sample gives the capacity of 44.1 kbps if a mono audio signal, sampled at 44.1 kHz is used. The obvious disadvantage is the method's extremely low robustness, due to fact that random changes of the LSBs destroy the coded watermark [12].

The increase in the embedding data rate is proportional to the number of the LSBs used for data hiding; two or more bits per sample could be used in order to enhance the bit rate of the hidden information. However, the increase of the number of samples used during LSB coding introduces a low power

additive white Gaussian noise. As already noted, HAS is very sensitive to the AWGN and this fact limits the number of LSBs that can be imperceptibly modified. In addition to subjective quality degradation, the probability of the statistical detection of the embedded watermark increases as well [13].

V. Experimental Results

We ran simulations to test obtainable data rates for LSB coding in diverse transform domains and in time domain, to be able to experimentally verify results from the analysis above.

In time domain the watermark encoder used all available host audio signal samples. The substitution operation on the LSBs was performed on each audio sample, with sampling frequency of 44.1 kHz and resolution of 16 bits per sample. The extraction process simply retrieved the hidden data from each sample by reading the value of these bits from LSBs.

Data hiding in the LSBs of the wavelet coefficients is practicable due to the near perfect reconstruction properties of the filterbank. The DWT decomposes the signal into low-pass and high pass components subsampled by two; the inverse transform performs the reconstruction. We decided to use the simplest quadrature mirror filter - Haar filter. The Haar basis is obtained with a multiresolution of piecewise constant functions [14]. The scaling function is equal to one. The Haar wavelet has the shortest support among all orthogonal wavelets, and it is the only quadrature mirror filter that has a finite impulse response [14]. Signal decomposition into the low-pass and high pass part of the spectrum is performed in five successive steps. After subband decomposition of 512 samples of host audio, using the Haar filter and decomposition depth of five steps, algorithm produces 512 wavelet coefficients. All 512 wavelet coefficients are then scaled using the maximum value inside the given subband and converted to binary arrays in the two's complement. A fixed number of the LSBs are thereupon replaced with bits of information that should be hidden inside the host audio. Coefficients are then converted and scaled back to the original order of magnitude and an inverse transformation is performed.

The similar scheme was implemented using the discrete Fourier transform (DFT) with 1024 samples as well. For the DFT decomposition we use only the magnitude of the DFT coefficients. In other words, the message signal added would change only the magnitude of the DFT coefficients and the phase is left intact. As no additional data is hidden in the phase, the phase is ignored during detection of the watermark.

Figure 7 shows that watermark channel data rates for all decompositions increase with decreased perceptual transparency, as expected. Transform domain watermark embedding outperforms significantly watermark embedding in time domain and wavelet domain embedding generally outperforms slightly DFT algorithm. Therefore, it is clear that the minimum watermark channel data rate is obtained when no signal decomposition is employed (in time domain), as expected. In addition, the experimental results demonstrated that signal decompositions with a high GTC, like the wavelet transform or DCT, are more suitable for high data rate steganography applications than decompositions with smaller GTC, like DFT and Hadamard transform.

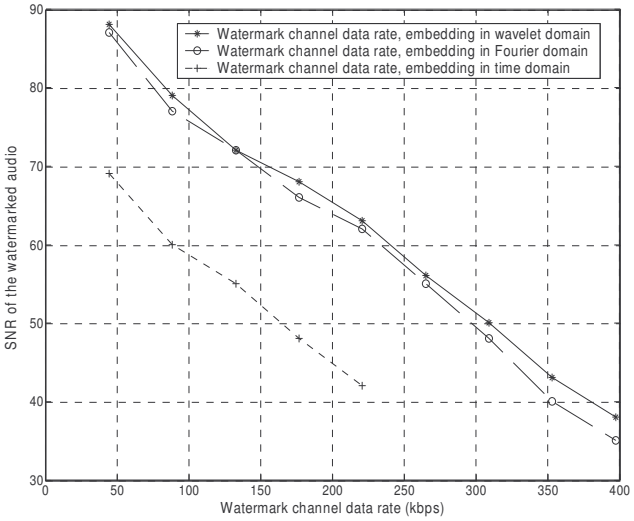

Figure 7. Watermark channel data rates for different transform domains

References

[1] I. Cox, M. Miller, J. Bloom "Digital Watermarking," Morgan Kaufmann Publishers, San Francisco, CA, 2003.

[2] J. Chou, K. Ramchandran, A. Ortega "High capacity audio data hiding for noisy channels," in *Proc. International Conference on Information Technology: Coding and Computing*, Las Vegas, NV, 2001, pp. 108–111.

[3] S. Servetto, C. Podilchuk, K. Ramchandran "Capacity issues in digital image watermarking," in *Proc. IEEE International Conference on Image Processing*, Chicago, IL, 1998, pp 445–449.

[4] S. Pradhan, J. Chou, K. Ramchandran "Duality between source coding and channel coding and its extension to the side information case," *IEEE Transactions on Information Theory*, Vol. 49, No. 5, 2003, pp. 1181–1203.

[5] J. Johnston "Estimation of perceptual entropy using noise masking criteria," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, New York, NY, 1998, pp. 2524–2527.

[6] J. Johnston "Transform coding of audio signals using perceptual noise criteria," *IEEE Journal on Selected Areas in Communications*, Vol. 6, No. 2, 1988, pp. 314–323.

[7] T. Cedric, R. Adi, I. McLaughlin "Data concealment in audio using a nonlinear frequency distribution of PRBS coded data and frequency-domain LSB insertion," in *Proc. IEEE Region 10 International Conference on Electrical and Electronic Technology*, Kuala Lumpur, Malaysia, 2000, pp. 275–278.

[8] M. Ramkumar, A. Akansu "Information theoretic bounds for data hiding in compressed images," in *Proc. IEEE Workshop on Multimedia Signal Processing*, Los Angeles, CA, 1998, pp 267–272.

[9] M. Ramkumar, A. Akansu "Capacity estimates for data hiding in compressed images," *IEEE Journal on Selected Areas in Communications*, Vol. 10, No. 8, 2001, pp. 1252–1263.

[10] Y. Lee, L. Chen "High capacity image steganographic model," *IEE Proceedings on Vision, Image and Signal Processing*, Vol. 147, No. 3, 2000, pp. 288–294.

[11] F. Hartung, B. Girod "Watermarking of uncompressed and compressed video," *Signal Processing*, Vol. 66, No. 3, 1998, pp. 283–301.

[12] B. Mobasseri "Direct sequence watermarking of digital video using m-frames," in *Proc. IEEE International Conference on Image Processing, Chicago*, IL, 1998, pp. 399–403.

[13] J. Fridrich, M. Goljan, R. Du "Lossless data embedding - new paradigm in digital watermarking," *Applied Signal Processing*, Vol. 2002, No. 2, 2002, pp 185–196.

[14] Mallat S "Wavelet Tour of Signal Processing," Academic Press, San Diego, CA, 2001.