HIGH CAPACITY ERROR FREE WAVELET DOMAIN SPEECH STEGANOGRAPHY

Sajad Shirali-Shahreza

Computer Engineering Department
Sharif University of Technology, Tehran, IRAN
shirali@ce.sharif.edu

M.T. Manzuri-Shalmani

Computer Engineering Department
Sharif University of Technology, Tehran, IRAN
manzuri@sharif.edu

ABSTRACT

Steganography is the art of hiding information in a cover media without attracting attention. One of the cover media which can be used for steganography is speech. In this paper, we propose a new speech steganography in wavelet domain. In this method, lifting scheme is used to create perfect reconstruction Int2Int wavelets. The data is hidden in some of the Least Significant Bits (LSB) of detail wavelet coefficients. The LSB bits for hiding are selected with a new adaptive algorithm. This algorithm does not hide information in silent parts, so there is no need for silent detection algorithms. This method has zero error in hiding/unhiding process, while normal wavelet domain LSB has about 0.2 % error in equal hiding capacity. This method is a high capacity steganography method which can hide information up to 20% of the input speech. The Signal-to-Noise Ratio (SNR) and listening tests show that the stegano audio is imperceptible from original audio.

Index Terms— Information Hiding, Lifting Scheme, Speech Steganography, Wavelet

1. INTRODUCTION

One of the concerns in the area of information security is the concept of hidden exchange of information. Steganography is one of the methods which have attracted more attention during the recent years. The word steganography is a Greek word that means 'writing in hiding'. The main purpose in Steganography is to hide data in a cover media so that other persons will not notice that such data is there. This is a major distinction of this method with the other methods of hidden exchange of data because, for example, in the method of cryptography, individuals see the encoded data and notice that such data exists but they cannot comprehend it. However, in steganography, individuals will not notice at all that data exists in the sources [1].

Among the methods of steganography, the most common one is to use images for applying steganography. In these methods, features such as pixels of image are changed in order to hide the information so as not to be identifiable by human users and the changes applied on the image are not tangible.

In audio steganography, the weaknesses of Human Auditory System (HAS) is used to hide information in the audio. Because the human auditory system has more precision than human visual system (HVS), audio steganography is more challenging than image steganography [2].

There are three important parameters in designing steganography methods: perceptual transparency, robustness and hiding capacity. These requirements are known as "the magic triangle" and are contradictory [3].

While robustness is usually is the most important factor for applications like watermarking, hiding capacity is more important for steganography applications because the goal of steganography is to transfer information.

Like steganography in other media, the simplest and most common method for steganography in audio is hiding information in Least Significant Bits (LSB) in the time domain. LSB steganography can be done on other domains such as wavelet domain [3] and Fourier domain [4]. One of the problems of steganography in non-time domains is their unhiding errors [3].

In this paper, we propose an adaptive wavelet domain method for steganography which has zero error in comparison to 0.2% error rate of normal LSB wavelet domain method. Listening tests results show that the stegano speech output of our method is imperceptible from original speech.

In section 2 we will discuss some of related works which has been done on audio steganography in wavelet domain. In section 3 we will explain our algorithm. Section 4 contains our experimental results. The final section is the conclusion.

2. RELATED WORKS

In this section we describe our algorithm for hiding data in an audio signal.

Time domain is one of the common domains used for steganography [5]. While time domain steganography is usually simple and fast [6], other domain can also used for steganography. For example wavelet domain [3] and

Fourier domain [4] are used to hide information in audio signals.

Different domains have special features which made them suitable for different application. Most time domain methods have zero error rates. But when the hiding and unhiding procedures are done in another domain such as wavelet or Fourier, some error are introduced. The source of this error is usually rounding error occurred during transform or when the signal is saved as a file or transferred via a communication channel. So it must be considered to know the error rate [7]. There are Error Correction Coding (ECC) methods which can be used to achieve zero error rates [8], but it is more desirable to have a method without any hiding or unhiding error.

2.1. Wavelet Domain LSB [3]

In this method, the wavelet coefficients of audio signal are calculated using Haar wavelets and discrete wavelet transform (DWT). Then the coefficients are scaled and converted to integers. Now the information is hided in the LSB of the coefficients. Then the coefficients are scaled back to original scale and the inverse DWT is done to generate the stegano audio. The unhiding procedure is similar to hiding procedure. This method has high capacity and also good SNR. The listening tests also show good perceptual transparency.

2.2. Audio-to-image Wavelet Domain Steganography [9]

In this method, first the wavelet transform of the audio is calculated. Then some of the details coefficients are chosen and sampled at a defined interval. The sampled coefficients are arranged to form an image. Then the data is hided in the image using an image steganography method. After that the stegano image is decomposed to a set of details coefficients. Then the stegano audio is constructed by performing the wavelet reconstruction transform.

This method has low hiding capacity (for example 256 bps), but has resistance to MP3 compression. The details of wavelet transforms such as the wavelet which is used, the level which wavelet transform is done and the selected coefficients can be used as a secret key to gain more security.

2.3. Comparing Wavelet and Fourier Transforms for High Rate Steganography [6]

In [6], a comparison is done between wavelet transform and Fourier transform for high capacity steganography. They use LSB method to hide data in an audio signal in Time domain, Wavelet domain and Fourier domain. Their results show that hiding in wavelet or Fourier domain has better perceptual transparency and SNR. And between Fourier and wavelet, wavelet has small advantages in SNR.

3. THE PROPOSED METHOD

In this section we describe our algorithm for hiding data in a speech signal.

Wavelet transform has special features which make it a good choice for compression algorithms. Based on the ability of wavelet in compression, we can construct methods to hide high amount of data in host signal.

In our method, we first calculate the wavelet coefficients. Previous wavelet domain steganography such as [3, 6, 9] use DWT with normal wavelets such as Haar or Daubechies. The problem which these wavelets have is that when applying them on an integer signal such as a speech (which each sample is 8bit or 16bit), the resulted coefficients are not integer. So those methods scaled the resulted coefficients and then convert them to a binary string.

To solve this problem, we use lifting scheme [10] to produce Int2Int wavelets. Int2Int means that if the input signal is integer, the wavelet coefficients are also integer. So there is no need to scale the coefficients and convert them to binary representation.

The main sources of errors which arise in non-time domains steganography are rounding errors and out of range errors. When we transform an audio signal to another domain, then change the signal and back to time domain, the resulting signal is not necessarily integer. In addition it is not in the range which the original signal was. For example if the input wav file samples are 16 bit, each sample has a value between 0-65536. But the resulted signal can have samples with any value such as 70000. But when the signal is saved in file or transferred via a channel, the samples must be in the defined range, so some information is lost.

Because we use Int2Int wavelets, we didn't have the problem of non-integer values. To reduce out of range errors, we did not hide equal information in all coefficients. If we change the value of a coefficient from a low value such as 1 to a large value such as 110, for example with LSB method and replacing 7 bits, the probability that in the stegano signal we have a value which is not in desired range is high. We choose the number of LSBs to replace with data according to the coefficient value. We hide more bits in bigger coefficients and fewer bits in smaller coefficients.

To calculate the number of bits to hold data in a coefficient with value c, we find the biggest power of 2 named p which is smaller than c, $2^p \le c < 2^{p+1}$. The number of bits which we hide in this coefficient is p - OBH. OBH (Original Bit to Hold) is a constant which shows that how many bits of the original signal is kept and how many bits of the signal are replaced with the data. The minimum value of OBH is 1. With different values of OBH, different hiding capacity and quality can be achieved.

When the number of bits to hold data in each coefficient is calculated, those bits are replaced with the data which must be hidden. In order to have good

perceptual transparency, we hide data only in details coefficients.

Finally the inverse (reconstruction) DWT is done and the stegano speech signal is created.

The unhiding procedure is identical to hiding procedure. First the wavelets coefficients are calculated. Then the number of bits which has information is calculated with the same method used during hiding. Then the information is extracted from bits which have information.

4. EXPERIMENTAL RESULTS

We implemented this method in Matlab. To hide data, the input signal is considered as a sequence of blocks with fix size, for example 1024 samples. Then each block is transformed to wavelet domain, then information is hided in it and converted back to time domain.

To test the method, we use 4 sample wav files from FARSDAT database [11]. These files have duration about 10 seconds, 22050 Hz sample rate and 16 bits per sample. The speeches are from 3 male and one female speaker.

Different wavelets can be used in our algorithm. To compare different wavelets, we use different wavelets to hide data with OBH=1 and the decomposition of wavelets done at level one. Table I shows the result of running algorithm with different wavelets. As can be seen, different wavelets have nearly similar results, so this method is not depending on a special type of wavelets. The error rate is zero for all of them. The SNR (Signal-to-Noise Ratio) value and hiding capacity did not differ a lot for different wavelets.

As mentioned in previous section, the OBH can be used as a parameter to determine hiding capacity and SNR. In the following tests, the Haar wavelet is used with level one decomposition. Figure 1 shows the relation between hiding capacity and OBH.

Figure 2 shows the relation between SNR of the stegano signal and the amount of information which is hidden in the signal. This figure shows that the SNR of output signal is nearly same for different voices.

To compare our method with normal wavelet domain LSB method, we hide data in sample signals with both methods. To get different hiding capacities for LSB method, different number of LSB is used to hide information. Data is only hided in detail coefficients. Haar wavelets and Daubechies2 wavelets are used to hide information in both methods. The tests were done for all of voices samples and the averages are compared in figures.

Table I. Different Wavelets Results

Wavelet Name	SNR	Hiding Capacity (bps)	Error (%)
Haar	56.1398	67200	0
Daubechies2	51.8032	58942	0
Daubechies3	47.9688	59080	0
Daubechies4	52.6348	40732	0
Daubechies5	50.8322	40894	0
Daubechies6	43.3143	62052	0
Daubechies7	38.3541	49029	0
Daubechies8	39.5448	63927	0
Symlets2	51.8032	58942	0
Symlets3	40.3605	36780	0
Symlets4	50.2209	52042	0
Symlets5	47.794	55527	0
Symlets6	32.2816	53020	0
CDF1.1	56.1398	67200	0
CDF1.3	56.008	67200	0
CDF1.5	55.8913	67200	0
CDF2.2	44.8443	49204	0
CDF2.4	44.907	49204	0
CDF2.6	44.9144	49204	0
CDF3.1	50.676	50097	0
CDF3.3	51.0779	50097	0
CDF3.5	51.1667	50097	0
CDF4.2	44.1027	51559	0
CDF4.4	44.5212	51559	0
CDF4.6	44.6957	51559	0
CDF5.1	48.665	55859	0
CDF5.3	47.6383	55859	0
CDF5.5	50.5233	55859	0
CDF6.2	43.0533	59008	0
CDF6.4	44.2636	59008	0
CDF6.6	44.8016	59008	0
Bior5.5	36.5429	45335	0

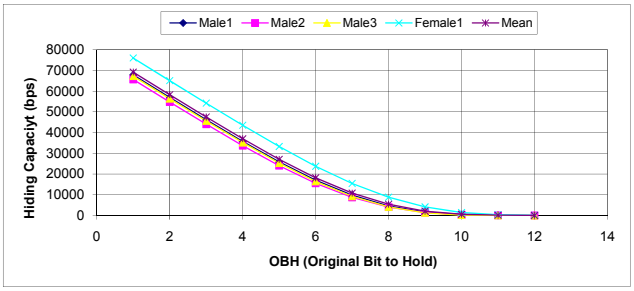

Figure 1. Relation between Hiding Capacity and OBH

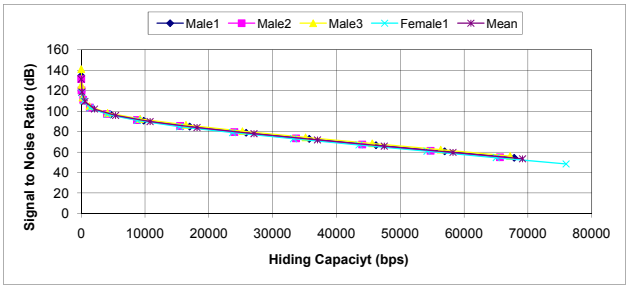

Figure 2. Relation between SNR and Hiding Capacity

Figure 3 compares output signal SNR of our method with normal wavelet domain LSB method. Figure 4 compares error rate in both methods. While normal wavelet domain LSB method has about 0.2% error, our method has 0% error rate which means error free data hiding and unhiding in our method.

For the listening test, we played both original and stegano speech for testers and ask them to select the stegano audio. The rate of selecting stegano speech was about 50% which shows the perceptual transparency is very good.

5. CONCLUSION

In this paper we propose a new method for steganography in speech signals. In this method lifting scheme is used to create Int2Int wavelets. The number of bits which is used to hide data in each coefficient is chosen based on the value of that coefficient to minimize the error. This method has high capacity for hiding data (up to 20% of the input speech).

The main advantage of this method in comparison to normal LSB is low error rate. This method has 0% error – means error free hiding and unhiding – in comparison to 0.2% error of normal LSB method.

Like method [9], the wavelet transform details such as wavelet name, the level which the transform is done and the selected coefficents can be used as secret key to gain more security.

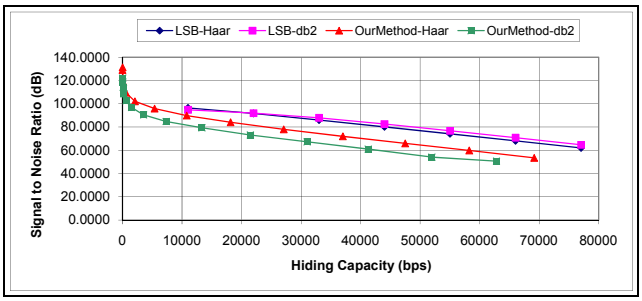

Figure 3. Comparison of SNR between our method and normal wavelet domain LSB method

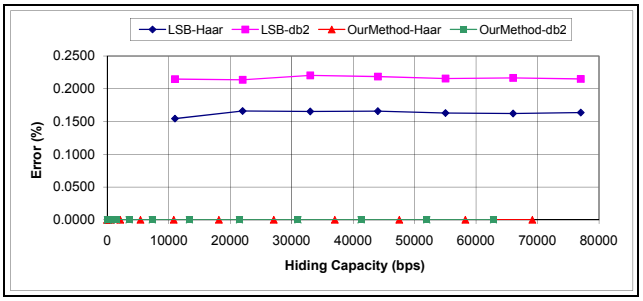

Figure 4. Comparison of error rate between our method and normal wavelet domain LSB method

6. REFERENCES

[1] J.C. Judge, "Steganography: Past, Present, Future," *SANS white paper,* November 30, 2001, http://www.sans.org/rr/papers/index.php?id=552, last visited: 31 March 2007.

[2] N. Taraghi-Delgarm, *Speech Watermarking*, M.Sc. Thesis, Comptuer Engineering Department, Sharif University of Technology, Tehran, IRAN, May 2006.

[3] N. Cvejic and T. Seppanen, "A wavelet domain LSB insertion algorithm for high capacity audio steganography," *Proc. of 10th IEEE Digital Signal Processing Workshop and 2nd Signal Processing Education Workshop,* October 2002, pp. 53- 55.

[4] L. Gang, A.N. Akansu, and M. Ramkumar, "MP3 resistant oblivious steganography," *Proc. of 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'01),* May 2001 vol.3, pp. 1365-1368.

[5] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, "Techniques for data hiding," *IBM Systems Journal,* vol. 35, issue 3-4, September 1996, pp. 313-336.

[6] N. Cvejic and T. Seppanen, "Channel capacity of high bit rate audio data hiding algorithms in diverse transform domains," *Proc. of 2004 Int. Symposium on Communications and Information Technology (ISCIT 2004),* October 2004, vol.1, pp. 84- 88.

[7] D. Kirovski and H.S. Malvar, "Spread-spectrum watermarking of audio signals," *IEEE Transaction on Signal Processing,* vol. 51, issue 4, April 2003, pp. 1020- 1033.

[8] P. Sweene, *Error Control Coding (An Introduction),* Prentice-Hall International Ltd., Englewood Cliffs, NJ, 1991.

[9] R.A. Santosa and P. Bao, "Audio-to-image wavelet transform based audio steganography," *Proc. of 47th Int. Symposium ELMAR,* June 2005, pp. 209- 212.

[10] W. Sweldens, "The Lifting Scheme: a Construction of Second Generation of Wavelets," *SIAM Journal on Mathematical Analysis,* vol. 29, issue 2, 1998, pp. 511-546.

[11] M. Bijankhan, J. Sheikhzadegan, "FARSDAT-The Farsi Spoken Language Database," *Proc. of the 5th Int. Conference of Speech Science and Technology*, vol. 2, 1994, pp. 826-831.