VISIÓN INVESTIGADORA

In-audio speech hiding based on spectrum shifts in the wavelet domain

Ocultamiento de voz en audio basado en el desplazamiento de espectro en el dominio *wavelet*

Catherine Mariño*

Ángel Suarez*

Dora M. Ballesteros***

Javier E. González****

Resumen

Este artículo describe un modelo de ocultamiento de voz (mensaje secreto) en audio (señal huésped) basado en la técnica de espectro desplazado, *Shift Spectrum Algorithm* (SSA), y la Transformada Wavelet Discreta (DWT). Las señales de voz y audio se descomponen utilizando la DWT multinivel. Los coeficientes del mensaje secreto se ocultan en los coeficientes de detalle de la señal huésped, utilizando un reordenamiento de sub-bandas basado en un criterio de similitud. La clave secreta contiene la información del reordenamiento de las sub-bandas del mensaje secreto. La reconstrucción de los coeficientes *wavelet* superpuestos de las dos señales corresponde a la señal estego, la cual tiene la misma escala de tiempo y rango dinámico de la señal huésped. La calidad de la señal estego se califica con la prueba de promedio de opinión, *Mean Opinion Score* (MOS) del estándar ITU-T P.835.

* Ingeniera en Telecomunicaciones. Universidad Militar Nueva Granada (Colombia). Grupo TIGUM. catherine7188@gmail.com; u1400393@unimilitar.edu.co

** Ingeniero en Telecomunicaciones. Universidad Militar Nueva Granada (Colombia). Grupo TIGUM. u1400365@unimilitar.edu.co

*** Ingeniera Electrónica, magíster en Ingeniería Electrónica y de Computadores. Docente Universidad Militar Nueva Granada (Colombia). dora.ballesteros@unimilitar.edu.co

**** Ingeniero Electrónico, magíster en Ingeniería Electrónica. Docente Universidad Santo Tomás (Colombia). javiergonzalezb@usantotomas.edu.co

Palabras clave

Mensaje secreto, señal huésped, señal estego, ocultamiento, recuperación, espectro desplazado.

Abstract

This work describes a model of in-audio speech hiding based on both a Shift Spectrum Algorithm (SSA) and the Discrete Wavelet Transform (DWT). The secret message (speech signal) and the background audio (host signal) are decomposed by using the multi-level DWT. The secret-message wavelet coefficients are hidden into the host signal detail coefficients through a selection based on the similitude between coefficient groups. The secret key is one the system's output and it contains information related to the position of the secret-message coefficients into the host-signal coefficients. The stego signal is the Inverse DWT of the

relocated secret-message coefficients plus the coefficients of the host signal; such signal has the same time scale and dynamic range of the host signal. The stego signal was tested using the Mean Opinion Score (MOS) conforming to ITU-T P.835 standard.

Key words
Secret message, host signal, stego signal, embedding, recovering, spectrum shift.

Introduction

Many algorithms have been developed to hide data into a host signal with the purpose to transmit information in a secure way. In covert communications, the stego signal does not generate suspicious about existence of the secret message and therefore the secret message can be transmitted into a secure channel. Text or voice signals have been hidden into audio signals based on classical techniques like Least Significant Bits (LSB) substitution, Frequency Masking (FM), Spread Spectrum (SS) and Shift Spectrum Algorithm (SSA).

In LSB, some significant bits of the host signal are replaced with the secret message; it can be in time domain, frequency domain or wavelet domain [1]-[3]. In FM, the secret message is hidden by using the masking property of the Human Auditory System (HAS) [4], [5]. In Spread Spectrum, the spectrum of the secret message is distributed on the spectrum of the host signal [6], [7]. In Shift Spectrum, the spectrum of the secret message is shifted up to the highest range of frequencies of the host signal [8].

Every method has strengths and weaknesses. For example, LSB is the simplest technique, but the robustness against signal manipulations is low; FM takes advantage of the Human Auditory System (HAS) with a good transparency but its maximum hiding capacity is lower than in LSB; the transparency of SSA is higher than in LSB and FM but its hiding capacity is lower than of the above schemes. Therefore, we propose a scheme based on SSA on the wavelet domain in which the strengths of the scheme are preserved and the weaknesses are overcome. Unlike SSA, in which the highest 4 kHz frequencies of the host signal are used to hide the secret data, in our scheme the highest 19.25 kHz frequencies are selected; it implies that the hiding capacity is higher in our proposal. Additionally, to increase the effort to discover the secret message, the sub-bands of the secret message are relocated before the hiding process; this step is based on the similitude between the coefficients of the host signal and the coefficients of the relocated secret message.

1. Background of the Discrete Wavelet Transform (DWT)

The Discrete Wavelet Transform (DWT) is a multi-resolution method which divides the bandwidth and the size of the input signal in two, level by level. It includes two steps: half band filters and subsampling. It is shown in figure 1.
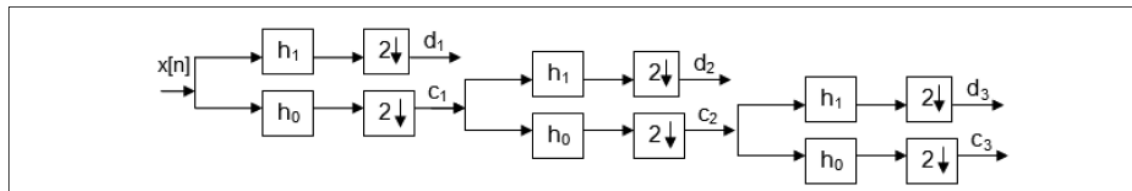
The coefficients are obtained according to eq. (1) and (2):

$$c_i[n] = \sum_k c_{i-1}[k-2n]h_0[k] \qquad (1)$$

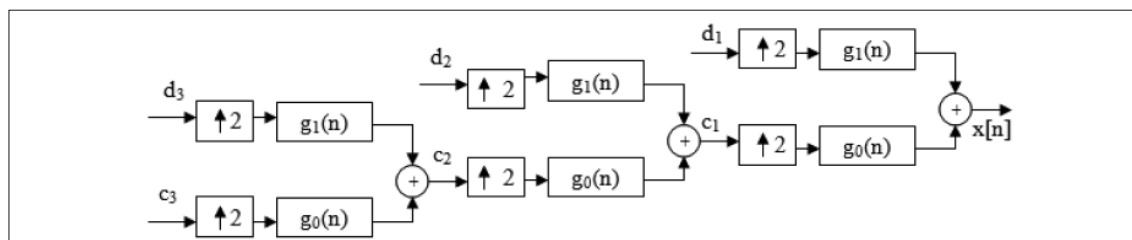$$d_i[n] = \sum_k c_{i-1}[k-2n]h_1[k] \qquad (2)$$

With h_0 and h_1 as the impulse response of the low pass and high pass filters of decomposition. The reconstruction of the IDWT is

Figure 1. Multi-level decomposition: detail (di) and coarse (ci) coefficients

Source: authors

Figure 2. Multi-level reconstruction

Source: authors

represented by oversampling and half band filters, according to figure 2 and equation (3):

$$c_i[n] = \sum_k c_{i-1}[n-2k] g_0[k] + \sum_k d_{i-1}[n-2k] g_1[k]$$

(3)

With g_0 and g_1 as the impulse response of the low pass and high pass filters of reconstruction.

2. Wavelet Transform in the Speech-in-Audio Hiding Model

The model proposed in this work is based on the multi-level decomposition of both the secret message and the host signal. The coefficients of the secret message are hidden into the detail coefficients of the host signal. The proposed model is presented in figure 3.

Multi-level DWT: it decomposes the secret message and the host signal in three levels with db1 wavelet base. The time-scale of the ho<st signal is the double of the time-scale of secret message, and then the size of the
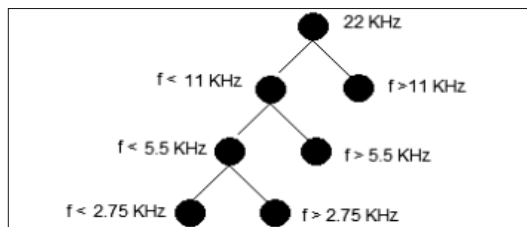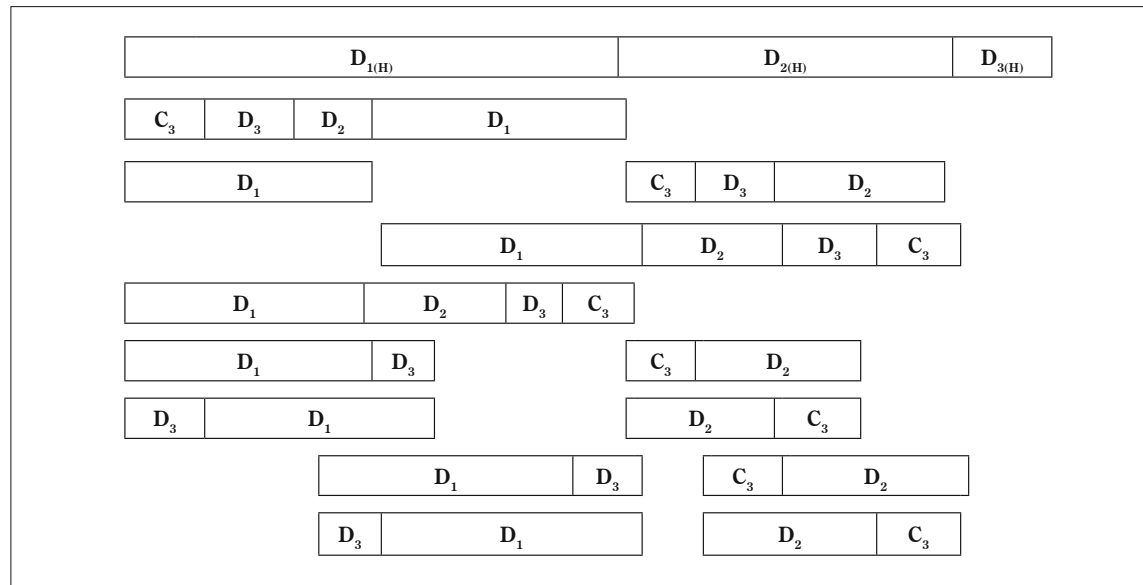
Figure 3. Hiding module of speech-in-audio
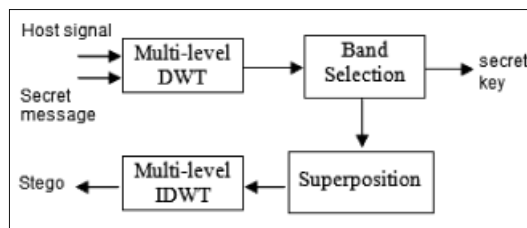
Source: autors

wavelet coefficients (coarse and detail) of the secret message is equal to the size of the detail coefficients of the host signal. Figure 4 illustrates the wavelet decomposition for three levels.

Figure 4. Wavelet tree for an audio signal

Source: authors

Figure 5. Options in band selection

Source: authors

Band selection: eight options are tested in order to find the best condition of hiding. Figure 5 plots the choices. The coefficients of the host signal are marked as $D_{i(H)}$ and the coefficients of the secret message as D_i (detail) and C_i (coarse). The coarse coefficients of the host signal are not modified. The selection is carried out according to the similitude between the coefficients of the host signal and the coefficients of the relocated secret message. Once the option has been selected, the secret key is related to the selected option; therefore, if the first option has been selected, the secret key is "1".

Superposition: the coefficients of the stego signal are obtained as the sum of the coefficients of the host signal and the coefficients of the secret message. It uses superposition.

Multi-level IDWT: the IDWT transform is applied to the coefficients of the stego signal.

To extract the secret message, the receiver needs to know: the original host signal, the stego signal and the secret key. The recovering module is shown in figure 6.

The blocks of the recovering module are: multi-level decomposition, extraction, multi-level reconstruction.

Multi-level DWT: the detail and coarse coefficients of the stego and host signals are calculated. It uses db1 and three levels of decomposition.

Figure 6. Recovering module of speech-in-audio

Source: authors

Universidad Distrital Francisco José de Caldas - Facultad Tecnológica
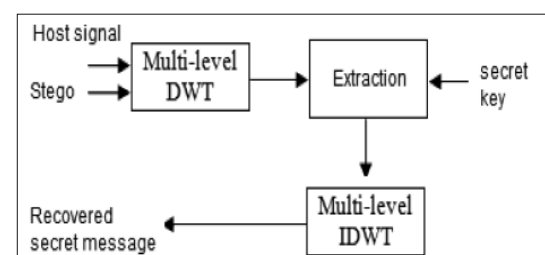
Extraction: this block performs the identification of the coefficients of the secret message and relocation.

Firstly, the coefficients of the secret message are obtained according to:

$$S(w) = G(w) - H(w) \qquad (4)$$

Where S(w), G(w), H(w) are the wavelet coefficients of the recovered-secret message, stego signal and host signal, respectively.

Secondly, the coefficients of the recovered-secret message are relocated according to the *key*.

Multi-level IDWT: the IDWT transform is applied to the coefficients of the relocated recovered-secret message.

3. Results

In this section, the results of one case of speech-in-audio hiding are shown. The secret message and the host signal are encoded with 16-bits and sampled at fs = 44 kHz. The time-scale of the host signal is 2-seconds, while the time-scale of the secret message is 1-second. The wavelet coefficients of the both the host and secret message are shown in figure 7.

Figure 7. Wavelet coefficients: host signal (up) and secret message (down)
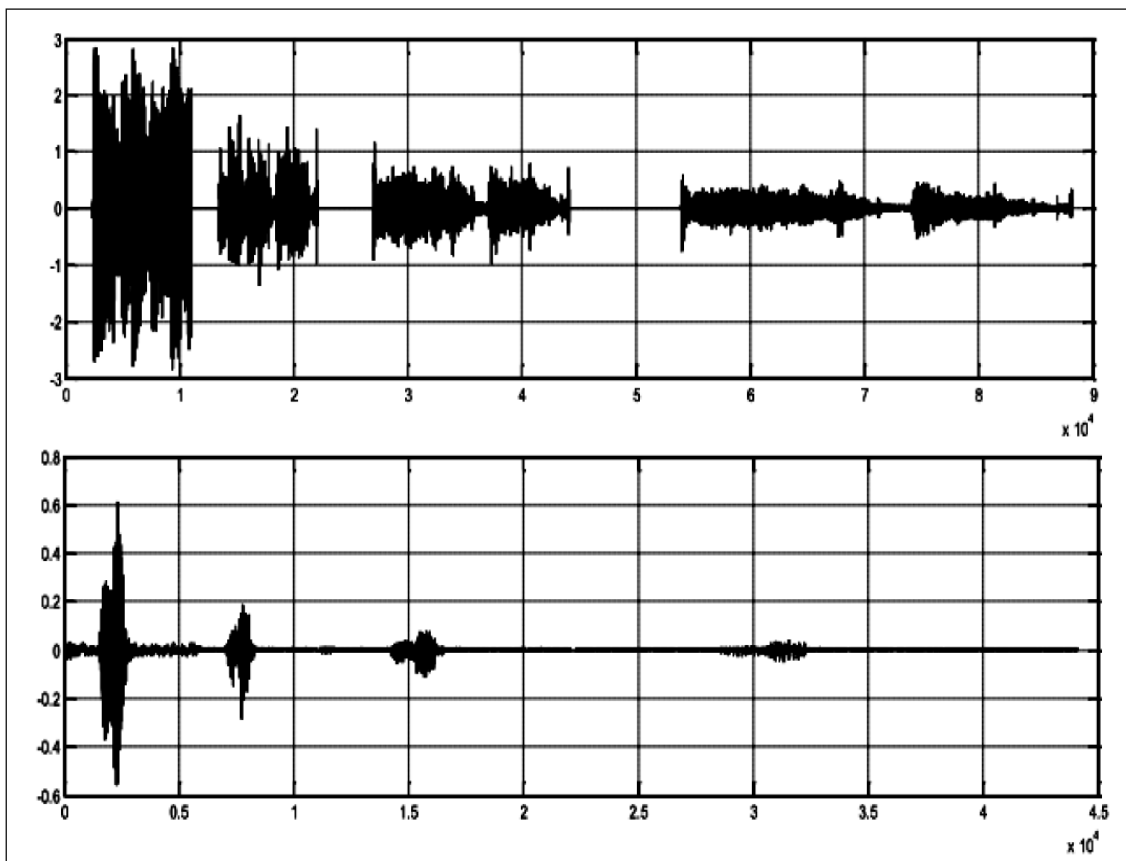
Source: authors

Figure 8. Wavelet coefficients: host signal (blue) and secret message (red). D2 (up); D1 (down)
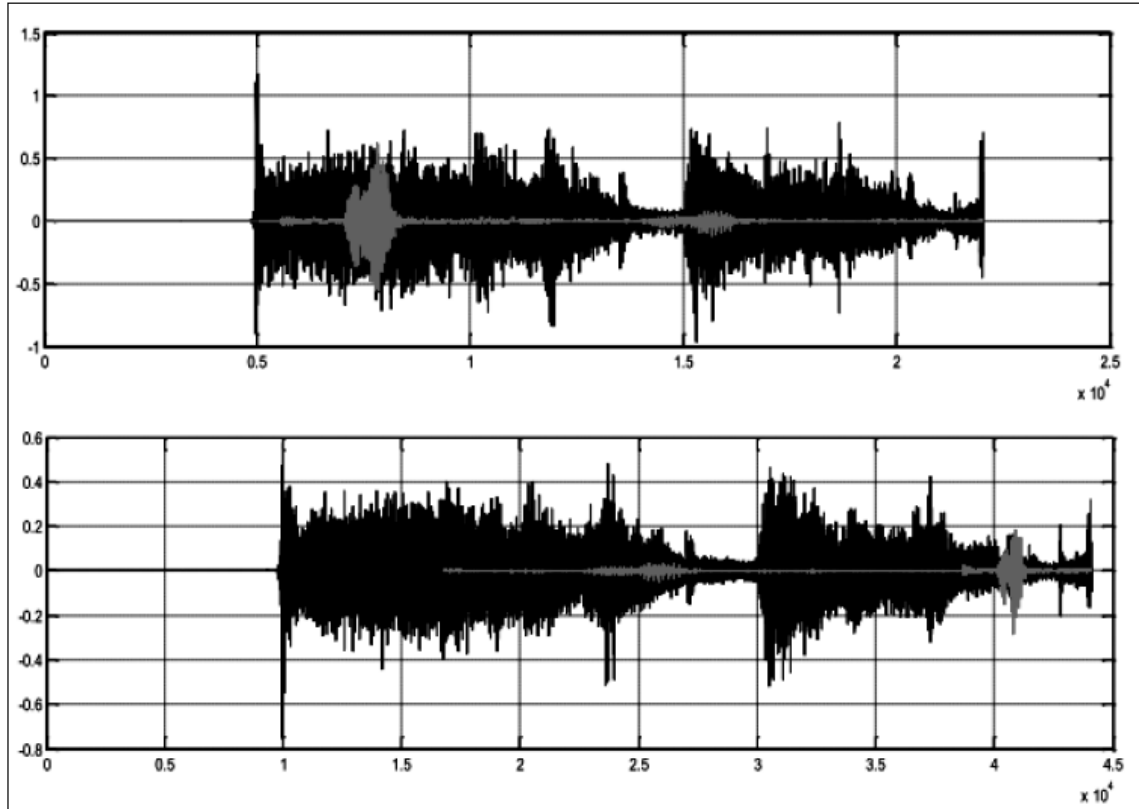
Source: authors

In figure 8, the coefficients of the secret message are re-located according to 7th option of figure 5. The coarse and detail coefficients of the secret message are masked by the detail coefficients of the host signal; therefore, the stego signal has a good quality.

With the purpose to have a good enough stego signal, the masking property must be verified. It means that the coefficients of the host signal must mask the coefficients of the relocated secret message. The masking property is tested in every one of the eight options of figure 5 and the best option is selected to hide the secret message. The correlation coefficient between the wavelet coefficients of the host signal and the relo-

cated secret message is taken into account. The higher the value of the correlation coefficient, the higher is the masking value.

To validate the selection of the option, the stego signal is qualified with the Mean Opinion Score (MOS) according to the ITU-standard [9]. In table 1, the correlation coefficient and the MOS are shown per every case.

According to table 1, there is a strong relationship between the coefficient correlation and the MOS. If the coefficients of the host signal and the coefficients of the stego signal are highly correlated, the stego signal should be of high quality.

Table 1. Validation of the band selection

Option	Correlation coefficient	MOS
1	0.83	2
2	0.81	2
3	0.67	1
4	0.92	4
5	0.84	2
6	0.81	2
7	0.99	5
8	0.89	3

Source: authors

4. Conclusions

We presented a model based on the multi-level Discrete Wavelet Transform and SSA for speech-in-audio hiding. Our proposal exploits the masking property of the Human Auditory System (HAS) by adding the re-located wavelet coefficients of the secret message to the wavelet coefficients of the host signal. The band selection is the core of our model and it consists in detecting the best hiding option based on the correlation coefficient between the coefficients of the host signal and the coefficients of the stego signal. We verified that the higher the value of the correlation coefficient, the higher is the MOS of the stego signal.

Since our scheme uses the half of the size of the wavelet coefficients to hide the secret message, the obtained hiding capacity is higher than in the classical Shift Spectrum Algorithm.

Acknowledgement

References

[1] N. Cvejic, T. Seppanen, "Channel capacity of high bit rate audio data hiding algorithms in diverse transform domains", IEEE International Symposium on Communications and Information Technology, ISCIT 2004, pp. 84-88 vol.81.

[2] N. Cvejic, T. Seppanen, "Reduced distortion bit-modification for LSB audio steganography", 7th International Conference on Signal Processing, 2004, pp. 2318-2321, vol.2313.

[3] K. Gopalan, "Audio Steganography using bit modification", IEEE International Conference on Acoustics, Speech, & Signal Processing, April, 2003.

[4] F. Djebbar, H. Hamam, K. Abed-Meraim, D. Guerchi, "Controlled Distortion for High Capacity Data-in-Speech Spectrum Steganography", Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2010, pp. 212-215.

[5] F. Djebbar, K. Abed-Meraim, D. Guerchi, H. Hamam, "Dynamic energy based text-in-speech spectrum hiding using speech masking properties", 2nd International Conference on Industrial Mechatronics and Automation (ICIMA), 2010, pp. 422-426.

[6] P. Dutta, D. Bhattacharyya, T. Kim, "Data Hiding in Audio Signal: A Review", International Journal of Database Theory and Application, Vol 2, N. 2, June 2009, pp. 1-8.

VISIÓN
INVESTIGADORA

VISIÓN
ELECTRÓNICA

[7]   D.E. Skopin, I.M.M. El-Emary, R.J. Ras-ras, R.S. Diab, "Advanced algorithms in audio steganography for hiding human speech signal", 2nd International Conference on Advanced Computer Control (ICACC), 2010, pp. 29-32.

[8]   T. Rabie, D. Guerchi, "Magnitude Spectrum Speech Hiding", IEEE International Conference on Signal Processing and Communications, ICSPC 2007, pp. 1147-1150.

[9]   ITU, "ITU-T P.835, Subjective Test Methodology for Evaluating Speech Communication Systems that Include Noise Suppression Algorithm", International Telecommunication Union 2003.