World Academy of Science, Engineering and Technology
International Journal of Electronics and Communication Engineering
Vol:3, No:5, 2009

Narrowband Speech Hiding using Vector Quantization

Driss Guerchi and Fatiha Djebbar

Abstract—In this work we introduce an efficient method to limit the impact of the hiding process on the quality of the cover speech. Vector quantization of the speech spectral information reduces drastically the number of the secret speech parameters to be embedded in the cover signal. Compared to scalar hiding, vector quantization hiding technique provides a stego signal that is indistinguishable from the cover speech. The objective and subjective performance measures reveal that the current hiding technique attracts no suspicion about the presence of the secret message in the stego speech, while being able to recover an intelligible copy of the secret message at the receiver side.

Keywords—Speech steganography, LSF vector quantization, fast Fourier transform

I. Introduction

In the last few years, steganography has witnessed a growing interest in applications aiming to provide digital data secrecy [1]. The arrival of digital camera and the growth of music industry have engendered a flood in both digital image and audio file available on the Web. While watermarking is motivated by the need to protect the copyright of the digital content of these data, steganography benefits greatly from this unsuspicious available digital information and uses it as a carrier to transmit secret digital image.

Image and speech steganography, both aver to be very promising since audio file and digital image are the most available and common carriers. In image steganography, the least-significant bit (LSB) approach is the most popular hiding approach. While the LSB hiding algorithm is very simple, the robustness of this technique against attacks is doubtful. New LSB steganalysis methods, based on statistical analysis techniques, were developed recently to not only recover the secret data but also to modify it [2]. Our alternative is a substitute carrier which could be used to hide speech messages.

Audio files as they are very popular and widely spread over the Internet constitute a very interesting cover for other multimedia signals. In this paper, we propose an efficient technique aiming to hide speech messages in narrowband speech. The hiding process takes place through the high-frequencies low-magnitudes part of the cover speech to generate a similar-quality stego-speech. We opted to work in the frequency domain [3], [4] and hide the digital information within the amplitude component. The resulting stego-speech is indistinguishable from the original cover speech, and consequently will not provoke any doubt on its authenticity.

In this work we present a drastic improvement to the technique presented in [5], where each secret speech frame is represented by 13 parameters that are embedded in the cover speech frame. To minimize the impact of the hiding process on the quality of the cover speech, we adopt in this paper vector quantization (VQ) concept to minimize the number of the secret speech parameters to only 4 parameters per frame. Vector quantization is extensively used in speech communication to reduce the coding rate since several parameters could be represented simultaneously by one index.

D. Guerchi is with the College of Information Technology, UAE University, Al Ain, 17551, UAE e-mail: guerchi@uaeu.ac.ae.

F. Djebbar is with the department of Informatics, Universite de Bretagne Occidentale, 29609 BREST cedex, Brest, email: fatiha.djebbar@etudiant.univ-brest.fr

II. Speech-in-speech hiding

Narrowband speech is a baseband signal with most of the relevant intelligibility-preserving frequency components in the [300:3400Hz] spectrum. In all vowels and most of the voiced consonants, the magnitude spectrum shows very weak components at high frequencies. In this paper, we will take advantage of these speech characteristics to design an efficient speech-in-speech hiding algorithm. Our speech steganography system consists of embedding the secret speech parameters in the high frequency locations of the magnitude spectrum of the cover speech. Theoretically, the resultant stego speech is expected to be perceptually indistinguishable from the cover speech since the pertinent low-frequency components will remain intact.

A. Cover speech decomposition

In general, speech signals are presented as a two-dimensional signals either in time domain or frequency domain. The time domain speech waveforms are more sensitive to modifications than the frequency domain counterparts. For this reason, the secret speech parameters are to be embedded in the magnitude spectrum of the cover speech. Hence, the need to convert the time domain speech frame $s_c(n)$, $n = 0, \dots M-1$, to frequency domain $S_c(k)$, $k = 0, \dots M-1$. The most popular tool to perform this conversion is called the fast Fourier transform (FFT).

$$S_c(k) = FFT(s_c(n), n = 0, \dots M-1) \;\; k = 0, \dots M-1 \tag{1}$$

Since the secret speech parameters will be hidden in the magnitude spectrum, the cover speech spectrum need to be decomposed first to phase spectrum $\varphi(k)$ and magnitude spectrum $|S_c(k)|$:

$$S_c(k) = |S_c(k)|e^{j\varphi(k)} \tag{2}$$

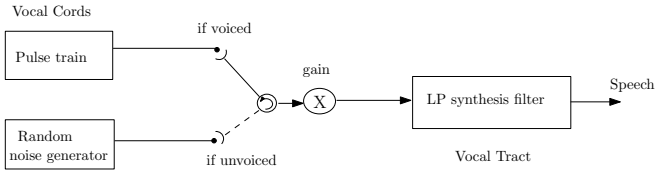

Fig. 1. Simplistic model of speech production.

B. Secret speech analysis

The necessity to parameterize the secret speech message before hiding is due to many factors. Among these factors, we mention the limited number of available hiding locations in narrowband cover speech. Secret speech must be represented by a very small number of parameters to accommodate the restricted number of available host locations. Speech parametrization known as speech analysis is widely used in many research areas, such as speech coding and speech recognition. In speech coding, the signal is subject to a speech analysis process to represent the original signal with the pertinent parameters. These parameters are coded and sent to the receiver where an inverse algorithm known as speech synthesis is used to reconstruct a copy of the original signal.

The most popular speech analysis algorithms are based on the human speech production model [6]. In this model, a speech signal is produced by the sequential excitation of two filters, a pitch filter, representing the periodicity in voiced segments (this periodicity is due to the vibration of the vocal cords), and a linear prediction (LP) filter modeling the vocal tract (this filter generates the short-term correlation present in all types of speech). Figure 1 shows a simplistic diagram of the speech production model. The linear prediction coding (LPC) model is based on this diagram. The LPC model is widely used in speech coding to represent the speech frames with a limited number of parameters for transmission. At the receiver, these parameters are used to reconstruct a synthetic-quality speech signal. Speech analysis consists of two phases: a pitch analysis to extract the pitch delay d and pitch gain g, and an LP analysis to get the 10 LP coefficients, a_i $(i = 1, \ldots, 10)$. The pitch and LP parameters are used to build the pitch filter and LP filter, respectively. In the LPC model, the pitch filter is used only for voiced segments. For unvoiced speech only the LP filter is used since there is no periodicity in this class of speech. In-depth details about the speech analysis steps are given in [5].

LP coefficients are very sensitive to errors. The direct quantization of these coefficients might produce an unstable LP filter. For this reason, the LP coefficients are often converted to a better representation before any processing. One of the popular representations is the line spectrum frequencies (LSF) [7]. In this work, we adopted this representation since the 10 LSF coefficients w_i $(i = 1, \ldots, 10)$ will be subject to vector quantization before hiding.

C. LSF vector quantization

Unlike scalar quantization which codes each LSF coefficient separately, vector quantization treats the 10 LSF coefficients as one vector V_{inp}. This vector is tested against all the LSF vectors of a codebook to select the closest match. For this purpose, a codebook of L LSF vectors is first designed after a training phase. Each codebook entry is represented by one index. For example, in this paper, we adopted a rich LSF codebook of 1024 entries V_l $(l = 1, \ldots, 1024)$. The index, I_{opt}, of the best match is determined by the minimization of the spectral distortion between the current frame LSF vector V_{inp} and each of the codebook vectors V_l.

$$I_{opt} = \underbrace{argmin}_{1 \le l \le 1024} SD(V_{inp}, V_l) \tag{3}$$

In the hiding process, the 10 LSF coefficients will be represented by one index I_{opt}. Compared to the scalar hiding (SH) technique in [5] (which embed all the 10 LSF ceofficients) , this new VQ hiding (VQH) approach saves nine cover speech frequency locations, hence lessening drastically the impact of the hiding process on the cover speech quality. The stego speech will look more similar to the cover speech, rendering any steganalysis attempt more difficult.

D. Hiding phase

In the hiding phase, each 10-ms secret speech frame will be hidden (in terms of its four parameters) in a 10-ms cover speech frame. The index I_{opt} as well as the pitch delay d, gain g and voiced/unvoiced (V/UV) bit vb of each secret speech frame will be embedded in the last frequency locations of the cover speech magnitude spectrum. Following is the hiding algorithm:

$|S_c(79)| = d$
$|S_c(78)| = g$
$|S_c(77)| = I_{opt}$
$|S_c(76)| = vb$
$|S_c(0:75)| = |S_c(0:75)|$

Combining the new magnitude spectrum with the unchanged cover speech phase spectrum gives the stego speech spectrum $S_s(k)$,

$$S_s(k) = |S_c(k)|e^{j\varphi(k)} \quad k = 0, \ldots, 79 \tag{4}$$

The time-domain stego speech, $s_s(n)$ is obtained by inverse FFT (IFFT) of the stego spectrum,

$$s_s(n) = ifft(S_s(k)) \quad n = 0, \ldots, 79 \tag{5}$$

Figure 2 illustrates the general steps of the VQ hiding technique. The stego signal can be made public. For example, it can be uploaded on the Internet. However, only those users having the reverse embedding algorithm can extract the secret message. To attract no suspicion about the presence of a secret message in the stego speech, a widespread speech signal that is available in thousands of copies on the Internet could be chosen as the cover signal.

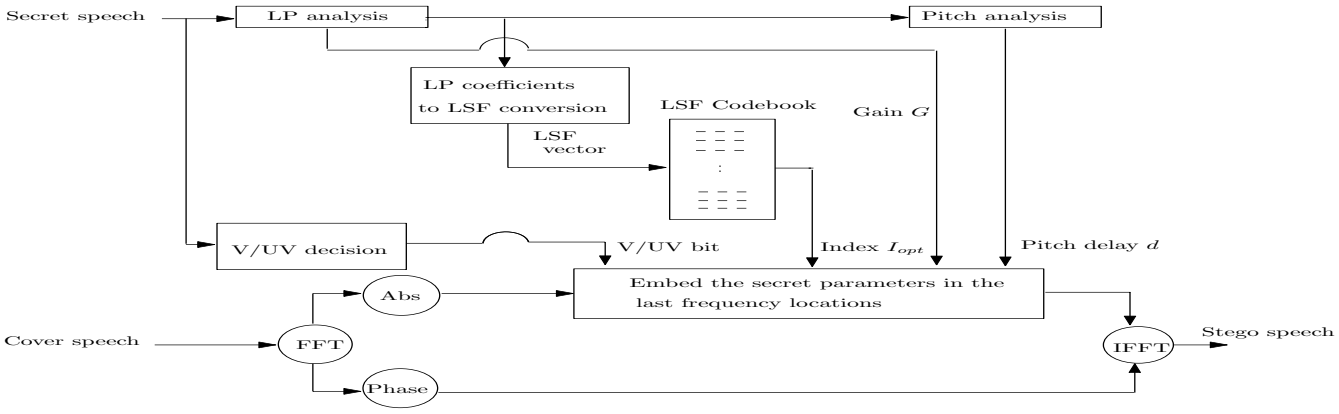

Fig. 2. Block diagram showing the general steps to hide the four secret speech parameters inside a cover narrowband signal.

III. Secret speech recovering

A. Secret speech parameters extraction

At the receiver, the secret speech reconstruction starts by extracting the four hidden parameters from the stego speech. A reverse order algorithm to the embedding process is used for this purpose. The stego speech $s_s(n)$ is first subject to a Fourier transform to convert it to the frequency domain $S_s(k)$. The stego speech spectrum $S_s(k)$ is then decomposed to magnitude and phase spectrum. The four secret speech parameters are then extracted from the same predefined magnitude spectrum locations.

- pitch delay $d = |S_s(79)|$
- gain $g = |S_s(78)|$
- LSF index $I_{opt} = |S_s(77)|$
- V/UV bit $vb = |S_s(76)|$

B. Secret speech reconstruction

Once extracted from the stego speech frame, the pitch delay d, gain g and the V/UV bit vb will be used directly to build the pitch filter for voiced speech. However, the index I_{opt} is applied to the LSF codebook to point to the optimal LSF vector $V_{I_{opt}}$. For this reason, the same copy of the LSF codebook must be available both at the transmitter and receiver. The LSF vector $V_{I_{opt}}$ is then converted back to a 10-dimensional LP vector $(a_1, \ldots, a_{10})$. The LP parameters are used to build the LP synthesis filter $H(z)$.

$$H(z) = \frac{1}{1 - \sum_{i=1}^{10} a_i z^{-i}} \tag{6}$$

A random generator produces a gaussian excitation signal $e(n)$ that is applied sequentially to the pitch and LP synthesis filters. The signal, $\hat{s}(n)$, at the output of the LP synthesis filter is a reproduction of the original secret message $s(n)$. Figure 3 illustrates the secret speech reconstruction process. Since the LPC-model parameter values that are extracted from the stego speech have the same exact values as the embedded ones, the reconstructed secret speech signal is not affected by the hiding algorithm. The minor degradations present in this signal, when compared with original secret signal, comes from the LPC model and the LSF vector quantization.

TABLE I
Objective performance of the LSF vector quantization. (In the Without LSF VQ: the secret speech is reconstructed from the LPC model 13 parameters.)

Speaker	SEGSNR (dB)	
	Without LSF VQ	With LSF VQ
Female	16.24	14.83
Male	16.12	14.75
Average	16.18	14.79

C. Impact of the LSF VQ on the secret speech

The LSF VQ has a positive impact on the cover speech since it reduces the amount of the information to be hidden. Better stego speech quality is achieved when using this technique. However, the negative impact of the LSF VQ is on the secret speech. The LSF vector $V_{I_{opt}}$ used at the receiver is just a closest match to the original LSF vector V_{inp}. Table I shows the impact of the LSF VQ on the secret speech in terms of the segmental signal to noise ratio (SEGSNR). It is apparent from this table that the LSF VQ introduces slight quality degradations into the secret speech. Informal listening tests to both the original and reconstructed secret speech signals approve the outcome of the objective measures. While some perceptual distortions are easily noticeable, the reconstructed speech $\hat{s}(n)$ remains perfectly intelligible.

IV. Evaluation

To assess the efficiency of the VQ hiding (VQH) method, we have conducted several comparative simulations in which both hiding techniques, scalar hiding (SH) and VQH are tested on the same database. We have used two assessment tools: 1) the segmental signal-to-noise ratio (SEGSNR), an objective criteria that measures the temporal discrepancy between the cover and stego signals, and 2) the comparative mean opinion score (CMOS), a subjective listening measure to spot any perceptual similarity between the cover and stego speech. The SEGSNR of a speech file is just the average of the SNRs of all file frames. For a each speech frame, the SNR in decibel (dB) is defined by

$$SNR(dB) = 10log_{10}\left(\frac{\sum_{n=0}^{79}[s_c(n)]^2}{\sum_{n=0}^{79}[s_c(n) - s_s(n)]^2}\right) \tag{7}$$

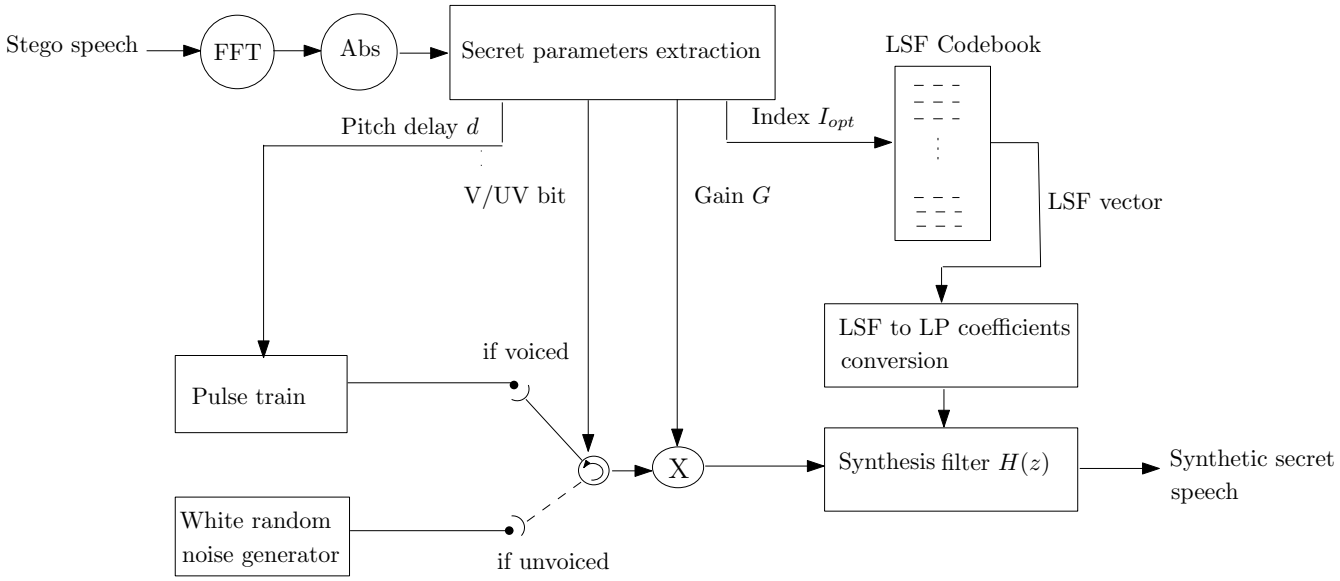

Fig. 3. Block diagram showing the steps to extract the secret parameters and reconstruct the secret speech $s(n)$ from the stego speech signal $s_s(n)$.

TABLE II
OBJECTIVE PERFORMANCE OF THE SCALAR AND VECTOR HIDING TECHNIQUES.

Speaker	SEGSNR (dB)	
	SH	VQH
Female	44.41	48.31
Male	44.32	48.28
Average	44.365	48.295

TABLE III
SUMMARY OF SUBJECTIVE TESTS (*CMOS*) WITH THE VECTOR QUANTIZATION HIDING (VQH) APPROACH

Cover speech	*CMOS*	
	SH	VQH
Female	0.14	0.42
Male	0.26	0.52
Average	0.20	0.47

The CMOS outcomes consist of a 3-level scale (-1,0,1). Each pair of a cover speech and its corresponding stego file is presented to each listener twice by reversing the order. Listeners have to announce the better quality signal between the cover and stego speech. Score 1 is marked if a listener chooses the cover speech, -1 if stego, and 0 if a listener couldn't notice any clear difference between both signals.
The evaluation speech database consists of 6 cover speech and 10 secret speech signals. The simulations are done in five rounds, in each round both SH and VQH systems are compared using one cover signal and one secret signal selected randomly from the secret speech database. In Table II, we present the SEGSNR for both hiding approaches. The VQH technique provides an important gain of almost 4 dB compared to the scalar hiding method. Table III shows that the similarity between the cover and stego speech is increased when using the VQ hiding technique.

V. CONCLUSIONS

This work has developed an efficient algorithm for hiding speech in speech without attracting any suspicion about the presence of the secret message in the stego signal. The technique presented used the LSF vector quantization to reduce the number of the secret speech parameters from 13 to 4. We have minimized the negative influence of the hiding process on the cover speech quality. Experimental results on real male and female voice segments have shown that our technique is capable of hiding one narrowband speech message inside another narrowband speech segment to produce a stego speech segment that is indistinguishable from the original cover speech, while being able to recover a perfectly intelligible copy of the secret speech message.

REFERENCES

[1] N. Johnson and S. Jajodia, "Exploring steganography: seeing the unseen," IEEE Computer, pp. 26-34, February 1998.
[2] Eugene T. Lin and Edward J. Delp "A review of data hiding in digital images", Proceedings of the Image Processing, Image Quality, Image Capture Systems Conference, PICS'99
[3] T. Rabie, A Novel Compression Technique for Super Resolution Color Photography, IEEE International Conference on Innovations in Information Technology (IIT2006), November 2006. Dubai, UAE, 1-5.
[4] D. Guerchi, H. M. Harmain, T. Rabie, and E. E. Mohamed, "Speech Secrecy: An FFT-based Approach," Special Issue on "Evolving Computer Science Applications" of the Journal of Mathematics and Computer Science, 3, n.2, pp. 107-125, 2008.
[5] D. Guerchi, "LPC-based Narrowband Speech Steganography," Journal of Multimedia. (To appear.)
[6] D. O'Shaughnessy, "Speech Communications: Human and Machine," Wiley-IEEE Press; 2 edition, 1999.
[7] F. Itakura, "Line spectrum representation of linear predictive coefficients," *Journal of Acoustics Society of America*, vol. 57, no. 1, pp. S35, 1975.