LPC-based Narrowband Speech Steganography

Driss Guerchi, Emad Mohamed
School of Engineering
Applied Sciences and Technology
Canadian University of Dubai
P.O. Box: 117781, Dubai, UAE
{guerchi, emad}@cud.ac.ae

ABSTRACT: *This paper proposes a new speech steganography system for secure sound massages sharing. This system exploits the advancements in speech processing to hide efficiently secret speech in narrowband cover speech. Linear predictive coding is used to represent the secret speech with reduced number of parameters. These parameters are embedded in selective perceptually-irrelevant frequency locations of the cover speech. Objective and subjective measures show that the resulting stego speech, which contains the secret message, is indistinguishable from the cover speech.*

Keywords: Speech Steganography, LPC Model, FFTbased Data Hiding

Received: 11 October 2011, Revised 9 December 2011, Accepted 20 December 2011

1. Introduction

Steganography is a young security engineering domain that offers a vital complement to encryption in order to achieve the ultimate in data protection [1], [2]. Speech steganography, which consists of hiding speech within speech [3], takes advantage of the recent advancements in speech compression and data hiding. In general, data hiding algorithms can be classified into two broad categories: substitution-based hiding and coding-based hiding. In substitution hiding, components of the cover signal are replaced by parameters of the secret message. In the coding-based steganography, the secret information is coded into some elements of the cover signal. Our original steganography system in [4] belongs to the first category since the last 30 frequency components of the cover speech are replaced by the parameters of the wideband secret speech. Even though this system has proven its efficiency in hiding a large amount of secret data, the replacement process, which takes place in a random manner, could be enhanced if the magnitude distribution of the cover magnitude spectrum is considered in the individual substitutions.

The objective in speech steganography is to hide as much information as possible while limiting the impact on the quality of the cover speech and making the steganalysis (the process of detecting the secret message by the opponent) more complex. Unlike wideband speech which tolerates more spectral changes because of its relatively larger bandwidth, hiding in narrowband speech is a particular task that should be done meticulously. The number of frequency locations available for embedding the secret parameters is very limited in narrowband cover speech, hence the need to compress the secret speech to a reduced number of parameters. Narrowband speech could be represented by few parameters but at the cost of degrading the synthetic speech quality. In military communications, low-rate compression is the purpose as long as the synthetic speech is intelligible. For these types of applications, parametric coding is the norm. For example, the US Government federal standard 1015 uses a 10^{th}-order linear prediction coding (LPC) vocoder operating at 2400 bits/s [5]. Speech frames are represented here by 13 parameters each.

The technique presented in this paper, which we name linear predictive coding-based narrowband speech steganography (LPC-based NSS), involves two major improvements on our original technique in [4]: firstly it combines both coding and substitution-based hiding techniques in an efficient hybrid data embedding system, and secondly, it utilizes a selective data embedding approach in which the secret parameters distribution will be shaped according to the magnitude of the cover frequency components. Our aim is to develop a steganography system that is more concerned with secure speech storage and multimedia applications than secure speech transmission.

The remainder of this paper is organized as follows. Section II presents some background on steganography. Section III describes the LPC-based narrowband speech steganography system. Section IV introduces the hybrid hiding algorithm and Section V illustrates the steps in recovering the secret message at the receiver. Section VI describes the simulation experiments and presents the evaluation results. Finally, the conclusions are summarized in Section VII.

2. Background

The performance of a steganography system is measured by the degree of meeting the constraints on its attributes. Among the main attributes of a steganography algorithm we list:

- The embedding capacity of the cover signal, measured by the amount of secret information to be hidden in the cover signal,
- The similarity between the cover and stego signals, evaluated by the impact of the hiding process on the cover speech quality,
- The resistance of the steganography system, related to the complexity of the hiding algorithm,
- The accuracy with which the hidden message are recovered at the receiver.

In essence, the fundamental objective of a steganography system is to optimize all the above attributes. For a given application, a tradeoff between these attributes must be achieved to satisfy the constraints for that application. In the particular case of speech steganography, the four attributes are related to the type of speech compression system. Speech compression is not only required in bandwidth-limited communication systems and digital voice storage but is today a fundamental component of any speech steganography system. For a fixed-size cover speech, secret speech compression increases the amount of hidden information (the embedding capacity) while producing a stego signal that is perceptually close to the cover signal. The secret sound could be represented with a reduced number of parameters that can be embedded in the frequency domain of the cover speech. However, such compression may lead to an artificial sounding but intelligible speech. As for military communications, one of the aspects of major importance in speech steganography is the intelligibility of the secret sound at the receiver. The secret message must be understood once extracted from the stego signal.

Compared to wideband cover speech, narrowband cover speech offers limited hiding capacity. For speech-inspeech hiding, we introduce a new parameter, named the hiding ratio and denoted by H_r, which refers to the embedding capacity of the cover signal. This ratio is defined by

$$H_r = \frac{\textit{Frame size of the secret speech}}{\textit{Frame size of the cover speech}} \times 100 \qquad (1)$$

The error between the cover and stego signals is tightly related to H_r. High values of the hiding ratio imply noticeable error, rendering the steganography system less efficient since the stego signal will arouse suspicions. This work tries to achieve a good tradeoff between all the above-mentioned performance criteria aided by recent advancements in speech compression and data hiding. Our goal is to develop a system with a hiding ratio of 100 % (embedding 20-ms secret speech frames into 20-ms cover speech frames) while limiting the impact on the cover speech.

3. LPC-Based Narrowband Speech Steganography

Speech analysis allows more efficient manipulation of the pertinent speech parameters than directly handling the time-domain speech signal. Speech hiding is one of these applications in which the secret speech signal must be first converted into a set of parameters. Hiding a time-domain secret message frame of 20 ms, for example, corresponds to embedding 160 samples in the cover speech frame.

For a narrowband cover speech, this will affect drastically the speech quality raising suspicions about the existence of a secret message. Hence, the necessity to compress the secret message before hiding. The selection of the adequate speech coder is crucial to the performance of the steganography system.

For the cover speech, hiding information in its timedomain representation will affect its quality producing a stego speech that is easily distinguishable from the cover speech. For these reasons, we introduced the Fourier transform-based steganography system which hides the parameters of the secret message in the magnitude spectrum of the cover speech. The following subsections present the preprocessing to which the secret and cover speech are subject before the hiding process.

3.1 Secret speech analysis

Compression has not only direct impact on the quality of the secret speech but also on the steganography performance. As one of the objectives of speech steganography is the capability of recovering an intelligible secret signal, we can adopt the LPC model [5] to compress the secret sound message to few parameters. At the receiver, the LPC model produces synthetic speech that is intelligible. In the analysis part of this model, each 20-ms speech frame $x(n)$, $n = 0, \ldots, 159$, is subject to: a linear prediction (LP) analysis to extract 10 linear prediction (LP) coefficients modeling the vocal tract, and a pitch analysis to determine the pitch delay for voiced speech. Sounds are classified here as voiced or unvoiced.

These two types of speech analysis removes most of the correlation present in $x(n)$ leaving a noise-like residual signal $r(n)$. This residual can be modeled by a white random noise $e(n)$.

1) LP analysis: For narrowband speech, ten LP coef- ficients are sufficient to model the first four formants. A tenth-order linear prediction analysis is performed once per 20-ms speech frame to estimate the LP filter coef- ficients $A(a_1, a_2, \ldots, a_{10})$. A 30-ms tapered rectangular window is used in this analysis [6]. This window picks 5 ms from past speech and another 5 ms from future speech. The autocorrelation method is adopted in the calculation of these spectral parameters. The LP gain G is obtained from the LP residual energy using the root mean square. Direct placement of the LP coefficients in the magnitude spectrum leads to negative frequency components. Hence the need for another representation of these parameters. One of the popular representations of the LP coefficients in speech coding is the line spectrum frequencies (LSF) [7]. The LSFs, w_i, are more suitable for hiding since they are monotonically increasing and are all positive.

$$0 < w_1 < w_2 < \ldots < w_{10} < \pi \tag{2}$$

2) Pitch analysis: Pitch analysis is performed only on voiced frames. A robust voiced/unvoiced (V/UV) speech classification is attained using zero crossing rate, normalized energy, and the LPC feature with linear discriminant analysis [8].

If the speech frame is voiced then an open-loop pitch analysis is applied on the LP filter residual $u(n)$.

$$u(n) = x(n) - \sum_{i=1}^{10} a_i x(n - i). \tag{3}$$

The optimum pitch period d is the one among all pitch lags (from 20 samples to 147 samples) that produces the highest autocorrelation of the LP filter residual [9].

The pitch analysis removes the long-term correlation leaving a noise-like signal $r(n)$

$$r(n) = u(n) - u(n - d). \tag{4}$$

3) Secret speech parameters: After both LP analysis and pitch analysis, each 20-ms speech frame is compactly represented by 13 parameters: 10 LSFs, gain G, pitch delay, and the V/UV bit. These parameters form the secret vector, H, to be hidden in the 20-ms cover speech frame. To reduce the impact of the data hiding on the magnitude components, the secret parameters, as well as the cover speech amplitudes, are normalized to unity (the U/UV bit doesn't need to be normalized since it has only two possible values, 0 or 1).

3.2 Cover speech adaptation

1) Where to hide: The ear is more tolerant to certain changes in speech spectral domain than in the time domain. This fact prompts us to exploit the speech spectral characteristics to hide some secret information in the magnitude spectrum while limiting speech quality degradation. In the LPC-based NSS system, the cover speech $c(n)$, $n = 0, \ldots, 159$, must be first transformed to frequency-domain $C(k)$ using the fast Fourier transform (FFT). The spectrum $C(k)$ could be decomposed into magnitude spectrum $|C(k)|$ and phase spectrum $\varphi(k)$.

$$C(k) = |C(k)|\, e^{j\varphi(k)} \qquad k = 0, \ldots, 159. \tag{5}$$

The spectrum of human speech production ranges approximately from 200 Hz to 3400 Hz. However, the standard 8 kHz narrowband speech sampling frequency is greater than the double of the highest frequency in narrowband speech, assuming a speech bandwidth of 4000 Hz. The LPC-based NSS technique will use the extra 600 Hz spectrum (3400-4000Hz) to hide a secret sound message exploiting the minor contribution of the high-frequency components to speech intelligibility. Lower frequencies are more important for speech intelligibility than higher frequencies [10].

2) Frequency locations/hidden data compromise: For each speech frame, the extra 600 Hz bandwidth added by the sampling process corresponds to $\frac{80+600}{4000}$= 12 frequency components $C(k)$, k = 68, . . . , 79, allowing the embedding of 12 secret parameters. However, the LPC model produces 13 parameters, hence the need to use one more frequency location (that of the component $C(67)$) from the initial speech bandwidth. A particular coding algorithm is used when hiding a parameter in this location. The objective is to minimize the impact of the hiding algorithm on this frequency component.

4. Hiding Algorithm

The hybrid hiding of the secret speech parameters consists of a combination of two embedding techniques:

Subvector	*Content*	*Embedding locations k*	*Embedding technique*
*H*1	Normalized LSF parameters	68 to 77	Substitution
*H*2	V/UV bit	67	Coding
*H*3	Normalized pitch delay and gain *G*	78 and 79	Coding or substitution

Table 1. Appropriate Embedding Technique the Different Secret Data Subsets

data coding and data substitution. The secret information is embedded in the spectrum range ($|C(67)|$ - $|C(79)|$) producing a stego magnitude spectrum ($|S(67)|$ - $|S(79)|$). In the hybrid hiding, the secret speech vector H is decomposed into three subvectors H_1, H_2, and H_3. Table I depicting this decomposition shows the appropriate embedding technique for the available frequency locations. For the LSF coefficients, a selective sequential substitution is performed to minimize the individual errors between each frequency component and its LSF coefficient substitute. Unlike the random substitution used in [4], the selective substitution modulates the LSFs distribution by the cover speech frequency components. The LSFs are here sorted according to the magnitude of the frequency components. This approach presents two advantages: 1) cover speech formants in the (3400-4000Hz) range are adequately modeled producing a stego speech that is perceptually close to the cover sound signal, and 2) hiding the LSF parameters, not in their ordered form, adds a level of complexity to the steganography algorithm, rendering the steganalysis more difficult. The selective placement of the LSF coefficients is summarized in the following algorithm:

$K = [68, 69, \ldots, 77]$
$W = [w_1, w_2, \ldots, w_{10}]$
for $i = 0 : 9$
  $k_{max} = \underbrace{argmax}_{k \in K} (|C(k)|)$
  $|C(k_{max})| = W(10 - i)$
  Update K by removing k_{max} *from this vector*
end
$|S(k)| = |C(k)|$ *for* $k = 68, \ldots, 77$

The V/UV bit, the pitch delay and LP gain are embedded into the remaining three spectrum locations using the following code:

```
if   V/UV bit = 1
        |S(67)| = |C(67)|
        |S(78)| = d
else
        |S(67)| = |C(66)|
        |S(78)| = |C(78)|
end
|S(79)| = G
```

The spectrum of the stego speech $s(n)$ is obtained by concatenating the modified part with the unchanged part of the cover speech spectrum.

$$S(k) = \begin{cases} C(k), & k = 0, \ldots, 66 \\ |S(k)|\, e^{j\varphi(k)}, & k = 67, \ldots, 79 \\ C(k), & k = 80 \\ S(160\ k), & k = 81, \ldots, 159 \end{cases} \quad (6)$$

The stego speech $s(n)$ is first acquired by inverse FFT of $S(k)$ then pulse code modulated and stored as a wave file. Figure 1 depicts the general steps of the LPC-based NSS algorithm.

5. Secret Speech Synthesis

The LPC-based NSS is a lossless hiding algorithm; the hidden parameters are extracted at the receiver from the stego speech following the hiding procedure in reverse order. Figure 2 depicts the steps in recovering the hidden parameters and reconstructing the secret sound signal. The secret parameters are retrieved from the magnitude spectrum of the stego speech and scaled to their original values (before the normalization). The LSFs *wi* are extracted from the magnitude spectrum of the stego speech then reordered according to the following algorithm.

```
S_lsf = [|S(68)| ,|S(69)| , . . . , |S(77)|]
for i = 0 : 9
  w_i = min(S_lsf )
  Update S_lsf by removing this relative minimum
  from this vector
end
```

The ordered LSFs are converted back to 10 LP coef- ficients a_i that are used to build the LP synthesis filter $H(z)$.

$$H(z) = \frac{1}{1 - \sum_{i=1}^{10} a_i z^{-i}} \quad (7)$$

The V/UV bit and the pitch delay (for voiced frames) are retrieved from the magnitude spectrum of the stego speech by the following decoding algorithm

```
if   |S(67)| = |S(66)|
        V/UV bit = 0 (frame is unvoiced)
else
        V/UV bit = 1 (frame is voiced)
        d = |S(78)|
 end
```

The LP gain G is obtained directly from location 79, $G = |S(79)|$.

According to the V/UV decision, the LP synthesis filter is excited by a pulse train with period d for voiced speech or a white random noise signal, $v(n)$, for voiced unvoiced sound. The LP excitation signal, $\hat{r}(n)$ is defined by

$$\hat{r}(n) = \begin{cases} \delta(n - d), & \text{if V/ UV bit} = 1 \\ v(n), & \text{if V/ UV bit} = 0 \end{cases} \quad (8)$$

where $\delta(n)$ is the dirac function. A copy of the secret speech, $\hat{s}(n)$ is obtained at the receiver by the following equation:

$$\hat{s}(n) = G\,\hat{r}(n) + \sum_{i=1}^{10} a_i\,\hat{s}(n - i) \quad (9)$$

Even though the temporal shape of the speech waveform is affected by the LPC analysis and synthesis, the informal listening tests reveal that the intelligibility of the secret message is preserved.

6. Evaluation and Discussions

We have examined the performance of the LPC-based NSS technique on a database of 10 cover speech and 5 secret speech. We conduced our simulations to study the main steganography attribute: the similarity between the cover and stego signals. Our comparative study consists of two types of measures: objective and subjective. As objective measures, we adopted the segmental signal-tonoise ratio (SEGSNR) and the average spectral distortion (AvgSD) [11]. The SEGSNR measures the impact of the hiding process on the time-domain waveform of the cover speech. It represents the average over the entire duration of the SNR for all speech frames. For a given speech frame, the SNR in decibel (dB) is defined by

$$SNR(dB) = 10 log_{10}\left(\frac{\sum_{n=0}^{159} [c(n)]^2}{\sum_{n=0}^{159} [c(n) - s(n)]^2}\right) \quad (10)$$

An informal listening test, represented by the comparative mean opinion score (CMOS) [12], has been performed as a subjective measure. For the CMOS, we rated, on a 3- level scale from -1 to 1, the listeners opinion on the better quality among the cover and stego signals. Score 1 is marked if a listener chooses the cover speech, -1 if stego, and 0 if a listener couldn't report any clear difference between both signals. Each pair of cover and stego speech is presented to each listener twice by reversing the order. We have also conducted other simulation experiments to compare the selective placement (SP) of the LSF parameters algorithm (adopted in the LPC-based NSS technique) with the random placement (RP) of these parameters (frequency components are sequently replaced by the ordered LSFs).

We provide results in Table II for the SEGSNR averaging values separately for female and male cover speech files. Figure 3 shows the cover and stego signals after using the LPC-based NSS algorithm to hide a female secret speech within a female cover speech. Both cover and stego signals looks almost similar producing a very high SEGSNRs. To evaluate the effect of the hiding process on the magnitude spectrum of the cover speech, we present in Table IV the average spectral distortion. The very small values of the spectral distortion reveal that the hiding of a secret message alters negligibly the magnitude spectrum of the cover speech. This result is supported by the resemblance between the cover and stego speech spectrograms in Figure 4.

Table 5 shows the CMOS for both LSFs placement algorithms. The tabulated results show that the LPC-based NSS technique is capable of hiding a secret signal in a cover signal of the same length while producing stego speech that is perceptually indistinguishable from the cover speech. The objective performance correlate with the subjective results since the computed SEGSNRs are very high. The processed speech files in this work are all in their raw format (wave format). In future work, we intend to study the impact of compression of the stego signal on the quality of the synthetic secret speech at the receiver. Our objective is to build a hiding system for secure transmission.

7. Conclusion

We presented in this paper a narrowband speech hiding technique for the purpose of secure storage of secret information. This

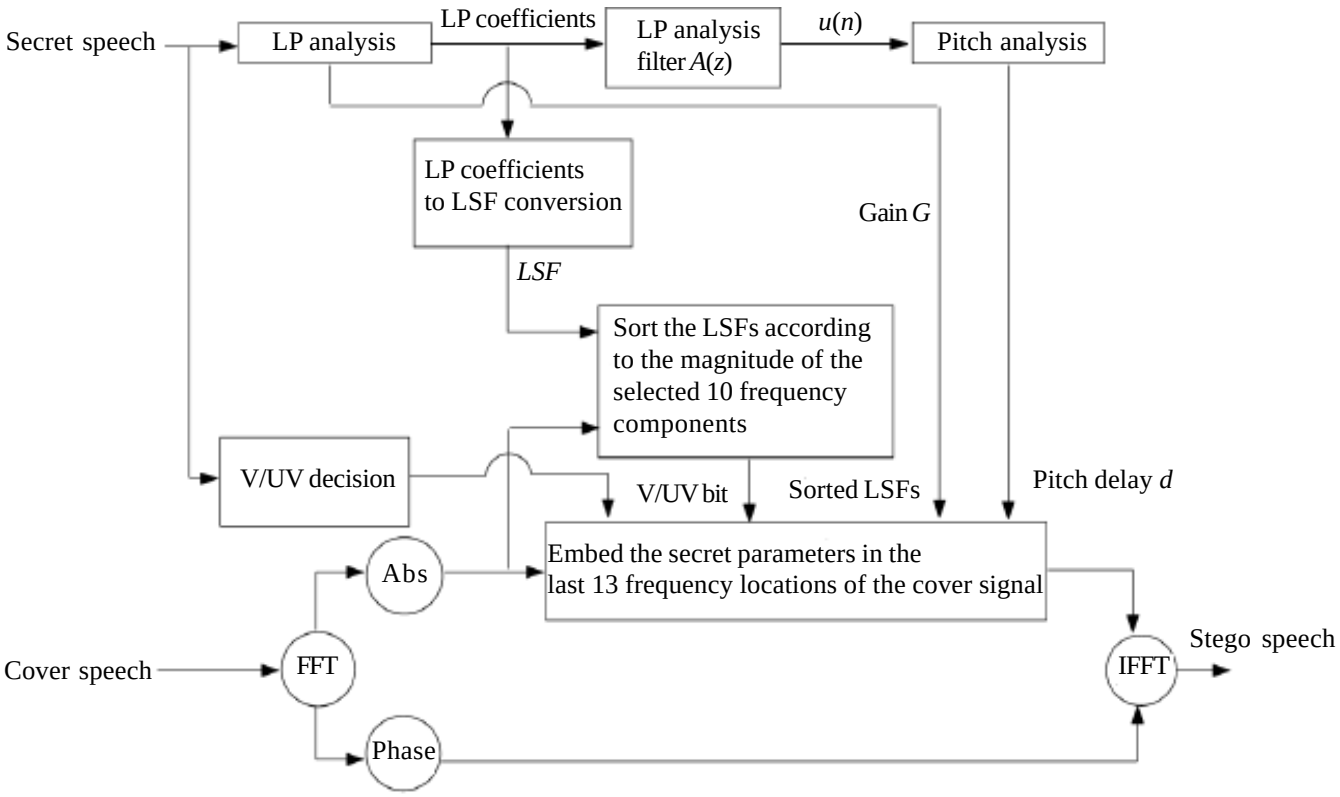

Figure 1. Block diagram showing the general steps to hide the secret speech parameters inside a cover narrowband signal $c(n)$

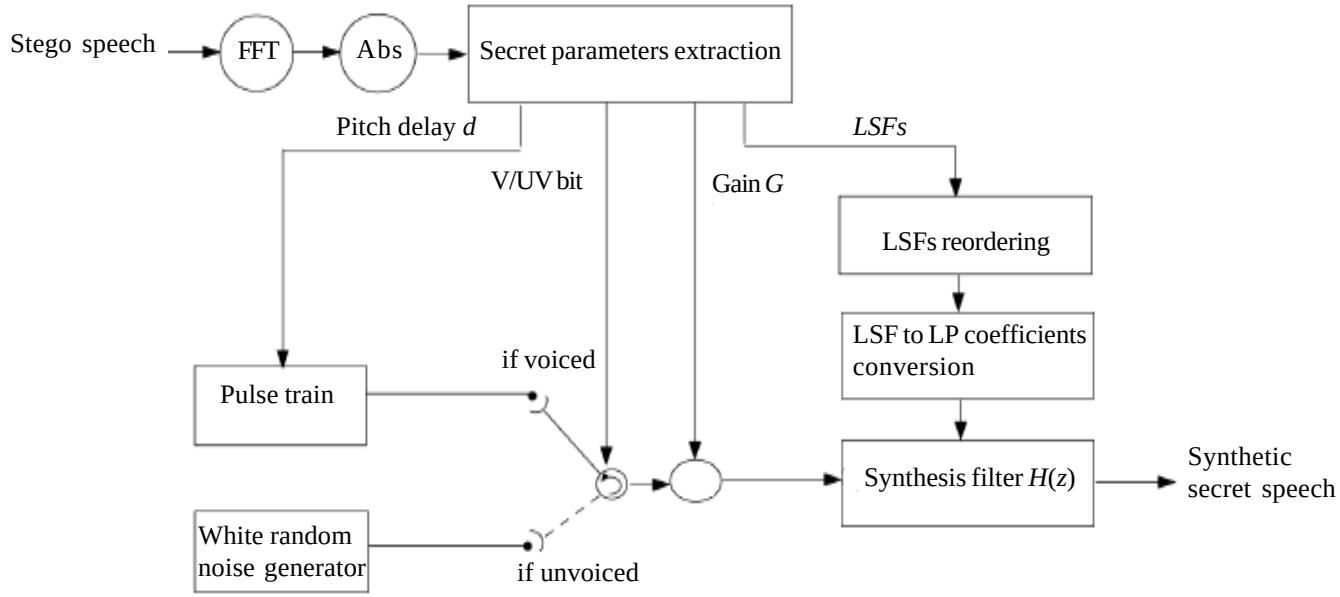

Figure 2. Block diagram showing the general steps to reconstruct the secret speech from the extracted LPC-model parameters

technique exploits the extra bandwidth added by the sampling process to embed a secret sound message in a narrowband cover speech without being detected. Linear predictive coding was used to model the secret speech by few parameters that were hidden in some perceptually-irrelevant frequency locations of the cover speech.

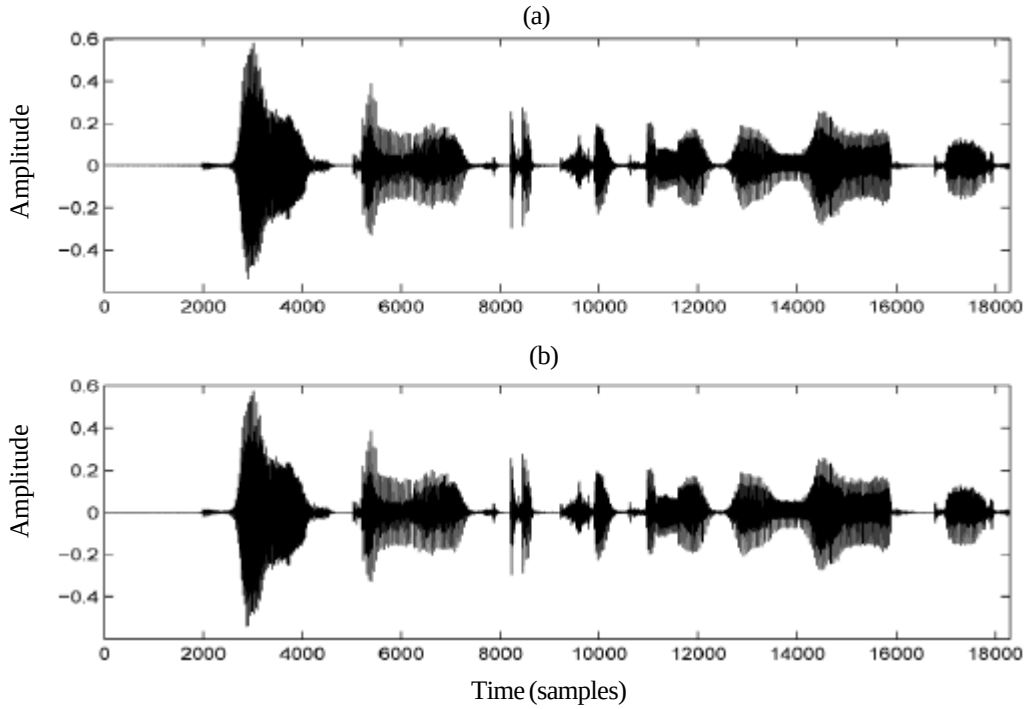

Figure 3. (a) Cover speech, (b) Corresponding stego speech after hiding a female secret message within the cover speech using the LPC-based NSS system

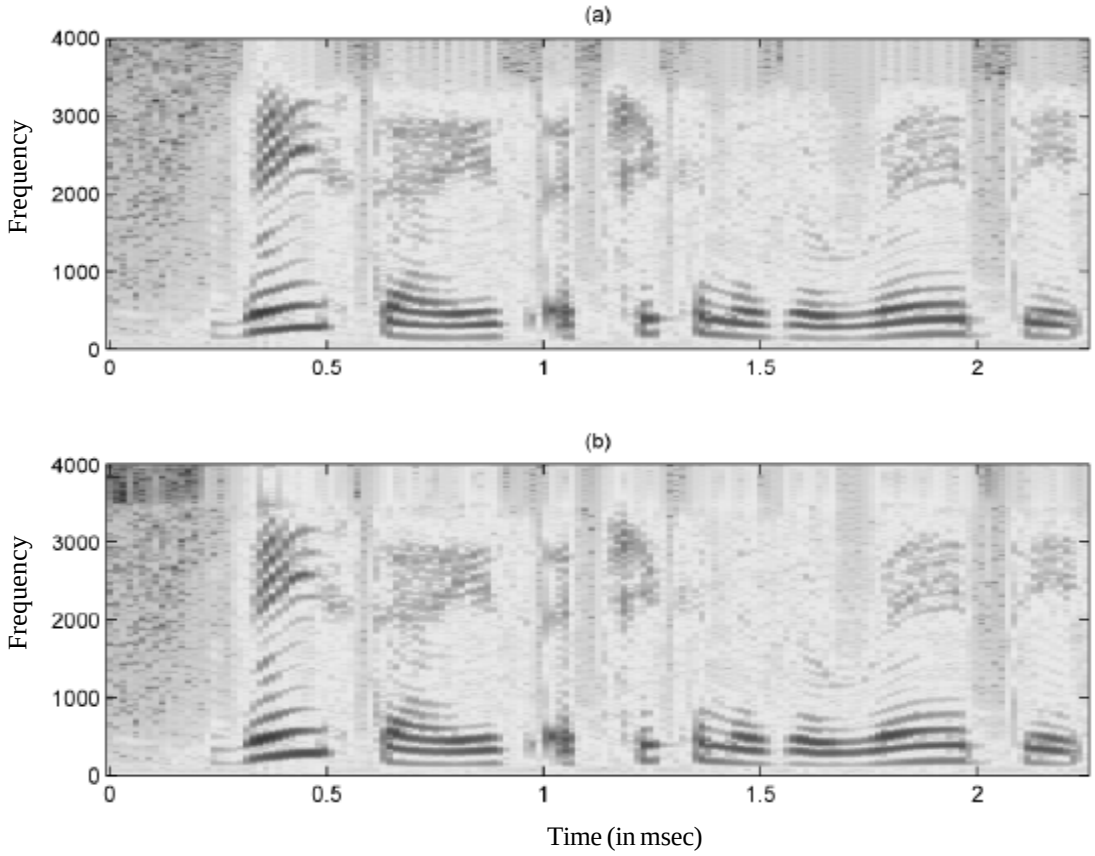

Figure 4. Spectrogram of (a) the cover speech, (b) the corresponding stego speech after hiding a female secret message within the cover speech using the LPC-based NSS system

Cover speech	*SEGSNR*(dB)	
	SP of LSFs	RP of LSFs
Female	48.24	44.43
Male	47.82	44.36
Average	48.03	44.395

Table 2. Summary of Objective Tests (*Segsnr*) With the LPC-Based NSS System

Cover speech	*Avg SD* (dB)	*Outliers (in %)*	
		2-4 dB	> 4 dB
Female	0.32	0.93	0.01
Male	0.35	0.94	0.01

Table 3. Average Spectral Distortion (For SP of LSFs)

Cover speech	*Avg SD* (dB)	*Outliers (in %)*	
		2-4 dB	> 4 dB
Female	0.58	0.94	0.01
Male	0.60	0.94	0.01

Table 4. Average Spectral Distortion (For RP of LSFs)

Cover speech	*CMOS*	
	SP of LSFs	RP of LSFs
Female	0.2	0.4
Male	0.3	0.6
Average	0.25	0.5

Table 5. Summary of Subjective Tests (*CMOS*) With the LPC-Based NSS System

8. Acknowledgment

The authors would like to thank Research Affairs at the Canadian University for funding.

References

[1] Potapova, R. O., Ponomar, M. O. (2006). Prospects of applications of speech steganography, *XVIII Session of the Russian Acoustical Society*, September 11-15.

[2] Artz, D. (2001). Digital steganography: Hiding data within data, *IEEE Internet Computing*, May-June 2001, p. 75-80.

[3] Aoki, N. (2006). A band extension technique for G.711 speech using steganography, *IEICE Transaction on Communication*, E89-B (6) June.

[4] Guerchi, D., Harmain, H., Rabie, T., Mohamed, E. (2008). Speech secrecy: An FFT-based approach, *International Journal of Mathematics and Computer Science*, 3 (2) 1-19.

[5] Tremain, T. E. (1982). The Government standard linear predictive coding algorithm, *Speech Technology*, p. 40-49.

[6] Guerchi, D., Qian, Y., Mermelstein, P. (2000). Pitchsynchronous linear-prediction analysis by synthesis with reduced pulse densities, *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Istanbul, Turkey, 3, 1491-1494, June.

[7] Itakura, F. (1975). Line spectrum representation of linear predictive coefficients, *Journal of Acoustics Society of America*, 57(1) S35.

[8] Shah, J. K., Iyer, A. N., Smolenski, B. Y., Yantorno, R. E. (2004). Robust voiced/unvoiced classification using novel features and gaussian mixture model, *IEEE International Conference on Acoustics, Speech, and Signal Processing*, May 17-21, Montreal, Canada.

[9] Salami, et al., R. A. (1998). Design and Description of CS-ACELP: A Toll Quality 8 kb/s Speech Coder, *IEEE Transactions on Speech and Audio Processing*, 6 (2) 116-130, March.

[10] Alku, P., Backstrom, T. (2004). Linear predictive method for improved spectral modeling of lower frequencies of speech with small prediction orders, *IEEE Transactions on Speech and Audio Processing*, 12 (2) 293-99, March.

[11] Paliwal, K. K., Atal, B. S. (1993). Efficient vector quantization of LPC parameters at 24 bits/frame, *IEEE Transactions on Speech, and Audio Processing*, 1(1), January.

[12] ITU-T P.800, Methods for subjective determination of transmission quality, (1996).