2014

Audio steganograpgy by phase modification

Fatiha Djebbar
United Arab Emirates University

Beghdad Ayad
University of Wollongong in Dubai, beghdadayad@uowdubai.ac.ae

Publication Details

Djebbar, F. & Ayad, B. 2014, 'Audio steganograpgy by phase modification', in R. Falk & C. Becker. Westphall (eds), SECURWARE 2014 - 8th International Conference on Emerging Security Information, Systems and Technologies, IARIA, Lisbon, Portugal, pp. 31-35.

Audio Steganograpgy by Phase Modification

Fatiha Djebbar
UAE University
College of Information Technology
AL Ain, UAE
Email: fdjebbar@uaeu.ac.ae

Beghdad Ayad
University of Wollongong in Dubai
Faculty of Engineering and Information Science
Dubai, UAE
Email: beghdadayad@uowdubai.ac.ae

***Abstract*—In this paper, we propose a robust steganographic system that embeds high-capacity data in phase spectrum. Our approach is based on the assumption that partial alteration of selected frequency bins in the phase spectrum leads to a smooth transition while preserving phase continuity. The frequency bins, in the phase, selected for data hiding are first defined in the magnitude through an election process and then mapped into the phase spectrum to embed data. Perceptual and statistical study results demonstrate that, in comparison with a recently proposed magnitude based audio steganography method, the phase based approach gains a considerable advantage against steganalysis attacks while giving similar or comparable hiding capacity and audio quality.**

Keywords–Information hiding; Phase Coding; Steganalysis.

I. INTRODUCTION

Digital audio steganography has emerged as a prominent source of data hiding across novel telecommunication technologies such as voice-over-IP and audio conferencing. Currently, three main methods are being used: cryptography, watermarking, and steganography. Encryption techniques are based on rendering the content of a message garbled to unauthorized people. In watermarking, data is hidden to convey some information about the cover medium such as ownership and copyright, where the hidden message could be visible or invisible. The primary goal of steganography consists of undetectably modifying a multimedia file to embed these data [1]. While steganography is about concealing the existence of 'hidden message', steganalysis is about detecting its existence [2]. Steganalysis, the counterpart of steganography, is regarded as "attacks" to break steganography algorithms by the mean of different audio processing and statistical analysis approaches.

Steganography in today's computer era is considered a sub-discipline of the data communication security domain. Lately, new directions based on steganographic approaches started to emerge to ensure data secrecy. Modern techniques of steganography exploit the characteristics of digital media by utilizing them as a carrier (cover) to hold hidden information. Covers can be of different types including image [4], audio [5], video [6], text [7], and IP datagram [8].

Several methods of audio data hiding have been proposed, whether in time or frequency domains, including low-bit coding, spread spectrum coding, phase coding, echo data hiding, etc [1]. To hide information within audio signals, [9][10] have designed a steganographic algorithm by manipulating higher LSB layers of the audio signal. Phase alteration and spread spectrum are used in [11][12]; wavelet coding is used in [13][14] and magnitude-based data hiding was proposed by [15]. Most of these methods take information hiding ratio as a major factor in evaluating the robustness of their algorithms. As it is generally expected, higher information-hiding ratio elevates the risk of detecting the presence of hidden data.

In this paper, we present a robust phase coding technique for digital audio steganography. The original contributions of the paper addresses mainly the undetectability issue of hidden data encountered in our previous work, where magnitude was solely considered [15]. The phase spectrum is explored, in particular, to benefit from the inherent advantages of phase data hiding, as it is commonly understood that, when phase coding can be used, it gives better signal to noise ratio [1]. Our work is supported by a thorough comparative study by steganalysis to judge the performance of our steganographic. The comparison is performed against our previously presented algorithm [15] and existing high capacity LSBs-based audio steganographic software: Steghide, S-Tools and Hide4PGP found respectively in [16]–[18]. Perceptual evaluation as well as the steganalysis study show that the resulting stego stream preserves the naturalness of the original signal and resists steganalysis attacks while achieving similar or comparable hiding capacity to that in [15].

The rest of the paper is organized as follows. Phase hiding algorithm is presented in Section II. Section IV describes the steps developed to recover the embedded message at the receiver's end. Section V presents the simulation experiments and subsequent evaluation results. Finally, we conclude our paper with a summary of our work and some future directions in Section VI.

II. MOTIVATION AND BACKGROUND

The particular importance of hiding data in audio files results from the prevailing presence of audio signal as an information vector in our human society. Data hiding in audio files is especially challenging because of the sensitivity of Human Auditory System (HAS). Alterations of an audio signal for data embedding purposes may affect the quality of that signal. However, data hiding in the frequency domain rather than time domain is of nature to provide better results in terms of signal to noise ratio [10]. In addition, Human auditory perception has certain particularities that must be exploited for hiding data efficiently. For example, our ability to resolve tones decreases with the increase of frequency of the tone. Thus, it is more effective for hiding data in the higher frequency regions than in low frequencies [19].

In audio signals sampled at 16 kHz and quantized at 16 bits, frequencies within the range of 50 Hz to 7 kHz are then

 ISBN: 978-1-61208-376-6

eligible to embed data. The cover audio is divided into M equal length frames. For a sampling frequency of 16 kHz, a 4 ms frame for example produces 64 samples. The resolution of each frequency component is equal to $16000/64 = 250Hz$. Thus, the first frequency component that could be used for hiding data will be 250 Hz instead of 50 Hz (the starting frequency of wide-band speech). If we consider the Fourier symmetry feature of the spectrum, the number of normalized frequencies or the number of locations to hide data within each frame will be from $F_{HDmin} = 1$ to $F_{HDmax} = 28$ in [0.25 7] kHz frequency band. In each selected energetic frequency component location, at least a bit from the payload is embedded.

III. Proposed hiding algorithm

In our scheme, the cover-signal is divided into M frames of 4 ms, each contains N samples, $s_c(m,n), \quad 1 \leq m \leq M$ and $1 \leq n \leq N$. The magnitude spectrum $|S_c(m,k)|$ is isolated by transforming each frame to frequency domain using Fast Fourier Transform (FFT), $S_c(m,k) = FFT(s_c(m,n))$. The hiding band is specified by $F_{HDmin} \leq k \leq F_{HDmax}$, where F_{HDmin} and F_{HDmax} are the minimum and the maximum hiding band locations. In our algorithm, we only select high energy frequency components in an attempt to minimize the embedding distortion. A $threshold$ value is set for that purpose where a frequency bin is selected for data hiding only if its energy is higher or equal to the threshold value. Data is embedded along a chosen LSB layer (CLSB) to $\Delta(m,k)_{dB}$. Where CLSB is the LSB layer lower-limit for hiding in a frequency bin. In our experiments, CLSB is chosen to be the 5^{th} LSB layer at minimum. The Δ value models the upper limit for data hiding in a selected bin. Δ value is set to impose a good quality on the stego-audio. The selection process of frequency bins done in the magnitude spectrum as well as the hiding locations are summarized in Figure 2. the details of the embedding process in a selected frequency bin is described in Figure 3. The value of $\Delta(m,k)_{dB}$ is set to $(|S_c(m,k)|)_{dB} - 13dB$. In doing so, we benefit from the fact that noise that is 13 dB below the original signal spectrum for all frequencies is inaudible [20]. Even though the frequency bins qualified for data hiding are selected in the magnitude spectrum, we believe that we will benefit also from mapping it to the phase spectrum for the following reasons:

1) As we partially alter selected frequency bins, only few bits in each selected frequency component are modified, which will give a smooth transition while preserving phase continuity.
2) When phase coding can be used, it gives better signal to noise ratio [20].
3) Opportunities to increase hiding capacity are worth to be explored

To embed in the phase spectrum, we map the exact selected frequency bins from the magnitude spectrum into the phase spectrum $\phi(m,k)$ and data is also embedded along CLSB layer to $\Delta(m,k)_{dB}$. Embedding data in phase spectrum is described as follows:

for $m = 1$ to M **do**
  for $n = 1$ to $N/2$ **do**
    $|\phi_s(m,n)| \leftarrow |\phi_c(m,n)|$
  end for
  for $k = F_{HDmin}$ to F_{HDmax} **do**
    if $10 * log_{10}(|S_c(m,k)|) \geq threshold_{dB}$ **then**
      if $\Delta(m,k)_{dB} \geq CLSB_{dB}$ **then**
        $|\phi_s(m,k)| \leftarrow |\phi_c(m,k)| + \delta(m,k)$
      end if
    end if
  end for
end for

Figure 1: Algorithm used to compute $|\phi_s(m,k)|$

In Figure 1, the value of $\delta(m,k)$ represents the modification in the phase value. A full description of the phase modification induced by embedded bits in a given frequency component is shown in Figure 3. The number of injected bits in a frequency component depends on its energy. In this manner, the embedding in a given frequency bin in $|\phi_s(m,k)| \leftarrow |\phi_c(m,k)| + \delta(m,k)$ is redefined as: $|\phi_c(m,k)| = (a_n 2^n + a_{n-1}2^{n-1} + a_{n-2}2^{n-2} + ...a_2 2^2 + a_1 2^1 + a_0 2^0)$

Where $a_n = \{0,1\}$ and $\delta(m,k) = (d_i 2^i + d_{i-1}2^{i-1} + ... + d_0 2^0)$. The value of stego-phase can be simply calculated using:

$$|\phi_s(m,k)| = (a_n 2^n + a_{n-1}2^{n-2} + d_i 2^i + ...d_1 2^1 + d_0 2^0 + a_1 2^1 + a_0 2^0)$$

Finally, the new phase is multiplied with its magnitude to produce the stego-spectrum such as: $S_s(m,k) = |S_c(m,k)|e^{j\phi_s(m,k)}$. The inverse $iFFT$ transformation is applied on the segment to get the new stego-audio segment $s_s(m,n)$.

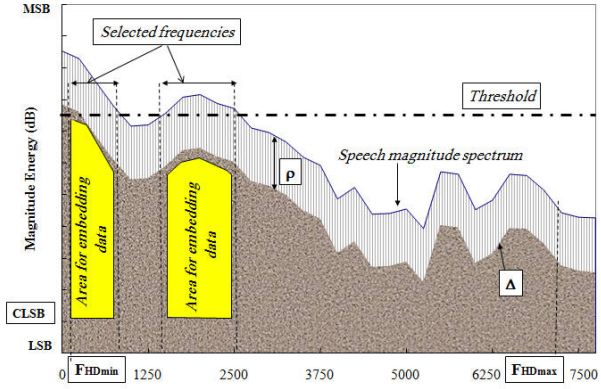


Figure 2: Spectral embedding area located in a frequency frame.

IV. Hidden data retrieval

To extract the hidden data from the phase spectrum, two main steps are followed: first, we locate the bins used for data embedding from the magnitude part $|S_s(m,k)|$. To do so, the parameters impacting the location of embedding in each selected bin such as $Threshold$, $\Delta(m,k)$, $CLSB$ are

 ISBN: 978-1-61208-376-6

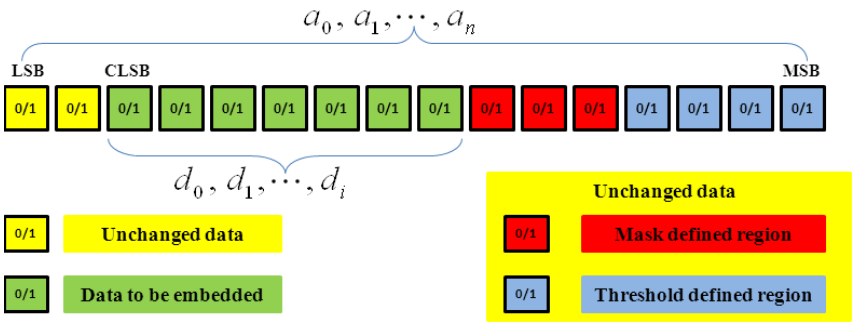


Figure 3: Embedding process in a selected frequency bin.

computed in the same way as done at the sender end. Second, we map the embedding locations found in the magnitude to the phase spectrum. Segments of the secret data are extracted and then reassembled as follows:

for $m = 1$ to M **do**
  for $n = 1$ to $N/2$ **do**
    $|\phi_s(m,n)| \leftarrow |\phi_c(m,n)|$
  end for
  for $k = F_{HDmin}$ to F_{HDmax} **do**
    if $10 * log_{10}(|S_c(m,k)|) \geq threshold_{dB}$ **then**
      if $\Delta(m,k)_{dB} \geq CLSB_{dB}$ **then**
        Extract $\delta(m,k)$ from $|\phi_s(m,k)|$
      end if
    end if
  end for
end for

Figure 4: Algorithm used to extract $\delta(m,k)$

V. Performance Evaluation

To evaluate the performance of the proposed algorithm, we conducted a comparative study between stego- and cover-audio signals. The study is based on (1) perceptual and (2) steganalysis undetectability.

A. Perceptual undetectability

In this section, we assess the quality of the stego-audio, when the hiding capacity is maximized. Tests have been conducted for magnitude [15] and the proposed phase configuration. Perceptual evaluation of speech quality ($PESQ$) measure defined in the ITU-T P862.2 standard combined with segmental SNR ($SegSNR_{dB}$) were utilized for the objective evaluation [21]. The hiding $Rate(Kbps)$ achieved is computed accordingly. Tests are carried out on a set of 100 audio waves, spoken in different languages by male and female speakers. Audio signals are 10s length each and sampled at 16 kHz and data is embedded within [0.25-7] kHz band with maximum hiding ratio of 23 kbps.

The PESQ test produces a value ranging from 4.5 to 1. A PESQ value of 4.5 means that the measured audio signal has no distortion: it is exactly the same as the original. A value of 1 indicates the severest degradation. The effectiveness of our algorithm is evaluated on audio frames sampled at 64. We set the algorithm parameters' value to maximize the hiding capacity while maintaining audio quality quality, i.e, Threshold = -20dB, ρ= 15dB, CLSB=1, F_{HDmin} and F_{HDmax} are set to 1 and 28 for 4ms frame length.

In our simulation, the distortion between stego and cover audio signals is calculated over several frames and by averaging the statistics, the overall measure is obtained. $SegSNR$ value for one modified audio frame of 4 ms is given by the following equation:

$$SegSNR_{dB}=10\log_{10}\left(\frac{\sum_{k=1}^{28}|S_c(m,k)|^2}{\sum_{k=1}^{28}|S_c(m,k)-S_s(m,k)|^2}\right) \quad (1)$$

The summation is performed over the signal per frame basis. To evaluate the results, the following criteria were used. First, the capability of embedding larger quantity of data (Kbps) is sought while naturalness of the stego-audio is retained. Second, the hidden data is fully recovered from the stego audio-signal.

TABLE I: PERFORMANCE EVALUATION AND COMPARISON

Hiding Method	SNR_{dB}	PESQ
[15]	26.86	4.32
Proposed	32.31	4.48

The values of SNR and PESQ registered in Table I are obtained from frames of 4 ms, hiding ration 23 Kpbs and 5^{th} LSB layer. They indicate clearly that stego-signals generated by the proposed phase embedding approach have experienced less distortion compared to [15]. Moreover, phase coding is robust to common linear signal manipulation such as: amplification, attenuation, filtering and resampling.

B. Comparative study by steganalysis

To further investigate our steganography algorithm performance, a comparative study by steganalysis is conducted based on a state-of-the-art reference audio steganalysis method [3]. The comparison is performed against our magnitude data hiding [15] and existing audio steganographic software: Steghide, S-Tools and Hide4PGP found respectively in [16]–[18]. The selected reference method was applied successfully in detecting the presence of hidden messages in high capacity LSBs-based steganography algorithms [3]. It is based on

 ISBN: 978-1-61208-376-6

extracting Mel-cepstrum coefficients (or features) from the second order derivative of audio signals. The features are then fed to a support vector machine (SVM) with RBF kernel [22] to distinguish between cover- and stego-audio signals.

For each studied steganography tool and algorithm, two datasets are produced: training and testing. Each dataset contains 270 stego and cover WAV audio signals of 10s length. All signals are sampled at 44.1 kHz and quantized at 16-bits. Each training and testing dataset contains 135 positive (stego) and 135 negative (cover) audio samples. We used online audio files from different types such as speech signals in different languages (i.e, English, Chinese, Japanese, French, and Arabic), and music (classic, jazz, rock, blues).

All stego-audio signals are generated by hiding data from different types: text, image, audio signals, video and executable files. To make a fair comparison between all assessed algorithms, the cover-signals were embedded with the same capacity of data. More precisely, S-Tools's (with hiding ratio of 50%) hiding capacity is used as a reference to embed the candidate steganographic algorithms and tools. The performance of each steganographic algorithm is measured through the levels by which the system can distinguish between stego and cover-audio signals (Table III). In order to analyze the obtained results, we first present the contingency table (see Table II).

TABLE II: THE CONTINGENCY TABLE

	Stego-signal	**Cover-signal**
Stego classified	True positives (tp)	False negatives (fn)
Cover classified	False positives (fp)	True negatives (tn)

The entries of the contingency table are described as follows:

- *tp*: stego-audio classified as stego-audio signal
- *tn*: cover-audio classified as cover-audio signal
- *fn*: stego-audio classified as cover-audio signal
- *fp*: cover-audio classified as stego-audio signal

In subsequent formulas, *all* represents the number of positive and negative audio signals. The value of the information reported in Table II is used to calculate the following measures:

$$Accuracy(AC) = \frac{tp + tn}{all} \quad (2)$$

The receiver operating characteristic (ROC) value is the fraction of true positive (TPR= true positive rate equivalent to Sensitivity) versus the fraction of false positive (FPR= false positive rate equivalent to 1-Specificity). Following the preparation of the training and testing datasets, we used the SVM library tool available at [23] to discriminate between cover- and stego-audio signals. The results of the comparative study are reported in Table III. The accuracy of each studied tool and method is measured by the values of AC and ROC.

In our second experimental work, we assess the performance evaluation of our algorithm and compare it to [15]–[18]. The values presented in Table III are the percentages of stego-audio signals correctly classified. Higher score values are interpreted as high-detection rate. Consequently, the proposed method show a significant improvement over the other, whereby, we were able to add a considerable accuracy to our steganographic algorithm against steganalysis attacks. The fact that the phase embedding scheme was able to perform better than the other algorithms, shows that the distortion amount resulting from embedding similar embedding ratios is much smaller.

TABLE III: OVERALL ACCURACY STEGANALYSIS RESULTS

Hiding methods	**AC**
Stools	0.725
Steghide	0.67
Hide4PGP	0.85
[15]	0.775
proposed	0.575

Further details on the behavior of each algorithm are represented in term of ROC curves in Figure 5. In each graph, a higher curve corresponds to more accurate detection rate while a lower curve corresponds to low accurate detection rate.

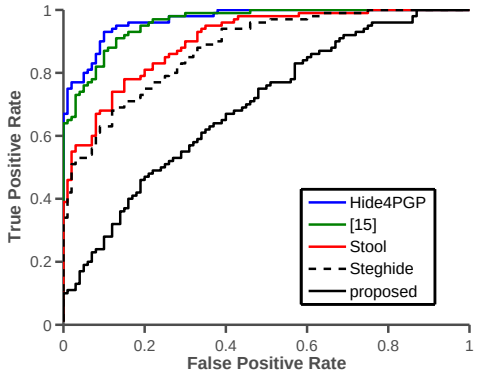


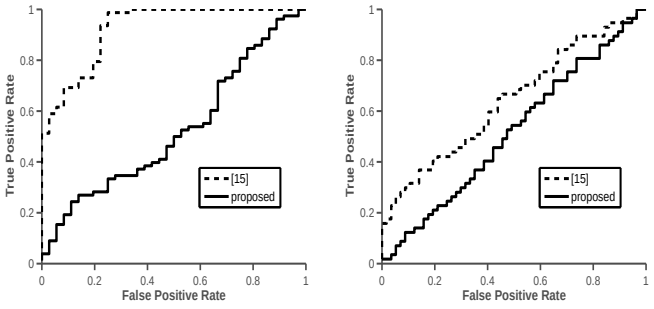
Figure 5: ROC curves for steganographic methods [15]–[18] and the proposed algorithm.

For the second experiment we further investigate the performance of our algorithm when the dataset contains only speech or music signals. The aim of this experiment is to put more emphasis on the behavior of the proposed algorithm when music audio-signals are used to convey hidden data versus those of speech audio-signals. We split the dataset into two sets A (130 speech signal) and B (130 music signal). Each set is further split to 65 stego- and 65 cover-signal to create a training and testing dataset for speech as well as for music. A set up similar to that described for experiment 1 was employed. The overall results in Table IV and Figure 6, show that our method performs better whether for speech- or music-signals. Our finding shows also that data-hiding in music Figure (6b) is less detectable than in speech-signals Figure (6a). In fact, the reference steganalysis method uses features extracted from high frequencies (lower in energy) while in our algorithm we target high energetic frequency components to embed data. In addition, the number of low-energy frequency components in music audio signals is smaller than that in speech signals.

 ISBN: 978-1-61208-376-6

TABLE IV: STEGANALYSIS RESULTS FOR DATA IN SPEECH AND IN MUSIC AUDIO SIGNALS

Hiding methods	Audio signal	AC	ROC
proposed	Music	0.504	0.502
	Speech	0.558	0.558
[15]	Music	0.6	0.598
	Speech	0.84	0.84



(a) (b)

Figure 6: ROC curves for [15] and the proposed method for data-hiding in speech (6a) versus music (6b) audio signals.

VI. CONCLUSION

In this paper, we presented a robust phase audio steganography. This work has a double aim. The first aim is to benefit from the fact that when phase coding can be used, it gives better signal to noise ratio. The second is to address the undetectability issue which is overlooked by most of the presented work in audio steganography. Perceptual and steganalysis study results reveal a great potential to hide large amounts of data, while ensuring their security and preserving the naturalness of the original signals. In the future, we plan to extend our work by investigating steganalysis of audio signals in codec domain.

REFERENCES

[1] F. Djebbar, B. Ayad, K. A. Meraim, and H. Hamam, "Comparative study of digital audio steganography techniques", EURASIP Journal on Audio, Speech, and Music Processing, Dec 2012, pp. 1-16.

[2] Avcibas, "Audio steganalysis with content independent distortion measures", IEEE Signal Process Letter, 2006, vol. 13, no. 2, pp. 92-95.

[3] Q. Liu, A. H. Sung, and M. Qiao, "Temporal derivative-based spectrum and mel-cepstrum audio steganalysis", IEEE Transactions on Information Forensics and Security, 2009, vol. 4, no. 3, pp. 359-368.

[4] A. Cheddad, J. Condell, K. Curran, and P. Mc Kevit, "Digital image steganography: Survey and analysis of current methods", Signal Processing, Marsh 2010, vol 90, issue 3, pp. 727-752.

[5] F. Djebbar, K. Abed-Maraim, D. Guerchi, and H. Hamam, "Dynamic energy based text-in-speech spectrum hiding using speech masking properties", 2nd International Conference on Industrial Mechatronics and Automation (ICIMA), May 2010, vol.2, pp. 422426.

[6] R. Balaji and G. Naveen, "Secure data transmission using video Steganography", IEEE International Conference on Electro/Information Technology (EIT), May 2011, pp. 1-5.

[7] M. Shirali-Shahreza and S. Shirali-Shahreza, "Persian/Arabic Unicode Text Steganography", SIAS Fourth International Conference on Information Assurance and Security, Sept. 2008, pp. 62-66.

[8] G. Handel Theodore and T. Maxwell Sandford II, "Hiding Data in the OSI Network Model", Information hiding: first international workshop, Cambridge, UK. Lecture Notes in Computer Science, 1996, vol. 1174, pp. 23-38.

[9] N. Cvejic and T. Seppanen, "Increasing Robustness of LSB Audio Steganography Using a Novel Embedding Method", Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC'04), 2004, vol. 2, pp. 533537.

[10] M. A. Ahmed, M. L. M. Kiah, B. B. Zaidan, and A. A. Zaidan, "A novel embedding method to increase capacity and robustness of low-bit encoding audio steganography technique using noise gate software logic algorithm", Journal of Applied Sciences, 2010, vol. 10, pp. 59-64.

[11] X. Dong, M. Bocko, and Z. Ignjatovic, "Data hiding via phase manipulation of audio signals", IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004. Proceedings (ICASSP'04), vol. 5, pp. 377-380.

[12] K. Gopalan, "Audio steganography using bit modification", Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP'2003), vol. 2, pp. 421-424.

[13] S. Shirali-Shahreza and M. Shirali-Shahreza, "High capacity error free wavelet domain speech steganography", Proc. 33rd Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP 2008), Las Vegas, Nevada, USA, pp. 17291732.

[14] N. Cvejic and T. Seppanen, "A wavelet domain LSB insertion algorithm for high capacity audio steganography", Proc. 10th IEEE Digital Signal Processing Workshop and 2nd Signal Processing Education Workshop, Georgia, USA, October 2002, pp. 5355.

[15] F. Djebbar, H. Hamam, K. Abed-Meraim, and D. Guerchi, "Controlled distortion for high capacity data-in-speech spectrum steganography", International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IEEE-IIHMSP), ISBN: 978-0-7695-4222-5, 2010, pp. 212-215.

[16] Steghide, http://steghide.sourceforge.net/. Retrieved 28 Sept, 2014.

[17] Stools Version 4.0, http://info.umuc.edu/its/online_lab/ifsm459/s-tools4/. Retrieved 28 Sept, 2014.

[18] Hide4PGP, http://www.heinz-repp.onlinehome.de/Hide4PGP.htm. Retrieved 28 Sept, 2014.

[19] G. S. Kang, T. M. Moran, and D. A. Heide, "Hiding Information Under Speech", Naval Research Laboratory, http://handle.dtic.mil/100.2/ADA443638, Washington, 2005.

[20] B. Paillard, P. Mabilleau, S. Morissette, and J. Soumagne, "PERCEVAL: Perceptual Evaluation of the Quality of Audio Signals", journal of Audio Engeneering Society, 1992, vol. 40, pp. 21-31.

[21] Y. Hu and P. Loizou, "Evaluation of objective quality measures for speech enhancement", IEEE Transactions on Speech and Audio Processing, 16(1), 2008, pp. 229-238.

[22] N. Cristianini and J. Shawe-Taylor, "An introduction to Support Vector Machines", Cambridge University Press; 2000.

[23] http://www.csie.ntu.edu.tw/$\sim$cjlin/libsvm. Retrieved 28 Sept, 2014.

 ISBN: 978-1-61208-376-6