Real Steganography in Non Voice Part of the Speech

S. Abdul kather
PG scholar
Department of CSE
National Engineering College Kovilpatti

K. Vimal
Assistant professor
Department of CSE
National Engineering College Kovilpatti

ABSTRACT
Embedding a secret message into a cover media without attracting any attention is known as steganography. Steganography is one of the methods used for hidden communication purposes. One of the cover media that can be used for steganography is speech. All the methods that we have found for audio steganography, changes the values of maximum number of samples from the audio signal. Usually change the sample values of the signal annoying the listener and reduce perceptual transparency. So the special methods are required for hiding the information in audio signal. In this paper, we propose a new steganographic method in speech signals. In this method, secret data are hidden in the silence part of the speech signal. The silence parts are identified by collaborative non voice detection algorithm. The secret data are hidden by reducing a small number of sample values from some samples of the silence part. The main features of our method is create the high perceptual transparent steganographic system with acceptable data hiding capacity This method can hide information in a speech stream with very low processing time that makes our method as a real-time steganography method.

Keywords
Steganograpgy, Information hiding, speech signal covert communication, data embedding

1. INTRODUCTION
Steganography is derived from the Greek words stegos which means covered and graphia which means writing. steganography is the art of covered or hidden writing. Hidden exchange of information is one of the important areas of information security which includes various methods like cryptography, steganography and water marking. Steganography is one of the methods that attract attention during the recent years. The steganography algorithm is to maintain the natural appearance of the cover media and to keep uninvolved people from even thinking about the information exists.

The main purpose of steganography is covert communication to hide a message from a third party. This is a major difference between steganography and other methods of hidden exchange of data. In cryptographic methods, individuals see the encoded data and notice the secret data exist but they cannot comprehend it. However in steganography individuals will not notice at all that data exist in the sources[11].

Watermarking is another popular technique to hide messages and it is usually used for providing ownership on copyrighted multimedia material and for detecting originators of illegally made copies. Therefore, an effective watermarking method must be robust against a variety of attacks. In contrast to watermarking, steganography prefers to hide information as much as possible and requires cover media with distortion as little as possible.

There are three important parameters in designing steganography methods. They are perceptual transparency, robustness and hiding capacity. Hiding capacity refers to the amount of information that can be hidden in the cover medium. perceptual transparency refers to an eavesdropper's inability to detect hidden information and robustness refers to the amount of modification the stego medium can withstand before an adversary can destroy hidden information [6]. These requirements are known as the magic triangle and are contradictory[3].

Most of the steganography methods are hiding information in images. Image steganography has been extensively explored and schemes based on bit modification of pixels and coefficients insensitive under the various psychovisual and statistical models such as spread spectrum generalized Gaussian in wavelet coefficients , qualified significant wavelet trees , variable-size model , and Quantization Index Modulation . All these schemes are characterized with high embedding capacity (large payload), excellent imperceptibility and accurate extraction, thanks to the low sensitivity of the human visual system (HVS) to luminance. 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 [3].

All the methods that we have found for audio steganography change the values of maximum number of samples from the audio signal. Usually change the sample values of the signal annoying the listener and reduce perceptual transparency. In this paper we propose a new steganographic method in speech signals. In this method, secret data are hidden in the silence part of the speech signal. The silence parts are identified by collaborative non voice detection algorithm. The secret data are hidden by reducing a small number of sample values from some samples of the silence part. A special feature of speech signals is silence frame. In speech signals there are the frames which the speaker is not speaking.In these frame the value of signal is low (ideally they must be zero).

One of the problems of applying general audio steganography methods such as LSB method to speech signals is that they usually hide information in all parts of the signal including silence intervals. But changing the values of samples of silence intervals for example by replacing LSB bits usually increases the values of the samples which make it annoying for listener and reduce perceptual transparency. So the special methods for hiding information in speech signals are required.

2. RELATED WORK

Audio steganography methods are classified into two major groups. They are temporal methods and transform domain methods. In temporal methods the hiding process such as least significant bit replacement is done in time domain. Time domain steganography methods are usually simple and fast [6].

In transform domain method the hiding process is done in wavelet domain [3]. One of the problems of steganography in transform domain is their unhidden errors [3]. Some methods like [3] use special strategies to reduce this error whereas other methods like [6] use error correction coding (ECC).

The method [6] presented a audio steganographic scheme specifically designed for high MP3 resistance where the covert data embedded in the signal properties which would be least affected by the compression process, DFT phase domain and noisy component. The method can achieve an embedding capability at 20-60 bps but with a poor imperceptibility. Gopalan et al [12] presented a steganography method specifically aimed at embedding covert voice messages in wireless communication using a spread spectrum based utterance embedding strategy Gopalan's scheme is LSB-based, generally not MP3 resistant thus inapplicable for the music-based covert communications.

The steganography methods like [5] try to achieve high hiding capacity whereas some other methods like [4] try to be robust against different attacks such as MPEG-1 layer III (MP3) compression. There are a number of audio steganography methods that did not directly hide the secret data in the waveform. The steganography method proposed in [7] hide the secret data during the MP3 compression algorithm. This is done by modifying the quantisation process of MP3 compression algorithm and select proper quantised values. The analysis by synthesis (ABS) based speech information hiding approach was adopted to embed the secret data in an original speech carrier with good efficiency in steganography and improve good quality of output speech [8].

The method proposed in [10] hides the secret data in inactive frames of low bit rate audio streams. The method like [10] can achieve larger hiding capacity only when it use ITUG.723.1 source codec to encode the audio streams.The scheme proposed in [9] hides the secret data in silence interval of the speech.The method [9] obtain high transparency only with low data hiding capacity.

The theoretical analysis foresaid suggest that steganography in nonvoice part of speech would attain a high data hiding capacity with high imperceptibility. The rest of this paper illustrate new steganographic method for hiding the secret data into the nonvoice part of the speech with high data hiding capacity and high transparency.

3. STEGNOGRAPHY IN NON-VOICE PART

Figure1 shows the block diagram of our proposed algorithm for steganography in Non voice part of the speech signal. The proposed process has three subprocess:Non Voice Detection, Data embedding and Data Extraction. The corresponding algorithms are detailed below.

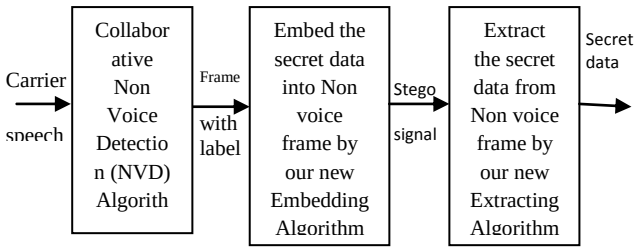

Figure1: Block Diagram of Our Proposed Algorithm For Steganography In Non Voice Part of the speech signal

3.1. COLLABORATIVE ADAPTIVE NON VOICE DETECTION (CNVD) ALGORITHM

The basic principle of a non voice detection algorithm is that it extracts measured features or quantities from the input speech signal and then compares these values with thresholds usually extracted from the selected feature of the signal. Non voice part (NVP) is declared if the measure values are lesser than the thresholds. Otherwise there is speech is present.

In collaborative based non voice detection, the energy and zero crossing rate of the signal is compared with their thresholds depending on the noise level. The non voice part is detected when the energy estimation lies below the threshold and the zero crossing rate beyond the threshold range.

If ($Ej<K.E_T$)&& $Nzcr(Fj)>R$ Frame is non voice (1)

Else Frame is voice

In the equation (1) Nzcr (Fj) is the number of Zero crosses detected in Fj . R is the set of values {5,6,7,….15} the number of zero crossings for the speech frame of 10ms.Ej represent the energy of the frames and E_T represent the threshold energy of the signal. Having a scaling factor K allows a safe band for adapting E_T.

The common way to calculate the energy of the speech signal is the root mean square energy (RMSE) which is the square root of the average sum of the squares of the amplitude of the signal samples.

$$E_j = 1/N \sum_{i=(j-1).N+1}^{j-N} x^2(i) \quad (2)$$

In the equation(2) Ej represent the energy of the frame. N denotes the number of samples in the frame, x(i) is the i^{th} samples of the carrier speech. The threshold energy (E_T) of the given carrier speech signal is calculated as,

$$\text{Threshold } (E_T)=(1-y)*E_{max}+y*E_{min} \quad (3)$$

$$y=(E_{max}-E_{min})/E_{max} \quad (4)$$

Here, Y is a scaling factor controlling the estimation process. E_{max} and E_{min} values are calculated by the equation (2). Collaborative Based Non Voice Detection Algorithm update the threshold value periodically by using the Following equations

$$E_{Tnew} = (1-p) \times E_{Told} + P.Esilence \quad (5)$$

Here, E_{Tnew} is the updated value of the threshold, E_{Told} is the previous energy threshold and Esilence is the energy of the most recent non voice frame. Parameter P is constant(0<P<1).

3.2. EMBEDDING ALGORITHM

Figure2 shows the detailed embedding process of our proposed embedding algorithm.

Step 1) To find non voice frame from the speech signal, First the speech signal is divided into equal size frames, F = {f1,…..fn}. N is the total number of frames in the carrier speech signal. These frames are inputted into the collaborative NVP algorithm above. The frame is marked with NF if it is determined to non voice part otherwise the frame is marked with voice frame (VF).

Step 2) We call the secret data that we want to hide in the non voice frame as S, then the value of S should be the range of $0 \leq S < 2^{BHF}$,BHF denotes the bit hide in each frame. We can hide only BHF bits in each frame.

Step 3) Encode the secret data S into the non voice frame. We can hide the secret data S into the non voice frame (NF) by making the number of non zero samples in the NF becomes congruent with S modulo 2^{BHF}, which is shown in (6).

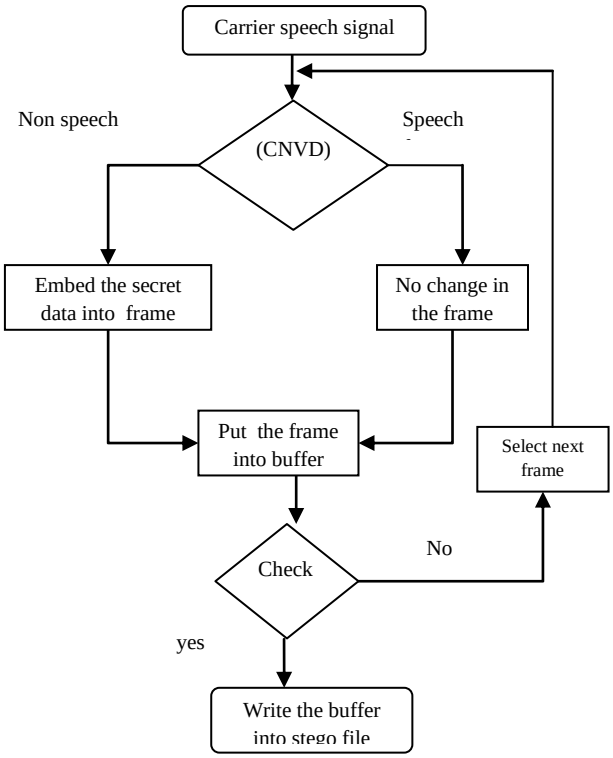

Figure 2: Flowchart for embedding process

$$New\ NZS \equiv S \bmod 2^{BHF} \quad (6)$$

After finding new number of non-zero samples (NEW NZS) value. We should reduce some non-zero samples into zero for making the number of non-zero sample in the frame is equal to NEW NZS. We can reduce only the sample which one is close to zero.

For Example, if BHF=4 and we want to hide the secret number 12 in the non-voice frame with 175 number of non-zero sample. We reduce 3 non-zero sample values into zero. Then the new number of non-zero sample is 172 which is congruent with 12 modulo 24.

3.3. EXTRACTION ALGORITHM

Figure 3 show the extraction procedure of secret information from the stego speech. The extraction procedure is divided into the following two steps.

Step1: Non voice frame detection from the stego signal. The stego signal is divided into frames (frame size must be the size set in the embedding process). The non voice frame from the stego signal must be detected by the collaborative NVP algorithm.

Step 2: Extract the secret data from the stego signal. We can extract the secret data by finding number of non-zero samples (NZS) in the non voice frame. The hidden information in each non voice frame is the remainder of NZS divided by 2^{BHF}.

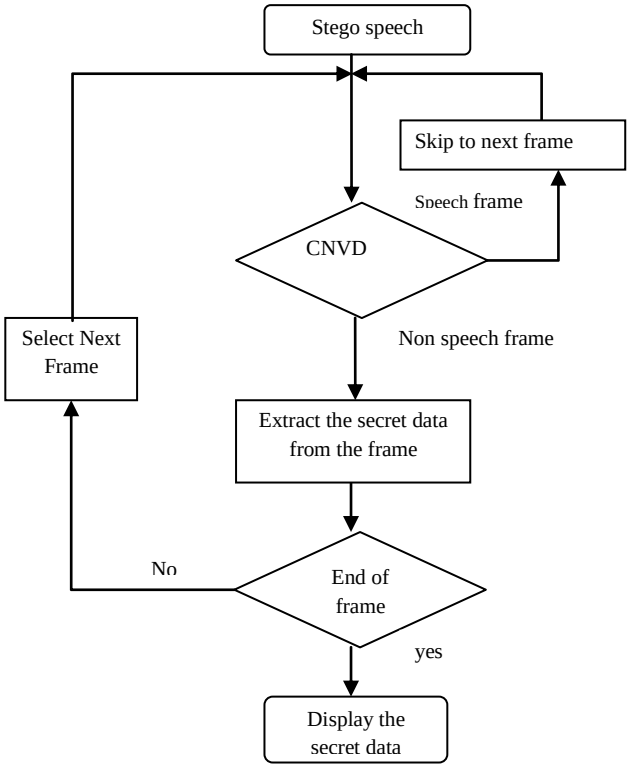

Figure 3: Flowchart For Extraction Algorithm

In this method the parameter BHF and frame size are acting like a key in the embedding and extraction process, because extracting the hidden data without them is impossible. A good feature of our method if we use the false values for this parameter during extraction then the meaningless data are extracted without any problem, which provides some kind of protection. The absence of a secure key, which enables anyone who knows the hiding parameters to extract the hidden data, is not a disadvantage of our

method. As we mentioned in the introduction section, the goal of stegonography is hiding the presence of data. For achieving high security, a cryptography method should be used to encrypt data before hiding it and decrypt it after extraction.

4. EXPERIMENTAL RESULTS AND PERFORMANCE ANALYSIS OF PROPOSED ALGORITHM

In this section , we will mention the experimental results and performance analysis of our proposed method. In our experiments, Non voice detection, data embedding and data extraction were conducted in the speech signal by means of the corresponding algorithms detailed in Section III.

We implemented this method in java. Two parameters, data embedding capacity and imperceptibility were used to evaluate the performance of the proposed steganography algorithm. We used five samples to test our method. In all the samples, the speakers say an identical sentence. Three of the speakers are men and two are women. These files have duration of about 10s, 22050hz sample rate , 16 bits per sample and are mono. Secret information was embedded in the non voice frame of the speech files, the data embedding capacity was estimated and the bit imperceptibility of the resulting stego files was evaluated accordingly for each sample. The experimental results are discussed in details below.

Table 1: Number of non voice frame of 5 speech sample

Speech File	Duration (s)	Total No of Frame	No of non voice Frame
Sample1	20	1375	189
Sample2	19	1306	244
Sample3	18.7	1284	247
Sample4	16.9	1160	151
Sample5	18.8	1295	206

4.1 Imperceptibility

To verify the imperceptibility of the proposed steganography system, the same secret information was embedded in the speech file mentioned in the Table 1. The imperceptibility was evaluated based on VSNR value and the sample modification rate of the resulting stego speech files. The VSNR is defined as the variation in signal to noise ratio(SNR) between the original speech and the stego speech , given by

$$VSNR = |SNR_o - SNR_s| \qquad (7)$$

where SNR_o and SNR_s are the SNR of the original speech and the stego speech, respectively.

The sample modification rate is the ratio between the number of sample modified (during hiding process) to the total number of the sample present in the original carrier speech file. We define the modification rate is

$$Sample\ Modificati\ on = \frac{Number\ of\ Sample\ Modified}{Total\ No.of\ Sample\ in\ Original\ Carrier\ File} \qquad (8)$$

The parameters BHF and frame size are controls the sample modification rate. The BHF is directly control the sample modification rate and the relation between BHF and sample modification rate is nearly linear. The reason is also clear, because if we double the number of bits that we hide in the non voice frame, then the total sample modification is also increased. So, more changes are done in the signal, which means more artifacts and lower quality.

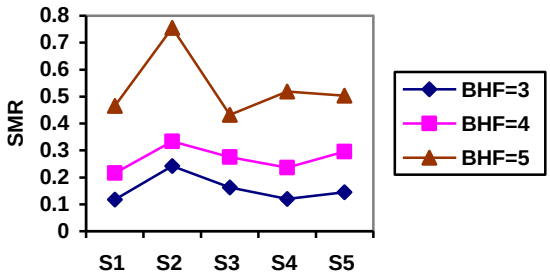

Figure 4: Sample Modification Rate

The acceptable sample modification rate for the selected frame size was calculated for different BHF value. Fig 4 shows the results of experiments on the 5 speech files listed in Table, horizontal axis represent the speech samples and the vertical axis represent the sample modification rate. Experiment indicates that in most instances the sample modification rate is small. The BHF value with sample modification rate is less than 0.4 sample is consider as optimal BHF value and it is chosen to embed secret information.

After finding the optimal BHF value then we calculate the VSNR values for all the samples mentioned in the Table1. The calculated VSNR values for all the samples are listed in the table. Experimental result in the table shows the VSNR value is so small between the original speech and stego speech, indicating that the proposed steganography algorithm for embedding information in the non voice frame achieved perfect imperceptibility.

Table 2: VSNR value for speech sample mentioned in table1

Speech File	VSNR
Sample1	0.53
Sample2	0.37
Sample3	0.32
Sample4	0.04
Sample5	0.38

4.2 Data Hiding Capacity

The hiding capacity is the ratio between the number of hidden bits to the total number of bits in the carrier signal. We define the hiding capacity as:

$$Hiding\ \ Capacity = \frac{Number\ of\ \ Hidden\ bits}{Size\ of\ \ Cover\ Signal\ in\ \ bit} \qquad (9)$$

The hiding capacity are depends on the parameter BHF and frame size selected in the embedding algorithm. If we choose larger BHF value and adaptable frame size then the hiding capacity is increased. If we choose the minimum BHF value and larger frames size then the hiding capacity is decreased. We use the BHF=4 and fix the frame size is equal to 320

samples during our tests. The hiding capacity of all the five samples with these parameters (BHF and frame size) is shown in Table1. The average hiding capacity is 44.42 bits per second (bps).

Table 3:Hiding capacity of speech samples listed in table

Speech File	Hidden Data size (bits)	Hiding capacity (bps)
Sample1	759	38
Sample2	979	51.42
Sample3	990	52.8
Sample4	605	35.88
Sample5	825	44

Several other algorithms such as variable low bit coding(VLBC) [5] and steganography in silence interval (SSI)[6] were previously suggested for embedding information in non voice part of the speech signal . For comparison purposes , these previously suggested algorithms and our proposed steganography algorithm were adopted to embed data in the speech sample files listed in Table1 respectively . Fig.5 shows the comparisons in data embedding capacity between our proposed algorithm (denoted as RSS) and the other algorithms.

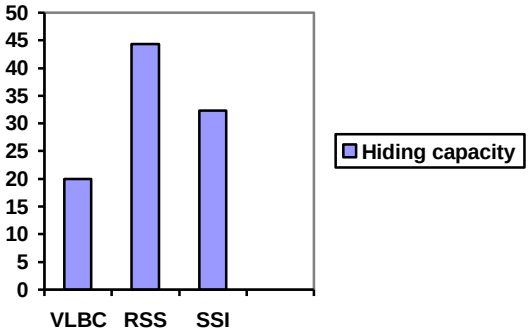

Figure 5:Comparisons in data embedding rates between the proposed algorithm "RSS" and other algorithms.

As Fig.5 shows the data embedding rate of our proposed algorithm RSS was much greater than those of the other algorithms. This is because the proposed algorithm identifies the maximum number of non voice frame from the speech signal by using the CNVD algorithms.

5. CONCLUSION

In this paper we propose a new steganography method for hiding data in speech signals. Our method hides data by changing the least number of samples in the non voice part of the speech signal. Our method has good perceptual transparency with acceptable high data hiding capacity. The main advantage of our method is a temporal domain. In addition, our method requires low computations. This enables our method to be a real-time method for streaming speech signals. In the encoder it modify a least number of samples from silence part which only requires a small buffer and can be done in real time. In the decoder, it only counts the non zero sample in the silence part to extract data and does not even require a small buffer.

6. REFERENCES

[1] E. Cole, Hiding in Plain Sight: Steganography and the Art of Covert Communication, Wiley Publishing, USA, 2003.

[2] N. Cvejic, Algorithms for Audio Watermarking and Steganography, Oulu University Press, Finland, 2004

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

[4] BAOP., MAX.: "MP3-resistant music steganography based on dynamic range transform". Proc. 2004 Int. Symp On Intelligent Signal Processing and Communication Systems (ISPACS 2004), 18–19 November 2004, pp. 266–271.

[5] HIRALI-SHAHREZA S., MANZURI-SHALMANI M.T.: 'High capacity error free wavelet domain speech steganography'. Proc. 33rd Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP 2008), April 2008, pp. 1729–1732.

[6] GANG L., AKANSU A.N., RAMKUMAR M.: 'MP3 resistant oblivious steganography'. Proc. of 2001 IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'01), 7–11 May 2001, vol. 3, pp. 1365–1368.

[7] PETITCOLAS F.: MP3Stego, 2006, http://www.petitcolas. net/fabien/steganography/mp3stego/, accessed August 2009.

[8] WU Z.-J., YANG W., YANG Y.-X.: 'ABS-based speech information hiding approach', Electron. Lett., 2003, 39, (22), pp. 1617–1619.

[9] SHIRALI-SHAHREZA S., SHIRALI-SHAHREZA M, 'Real-time and MPEG-1 layer III compression resistant steganography in speech'.IET Information security 2010,vol.4,Iss.1,pp. 1-7.

[10] CVEJIC N., SEPPA¨NEN T.: 'Increasing robustness of LSB audio steganography by reduced distortion LSB coding', J. Univers. Comput. Sci., 2005, 11, (1), pp. 56–65.

[11] JUDGE J.C.: 'Steganography: past, present, future', SANS White Paper, 30 November 2001, http:// www.sans.org/ rr / papers/ index.php?id¼ 552, accessed August 2009.

[12] K. Gopalan, Audio steganography using bit modification, Proceeding, of International Conference an Multimedia and expo, July 2003.