Vector Quantization based Steganography for Secure Speech Communication System

Bekkar Laskar[1] and Merouane Bouzid[2]
[1]*Tamanghasset University, Tamanghasset, Algeria*
[2]*Dept. Telecommunication, Electronics Faculty, USTHB University, Algiers, Algeria*

Keywords: Data Hiding, Steganography, Vector Quantization, Binning Scheme, ISF Parameters, Secure Speech, Wideband Speech Coder, MELP, AMR-WB.

Abstract: Data hiding (steganography or watermarking) involves embedding secret data into various forms of digital media such as text, audio, image and video. In this paper we propose two variants of vector quantization (VQ) based steganography method to hide secret speech signal in host public speech coded by the AMR-WB (Rec. G.722.2). The secret bit stream is hidden by using the basic principle of binning scheme which is carried out in the split-multistage vector quantization of G.722.2 immittance spectral frequencies (ISF) parameters.

1 INTRODUCTION

Steganography is the art of hiding secret information in a cover media without attracting attention. Indeed, modern steganography techniques exploit the characteristics of digital media by using them as carriers (covers) to hold hidden information. Covers can be of different types including text, speech/audio, image and video. Thus, the sender embeds secret data in a digital cover file using a key to produce a stego-file, in such a way that an observer cannot detect the existence of the hidden message (Cox et al., 2008). In this work, we focalize particularly on speech steganography techniques which consist in hiding a secret speech signal into a cover (host) signal.

A variety of speech/audio steganography methods have been proposed in the past, where most of them are based on the temporal domain, the transform domain and the compression domain. An extended review of the current state-of-art literature in digital audio/speech steganography techniques in each domain is given in (Djebbar et al., 2012). In compression domain, speech steganography techniques based on vector quantization (VQ) have been getting more and more popular, since they enhance the traditional VQ compression by adding the ability of data hiding.

In (Geiser and Vary, 2008), Geiser and Vary proposed a method to embed digital data in the bitstream of an ACELP speech codec. In (Yargicoglu and Ilk, 2010), Yargicoglu and Ilk proposed a data hiding methods that embed secret data during MELP coding of the speech signal. The secret bits are hidden by using quantization index modulation (QIM) with the multistage vector quantization (MSVQ) of line spectral frequencies (LSF) parameters.

In this paper, two variant of the binning scheme approach are developed for secure speech communication. It is about a two steganographic binning scheme (SBS) methods by VQ codebook division called balanced and unbalanced codebook partitioning. Our steganographic speech system consists in embedding secret speech signal into host public speech coded by the Adaptive Multi-rate Wideband (AMR-WB, ITU-T G.722.2) (Bessette et al., 2002) speech coder. For the compression of the secret speech, we used the 2.4 kbits/s Mixed-Excitation Linear Predictive (MELP) (McCree, 1996) speech coder. The embedding process of the MELP secret bit stream is carried out into the split-multistage vector quantization (S-MSVQ) indices of G.722.2 immittance spectral frequencies (ISF).

Laskar, B. and Bouzid, M.
Vector Quantization based Steganography for Secure Speech Communication System.
DOI: 10.5220/0006398304070412
In *Proceedings of the 14th International Joint Conference on e-Business and Telecommunications (ICETE 2017) - Volume 4: SECRYPT*, pages 407-412
ISBN: 978-989-758-259-2

2 VQ-BASED DATA HIDING: BINNING SCHEME APPROACH

Several data hiding methods, based on vector quantization, have been proposed in literature (Cox et al., 2008), (Moulin and Koetter, 2005). One of the most popular quantization-based data hiding method is probably the quantization index modulation (QIM) (Moulin and Koetter, 2005). Before presenting the basic idea of our approach, based on the QIM binning scheme, let us first review briefly the basics of the conventional VQ system.

2.1 Vector Quantization

A k-dimensional VQ of rate R bits/sample is a mapping of k-dimensional Euclidean space $\Re^k$ into a finite codebook $Y = \{y_0, \ldots, y_{L-1}\}$ composed of $L = 2^{kR}$ codevectors (Gersho and Gray, 1992). The design principle of a VQ consists of partitioning the k-dimensional space of source vectors x into L non overlapping cells $\{R_0,..., R_{L-1}\}$ (partition) and associating with each cell R_i a unique codevector y_i such that the total average distortion D is minimized (Gersho and Gray, 1992).

Various algorithms for the optimal design of VQ have been developed in the past. The most popular one is certainly the LBG algorithm (Gersho and Gray, 1992). This algorithm is an iterative application of the two optimality (nearest neighbor and centroid) conditions such as the partition and the codebook are iteratively updated.

2.2 Binning Scheme based on VQ Codebook Partition

The considered steganographic binning scheme (SBS) in this work is the one that modify the VQ indices by codebook partitioning to hide secret bits sequence (Geiser and Vary, 2008).

To embed a message of n bits per input cover vector, the basic idea of the SBS is to partition first the main VQ codebook Y into 2^n disjoint sub-codebooks Y_i ($i = 0,\ldots, 2^n - 1$), by referring to a user-key K called also "stego-key". Then, for each input cover vector, a sub-codebook is selected according to the steganographic bit to be embedded. The traditional VQ search procedure is then done using the sub-codebook for the input vector. Notice that in the one bit embedding case (n = 1 bit), the VQ codebook Y is partitioned in two sub-codebooks Y_0 and Y_1. Then, the input vector is coded using the nearest codevector from Y_0 and Y_1 according to whether the secret bit is 0 or 1, respectively.

Figure 1 presents an example of SBS codebook partitioning for embedding one bit per VQ index according to stego-key $K = \{1, 0, 0, 1, 1, 0, 1, 0\}$.

Many codebook partitioning techniques for VQ-based data hiding have been proposed in the past. In (Kim, 2002), Jo and Kim proposed a method to improve imperceptibility by partitioning each pair of the codevectors with the shortest Euclidean distance into three sub-codebooks, according to a given threshold. They modified the VQ indices to embed secret bits, while only two sub-codebooks are used in the data hiding procedure. In (Wang, 2007), Wang et al. developed an efficient method which also partitions the VQ codebook into sub-codebooks to modify the VQ indices in order to carry secret bits.

In the next section, we present our two SBS codebook partitioning methods inspired from the basic idea of Wang, Jain and Pan (Wang, 2007) VQ-based data hiding approach applied for image watermarking.

3 PROPOSED VQ-BASED DATA HIDING METHOD

Our SBS system can be divided into three phases: pretreatment (VQ codebook partitioning), message embedding and message extraction. For simplicity, the description below is limited to embedding only one bit into each input cover vector.

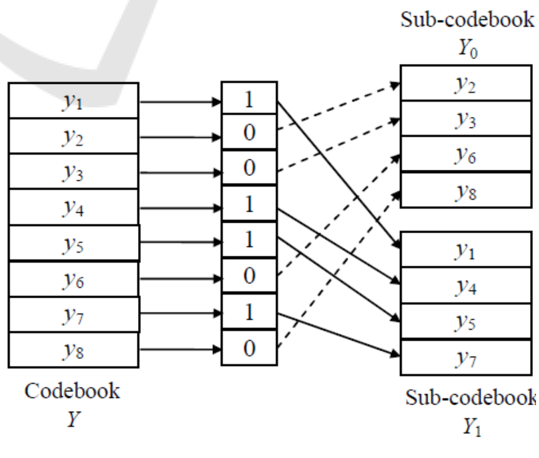

Figure 1: SBS codebook (L = 8) partitioning in the case of one bit embedded per VQ index.

3.1 Pretreatment Phase

The pretreatment phase consists in partitioning the k-

dimensional VQ codebook Y into two disjoint sub-codebooks Y_0 and Y_1 according to a secret key $K = \{k_1, \ldots, k_L\}$, $k_i \in \{0, 1\}$ ($i = 1,\ldots, L$). This division is carried out implicitly where each codevector y_i of Y is assigned to sub-codebook Y_{k_i}.

To ensure that the partition has a minimum effect on the imperceptibility, an object function must be minimized. It is about the total distortion caused by the embedding process, formulated in our work as:

$$D_T = \sum_{i=1}^{L} D_{\min}(y_i, Y_0) + D_{\min}(y_i, Y_1) \qquad (1)$$

where $D_{\min}(y_i, Y_m)$ is the minimal squared Euclidean distance between the i^{th} codevector of Y and all the vectors of a sub-codebook Y_m ($m = 0$ or 1).

3.1.1 Balanced Codebook Partitionning

According to a stego-key $K = \{k_1,\ldots, k_L\}$, the codebook Y of L codevectors is split into two sub-codebooks Y_0 and Y_1 containing the same number of codevectors. In this method, named Balanced Codebook Partitioning (BCP), the key K must contain the same number of "0" and "1" which is $L/2$. Each codevector y_i ($1 \le i \le L$) of Y will then be assigned to sub-codebook Y_{ki}. The basic steps of our BCP of SBS system are given below.

***<u>Input data</u>*:**

- Database of cover vectors $X = \{x_1,\ldots, x_N\}$.
- Sequence of secret bits $M = \{m_1, m_2,\ldots\}$.

<u>Step 1:</u>

- Generate randomly S keys $\{K_1,\ldots, K_S\}$ such as the numbers of "0" and "1" are identical in each key.

<u>Step 2:</u>

- Embed the given secret bits $M = \{m_1, m_2,\ldots\}$ into VQ indices of the nearest codevectors of cover vectors X by using respectively the S stego-keys.

<u>Step 3:</u>

- Evaluate the performance of each key K_i ($i = 1,\ldots, S$) according to the embedding total distortion D_T.

<u>Step 4:</u>

- Select and save the best stego-key which generated the smallest total distortion.

Notice that the embedding procedure used in Step 2 is given below in sub-section 3.2.

The BCP method was developed to ensure a minimum overall distortion, however it does not ensure minimal degradation for each codevector of the codebook Y. To add this characteristic, we proposed a codebook partitioning method which ensures a minimum overall distortion while ensuring minimal degradation for each codevector of Y. This method was named Unbalanced Codebook Partitioning (UCP).

3.1.2 Unbalanced Codebook Partitionning

The UCP method splits the codebook Y into two sub-codebooks which does not have the same number of codevectors. The basic idea is to find a minimal degradation for each codevector of the codebook Y. Thus, for each index i ($1 \le i \le L$), we must find the index j ($1 \le j \le L$, $i \ne j$) such as the distance between the pair of codevectors y_i and y_j ($d(y_i, y_j)$) is minimal. The two codevectors of indices i and j will then be assigned to two different sub-codebooks. Each codevector must belong to only one sub-codebook. These two sub-codebooks will permit to generate the stego-key K which will be used in the embedding and extraction procedures. The steps of our UCP approach are as follow.

***<u>Input data</u>*:**

- VQ codebook of L codevectors.

<u>Step 1:</u>

- Find all pairs of indices i, j ($1 \le i, j \le L$, $i \ne j$) such as the distances $d(y_i, y_j)$ are minimal.

<u>Step 2:</u>

- For each index pair (i, j), put the index i in a group ("0") and the index j in the other group ("1").

- Label each index by its group number:
  Label(i) = 0 and Label(j) = 1.

<u>Step 3:</u>

- Generate the stego-key $K = \{k_1,\ldots, k_L\}$ by using the labels of the L indices : k_i = Label(i), for $i = 1,\ldots, L$.

3.2 Embedding Phase

Similar to (Kim, 2002) and (Wang, 2007), we use the codevectors indices in Y_0 or Y_1 to hide bit "0" or bit "1", respectively. The embedding procedure steps are given below.

***<u>Input data</u>*:**

- VQ codebook of L codevectors.
- Secret stego-key $K = \{k_1, k_2, \ldots, k_L\}$.
- Database of cover vectors $X = \{x_1,\ldots, x_N\}$.
- Sequence of secret bits $M = \{m_1, m_2,\ldots\}$.

Step 1:

- For the vector x_i and the secret bit m_i, find the nearest codevector y for x_i from Y with the condition $k_i = m_i$ (i.e., use Y_0 if $m_i = 0$ or Y_1 if $m_i = 1$).

- Send the stego-index of the selected codevector y to the reception side.

Step 2:

- Repeat Step 1 until all the secret bits have been treated.

3.3 Extraction Phase

To extract the bit hidden in a received stego-index i, one have only to know the value of k_i in the key K. The extraction procedure steps are given below.

***Input data*:**

- Secret stego-key $K = \{k_1, k_2, \ldots, k_L\}$.

- Quantization stego-indices : $I = \{i_1, i_2, \ldots\}$.

Step 1:

- For the j^{th} index i_j of the obtained carrier codevector y_{i_j}, determine the hidden bit: $m_j = k_{i_j}$

Step 2:

- Repeat Step 1 until all the hidden bits $M = \{m_1, m_2, \ldots\}$ have been extracted.

4 SPEECH STEGANOGRAPHIC SYSTEMS: APPLICATION OF THE BCP AND UCP METHODS

In this section, we evaluate the performance of our speech steganographic systems, designed based on SBS by codebook partitioning and called "SBS-CP". These systems were developed separately by the BCP and UCP methods presented above.

In our applications, the main purpose is to hide a secret speech signal coded by the 2.4 kbps MELP into a host public speech coded by the AMR-WB Rec. G.722.2 (Bessette et al., 2002). The embedding is done during the S-MSVQ quantization of the G.722.2 ISF parameters (ISFs) by using a secret stego-key K. Notice that in all simulations, we used the G.722.2 in mode 12.65 kbits/s where the ISFs are coded by an S-MSVQ of 46 bits/frame.

Recall, that the G.722.2 ISF parameters are quantized by a split multistage vector quantizer (S-MSVQ) with 1st order MA predictor. The G.722.2 S-MSVQ uses 7 codebooks, where 2 codebooks at the first stage (named here CB_{11} and CB_{12}) and 5 codebooks (named CB_{21}, CB_{22}, CB_{23}, CB_{24}, CB_{25}) at the second stage.

4.1 Performance Evaluation Criteria

Performance evaluation of the implemented speech steganographic systems will be done according to two criteria: the hiding capacity represented by the embedding rate of the secret speech and the transparency (imperceptibility) represented by the perceptual quality of the speech stego-signal synthesized by the G.722.2 with embedding procedure.

The total embedding rate R is given by the ratio of the number of hidden secret bits and the length of the host speech coder frame:

$$R = n \times \frac{1}{20 \cdot 10^{-3}} = n \times 50 \ \ bits/s \qquad (2)$$

where n is the total number of ISF S-MSVQ quantization stego-indices used in all the embedding process. Note that the highest embedding rate, which can be obtained by this SBS-CP method, is reached when we hide one bit in each of the 7 S-MSVQ codebook quantization indices, i.e., 7 bits hidden per frame. The maximum total embedding rate would then be: $R = n \times 350$ bits/s.

On the other hand, for imperceptibility, we use the ITU-T Rec. P.862.2 known under the abbreviation WB-PESQ (Wide Band extension of Perceptual Evaluation Speech Quality) (ITU-T Rec. P.862.2., 2005) to evaluate the coded cover/stego speech signals quality. The hidden speech signal is imperceptible if a listener is unable to distinguish between the cover and the stego speech signals; which means that the WB-PESQ difference between the two cover/stego signals is negligible.

The performance of the steganographic S-MSVQ quantizer will be also evaluated by the well-know average spectral distortion (SD) measure. The spectral distortion of each frame i is given, in decibels, by (Paliwal and Atal, 1993), (Cheraitia and Bouzid, 2014):

$$SD_i = \sqrt{\frac{1}{n_1 - n_0} \sum_{n=n_0}^{n_1-1} \left[10 \log_{10} \frac{S(e^{j2\pi n/N})}{\hat{S}(e^{j2\pi n/N})} \right]^2}, \qquad (3)$$

where $S(e^{j2\pi n/N})$ and $\hat{S}(e^{j2\pi n/N})$ are respectively the original and quantized power spectra of the LPC synthesis filter, associated with the i^{th} frame of speech signal.

Generally, we can get transparent quantization quality if we maintain the three following conditions (Paliwal and Atal, 1993): 1)- The average spectral distortion (SD) is about 1 dB, 2)- No Outliers frames with SD greater than 4 dB, 3)- The percentage of Outlier frames having SD within the range of 2-4 dB must be less than 2%.

4.2 Performances of Steganographic SBS-CP Systems Implemented in G.722.2 S-MSVQ Quantizer

For a given embedding rate, we performed an optimization procedure of our steganographic systems which consists in finding the best choice of S-MSVQ codebooks that can be used in the hiding process to obtain the best possible performance.

The speech database used in the following experiments consists of 60 minutes of speech taken from the international TIMIT database (f_s = 16 kHz) (Garofolo et al., 1988). To construct the ISF database, we used the same LPC analysis function of the G.722.2, where a 16-order LPC analysis, based on the autocorrelation method, is performed every analysis frame of 20 ms. Thus, a database of 180000 ISF vectors was constructed.

For embedding rates varying between 1 and 7 bits/frame, the SD performances of speech steganographic SBS-CP systems implemented in G.722.2 S-MSVQ are shown in Table 1. The SBS-CP systems were designed respectively by our BCP and UCP methods. For comparative evaluation, the performance of a conventional steganographic system designed by a random codebook partitioning (RCP) method was also included in Table 1. For a secret bits sequence, the wideband ISF vectors of dimension 16 are quantized by the same G.722.2 S-MSVQ quantizer of seven codebooks, denoted in the Table as follows (CB_{11}, CB_{12}, CB_{21},…, CB_{25}). For example, the notation "**2** (0-0-0-**1**-**1**-0-0)" means that for an embedding rate of 2 bits/frame, the codebooks CB_{22} and CB_{23} of the S-MSVQ 2nd stage were selected as best choice to be used in hiding 2 bits per each frame.

These comparative results show that the performances of steganographic SBS-CP systems designed by the UCP method are slightly better than those designed by the BCP. On the other hand, the systems designed by UCP and BCP methods outperform the systems designed by the traditional RCP. Indeed, they can both achieve the transparent quantization quality until an embedding rate of 3 bits/frame.

4.3 Performance Evaluation of G.722.2 with SBS-CP Systems Implemented in S-MSVQ Quantizer

The cover public speech database used in the following evaluations is composed of 10 speech sequences of 32s extracted from the same TIMIT database. The secret bit stream was generated by the 2.4 kbps MELP from a speech sequence of f_s = 8 kHz extracted from a phonetically balanced Arabic speech database (Boudraa et al., 1992).

Table 2 presents WB-PESQ performance comparative evaluation of the global G.722.2 where its ISF parameters were quantized by the 46 bits/frame steganographic S-MSVQ with SBS-CP systems designed respectively by BCP and UCP. Notice that an embedding rate of 0 bits/frame means the original standard G.722.2 without steganography (i.e assessment of the cover speech signal).

Table 1: Performance of steganographic SBS-CP systems designed respectively by UCP, BCP and RCP methods.

Embedding rate (Bits/frame)	SBS-CP systems by UCP			SBS-CP systems by BCP			SBS-CP systems by RCP		
	Av. SD (dB)	Outliers (in %)		Av. SD (dB)	Outliers (in %)		Av. SD (dB)	Outliers (in %)	
		2 - 4 dB	> 4 dB		2 – 4 dB	> 4 dB		2 – 4 dB	> 4 dB
1 (0-0-0-**1**-0-0-0)	0.95	1.13	0.0005	0.95	1.11	0.001	0.96	1.32	0.001
2 (0-0-0-**1**-**1**-0-0)	0.99	1.45	0.002	1.00	1.42	0.001	1.03	1.93	0.005
3 (0-0-0-**1**-**1**-**1**-0)	1.03	1.80	0.005	1.05	1.83	0.007	1.09	2.66	0.010
4 (0-0-**1**-**1**-**1**-**1**-0)	1.09	2.48	0.008	1.11	2.55	0.007	1.17	4.43	0.005
5 (0-0-**1**-**1**-**1**-**1**-**1**)	1.15	3.19	0.009	1.17	3.35	0.007	1.24	5.84	0.015
6 (0-**1**-**1**-**1**-**1**-**1**-**1**)	1.23	5.57	0.017	1.25	6.14	0.027	1.37	11.03	0.120
7 (**1**-**1**-**1**-**1**-**1**-**1**-**1**)	1.30	7.61	0.035	1.33	8.85	0.041	1.51	17.46	0.465

Table 2: Performance of the global G.722.2 with steganographic SBS-CP implementation.

Embedding rate (Bits/frame)	G.722.2 with SBS-CP by BCP	G.722.2 with SBS-CP by UCP
	WB-PESQ	WB-PESQ
0	**3.790**	**3.790**
1	3.798	3.705
2	3.823	3.744
3	3.687	3.814
4	3.719	3.680
5	3.756	3.747
6	3.766	3.720
7	3.676	3.720

For all embedding rates, these simulation results show that the overall quality of stego-speech is almost identical to quality of cover public speech; which means that our proposed steganographic techniques are practically imperceptibles. Most WB-PESQ scores of the stego-signals are between 3.67 and 3.82. Hence, a good speech quality was obtained and no degradation was caused by the embedding process. On the other hand, steganographic SBS-CP systems designed by the UCP yields slight improvement to the G.722.2 WB-PESQ performance compared to SBS-CP with balanced partitioning.

5 CONCLUSION

In this paper, we proposed two variants of VQ-based speech steganography binning schemes for G.722.2 secure speech communication system. The simulation results showed that the two steganographic SBS-CP methods by UCP and BCP can generate stego-speech signals with similar quality to cover speech signals; which means that the resulting stego-speech is indistinguishable from the original cover speech. Hence, the two proposed variants of SBS-CP method can ensure a high transparency with a maximal embedding rate of 7 bits/frame (350 bits/s).

Robustness against intentional and non-intentional attacks has not been investigated in this work; it will be studied in future research.

REFERENCES

Bessette, B., Salami, R., Lefebvre, R., Jelínek, M., Rotola-Pukkila, J., Vainio, J., Mikkola, H., Järvinen, K., 2002. The adaptive multirate wideband speech codec (AMR-WB), *IEEE Transactions on Speech and Audio Processing*, vol. 10, no. 8, pp. 620-636.

Boudraa, M., Boudraa, B., Guerin, B., 1992. Mise en place de phrases arabes phonetiquement equilibrées. In *JEP'92, XIXèmes Journées d'Etude sur la Parole*. Bruxelles.

Cheraitia, S., Bouzid, M., 2014. Robust coding of wideband speech immittance spectral frequencies. *Speech Communication, Elsevier*, vol. 65, pp. 94-108.

Cox, I. J., Miller, M. L., Bloom, J. A., Fridrich, J., Kalker, T., 2008. *Digital Watermarking and Steganography*, Second Edition, Morgan Kaufmann Publishers, USA.

Djebbar, F., Ayad, B., Meraim, K. A., Hamam, H., 2012. Comparative study of digital audio steganography techniques, *EURASIP Journal on Audio, Speech, and Music Processing*, Springer, vol. 25, pp. 1-16.

Garofolo J. S., et al., *DARPA TIMIT Acoustic-phonetic Continuous Speech Database.* National Institute of Standards and Technology (NIST), Gaithersburg, October 1988.

Geiser, B., Vary, P., 2008. High rate data hiding in ACELP speech codecs. In *ICASSP'2008, IEEE International Conference on Acoustics, Speech and Signal Processing*. pp. 4005-4008. USA.

Gersho, A., Gray, R. M., 1992. *Vector quantization and Signal compression*, Kluwer Acad. Publishers, USA.

ITU-T Recommendation G.722.2. *Wideband coding of speech at around 16 kb/s using Adaptive Multi-rate Wideband (AMR-WB)*, 2003.

ITU-T Recommendation P.862.2. *Wideband Extension to Recommendation P.862 for the Assessment of Wideband Telephone Networks and Speech Codecs*, Geneva, 2005.

Kim, Jo. M., 2002. A digital image watermarking scheme based on vector quantisation. *IEICE Trans. on Inf. and Systems*, vol. E85-D, pp. 1054-1056.

McCree, A., Truong, K., George, E. B., Barnwell, T. P., Viswanathan, V., 1996. A 2.4 kbits/s MELP Coder Candidate for the New U.S. Federal Standard. In *ICASSP'96, IEEE International Conference on Acoustics, Speech and Signal Processing*. pp. 200-203.

Moulin, P., Koetter, R., 2005. Data-Hiding Codes. *Proceedings of The IEEE*, Vol. 93, pp. 2083-2126.

Paliwal, K. K., Atal, B. S., 1993. Efficient vector quantization of LPC parameters at 24 bits/frame. *IEEE Transactions on Speech and Audio Processing*, vol. 1, no. 1, pp. 3-14.

Wang, F. H., Jain, L. C., Pan, J. S., 2007. A novel VQ-based watermarking scheme with genetic codebook partition. *Journal of Network and Computation Applications (JNCA)*, Elsevier, vol. 30, no. 1, pp. 4-23.

Yargıcoglu, A. U., Ilk, H. G., 2010. Hidden data transmission in mixed excitation linear prediction coded speech using quantisation index modulation, *IET Information Security*, vol. 4, Issue 3, pp. 158–166.