# Quality-Aware GSM Speech Watermarking

Koh Jun-Li Christabel, Sabu Emmanuel, *Mohan S. Kankanhalli

*School of Computer Engineering, Nanyang Technological University, Singapore 639798*

**School of Computing, National University of Singapore, Singapore 117590*

kohj0006,asemmanuel@ntu.edu.sg, mohan@comp.nus.edu.sg

***Abstract*—Use of watermarking techniques to provide authentication and tamper proofing of speech in mobile environment is becoming important. However, the current efforts do not allow for user-specifiable quality for the watermarked speech. This paper proposes a watermarking algorithm that allows user-customizable quality for watermarked GSM (Global System for Mobile) speech. Sensitivity (in terms of quality degradation) of each GSM coefficient bits against bit watermark embedding was investigated first, which is then used to select the coefficient bits in a secure manner for watermarking. The proposed algorithm's execution time requirement was studied to draw conclusions on the real-time usability of the algorithm. The embedding capacity and the quality awareness of the algorithm were also investigated.**

## I. INTRODUCTION

With the introduction of wireless networks and mobile commerce, mobile phones are not only used in speech conversation but also data streaming. It is often necessary to verify the authenticity of the speaker or digital contents that are output from the mobile phones especially in phone banking and stock market trading scenarios. Speech watermarking can be used in mobile environment for proving the authenticity of the speaker and integrity of the speech or other data streams. However, as more watermark information is embedded, the speech quality will tend to degrade. And the users often want to specify the quality of the watermarked speech to be above a certain quality, so that the received watermarked speech is intelligible and easy to understand. Since the watermarking has to be performed in GSM compressed domain where every frame consisting of one hundred sixty 13-bit speech samples is converted to a fixed 260 GSM coefficient bits, maintaining watermarked speech quality above certain specified quality value after GSM decoding is an interesting and challenging problem. This paper proposes a quality-aware speech watermarking technique that is able to preserve the intelligibility of the speech and meet the targeted watermark information to be embedded per frame.

Several techniques [1-8] have been proposed in the field of speech watermarking. Celik et.al. [1] proposes a speech watermarking technique based on altering the pitch and duration of the periodic speech segments. The technique makes use of quantization index modulation (QIM), and pitch synchronous overlap and add (PSOLA) method to carry out the modification of pitch values. This watermarking technique is shown to be robust against low data rate codecs such as GSM 06-10, Adaptive Multi-Rate coder (AMR) etc. However, the method only allows a relatively low embedding capacity (approx. 3 bits/sec). Wu et.al. propose two fragile speech watermarking techniques—exponential-scale odd/even modulation technique [2][3] and linear additive watermarking [3] for content integrity verification. Exponential-scale odd/even modulation is better in detecting localized content alteration, but its limitations include inability to tolerate low bit-rate speech coders such as the code excited linear predictive (CELP) coders and the initial latency needed to produce detection results. The latter algorithm overcomes the limitations, but it may not detect malicious alteration by the hacker. Arora et.al. [4] proposes a signal adaptive watermarking technique for GSM speech using spread spectrum technique. Field programmable gate array (FPGA) implementation of the algorithm was carried out to study the hardware and time complexity. Gurijala et.al. [5] proposes to perform watermarking by perturbing the linear prediction (LP) coefficients of the speech. Set membership filtering (SMF) technique is used to obtain sets of allowable parameter perturbations. In [6], the authors apply constrained optimization technique to obtain the best perturbations that are most robust against filtering and quantization attacks.

In our proposed scheme, we allow the user to specify the watermarked speech quality which helps each user to customize the watermarking level to their needs. The proposed algorithm also allows the user to specify the number of bits (to the maximum embedding capacity level or any level) to be embedded per frame. The number of bits that can be embedded per frame depends on the watermarked speech quality specified by the user. If the specified number of bits to watermark per frame exceeds the maximum embedding capacity for the specified watermarked speech quality, the algorithm declares that the watermarking cannot be performed. Watermarking is carried out on the full rate GSM 06.10 coefficient bits. The proposed scheme randomizes the selection of GSM coefficient bit positions (keeping in mind the specified watermark quality and the number of bits for embedding) to watermark and use the bit position index as the embedding and detection keys.

The rest of the paper is organized as follows. In Section II, GSM 06.10 RPE-LTP Codec, in Section III, proposed quality aware watermarking technique, in Section IV, results and discussion and in Section V, conclusion.

## II. GSM 06.10 RPE-LTP CODEC

This codec operates on a 20ms frame composed of 160, 13-bit speech samples and outputs 260 GSM coefficient bits per frame. Each frame consists of 8 log area ratio (LAR) parameters totalling 36bits. The rest of the frame is divided into 4 sub-frames. Each sub-frame consists of 1 long term prediction (LTP) lag parameter (7 bits), 1 LTP gain parameter (2 bits), 1 regular pulse excitation (RPE) grid position parameter (2 bits), 1 block amplitude parameter (6 bits) and 13 RPE pulses (3x13=39 bits) totalling 56bits. Thus it adds up to a total of 36+4x56=260 bits per frame. We first investigate the sensitivity of each GSM coefficient bits to watermark bit embedding. The sensitivity information is then used by the GSM bit selection algorithm for embedding the watermark.

### *A. Sensitivity of GSM Coefficients to Bit Embedding*

In order to quantitatively measure the impact/sensitivity of each GSM coefficient bits to bit watermark embedding, we choose peak signal to noise ratio (PSNR) value computed using the speech and watermarked speech samples as a measure. Each GSM coefficient bit is individually watermarked using the bit watermarking technique explained in Section III. And then the watermarked GSM coefficient frame is decoded to measure the PSNR value in dB compared to the non watermarked version. GSM coefficient bits are watermarked one at a time to investigate the sensitivity of watermarking the GSM coefficient bit. The sensitivity in decibels (dB) per frame can be computed using the equation (1):

$$\text{Sensitivity} = \text{PSNR} = 20\log_{10}\left(\frac{8191}{(\boldsymbol{S}-\boldsymbol{S}^{w})/160}\right) \qquad (1)$$

where $\boldsymbol{S}$ is the non-watermarked speech frame and $\boldsymbol{S}^{w}$ is the watermarked speech frame at the decoder output. For 13-bit speech samples, we use the value 8191 in the equations. The number of 13-bit speech samples per frame is 160. Table I shows the sensitivity of each GSM coefficient bits to the bit embedding. The sensitivity shown is the average over several frames from different speech files and quantized within a 1dB step size. The lower level of the quantization step is taken as the sensitivity level for the group. This allows many bits to be grouped into one group so random group and bit selection algorithms can be designed to improve the security and efficiency. Since there are only distinct 56 bits out of 260 bits, (other 204 bits being replications of the LTP lag and gain parameters, RPE grid positions, block amplitude and RPE pulse bits in different/same sub frames), the sensitivity need only be computed for the 56 distinct bits.

The "*gpWmValue*" in Table I is computed using the equation (2):

$$gpWmValue = 160 \times \left(8191/10^{PSNR/20}\right)^2 \qquad (2)$$

Table I shows the sensitivity of the coefficient bits which is used in the proposed watermarking technique in the next section.

TABLE I
SORTED WATERMARK BIT SIGNIFICANCE AT DIFFERENT POSITIONS

| *Group Number* | **Sensitivity** | *gpWmValue* | **Total Bits** | **Bit Position & No. of Bits Available** |
|---|---|---|---|---|
| 0 | 87dB | 21.41873583 | 114 | Xmaxc-b1(4); xmc-b1(52); |
| | | | | Xmaxc-b2(4); LARc5-b1(1); |
| | | | | xmc-b2(52); LARc-7b1(1); |
| 1 | 86dB | 26.96459083 | 61 | Xmaxc-b3(4); xmc-b3(52); |
| | | | | Mc-b1(4); LARc6-b1(1); |
| 2 | 85dB | 33.94640861 | 10 | LARc3-b1(1); bc-b1(4); |
| | | | | Mc-b2(4); LARc1-b1(1); |
| 3 | 84dB | 42.73599644 | 9 | LARc4-b1(1); Nc-b1(4); |
| | | | | Xmaxc-b4(4); |
| 4 | 83dB | 53.80143192 | 6 | LARc7-b2(1); LARc2-b1(1); |
| | | | | bc-b2(4); |
| 5 | 82dB | 67.73198983 | 14 | LARc5-b2(1); Nc-b5(4); |
| | | | | Nc-b2(4); LARc0-b1(1); |
| | | | | Nc-b6(4); |
| 6 | 81dB | 85.26952319 | 14 | Nc-b7(4); Nc-b3(4); |
| | | | | Nc-b4(4); LARc6-b2(1); |
| | | | | LARc3-b2(1); |
| 7 | 80dB | 107.3479696 | 2 | LARc1-b2(1); LARc4-b2(1); |
| 8 | 79dB | 135.1430868 | 3 | LARc2-b2(1); LARc5-b3(1); |
| | | | | (msb)LARc7-b3(1); |
| 9 | 78dB | 170.1350662 | 11 | LARc3-b3(1); LARc0-b2(1); |
| | | | | LARc1-b3(1); Xmaxc-b5(4); |
| | | | | LARc2-b3(1); (msb)LARc6-b3(1); |
| | | | | (msb)LARc3-b5(1); (msb)LARc5-b4(1); |
| 10 | 77dB | 214.1873583 | 3 | LARc4-b3(1); LARc3-b4(1); |
| | | | | LARc0-b3(1); |
| 11 | 76dB | 269.6459083 | 2 | (msb)LARc4-b4(1); LARc0-b5(1); |
| 12 | 75dB | 339.4640861 | 2 | LARc1-b4(1); LARc2-b4(1); |
| 13 | 72dB | 677.3198983 | 1 | (msb)LARc2-b5(1); |
| 14 | 71dB | 852.6952319 | 1 | LARc0-b4(1); |
| 15 | 64dB | 4273.599644 | 1 | LARc1-b5(1); |
| 16 | 62dB | 6773.198983 | 4 | Xmaxc-b6(4); |
| 17 | 55dB | 33946.40861 | 1 | (msb)LARc1-b6(1); |
| 18 | 50dB | 107347.9696 | 1 | (msb)LARc0-b6(1); |

Note: LARc refers to LAR parameter; Nc refers to LTP lag; bc refers to LTP gain; Mc refers to RPE grid position; Xmaxc refers to Block Amplitude; xmc refers to RPE pulses. Bit position representation is of the format "bit position name (number of bits)"

## III. PROPOSED QUALITY AWARE WATERMARKING TECHNIQUE

In this section we propose a quality aware watermarking technique which works as follows (shown in Figure.1):

1. User specifies the required speech quality '*Q*' in dB and the number of watermark bits to be embedded per frame '*M*'.

2. Given '*N*' total number of watermark bits to be embedded; we split *N* watermark bits into several segments of *M* bits for embedding per frame. The *M* chosen should be such that $N/M \notin Z$. This condition makes the last segment to have only less than *M* bits to be embedded and is necessary to defeat the attack explained in Section IV.D.
3. The proposed algorithm will compute *userWm* (watermark value of '*Q*') using the expression.

$$userWm = 160 \times \left(8191/10^{Q/20}\right)^2 \tag{3}$$

4. The algorithm then randomly picks *M* group positions, $randGP_i \in \{Group\ Number\}$, $i = 0...M-1$, from Table I, in such a way that the *gpWmValue* of the respective groups add up to *userWm*. Each round of random picking must satisfy two conditions. Firstly,

$$\begin{aligned} resWm_0 &= userWm \\ resWm_i &= resWm_{i-1} - gpWmValue_i, \quad i = 1...M-1 \end{aligned} \tag{4}$$

the $resWm_i$ for every round should be dividable by the smallest *gpWmValue* in Table I to ensure that there are sufficient rounds for embedding *M* watermark bits. Secondly, we have to ensure that the number of times the same *Group Number* being picked does not exceed the number of bits in the quantized group (*bitLimit*). Repeat the process until *M* watermark bits are fully allocated.
5. The algorithm allows an error range of $Q + \delta$, where $\delta$ is the 1dB step size.
6. Now we know how many watermark bits are allocated for each quantization groups and the algorithm will then randomly pick the actual bit positions from the groups for watermark embedding. These bit positions form the key, ***K*** for embedding and detection.
7. The embedding process explained in Section III.A will embed the watermark bits into the GSM coefficient bits at the positions specified by ***K*** only and the other positions are not embedded. The watermarked speech frame and ***K*** are sent to the receiver. The ***K*** is sent through a secure channel.
8. When the receiver receives ***K*** and the watermark speech frame, it detects the watermark bits according to the bit positions specified by ***K***. The detection algorithm is explained in Section III.A.

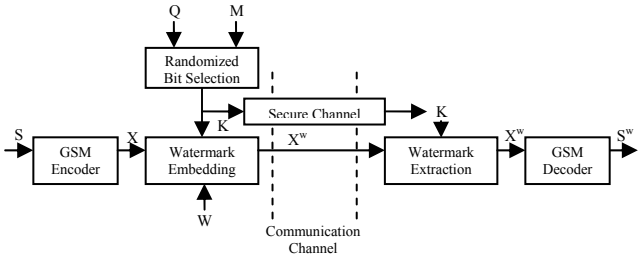



Figure 1. Watermarking embedding and extraction

The embedding process used in Section III.A is simple however any other bit embedding algorithm could be used in place. The '*N*' number of watermark bits to be embedded need to be split into several segments of *M* bits each and then embed the *M*-bit segments into different frames. The proposed algorithm will meet the user's speech quality by satisfying the two conditions for each round and limiting the error range to $\delta$. The key has to be transmitted through secure channel for security purposes. Figure.2 shows the pseudo code of the algorithm.

```
PICKING OF BITS:
Begin with round 1
WHILE the rounds don't sum up to M
    Generate a randGP_i from gpWmValue
    Compute resWm=userWm-randGP_i
    IF resWm/smallest gpWmValue >= M-round OR randGP_i > bitLimit
            IF (M-round!=0) OR (M-round==0 AND 0≤resWm≤ δ )
            resWm=userWm
            round++;
            ENDIF
    ENDIF
ENDWHILE
```

Figure 2. Pseudo Code

This algorithm provides randomness in two areas: allocation of bits to quantization groups and selection of actual bit positions in the group which leads to better security.

### *A. Watermark Embedding and Detection Processes*

Let ***X*** be the GSM coefficient bit vector of speech, $\boldsymbol{X}_r$ be the $r^{th}$ GSM coefficient bit frame of ***X***. Let ***K*** be a 260-bit, bit vector denoting the key, indicating the bit positions to watermark. A 260-bit $\boldsymbol{W}_r$ vector is made from the M-bit watermark vector for the $r^{th}$ frame by placing the watermark bit in the bit positions for embedding. The watermarked $r^{th}$ GSM coefficient bit frame $\boldsymbol{X}_r^w$ can be obtained by the equation (5).

$$\boldsymbol{X}_r^w(l) = \begin{cases} X_r(l) & \text{for } \boldsymbol{K}(l) = 0, \quad 0 \le l < 260 \\ 1 & \text{for } \boldsymbol{K}(l) = 1 \,\&\, \boldsymbol{W}_r(l) = 0, \quad 0 \le l < 260 \\ 0 & \text{for } \boldsymbol{K}(l) = 1 \,\&\, \boldsymbol{W}_r(l) = 1, \quad 0 \le l < 260 \end{cases} \tag{5}$$

The watermarked GSM coefficient bit vector of speech $\boldsymbol{X}^{\mathrm{w}}$ can be obtained by concatenating the $\boldsymbol{X}_r^w$ for all *r*.

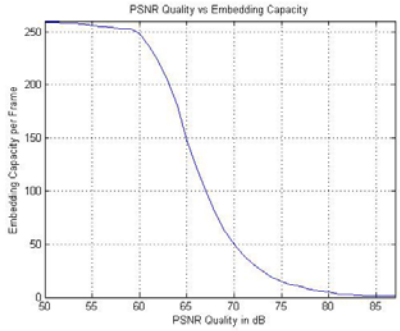
The watermark detection process uses the equation (6). The watermark bit detected from the $l^{th}$ bit position is given by.

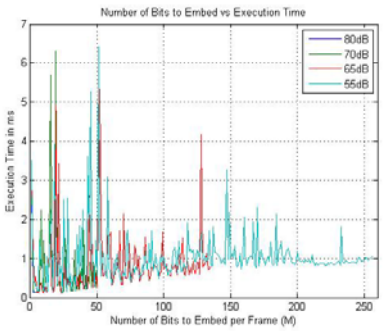
$$\text{Watermark bit} = \begin{cases} 0 & \text{if } \boldsymbol{K}(l) = 1 \,\&\, \boldsymbol{X}_r^w(l) = 1 \\ 1 & \text{if } \boldsymbol{K}(l) = 1 \,\&\, \boldsymbol{X}_{\mathrm{r}}^{\mathrm{w}}(l) = 0 \end{cases} \tag{6}$$

These embedding and detection processes use low complexity bit manipulation functions, thus reduces the computation time for embedding and extraction of watermark bits.

## IV. RESULTS AND DISCUSSION

Several speech files were used to measure the embedding capacity for various PSNR quality, execution time, quality-awareness of the proposed algorithm. We used Intel Pentium® 4 CPU 3.20GHz processor, 1GB RAM of memory and Microsoft Windows XP Professional for the experiments.



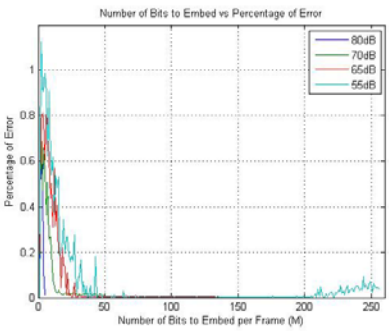




Figure 3. (a) PSNR Quality vs Embedding Capacity (b) Execution Time (c) Quality-Awareness

### A. *PSNR Quality vs Embedding Capacity*

Figure.3(a) shows the plot of PSNR in dB versus embedding capacity. And it is noted that the higher PSNR has lower embedding capacity. The increase in the number of bits embedded will thus lead to a lower quality of speech. However, the focus of our algorithm is on quality awareness. Therefore, the degradation of speech quality is not of importance as long as we fulfil number of bits to be embedded without exceeding the embedding capacity at a PSNR level requested by the user. Figure.3(a) provides information about the maximum embedding capacity for a given PSNR level.

### B. *Execution Time*

Figure.3(b) illustrates the graph of execution time versus number of bits embedded. The execution time for the watermarking algorithm is collected using the native Java System function, nanoTime(). The spikes in the graph shows the specific number of bits for embedding requires more time than other bits to get the random set. Additionally, the watermarking algorithm is capable of embedding bits with less than 7ms which is acceptable in real-time environment as the time does not exceed single speech frame duration of 20ms.

### C. *Quality-Awareness*

Figure.3(c) shows the plot of percentage error versus number of bits embedded. The percentage error is computed using the equation (7):

$$Percentage\ \ Error = [(Achieved\ \ PSNR - Q)/Q] * 100 \quad (7)$$

The graph demonstrates that there is higher variation in percentage error for lower number of embedding bits. This is expected as low number of bits indicates smaller pool of values to randomize and it will be much more difficult to attain the exact PSNR value. However as shown in the graph, we are able to stay within 1 dB variation which means that we are being quality aware.

### D. *Discussion*

For watermarking of real-time speech conversation the $N$ number of watermark bits must be concatenated one after the other to make a continuous watermark bit stream for embedding. Consecutive $M$ bits from this continuous watermark bit stream are taken at a time (non overlapping manner) to be embedded in a speech frame. Since $N$ is not a multiple of $M$, there may not be any repetition of watermark bit frames embedded in various frames, which would prevent the attacker from using the periodicity information to find out the watermarked bit positions and thus to obtain the watermark bits using the several speech frames collected.

## V. CONCLUSION

We have proposed a quality aware, user-customizable watermarking scheme for mobile full rate GSM speech 06-10 RPE-LTP codec. Each GSM coefficient bit's sensitivity to bit embedding was investigated first, which was then used for deciding the bit positions to watermark. The experimental results showed that that the algorithm is time efficient and is able to meet the speech quality and number of watermark bits per frame specified by the user. The algorithm is suitable for real-time watermarking of GSM speech.

## REFERENCES

[1] Celik M., Sharma G. and Murat Tekalp A., *Pitch and Duration Modification for Speech Watermarking*, IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), March 2005.

[2] Wu C.P. and Kuo, C.-C.J., *Fragile speech watermarking for content integrity verification*, IEEE International Symposium on Circuits and Systems (ISCAS), Vol. 2, May 2002.

[3] Wu C.P. and Jay Kuo C.C., *Fragile Speech Watermarking based on Exponential Scale Quantization for Tamper Detection,* 2002 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Ma.y 2002.

[4] Arora S. and Emmanuel S., *Real-time adaptive speech watermarking scheme for mobile applications,* Pacific Rim Conference on Multimedia (PCM), Vol. 2, Dec. 2003.

[5] Gurijala A.R. and Deller J.R. Jr., *Speech watermarking with objective fidelity and robustness criteria*, Thirty-Seventh Asilomar Conference on Signals, Systems and Computers (ACSSC), Vol. 2, Nov. 2003.

[6] Gurijala A.R., Deller J.R. Jr. and Joachim D., *Robustness optimization of parametric speech watermarking*, IEEE International Symposium on Circuits and Systems (ISCAS), May 2006.

[7] Qiang Cheng and Sorensen J., *Spread spectrum signalling for speech watermarking*, 2001 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 2001.

[8] Yan B., Zhen-Ming Lu and Sheng-He Sun, *Security of autoregressive speech watermarking model under guessing attack,*, IEEE Transactions on Information Forensics and Security, Sept 2006.