A Lossless Steganography Technique for G.711 Telephony Speech

Naofumi Aoki*
*Graduate School of Information Science and Technology, Hokkaido University, Sapporo 060-0814 Japan
E-mail: aoki@nis-ei.eng.hokudai.ac.jp Tel: +81-11-706-6532

***Abstract*— Steganography may be employed for secretly transmitting side information in order to improve the performance of signal processing such as packet loss concealment and band extension of telephony speech. The previous studies employ LSB replacement technique for embedding steganogram information into speech data. Instead of such a lossy steganography technique, this study proposes a lossless steganography technique for G.711, the most common codec for the telephony speech based on VoIP. The proposed technique in this study exploits the characteristic of G.711 for embedding steganogram information into speech data without degradation. This paper also proposes a semi-lossless steganography technique for increasing the capacity of the proposed technique.**

I. INTRODUCTION

Recently, several applications that employ steganography, an information hiding technique, have been investigated for enhancing the quality of speech communications [1], [2]. For such applications, side information, referred to as steganogram, is embedded into G.711 telephony speech in order to improve the performance of signal processing such as packet loss concealment and band extension.

G.711 is an ITU (International Telecommunication Union) standard codec that encodes either 14 bit or 13 bit speech data into 8 bit speech data at an 8 kHz sampling rate [3]. It is known as the most common codec for the telephony speech based on VoIP (Voice over IP).

The previous studies employ LSB (Least Significant Bit) replacement technique for embedding steganogram information into speech data [1], [2]. This is the simplest steganography technique, in which the LSB of speech data is just replaced with steganogram information.

Since LSB replacement technique changes speech data itself, it may degrade the quality of speech data. In order to avoid such a problem, this study proposes a lossless steganography technique for G.711. The proposed technique may embed steganogram information into speech data without degradation [4].

II. CHARACTERISTICS OF G.711

G.711 consists of μ-law and A-law schemes designated PCMU and PCMA, respectively. PCMU is mainly employed in North America and Japan. PCMA is mainly employed in Europe.

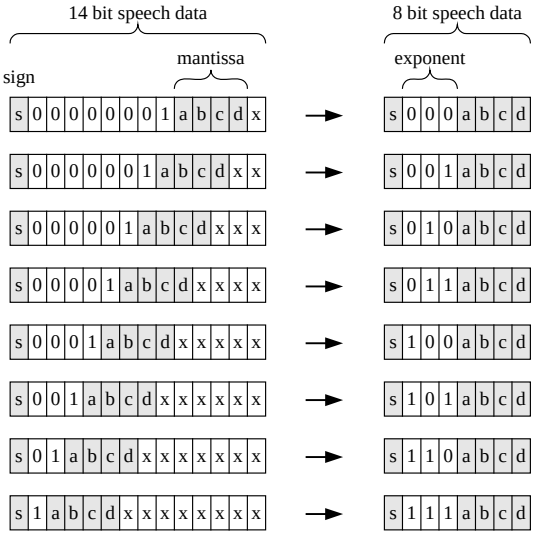

Fig.1 Encoding procedure of PCMU.

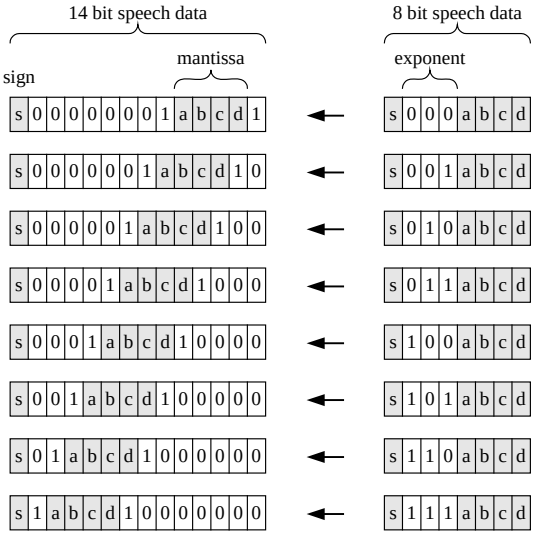

Fig.2 Decoding procedure of PCMU.

This paper only focuses on PCMU. PCMU encodes 14 bit speech data into 8 bit speech data at an 8 kHz sampling rate [3].

Table I 2's complement and folded binary code.

decimal	2's complement	folded binary code
+127	0 1 1 1 1 1 1 1	0 1 1 1 1 1 1 1
~		
+3	0 0 0 0 0 0 1 1	0 0 0 0 0 0 1 1
+2	0 0 0 0 0 0 1 0	0 0 0 0 0 0 1 0
+1	0 0 0 0 0 0 0 1	0 0 0 0 0 0 0 1
+0	0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0
-0	0 0 0 0 0 0 0 0	1 0 0 0 0 0 0 0
-1	1 1 1 1 1 1 1 1	1 0 0 0 0 0 0 1
-2	1 1 1 1 1 1 1 0	1 0 0 0 0 0 1 0
-3	1 1 1 1 1 1 0 1	1 0 0 0 0 0 1 1
~		
-127	1 0 0 0 0 0 0 1	1 1 1 1 1 1 1 1
-128	1 0 0 0 0 0 0 0	

Figure 1 and 2 show the encoding and decoding procedure of PCMU. As shown in these figures, PCMU defines 1 bit sign, 3 bit exponent, and 4 bit mantissa for representing 8 bit speech data.

PCMU employs folded binary code for representing 8 bit speech data [5]. Table I shows how 8 bit speech data is represented by folded binary code.

This table also shows how 8 bit speech data is represented by 2's complement, the general binary code used for representing signed integers.

As shown in this table, 8 bit speech data represented by 2's complement range from -128 to +127. On the other hand, 8 bit speech data represented by folded binary code range from -127 to +127. Although the folded binary code cannot represent -128, it may represent +0 and -0 instead.

III. LOSSLESS STEGANOGRAPHY FOR PCMU

Exploiting the characteristic of PCMU in which speech data whose absolute amplitude is 0 may be represented by two overlap codes, namely +0 and -0, either 0 or 1 steganogram information may be embedded into speech data without degradation.

The proposed technique takes advantage of this redundancy of PCMU. If 0 is required to be embedded into 8 bit speech data, speech data whose absolute amplitude is 0 is set to be +0. On the other hand, if 1 is required to be embedded into 8 bit speech data, speech data whose absolute amplitude is 0 is set to be -0.

The embedding procedure of the proposed technique is defined as follows.

$$c(n) = \begin{cases} +0 & (|c(n)| = 0, b = 0) \\ -0 & (|c(n)| = 0, b = 1) \end{cases} \quad (1)$$

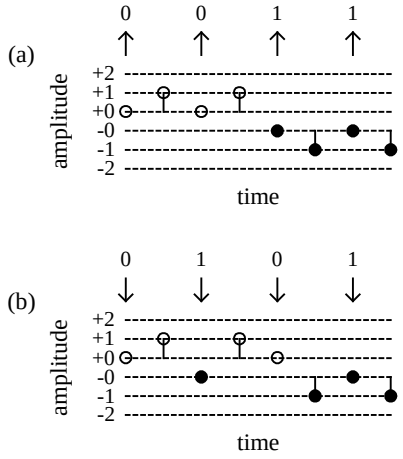

Fig.3 Embedding procedure of the lossless steganography technique: (a) speech data before embedding, (b) speech data after embedding.

where $c(n)$ represents 8 bit speech data, and b represents steganogram information that is to be embedded into the 8 bit speech data.

Figure 3 shows how the proposed technique embeds steganogram information. In this figure, the blank circles are the speech data whose sign bits are 0, and the filled circles are the speech data whose sign bits are 1.

In this example, there are four speech data whose absolute amplitudes are 0. As shown in this figure, these speech data originally contain 4 bit steganogram information represented as (0, 0, 1, 1).

If the steganogram information represented as (0, 1, 0, 1) is required to be embedded into these speech data, the speech data is changed as shown in this figure.

IV. CAPACITY OF THE PROPOSED TECHNIQUE

The capacity of the proposed technique is defined by the number of speech data whose absolute amplitude is 0.

In general, most of speech data show exponential distribution in amplitude [6]. Therefore, it is expected that there is substantial capacity for most of speech data.

This study evaluated the capacity of the proposed technique by using a speech database [7].

Since most of telephony speech generally show half duplex structure due to alternate conversation process, the voice activity ratio of such speech data is statistically at around 50 % [8].

In order to take account of such a characteristic for the evaluation, the speech data were concatenated with silent pauses as shown in Fig.4. According to this manner, 5 male and 5 female speech data were prepared from 10 male and 10 female speech data obtained randomly from the speech database. The total length of the speech data ranged from 15 s to 24 s.

In the evaluation, the background noise of telephony speech was also taken into consideration by mixing white Gaussian noise into the speech data.

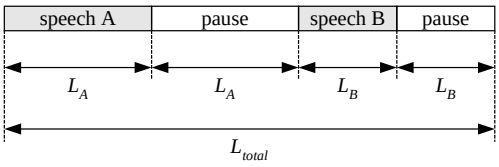

Fig.4 Speech data in the evaluation: L_A and L_B are the length of speech data A and B obtained from speech database. L_{total} is the total length of the speech data.

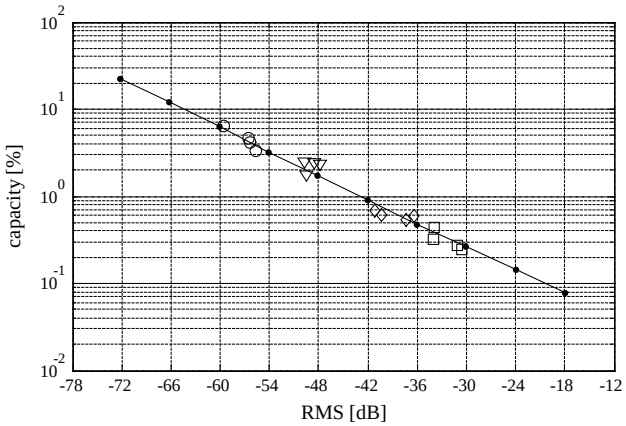

Fig.5 Capacity of the proposed technique evaluated from speech database and actual environment: Solid line represents the capacity obtained from speech database. Circles, triangles, diamonds, and squares represent the capacity obtained from a private room, an office room, a cafeteria, and a railroad station, respectively.

The solid line shown in Fig.5 is the average capacity calculated from all the speech data. It is indicated that the capacity changes according to the power of the background noise. When the power of the background noise becomes small, the capacity tends to be large, and vice versa.

In addition, this study evaluated the capacity of the proposed technique by using telephony speech obtained from actual environment such as a private room, an office room, a cafeteria, and a railroad station. In each condition, 2 male and 2 female speech data were collected. The total length of the speech data ranged from 123 s to 136 s.

The results are shown in Fig.5. The tendency of the capacity is similar to the result obtained from the speech database. The capacity is around 6 % for the telephony speech obtained from a private room. On the other hand, the capacity is around 0.2 % for the telephony speech obtained from a railroad station.

Consequently, it is indicated that the proposed technique may potentially convey steganogram information at around 480 bps when the background noise level is relatively small such as in a private room. On the other hand, the proposed technique may potentially convey steganogram information at around 16 bps when the background noise level is relatively large such as in a railroad station.

V. SEMI-LOSSLESS STEGANOGRAPHY FOR PCMU

In order to increase the capacity of the proposed technique, this study newly proposes a semi-lossless steganography technique.

Figure 6 shows how the proposed technique embeds steganogram information. In the embedding procedure, this technique modifies 8 bit speech data as follows.

$$c'(n) = \begin{cases} c(n)+1 & (c(n) \geq +0) \\ c(n)-1 & (c(n) \leq -0) \end{cases} \quad (2)$$

This amplitude modification may cause undesirable clipping in the 8 bit speech data whose absolute amplitude is 127. Consequently, this technique may recover the original speech data only when the amplitude of the 8 bit speech data ranges from -126 to +126.

Most of speech data meet this condition, since most of speech data show exponential distribution in amplitude [6]. In general, there is only a few speech data whose absolute amplitude is 127.

However, if the amplitude of the original speech data is out of this range, this technique cannot recover the original speech data any more. Therefore, this technique is named semi-lossless steganography technique in this study.

The embedding procedure of the proposed technique is defined as follows.

$$c'(n) = \begin{cases} +1 & (|c'(n)| = 1, b = 0) \\ +0 & (|c'(n)| = 1, b = 1) \\ -0 & (|c'(n)| = 1, b = 2) \\ -1 & (|c'(n)| = 1, b = 3) \end{cases} \quad (3)$$

Since this technique may embed 2 bit steganogram information, the capacity of this technique is twice as large as that of the lossless steganography technique.

Note that the capacity of this technique could be increased when the amplitude modification level is increased. However, undesirable clipping is expected to occur more frequently in such a condition.

After the extraction of the steganogram information, this technique recovers the 8 bit speech data as follows.

$$c(n) = \begin{cases} c'(n)-1 & (c'(n) \geq +1) \\ c'(n)+1 & (c'(n) \leq -1) \end{cases} \quad (4)$$

This procedure is necessary for the recovery of the original speech data. However, this procedure is not very sensible for speech perception when the amplitude modification level is small enough. Even if this procedure is omitted, the speech quality is almost the same as that of the original speech data.

This may potentially guarantee the compatibility of the speech communications. It means that conventional telephony systems that do not implement the proposed technique can normally play the speech data that is modified with the proposed technique.

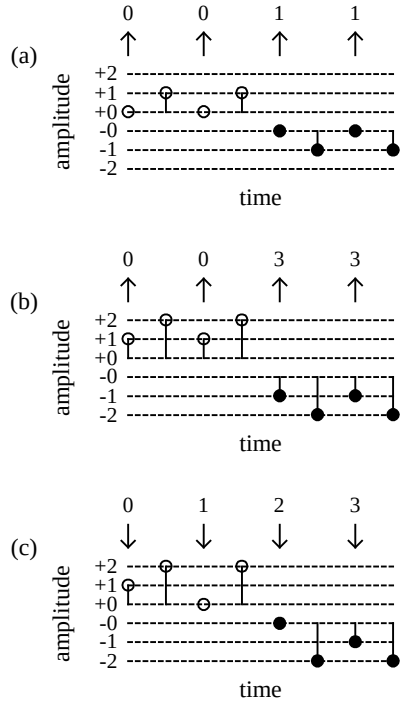

Fig.6 Embedding procedure of the semi-lossless steganography technique: (a) speech data before embedding, (b) amplitude modification, (c) speech data after embedding.

VI. CONCLUDING REMARKS

The proposed techniques take advantage of the redundancy of PCMU in which speech data whose absolute amplitude is 0 may be represented by two overlap codes.

The capacity of the proposed techniques is defined by the number of speech data whose absolute amplitude is 0. It should be noted that the capacity could be decreased by the silence compression, an optional function of VoIP for the efficient communications by reducing the transmitted speech data during silent intervals [8].

It should be noted that the proposed technique is not robust in malicious scenarios. Since the embedding position of the steganogram information is easily detected, the steganogram information is easily replaced with adversarial other information.

The idea of lossless steganography techniques described in this paper may be applicable to other codecs that employ folded binary code. However, it should be noted that it is not always possible to perform lossless steganography techniques even if the codecs employ folded binary code.

For example, PCMA also employs folded binary code for representing 8 bit speech data. However, it is impossible to perform a lossless steganography technique in the same manner described in this paper, since there is no overlap code in PCMA. The amplitude of the 8 bit speech data in PCMA ranges -128 to -1 and +1 to +128 [3].

It is of interest to find out other codecs to which the proposed technique may be applicable. This is one of the future works of this study.

ACKNOWLEDGMENT

The author would like to express the gratitude to the Ministry of Education, Culture, Sports, Science and Technology of Japan for providing a grant (no.21760270) toward this study.

REFERENCES

[1] N. Aoki, “A packet loss concealment technique for VoIP using steganography based on pitch waveform replication,” IEICE Transactions on Communications, vol.J86-B, no.12, pp.2551-2560, 2003.
[2] N. Aoki, “A band extension technique for G.711 speech using steganography,” IEICE Transactions on Communications, vol.E89-B, no.6, pp.1896-1898, 2006.
[3] ITU-T G.711, Pulse code modulation (PCM) of voice frequencies, 1988.
[4] N. Aoki, “A technique of lossless steganography for G.711,” IEICE Transactions on Communications, vol.E90-B, no.11, pp.3271-3273, 2007.
[5] ITU-T G.191, Software tools for speech and audio coding standardization, 2005.
[6] L.R. Rabiner and R.W. Schafer, Digital Processing of Speech Signals, Prentice-Hall, 1978.
[7] The Acoustical Society of Japan, Continuous Speech Corpus for Research, 1991.
[8] D.J. Wright, Voice over Packet Networks, John Wiley & Sons, 2001.