Research on covert communication channel based on modulation of common compressed speech codec

Fufang Li1 · Binbin Li1 · Yongfeng Huang2 · Yuanyong Feng1 · Lingxi Peng3 · Naqin Zhou4

Received: 11 January 2020 / Accepted: 20 March 2020
© Springer-Verlag London Ltd., part of Springer Nature 2020

Abstract
As is well known, multimedia has been widely used in VoIP and mobile communications. Research on how to establish covert communication channel over the above popular public applications has been flourishing in recent years. This paper tries to present a novel and effective method to construct a covert channel over common compressed speech stream by embedding sense information into it. In our method, after analysing the characteristic features of the excitation pulse positions of the ITU-T G.723.1 and G.729A speech codec, we design a novel and effective covert communication channel by finely modulating the codes of excitation pulse positions of the above two codecs in line with the secret information to be hidden. To improve the embedding capacity of the proposed method, we also use all the odd/even characteristics of pulse code positions to conduct information hiding. To test and verify the proposed approach, experiments are conducted on several different scenarios. Experimental results show that our methods and algorithms perform a higher degree of secrecy and sound information embedding efficacy compared with exiting similar methods.

Keywords G.723.1 and G.729 speech codec · Information hiding · Covert communication channel · Codebook · Pulse position code

1 Introduction and related work
Covert communication uses human perceptual defects and fuzziness to embed sensitive information into digital carrier media such as text, sound, images, and video signals in a manner that is not easy to be perceived. Voice communication is one of the most popular ways for people’s daily information exchange. It has a very wide application base and a large amount of communication capacity. Therefore, information hiding based on voice carrier has a very wide range of application requirements in the military, security, business, and other fields. The research on information hiding of carriers has very important research significance and broad application prospects.

Steganography entails the embedding of secret information in digitalized carriers to conceal both its transmission behaviour and contents, thus ensuring its secrecy and security during transmission, which has become an important technical tool for secure secret information transferring [1–3]. Voice-over-Internet protocol (VoIP)-based low-bit rate voice services have developed rapidly with the widespread and popularization of Internet
technologies. Therefore, studies on steganographic techniques based on low-bit rate VoIP are highly significant in theory and in practice [2]. VoIP-based steganography had been developed most rapidly among multitudes of steganographic methods that use streaming media as carriers. This is because the secret information embedded within VoIP communications is highly dynamic, which renders these steganographic methods robust against tampering of attackers. Furthermore, VoIP speech codec can provide excellent levels of imperceptibility and hiding capacity [4, 5].

In the research of information hiding based on speech coding, many researchers have carried out fruitful work. Reference [6] proposed a novel QIM (quantization index modulation) steganography based on the replacement of quantization index set in linear predictive coding (LPC), by treating each quantization index set as a point in quantization index space. Their algorithm had significantly improved the embedding efficiency. Tian [7] tried to present a novel steganalysis method based on statistic characteristics of fundamental frequency, so as to prevent the illegitimate use of voice-over-IP-based information hiding. They introduced the statistics for zero-crossing count (ZCC) and the average values of mel-frequency cepstral coefficients (MFCCs) to characterize inactive frames, and then, they propose a support vector machine-based steganalysis for inactive speech frames. Experiment results showed that their method significantly outperforms the previous ones. Wang [8] put forward a new steganography algorithm by embedding data while pitch period prediction is conducted during low-bit rate speech encoding. Their approach has great compatibility and can achieve high quality of speech and prevent detection of steganalysis without causing further delay by data embedding and extraction. Article [9] proposed a method of text emotional modulation steganography based on machine learning. By intelligently expanding the text emotional lexicon and using cosine similarity algorithm to conduct the steganography algorithm, their method obtained sound efficiency, security, concealment, and robustness. By revealing the inactive frames of VoIP streams which are more suitable for data embedding than the active frames of the streams, paper [10] presented a novel high capacity steganography algorithm for embedding data in the inactive frames of low-bit rate audio streams. Their algorithm had achieved perfect imperceptibility and high data embedding rate up to 101 bits/frame, in which the data embedding capacity is very much larger than most of the existing algorithms. Yang [11] proposed a highly imperceptible, high capacity steganographic algorithm based on the noise coding frames of the G.723.1 codec. Literature [12] studied a novel method to embed sensitive data into G.723.1 speech codec by using the characteristic of the codec. Their algorithm can obtain good efficiency and performance. Guo [13] put forward an imperceptible information hiding approach on 5.3 Kbit/s codec of G.723.1. To increase the data embedding capacity, they also propose a voice activity detection method that uses the residual signal energy of the speech signal.

In recent years, research on information hiding based on 729A speech coding has also attracted the attention of many researchers. Literature [14] demonstrates the respective performance advantages of the two standards in the error concealment process through the frame loss experiment and subjective test on G.729.1 and AMR_WB (adaptive multi-rate wideband). El-Khamy et al. [15] give out a new robust audio steganography technique based on optimum two-dimensional complete complementary codes (CCC). When adopted their method to embed colour images data, and with various attacks applied to the carrier audio signals, simulation experimental results showed that the extracted image had good quality, which proves high robustness and capacity of their proposed algorithm. Literature [16] counts the unimportant bits of the compressed low-bit speech stream, proposes embedding hidden information in the speech stream, and designs a voice concealed communication system of IP network using this algorithm to improve the security of secret information transmission. Literature [17] proposes an algorithm based on code word characteristics, which enables G.729A to carry out QIM steganography to detect effectively in a short time. Paper [18] discussed the application of speech-based information hiding in the area of medical scenario. Literature [19] carries out a comprehensive test for the frame-by-frame bitwise negation of G.729A speech and finds that the speech has hidden bits, and a method of information hiding combined with matrix coding is proposed.

As is discussed above, steganography based on compressed speech codec has been deeply studied by many researchers. However, the robustness, concealment, and hiding capacity of existing algorithms are still difficult to meet the needs of practical applications. Therefore, it is urgent to study new information hiding methods that have both high privacy and high hiding capacity. To explore a more effective way to establish a novel convert communication channel over compressed speech stream, this paper presents a steganography method based on modulation of the excitation pulse positions of the G.723.1 low-bit rate codec and G.729A speech codec as the research scenario. In the proposed method, the secret sensitive information is hidden in the above speech stream by finely modulating the codes of excitation pulse positions. Experimental results show that the proposed steganography algorithm of G.723.1 speech codec with an embedding rate of 3.1% and secret information transfer rate of 166 bits/s results in $< 5.7\%$ degradations in perceptual evaluation of speech.
quality (PESQ) scores, which indicates a high degree of imperceptibility. As to the proposed steganography algorithm of G.729A speech codec, our method can achieve 400 bite/s secret information communication with 3% embedding rate, while the capacity and the degradation rate of the PESQ score for the carrier speech is less than 7%, indicating that the hidden method studied has good privacy and efficacy.

This paper consists of six sections: Sect. 1 (that is, this chapter) briefly introduces the original purpose and related work of the research; Sect. 2 introduces the related G.723.1 and G.729A speech codec; Sect. 3 discusses the model of information hiding and extraction construction method of emotional lexicon based on deep learning; Sect. 4 gives the information hiding algorithm based on matrix encoding and emotional word replacement; Sect. 5 shows the experiment results and analysis; Sect. 6 shows the summary and prospect part.

2 Brief introduction of low-bit rate codec of G.723.1 and G.729A

2.1 Brief introduction of G.723.1 codec

The G.723.1 speech coding standard is a low-bit rate codec scheme that was formulated by ITU-T for visual telephony. The excitation signals corresponding to the high (6.3 Kbit/s) and low (5.3 Kbit/s) bit rates of G.723.1 are multipulse maximum likelihood quantization (MP-MLQ) and algebraic codebook excited linear prediction (ACELP), respectively. The G.723.1 recommendation uses the ACELP (with 5.3 Kbit/s low bit rates) and MP-MLQ (with 6.3 Kbit/s high bit rates) algorithms to minimize the perceptually weighted error signal [20].

According to G.723.1 speech recommendation, the analogue input signal is firstly analysed and processed by using the rules of G.723.1 standard, and then, the parameters, such as LSP parameters, codebook indices, and gains, are calculated. Secondly, the parameters are digitalized so as to facilitate transmission in the next step. At last, the digital parameters are packed and wrapped according to G.723.1 standard and then send to the transmission channel. By this way, the encode process of G.723.1 speech codec has been finished. At the receiving end, firstly, the wrapped parameters are unpacked, and thus, the parameters can be obtained by the decoder. Secondly, the decoder uses these regained parameters to reconstruct the analogue speech signal [21]. The encode and decode process of G.723.1 speech codec is shown in Fig. 1.

2.2 Brief introduction of G.729A codec

The G.729A protocol, which proposed by International Telecommunication Union, Telecommunication Standardization Sector (ITU-T), is a 8-Kbps speech coding standard and is a simplified version of the speech coding standard G.729. The G.729A protocol uses conjugate structured arithmetic codebook excited linear prediction (CS-ACELP) as its coding principle. The CS-ACELP encoder is based on a codebook excited linear predictive coding (CELP) model. The encoding procedure of G.729A speech codec is: (1) The analogue speech signal is sampled at 8 kHz after being band-pass filtered by the speech channel, quantized into a 16-bit linear PCM digital signal, and then, they are inputted to the G.729A encoder. (2) The G.729A encoder first processes the 10-ms-long speech frame as follows: linear prediction analysis and quantization of LPC coefficients; open-loop pitch period estimation; adaptive codebook search; fixed codebook search. (3) Then, calculate the CELP model parameters (LP filter coefficients, adaptive codebook, and fixed codebook index and gain) frame by frame. (4) Finally, encode and transmit these parameters, and thus, the encoding process is finished. The detailed bit allocation of the 80 bits of G.729A coding parameters is shown in Table 1.

The G.729A decoder decodes the G.729A code stream and extracts the corresponding line spectrum pair (LSP) coefficients and two fractional pitch delays, two fixed codebook vectors, and two sets of adaptive codebook and fixed codebook gains. These parameters are used to recover the excitation signal and synthesis filter parameters. According to the mechanism of speech generation, linear prediction (LP) filter filtering is used to implement speech synthesis (i.e. speech reconstruction). After the synthesized speech is obtained, post-filtering is performed to improve the speech quality.

3 Information embedding and extraction model based on common compressed speech codec

The method of embedding and extraction secret information based on modulation of the excitation pulse positions of the common compressed speech codec is illustrated in Fig. 2. Figure 2a describes the method for embedding secret information. Firstly, speech analysis and processing are performed, followed by the calculation of LPC indices, adaptive codebook lags, the combinatorial codes of all gains, pulse signs, and pulse positions. The pulse positions are then finely modulated according to the secret information; finally, the data are packaged for transmission via ordinary
communication channels. The extraction of secret information is illustrated in Fig. 2b. After the data packets have been received from a communication channel, they are unpacked to extract all of the relevant parameters. The secret information hidden in the excitation pulse positions is then extracted according to its embedding rules, before the data are passed on to the decoder for speech reconstruction.

4 Steganography algorithms based on the excitation pulse positions of compressed speech codecs

The proposed steganography method includes four algorithms according to three different scenarios of G.723.1 speech codec: (1) steganography algorithm based on excitation pulse positions of ACELP of G.723.1 speech codec (Algorithm 4_1); (2) the first and second steganography algorithm based on excitation pulse positions of MP-MLQ of G.723.1 speech codec (Algorithm 4_2 and Algorithm 4_3); and (3) steganography algorithm for scenario of G.729A compressed speech codec (Algorithm 4_4).

4.1 Steganography algorithm for scenario of ACELP of G.723.1 (Algorithm 4_1)

The low-bit rate ACELP of G.723.1 codec uses adaptive codebook searches. A G.723.1 ACELP frame is composed of four subframes, and each subframe has four excitation pulse positions. ACELP codec uses 3 bits to identify every excitation pulse position in each subframe. To embedding sensitive secret information into the ACELP codec bitstream, we finely modulate the excitation pulse positions’ code according to the bit of secret data to be embedded. Let’s suppose the secret information to be embedded is $S_t = [S_t(0), S_t(1), \ldots, S_t(i), \ldots, S_t(L - 1)]$ ($L$ is the length of $S_t$). The information embedding subalgorithm of Algorithm 4_1 (named as Algorithm 4_1_1) and the extraction subalgorithm (named as Algorithm 4_1_2) are as follows:
4.2 First steganography algorithm for scenario of MP-MLQ of G.723.1 (Algorithm 4_2)

In high bit rate G.723.1 MP-MLQ codec, each subframe has four positions that generate excitation pulses, while even and odd subframes have six and five nonzero excitation pulses, respectively, and the positions of all nonzero pulses are either all odd or all even. Therefore, there are fewer than 30 pulse positions in each subframe. The algorithm of performing steganography based on the high bit rate MP-MLQ codec of G.723.1 is described as below: Let’s suppose the bitstream of the secret information to be embedded is $S_i = [S_i(0), S_i(1), \ldots, S_i(i), \ldots, S_i(L - 1)]$ ($L$ is the length of $S_i$). The information embedding subalgorithm of Algorithm 4_2 (named as Algorithm 4_2_1) and the information extraction subalgorithm of Algorithm 4_2 (named as Algorithm 4_2_2) are as follows:

**Algorithm 4_2_1:**

**Input:** the normal G.723.1 ACELP speech codec flow data.

**Output:** the G.723.1 ACELP speech codec flow data with secret information embedded in.

BEGIN

for $i=0, i<n, i++$ {
    read a g.723.1 ACELP frame code data, store $S_i(i)$ into the reserved bit of the current frame;
    $i++$;
    if ($i<n$) {
        read the pulse position-code of each subframe;
        modulate the odd and even properties of the pulse position-code be the same as $S_i(i)$, by adding or subtract 1 to the pulse position-code, and thus the secret information bit has been embedded;
        $i++$;
    } else {
        break;
    }
} //end of for loop

END.

**Algorithm 4_2_2:**

**Input:** the G.723.1 ACELP speech codec flow data with secret information embedded in.

**Output:** the normal G.723.1 ACELP speech codec flow data and the extracted secret information from the carrier.

BEGIN

for $i=0, i<n, i++$ {
    read every g.723.1 frame code data in sequence;
    extract the secret information in the reserved bit and store it to $S_i(i)$;
    $i++$;
    extract the information hidden in each subframe of the current frame in sequence according to formula $S_i(i) = (ipx) & (001Hex)$;
    $i++$;
} //end of for loop

END.
Algorithm 4.2.1:
**Input:** the normal MP-MLQ of G.723.1 speech codec flow data and the secret information to be embedded.
**Output:** the MP-MLQ of G.723.1 speech codec flow data with secret information embedded in.

BEGIN
for \(i=0, i<_L, i++\) {
read the 1-bit information being hidden, \(S_1(i)\);
read a frame of the MP-MLQ encoded data and store \(S_1(i)\) in the reserved bit of the current frame;
i++;
if \(i>=L\) { break; }
}

For (the subframes in the current frame have not been completely processed) {
read the pulse position-codes of each subframe in the current frame in sequence;
if (this subframe is an odd frame) {
assign all of the extracted pulse position-code values to the PO[0..5] array;
sort the PO[0..5] array in ascending order;
For \(j=0, j<5, j++\) {
P_t = PO[j]; P_t>>1;
read the 1-bit data being hidden, \(S_1(i)\);
if \(P_t\) has the same parity as \(S_1(i)\) {} else {
perform a +2 or -2 operation on \(P_t[j]\); // please refer to the comments
}
i++;
if \(i>=L\) { break; }
} else {
assign all of the extracted pulse position-code values to the PE[0..4] array;
sort the PE[0..4] array in ascending order;
For \(j=0, k<=4, k++\) {
P_t = PE[k]; P_t>>1;
read the 1-bit data being hidden, \(S_1(i)\);
if \(P_t\) has the same parity as \(S_1(i)\) {} else {
perform a +2 or -2 operation on \(P_t[j]\); // please refer to the comments
}//if
i++;
if \(i>=L\) { break; }
}//for
}//if
}//for

END.

Algorithm 4.2.2:
**Input:** the MP-MLQ of G.723.1 speech codec flow data with secret information embedded in.
**Output:** the original MP-MLQ of G.723.1 speech codec flow data and the extracted original secret information from the carrier.

BEGIN
for \(i=0, i<_L, i++\) {
read each frame of the ACELP-encoded data in sequence; extract the information in the reserved bit and store this as \(S_1(i)\);
ext the information hidden in each subframe of the current frame in sequence according to \(S_1(i) = ((ip) > 1) \& 0011_{hex}\);
}

END.

Note: (a) Since the positions of all non-zero pulses in each subframe of the high bit rate MP-MLQ codec specified by the G.723.1 standard are all odd or all even, the least significant bit cannot be used to hide information; (b) the pulse positions are prevented from exceeding the upper and lower boundaries by changing the addition or subtraction operations when approaching the boundaries of the pulse code positions; (c) using the algorithm described above, the data that is being hidden will be embedded in the second most significant bit of the pulse position-codes.

END.
4.3 Second steganography algorithm for scenario of MP-MLQ of G.723.1 (Algorithm 4_3)

Since the pulse positions of every subframe in the G.723.1 high bit rate MP-MLQ codec must be integers in the [0–59] interval, and the pulse sequence numbers of each subframe are either all even or all odd, there are only 30 (fewer than 25–1) possible variations in position for any pulse in a subframe. As the code of pulse positions has less than 5 bits, each pulse may then hide a maximum of 4 bits of secret information. However, not every pulse position code has 30 bits, so it is more appropriate to hide only 3 bits of secret data in each position. This is consistent with the findings of our preliminary experiments. Suppose again the bitstream of secret information to be embedded is \( S_i = [S_i(0), S_i(1), \ldots, S_i(i), \ldots, S_i(L - 1)] \) (\( L \) is the length of \( S_i \)). The secret information embedding subalgorithm of Algorithm 4_3 (named as Algorithm 4_3_1) and the information extraction subalgorithm of Algorithm 4_3 (named as Algorithm 4_3_2) are as follows:

**Algorithm 4_3_1:**
**Input:** the normal MP-MLQ of G.723.1 speech codec flow data and the secret information to be embedded.
**Output:** the MP-MLQ of G.723.1 speech codec flow data with secret information embedded in.

**BEGIN**
for \( i=0, i< L-2, i++ \) {  
  read the 1-bit information being hidden, \( S_i(i) \);
  read the first frame of the MP-MLQ coded data and store \( S_i(i) \) in the reserved bit of the current frame;
  \( i++ \); if \( i= L \) { break; }
  for (the subframes in the current frame have not been fully processed) {
    extract all of the pulse position-codes of each subframe in sequence;
    read the 3-bit stream of information being hidden and assign a value to a binary number, \( N_p \), such that \( N_p = S_i(i)S_e(i+1)S_e(i+2) \);
    \( N_p \leftarrow 1 \);
    if (the parity of \( N_p \) is the same as that of the numerical value of the first pulse position-code of this subframe) {
      replace the 4 least significant bits of this pulse position-code with \( N_p \);
    } else {
      replace the 4 least significant bits of this pulse position-code with \( N_p \) and add 1 to the least significant bit;
    }
  }
  \( i = i+3 \);
  if \( (i= L) \) { break; }
}  
**END.**

**Algorithm 4_3_2:**
**Input:** the MP-MLQ of G.723.1 speech codec flow data with secret information embedded in.
**Output:** the original G.723.1 speech codec flow data and the extracted secret information from the carrier.

**BEGIN**
for \( i=0, i< L-2, i++ \) {
  read each frame of the MP-MLQ coded data in sequence;
  extract the information in the reserved bit and store this as \( S_i(i) \);
  extract the hidden information in each subframe of the current frame in sequence, according to \( S_i(i+1) = (\{(i) \gg 1\}) \& (0001 \text{Hex}) \);
}  
**END.**

4.4 Steganography algorithm for scenario of G.729A codec (Algorithm 4_4)

The fixed code book encoded by G.729A is an algebraic code book structure. Each code book vector contains four nonzero pulse signals, and the amplitude of each pulse is positive or negative. The positions of each pulse are shown in Table 2. From Table 2, it can be seen that the G.729A code has 4-pulse excitation for each frame of speech signal (80 samples), the first, second and third pulses have eight possible positions, and each pulse position is represented by 3 bits. The fourth pulse has two groups of odd and even numbers, with 16 possible positions. The position of the pulse is represented by 4 bits. Therefore, the positions of the four pulses in a frame in G.729A coding are presented by 13 bits. The detailed distribution of the position of the pulses in G.729A fixed code book is shown in Table 2.

The ideas of secret information embedding method in information hiding algorithm of G.729A coding based on excitation pulse position modulation are shown in Fig. 3.

The proposed information embedding subalgorithm of Algorithm 4_4_1 is as follows:
Step 1 Read the information that needs to be hidden and transforms the information into a binary data form of 01.

Step 2 The hidden information of 1 bit (Either 0 or 1) is read and compared with the parity of the current pulse position value (Either 0 or 1). If they are also even or odd, the information can be hidden directly without numerical processing of pulse position; otherwise, the position of pulse will be added or subtracted by 1, so that the 1 bit information to be hidden can be the same as the parity of pulse position value. At this time, the parity of the lowest bit of the pulse position value (binary 0 or 1) indicates that 1 bit of information has been hidden.

The bit of pulse excitation is found in all parameter bit streams decoded by the G.729A decoder. Comparing the parity of the hidden information in each pulse coding, i.e. the lowest bit’s value of the binary form of the pulse position, the hidden bit of binary information is extracted from the corresponding pulse sequence by (ip) & (0001Hex). The secret information extraction method in information hiding algorithm of G.729A coding based on excitation pulse position modulation is shown in Fig. 4.

5 Experimental results and analysis

5.1 Design of experimental scheme

To evaluate the proposed speech algorithms, we examine the changes in objective speech quality of PESQ scores in different types of voices, and under condition of before and after the embedding of secret information, at different levels of hiding capacity. In the experiments, we used self-edited clips of Chinese and foreign male and female voices that were sampled at 8 kHz to perform the PESQ tests. For each one of the four kinds of voices, we created 200 edited clips of each voice and perform 200 times of tests; we then record the average PESQ scores of these 200 times of tests of each kind of voice. To provide an intuitive representation of the steganographic efficacy of the proposed algorithms, the reduction percentage of the PESQ scores are also listed, in addition to the raw PESQ scores obtained in each test. The effectiveness of our algorithms is then further examined in detail by performing comparative experiments with similar algorithms. Detailed descriptions of the experiments and their results are provided in the following.

5.2 Experimental results of steganography Algorithm 4_1

By using the experimental scheme described in Sect. 5.1, objective speech quality tests of PESQ values are carried out. The average PESQ scores of these self-edited clips of voices (that do not contain any hidden data) are slightly lower than that of standard G.723.1 ACELP speech codec. To facilitate an intuitive observation of steganographic efficacy, the results of this experiment are represented as reduction percentage of PESQ scores. The average PESQ scores and their reduction percentages of before and after steganography by conducting Algorithm 4_1 are shown in Table 3, while Table 4 illustrates how PESQ scores are changed when hiding capacity and embedding rate increase.

In Table 3, the first row of data field shows the original average PESQ scores of the 200 edited clips of each kind of voice in which no sensitive data have been embedded in. The other rows in the data field of Table 3 indicate the average PESQ values and their percentages compared to the original PESQ values under different embedding rates of sensitive information by using Algorithm 4_1. The corresponding rows in the data field of Table 4, respectively, indicate the average PESQ value of the carrier voice and its decrease percentage compared to the original PESQ value under different embedding rates and data transmission rates. From Tables 3 and 4, we can also see that, even when the embedding rate goes up to 10.63% and the hiding capacity was 566 bits/s, the reduction percentage in PESQ score was less than 12.5% and thus indicates an excellent level of imperceptibility and efficiency can be achieved by Algorithm 4_1. And Table 4 tells us the average PESQ scores of the speech carrier’s decrease with increasing hiding capacity and embedding rate when steganography is performed by using Algorithm 4_1.

5.3 Experimental results of steganography Algorithm 4_2

Experiment of Algorithm 4_2 is also based on the experimental scheme described in Sect. 5.1. The average PESQ scores of these voices without information hidden are also slightly lower than the standard PESQ score of G.723.1 MP-MLQ speech codec. Again, to facilitate an intuitive observation of steganographic efficacy, the results of this experiment are represented as percentage reductions in
PESQ scores. The results of average PESQ scores and reduction percentages of before and after steganography by using Algorithm 4_2 are shown in Tables 5 and 6.

The first row of the data field of Table 5 is as the same as the first row of Table 3. The other rows in the data field of Table 5 indicate the average PESQ values and their percentages compared to the original PESQ values under different embedding rates of sensitive information by using Algorithm 4_2. The data rows in Table 6, respectively, indicate the average PESQ value of the carrier voice and its reduction percentage compared to the original PESQ value under different embedding rates and data transmission.
rates. Tables 5 and 6 show that the degradation of PESQ score is less than 12.1% even when the embedding rate is 8.8% and the hidden information transmission rate is 566 bits/s, which indicates an excellent level of imperceptibility. And in Table 6, we can see that the average PESQ score of the speech carrier decreases with increasing hiding capacity and embedding rate when steganography Algorithm 4_2 is performed.

5.4 Experimental results of combination of steganography Algorithm 4_2 and 4_3

To further increase data hiding capacity, we perform an experiment with Algorithms 4_2 and 4_3 being used in combination. These experiments are also implemented using the scheme described in Sect. 5.1. To investigate different combinations of these algorithms, we perform experiments based on the four scenarios that are described below: (1) 3 bits of information is hidden in the first pulse of each subframe, except 1 bit of information that is hidden in the reserved bit; hence, each frame (which contains 4 subframes) can hide 13 bits of information in total. (2) 3 bits of information is hidden in the first pulse of each subframe, and 1 bit of information is hidden in the last bit of the last subframe, while 1 more bit of information is hidden in the reserved bit; hence, each frame hides 17 bits of information in total. (3) In each subframe, 3 bits of information is hidden in the first pulse and 1 bit of information is hidden in each of the last two pulses, and 1 bit of information is hidden in the reserved bit. Hence, each frame is hiding 21 bits of information in total. (4) In each subframe, 3 bits of information is hidden in the first pulse while the 1 bit of information is hidden in the last bit in each of the remaining pulses, and 1 more bit of information is hidden in the reserved bit. Since the first and third subframes have 6 pulse positions while the second and fourth subframes have 5 pulse positions; each frame is therefore hiding 31 bits of information in total. For the PESQ reduction and its reduction percentage of the above fourth scenario is too sharp, the fourth scenario is not suitable or practicable for information hiding.

The experimental results of the three practicable aforementioned scenarios are shown in Tables 7 and 8. Table 7 shows the average PESQ scores and their reduction percentage according to schema described in Sect. 5.1 by using combination of Algorithms 4_2 and 4_3. Table 8 describes how the reduction percentage of the average PESQ scores is affected to hiding capacity and embedding rate increases when Algorithms 4_2 and 4_3 are used in combination.

In Table 7, the first data row is as the same as the first row of Table 3. The other data rows of Table 5 indicate the average PESQ scores and their percentages compared to the original PESQ scores under different embedding rates of sensitive data by using combination of Algorithms 4_2 and 4_3. The data rows in Table 8, respectively, show the average PESQ values of the carrier voice and their reduction percentage compared to the original PESQ values under different embedding rates and data transmission rates. In Tables 7 and 8, it is shown that a combination of Algorithms 4_2 and 4_3 greatly increases the hiding capacity while the PESQ reduction percentages are acceptable. The PESQ scores of the speech carrier decreased with increasing hiding capacity and embedding rate. Nonetheless, when the hiding capacity reaches 700 bits/s and embedding rate goes up to 10.9%, the reduction percentage of the PESQ score was still <16.7%, which shows that the proposed scheme is rational, sound, and effective.

5.5 Experiment results of steganography algorithm for scenario of G.729A codec (Algorithm 4_4)

The experiment schema of this scenario is the same as Sect. 5.1. At first, the fourth pulse was used as the basis of information hiding, and the amount of hidden information was gradually increased on other pulses in succession. The

| Steganography schema | China male | China female | Foreign male | Foreign female |
|----------------------|------------|--------------|--------------|----------------|
| No data hidden       | 3.503025   | 3.356881     | 3.653155     | 3.243805       |
| 5 bits of data is hidden in each frame (including 1 bit in the reserved bit) | 3.312811 | 3.181987 | 3.479995 | 3.101402 |
| 94.57% | 94.79% | 95.26% | 95.61% |
| 9 bits of data is hidden in each frame (including 1 bit in the reserved bit) | 3.206319 | 3.070203 | 3.345924 | 3.011224 |
| 91.53% | 91.46% | 91.59% | 92.83% |
| 17 bits of data is hidden in each frame (including 1 bit in the reserved bit) | 3.070051 | 2.944992 | 3.212219 | 2.872065 |
| 87.64% | 87.73% | 87.93% | 88.54% |
average PESQ scores and its’ reduction percentages before and after steganography is performed by using the presented algorithm are shown in Table 9. The first row of Table 9 is similar to the first row of Table 3, which shows, under the G.729A codec being conducted, the original average PESQ scores of the 200 edited clips of each kind of voice in which no sensitive data have been embedded in. The other data rows in Table 9, respectively, indicate the average PESQ value of the carrier voice and its’ percentages (or reduction percentages) compared to the original average PESQ values under different embedding rates and data transmission rates. As is shown in Table 9, taking the fourth pulse position as the basis of hiding information, when each subframe only hides 1 bit of information, the hidden speech quality is about 94.5% of the speech without hidden information, showing that the hiding effect is very good. When each fourth pulse of every subframe embeds 2 bits of sensitive information, the percentage of PESQ scores is still 91%, showing the speech quality is also very sound. While more than one pulse of each subframe or all the four pulses are used for embedding information (as is shown in the last row of the table), the quality of PESQ scores of the carrier speech decreases sharply, showing that these occasions are unsuitable for information hiding.

5.6 Comparative experiment between the algorithms developed in this work and similar algorithms

To examine the steganographic efficacy of our algorithms in further depth, we choose our Algorithm 4_2 and 4_4 to
conduct comparative experiments with existing similar methods. We compared our Algorithm 4_2 with algorithms in Refs. [8, 11] and Algorithm 4_4 with algorithm in Ref. [13], which are more similar with our methods. The experiments are conducted using the scheme described in Sect. 5.1. The results of the comparative experiments in terms of PESQ scores and their reduction percentages when steganography was performed using Algorithm 4_2 and the algorithms described in Refs. [8, 11] are shown in Table 10. And the comparative experiment results of our Algorithm 4_4 and the algorithms described in Ref. [13] are shown in Fig. 5.

In Table 10, the first column of data field shows the original average PESQ scores of the 200 edited clips of each kind of voice in which no sensitive data have been embedded in (note: the G.723.1 codec being conducted). The other columns in the data field of Table 10, respectively, indicate the average PESQ values and their decreased percentages are compared to the original PESQ values under corresponding embedding algorithms of our proposed Algorithm 4_2 and similar algorithm of Refs. [8, 11] being conducted. As is shown in Table 10, as compared with the algorithms described in Refs. [8, 11], the PESQ scores and their reduction percentages are lower, when our steganographic algorithm is used. It is thus shown that the proposed algorithm has a greater degree of imperceptibility or in other words higher levels of steganographic efficacy and secrecy and robustness, comparing with other two contrast algorithms.

In Fig. 5, the horizontal axis represents the four kinds of edited clips of speech voices, while vertical axis represents the reduction percentage of the PESQ scores when different steganography algorithms are used. From Fig. 5, we can see that the PESQ scores’ reduction percentage of Algorithm 4_4 is lower than that of the similar algorithm described in Paper [13], which shows that our algorithm has a greater degree of secrecy, imperceptibility, and robustness.

### 6 Design of application scenario system for sensitive information covert transmission

With the widespread popularity of information and communication applications, the secure and confidential distribution and transmission of sensitive information have a wide range of practical value. Obviously, covert communication based on information hiding can be widely used for the secret transmission of confidential information of institutions, organizations, enterprises, and individuals. Especially in the recent application scenario of the key data sharing of the novel coronavirus (i.e. COVID-19) outbreak information, covert communication based on information hiding can not only realize the safe sharing of related information, but also effectively protect the privacy of related personnel and even personal safety. Next, we discuss the design of a stealthy communication experiment system for the safe and secret sharing of key sensitive data. The structure diagram of the hidden communication experimental application system based on information hiding is shown in Fig. 6.

It can be seen from Fig. 6 that the system is composed of a sending terminal, a sending agent, a virtual covert channel, a receiving agent, and a receiving terminal (similar to the receiving terminal). The sending terminal and the receiving terminal are served by various types of devices participating in the sharing of confidential information,
such as personal/portable computers, mobile phones, and handheld digital devices. These devices need to have the conditions or capabilities to deploy sending/receiving agents. The sending and receiving agents implement the processing procedures of the information hiding algorithm proposed in this paper. They are responsible for embedding the confidential information that needs to be sent or shared into the voice carrier, or extracting the confidential information embedded in the voice carrier. The virtual hidden channel is realized by transmitting the encrypted voice data on the public shared channel.

The process of sharing or transmitting secret sensitive information is as follows: First, the sender terminal calls the sending agent to embed the confidential and sensitive information in ordinary speech coded data (such as G723.1 or 729A); then, the sending agent hands over the encrypted speech coded data The public channel is then sent to the receiver; finally, the receiver terminal calls the receiving agent to present the confidential information hidden in the voice-encoded data, thereby completing the secure sharing of confidential and sensitive information.

### Table 10 Results of the comparative experiment of Algorithm 4.2 compared to similar algorithms of Refs. [8, 11]

| Voice type | Our algorithm | Algorithm [8] | Algorithm [11] |
|------------|---------------|---------------|----------------|
|            | PESQ score    | Reduction percentage (%) | PESQ score    | Reduction percentage (%) | PESQ score    | Reduction percentage (%) |
| China male | 3.503025      | 12.36         | 3.042377      | 13.15            | 3.175937      | 13.73            |
| China female | 3.356881 | 12.27         | 2.905045      | 13.46            | 3.032576      | 13.99            |
| Foreign male | 3.653155 | 12.07         | 3.151577      | 13.73            | 3.289931      | 14.33            |
| Foreign female | 3.243805 | 11.46         | 2.809459      | 13.39            | 2.792267      | 13.92            |

7 Conclusion and outlook

In this paper, we have developed a covert communication channel by steganography on common compressed speech codec such as G.723.1 and G.729A. We present four steganography algorithms based on different strategies. It is experimentally demonstrated that the proposed steganographic method can obtain high levels of imperceptibility, security, and robustness with sound hiding capacities. As for G.723.1 speech codec, we can achieve a data hiding capacity of 3.1%, 5.6%, or 10.6% (secret information transmission rate of 166 bits/s, 300 bits/s or 566 bits/s) results in 5.7%, 9.4%, or 12.5% reductions in PESQ scores, respectively, while for G.729A, we can also obtain a data hiding capacity of 2.5%, 3.75%, or 5% (secret information transmission rate is 200 bits/s, 300 bits/s or 400 bits/s) results in < 5.4%, < 8.1%, or < 16.8% reductions in PESQ scores. The above experimental results show our algorithms can obtain high degree of imperceptibility, security, and robustness. The comparative experiment also demonstrated that our algorithm has a higher level of overall efficacy than similar algorithms in the related literature. The information hiding algorithms studied in this paper can be widely used in various application scenarios for the confidential transmission of confidential information, such as protecting the important sensitive information of various institutions or the privacy of personal financial and health information. Although the algorithms proposed in this paper have good comprehensive performance and efficiency, the capacity of information hiding still does not meet the needs of high-speed covert communication, and the contradiction of increasing hidden capacity and increasing privacy has not yet been fully resolved. In the future, we will focus on applications of the methods developed in this work. In addition, we will further investigate the possibility of performing steganography based on other speech codec parameters so as to discover even more effective steganography approach in the area.
Acknowledgements  The study was supported by the Natural Science Foundation of China (NSFC) under Grant Nos. 61472092, 61977018, and 11271097 and the Guangdong Provincial Science and Technology Plan Fund under Grant No. 2013B010401037.

References

1. Zhang Q-Y, Yu S, Chen N, Hu W-J, Qiao S-B (2016) An information hiding algorithm for speech perceptual hashing authentication system in G.729 bitstream. J Inf Hiding Multimed Signal Process 7(5):1031–1040
2. Mazurczyk W (2013) VoIP steganography and its detection—a survey. ACM Comput Surv 46(2):1–21. https://doi.org/10.1145/2543581.2543587
3. Dutta H, Das RK, Nandi S, Prasanna SM (2019) An overview of digital audio steganography. IETE Tech Rev. https://doi.org/10.1080/02564602.2019.1699454
4. Liu P, Li S, Wang H (2017) Steganography in vector quantization process of linear predictive coding for low-bit-rate speech codec. Multimed Syst 23(4):485–497. https://doi.org/10.1007/s00530-015-0500-7
5. Wei Z, Zhao B, Liu B (2014) A novel steganography approach for voice over IP. J Ambient Intell Humaniz Comput 5(4):601–610. https://doi.org/10.1007/s12652-013-0212-9
6. Huang YF, Tang S, Yuan J (2011) Steganography in inactive frames of VoIP streams encoded by source codec. IEEE Trans Inf Forensics Secur 6(2):296–306. https://doi.org/10.1109/tifs.2011.2108649
7. Tian H, Liu J, Chang C-C, Huang Y, Cai Y (2020) Detecting steganography in inactive voice-over-IP frames based on statistic characteristics of fundamental frequency. IEEE Access 8:6117–6129. https://doi.org/10.1109/ACCESS.2019.2962009
8. Wang C, Huang Y, Wang W, Deng B (2010) A large-capability self-adaptive steganography algorithm based on G.723.1. J Wuhan Univ Technol (Inf Manag Eng) 32(4):522–525. https://doi.org/10.3969/j.issn.1007-144x.2010.04.002 (in Chinese)
9. Li F, Tang H, Zou Y, Huang Y, Peng Y, Peng L (2020) Research on information security in text emotional steganography based on machine learning. Enterp Inf Syst. https://doi.org/10.1080/17517575.2020.1720827
10. Lin R-S (2016) A synchronization scheme for hiding information in encoded bitstream of inactive speech signal. J Inf Hiding Multimed Signal Process 7(5):916–929
11. Yang S, Zhou JL, Yu SS (2001) Clustering optimization strategy for G.723.1 speech coder and its application. J Commun 22(2):113–117. https://doi.org/10.1117/12.420835 (in Chinese)
12. Li F, Li B, Peng L, Chen W, Zheng L, Xu K (2018) A steganographic method based on high bit rates speech codec of G.723.1. In: Sun X, Pan Z, Bertino E (eds) Cloud computing and security. ICCCS 2018. Lecture notes in computer science, vol 11068. Springer, Cham, pp 312–322. https://doi.org/10.1007/978-3-030-00021-9_29
13. Guo S, Tian H (2014) An information hiding method based on G.729a speech. J Huaihua Univ (Nat Sci) 35(3):277–282. https://doi.org/10.111830/issn.1000-5013.2014.03.0277 (in Chinese)
14. Cao Y, Bai S, Cai K, Li WD (2013) Study on covert communication system based on G.729 compressed voice stream. Mod Electron Tech 36(17):68–70. https://doi.org/10.16652/j.issn.1004-373x.2013.07.035 (in Chinese)
15. El-Khamy SE, Korany NO, El-Sherif MH (2019) Highly secured image hiding technique in stereo audio signal based on complete complementary codes. Multimed Tools Appl 78(24):34373–34395. https://doi.org/10.1007/s11042-019-08122-x
16. Li FF, Xie D, Qi D, Xie G (2014) Research on effective and intelligent resource management in internet computing. Appl Math Inf Sci 8(2):625–631
17. Huang Y, Liu C, Tang S (2012) Steganography integration into a low-bit rate speech codec. IEEE Trans Inf Forensics Secur 7(6):1865–1875. https://doi.org/10.1109/TIFS.2012.2218599
18. Peng L, Chen W, Zhou W, Li F (2016) An immune-inspired semi-supervised algorithm for breast cancer diagnosis. Comput Methods Programs Biomed 134:259–265. https://doi.org/10.1016/j.cmpb.2016.07.020
19. Yong-feng H, Jiang-ling Z (2000) Implementation of ITU-T G.729 speech codec in IP telephony gateway. Wuhan Univ J Nat Sci 5(2):159–163. https://doi.org/10.1007/s10728-001-0027-1 (in Chinese)
20. Cheng H, Yang Y, Hu R, Wang S (2011) Research of error concealment and control based on G.729.1 and AMR-WB. Audio Eng 35:68–72 + 83. https://doi.org/10.16311/j.audioe.2011.12.018 (in Chinese)
21. Li SB, Sun DH, Yuan J, Huang YF (2012) A steganalysis method for G.729 A compressed speech stream based on codeword distribution characteristics. Acta Electron Sin 40(4):842–846. https://doi.org/10.3969/j.issn.0372-2112.2012.04.036

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.