# VoIP Steganography using iLBC Start State Residuals

Ting-Xiao Miaw
*Dept. of Computer Science and*
*Information Engineering*
*National Chi Nan University*
Nantou, Taiwan
email: s109321053@ncnu.edu.tw

Quincy Wu
*Dept. of Computer Science and*
*Information Engineering*
*National Chi Nan University*
Nantou, Taiwan
ORCID: 0000-0002-4867-5627
email: solomon@ncnu.edu.tw

***Abstract*—In this paper, we propose a novel approach for information concealment within speech signals encoded using the iLBC codec. Our method leverages the start state's residual samples. We explore the utilization of the 3-bit quantized start state residual signal as an information hiding field. By strategically selecting specific quantization table indices based on the hidden bits, we achieved a remarkable four-fold enhancement in information hiding capacity while maintaining a high level of Perceptual Evaluation of Speech Quality (PESQ). This method offers a flexible balance between capacity and imperceptibility, allowing adaptation according to varying real-world scenarios. This adaptability empowers customization of information hiding techniques, making it highly applicable in diverse contexts.**

***Keywords—iLBC, speech coding, information hiding, steganography, steganographic capacity, imperceptibility***

## I. Introduction

In an age dominated by digital interactions, safeguarding the confidentiality of information has emerged as a critical challenge. While cryptographic technologies have strengthened data protection during transmission, the realm of secure communication expands beyond these conventional encryption/decryption methods. In addition to keeping the data contents secret, the study of *information hiding* tries to further hide the fact that some data are being transmitted. Traditional methods of information hiding have largely focused on text and image data. Nowadays, with the widespread utilization of voice-based applications and the growing reliance on voice communication in various domains, innovative approaches are proposed to utilize audio data as the carrier media.

This study delves into the innovative fusion of the Internet Low Bitrate Codec (iLBC) [1] with advanced information hiding techniques, aiming to strengthen voice communication channels against various security threats. The open source codec iLBC, which is designed for efficient speech transmission in bandwidth-constrained environments, provides an ideal platform for embedding covert information within voice signals. By utilizing on iLBC's unique characteristics, this study endeavors to propose an innovative method for embedding covert information within voice data streams.

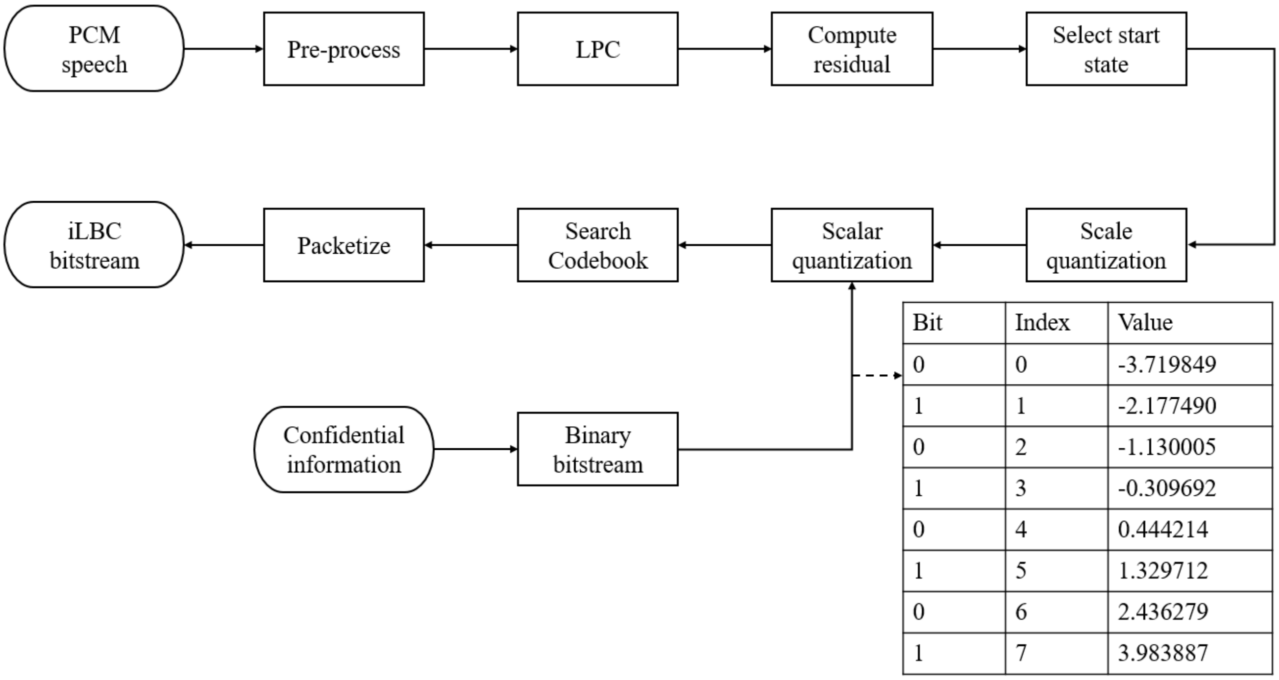
One of the primary challenges is optimizing the limited transmission capacity while maintaining voice quality. Additionally, synchronizing the embedding and extracting processes without introducing noticeable latency presents a significant obstacle.

Our primary consideration revolves around optimizing transmission capacity while maintaining high voice quality. By exploring the convergence of advanced information hiding methodologies with the unique characteristics of iLBC, this study is dedicated to developing practical solutions to overcome the challenges faced in voice-based information hiding.

## II. Related Work

According to the diverse embedding areas within the field of covert communication, methods of steganography based on Voice over Internet Protocol (VoIP) can be categorized into two types [2], namely, voice payload steganography and protocol steganography.

In [3], the adaptive codebook from iLBC was effectively employed and integrated with the QIM (Quantization Index Modulation) [4] principle to achieve information concealment. This approach exhibited impressive hiding capacity while maintaining a high level of imperceptibility. In [5], previous research based on fixed codebook method was enhanced and resulted in a significant hiding capabilities. In [6], a novel method for embedding secret information within the linear predictive coding (LPC) procedure was proposed. The process began with the creation of a mapping table based on the minimum distance between the LPC vectors before and after embedding the secret message. The codeword to be modified and embedded was determined according to the secret bits and Matrix Embedding (ME) algorithm. The proposed approach leads to a less speech distortion and great security. In [7], a steganography algorithm utilizing QIM control was proposed. The concept behind this approach involved creating a graph model of the quantizer's codebook area. This method can minimize signal distortion while steganography is being applied, so it provides enhanced security and robustness compared to traditional QIM techniques. An adaptive steganographic scheme based on the principles of AMR (Adaptive Multi-Rate) fixed codebook search was introduced [8]. Optimal pulse probabilities and pulse correlations from the same track are integrated into the cost function, enhancing the statistical security of the proposed algorithm. This approach demonstrates excellent imperceptibility and robustness against existing steganalysis algorithms. An innovative steganography approach centered around decimal pitch delay search was presented in [9]. The partial similarity between the secret message and the decimal pitch delay was computed, deciding whether to embed the confidential



| Bit | Index | Value |
|---|---|---|
| 0 | 0 | -3.719849 |
| 1 | 1 | -2.177490 |
| 0 | 2 | -1.130005 |
| 1 | 3 | -0.309692 |
| 0 | 4 | 0.444214 |
| 1 | 5 | 1.329712 |
| 0 | 6 | 2.436279 |
| 1 | 7 | 3.983887 |

Fig. 1. Embedding methods for confidential information.

data based on a predetermined decision threshold. All fractional pitch delays were treated as replaceable cover bits, which maximizes the embedding capacity.

## III. Steganography Based on iLBC

iLBC is an audio codec developed by the Internet Engineering Task Force (IETF). iLBC aims to provide high-quality speech communication even under bad network conditions, such as low bandwidth or high packet loss rate. It achieves these goals by employing a variety of techniques, including a non-uniform quantizer, frame-independent coding, and built-in error concealment methods. Because of the automatic error concealment mechanism, iLBC is optimized for applications over IP networks and is commonly used in voice over IP (VoIP) applications, video conferencing systems, and other real-time communication platforms. Overall, iLBC is widely recognized for its efficiency in handling network limitations, making it a popular choice in various internet-based communication services. In this section, we briefly describe these techniques and how they are utilized in data hiding.

### A. Residual Signal Quantization in iLBC Encoder

*1) Calculating the residual signal:* The residual signal represents the difference between the original audio frame and the predicted frame. Speech sample prediction employs LPC analysis filters specific to the corresponding sub-block to yield the residual signal.

*2) Selecting and encoding start state:* Selecting the start state in the context of iLBC is a critical step in the encoding process. It involves determining an appropriate initial state for the encoder, which is used as a reference point for predicting and encoding subsequent audio samples.

*a) Start state estimation:* The two sub-blocks containing the start state are determined by finding the two consecutive sub-blocks in the block having the highest power. To ensure efficient transmission over networks while maintaining a low bitrate, a starting state of 57 samples is selected for 20ms frames, and 58 samples are chosen for 30ms frames. (In the following, we shall use the notation “57/58 residual samples” stands for “57 residual samples for 20ms frames, and 58 residual samples for 30ms frames.”)

*b) All-Pass filtering and scale quantization:* Firstly, the residual samples in the start state is first filtered by an all-pass filter. Following the all-pass filtering, the filtered block of residual samples is carefully analyzed to identify the sample with the largest magnitude. This sample is quantized with a 6-bit quantizer to get a quantized value to yield normalized samples of the all-pass filtered residual samples in the block.

*c) Scalar quantization:* The normalized samples are now poised for perceptually weighted scalar DPCM quantization. Each normalized sample undergoes a weighted filtering operation, these weighted samples are utilized to construct the target sample, which is calculated by subtracting a predicted sample. Subsequently, the coded state sample is derived by quantizing this target sample using a 3-bit quantizer. This quantization involves referencing a specific quantization table for precision.

*3) Encoding the remaining samples:* A dynamic codebook is used to encode the 23 or 22 remaining samples in the two sub-blocks containing the start state and the sub-blocks which are not in the start state. (There are 23

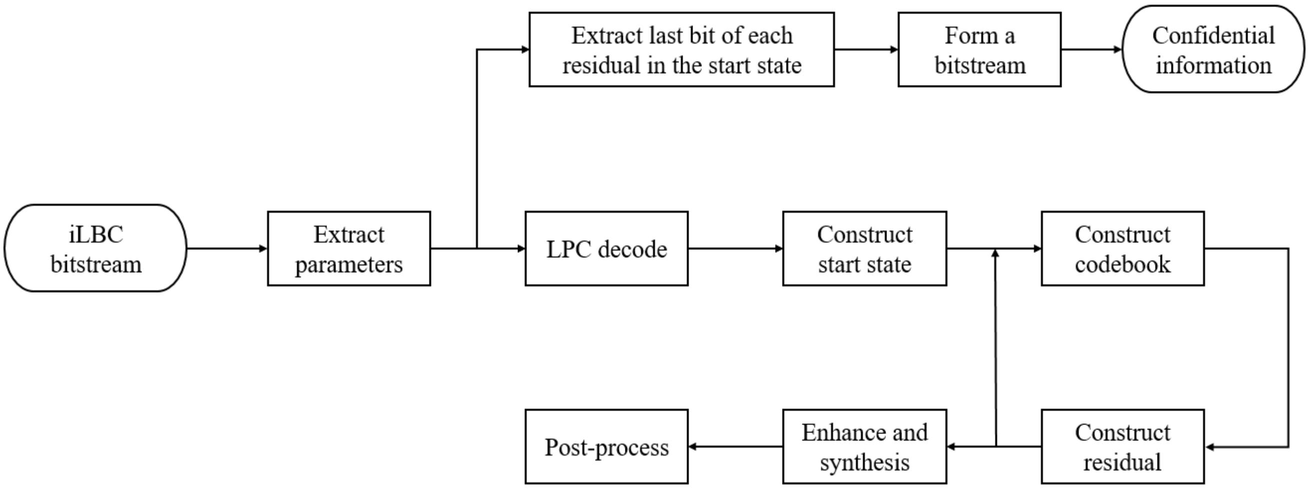


Fig. 2. Extracting methods for confidential information.

remaining samples for the 20ms frames, and 22 samples for the 30ms frames. These values are derived from 80 samples, corresponding to two sub-blocks, by subtracting 57/58 samples. As mentioned above, we shall use the notation “23/22” to represent “23 samples for 20ms frames, and 22 samples for 30ms frames.”) The coding relies on an adaptive codebook constructed from a codebook memory containing decoded LPC excitation samples obtained from the already encoded segment of the block.

### B. *Residual Signal-Based Information Hiding Method*

Focusing on the quantization process of start state residuals, we developed an information hiding method utilizing the QIM principle. By strategically selecting specific quantization table indices based on the hidden bits, our approach ensures that the values closest to the residuals are minimally altered. Leveraging this technique, we explored the inherent quantization process of iLBC, taking advantage of the fact that approximately half of the residuals remain unchanged. Additionally, with a total of 57/58 residual samples available, our method significantly amplifies the information hiding capacity.

*1) Embedding secret information:* In the quantization process of residual samples, the selection of the nearest quantized value from odd or even indices is determined by the embedded secret message bit, as depicted in Fig. 1. This selection applies to 57/58 samples chosen as the start state. The specific embedding algorithm is described as follows:

Step 1. The confidential information is converted to a binary secret message bitstream.

Step 2. During start state residual quantization, the corresponding quantized value is selected based on the current confidential information to be embedded. If the confidential information is 1, it matches with the quantized value at an odd index. If the confidential information is 0, it matches with the quantized value at an even index.

Step 3. The closest quantized value is then selected, and its index is encoded. This encodes the secret information, and appends three bits at the end of the bitstream.

Step 4. If the secret information has been completely embedded or there is no carrier left, the process terminates. Otherwise, it goes to Step 2.

*2) Extracting secret information:* The receiver receives a bitstream containing the secret information, it undergoes iLBC decoding process to extract the secret information, as shown in Fig. 2. The extraction algorithm is described as follows:

Step 1. The receiver parses the received audio packets and extracts various parameters.

Step 2. The quantized residual is analyzed to obtain the quantization index for each sample. For a 20ms frame, the obtained quantization index contains 57 bits of secret information. For a 30ms frame, the quantization index contains 58 bits of secret information.

Step 3. The parity of the index value determines whether the current secret information bit is 0 or 1. Note that since each sample's quantization index comprises three bits, the last bit represents the secret information.

Step 4. The extracted secret information is merged to form the confidential information bitstream.

## IV. Experiments and Performance Evaluation

In the context of information hiding, evaluating performance typically involves assessing both hiding capacity and imperceptibility. However, recognizing the diverse nature of secret information and its subtle influence on concealment performance, our analysis delves deeper into this complexity. We examined concealment capabilities across different confidential information, assessing how varied secret messages influence concealment efficiency. Furthermore, our study involved a correlation analysis between concealment and hiding capacity, aiming to identify the optimal balance between the two.

### A. *Hiding Capacity Analysis*

By embedding information within the residual samples of each block, our approach allows us to hide a substantial 57/58 bits of secret information within every block. The 20ms frame conceals 57 bits within a total of 192 bits,

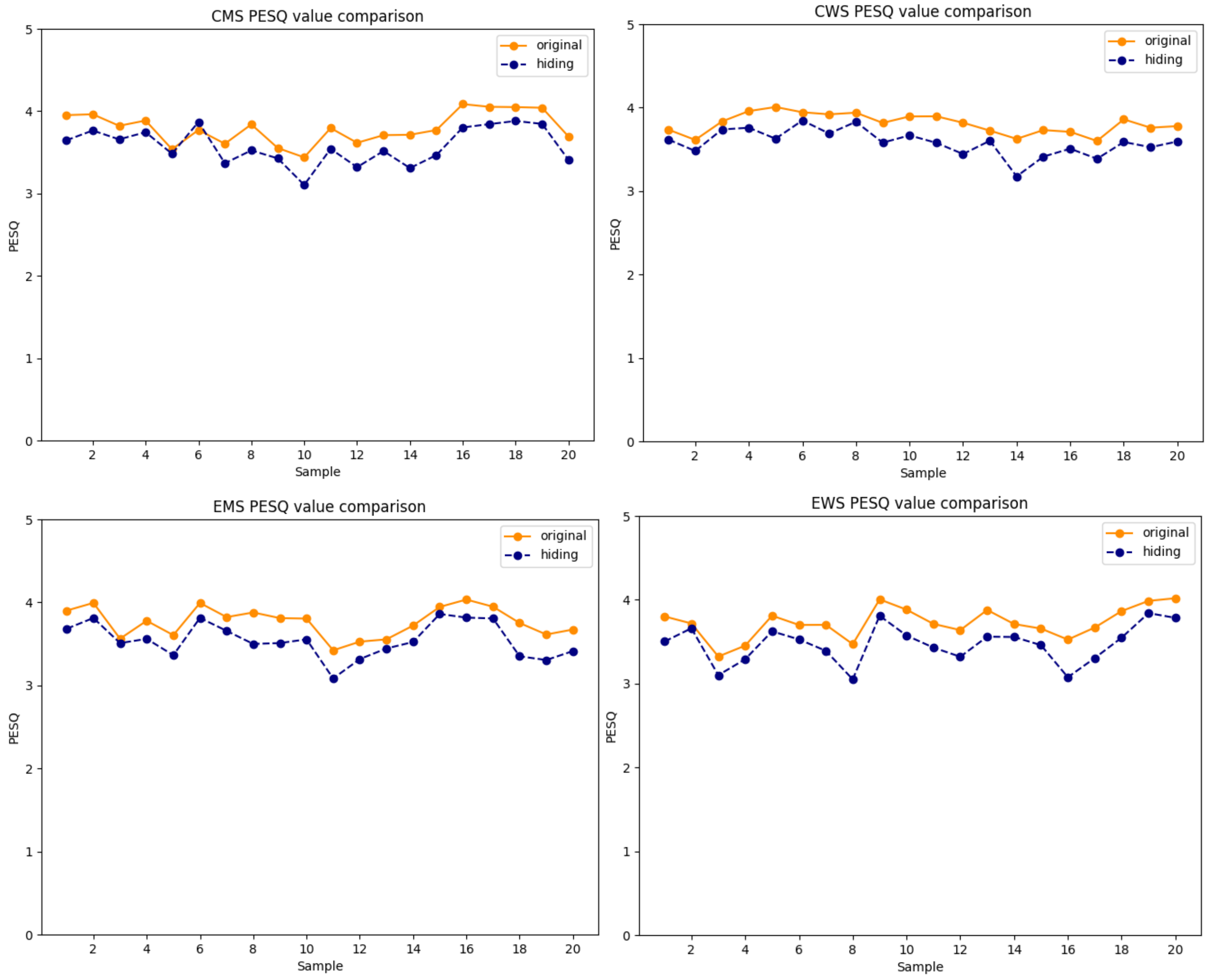


Fig. 3. PESQ values comparison under 20ms frames.

resulting in a capacity of 2850bps, as shown in formula (1). Meanwhile, the 30ms frame hides 58 bits within 240 bits, leading to a capacity of 1933bps, as shown in formula (2). In comparison to the method presented in [3], they achieved a hiding capacity of only 450bps for 20ms frames and 500bps for 30ms frames. This breakthrough result in impressive hiding capacities. Compared with previous techniques, our method significantly surpasses existing capabilities, providing a significant leap in information hiding capacity.

$$\text{Capacity} = \frac{1000}{20} * 57 \ = 2850 \tag{1}$$

$$\text{Capacity} = \frac{1000}{30} * 58 \ \approx 1933 \tag{2}$$

This substantial enhancement not only demonstrates the innovation of our approach but also indicates its potential to revolutionize secure communication methods. This definitely opens new avenues for data concealment applications.

### *B. Speech Quality and Imperceptibility Analysis*

In the realm of information hiding within speech signals, the evaluation of concealment effectiveness is crucial and revolves around comparing speech quality in scenarios with and without embedded information. To assess our hiding algorithm's concealment efficacy, we employed the Perceptual Evaluation of Speech Quality (PESQ) method.

For our experiments, we utilized the Common Voice dataset [10], ensuring a diverse and comprehensive selection of speech samples. We chose four distinct voice categories from the dataset: Chinese male, Chinese female, English male, and English female, each comprising 20 speech segments, totaling 80 speech segments. This deliberate selection process aimed to ensure linguistic diversity as well as the subtle nuances in pronunciation and intonation that are vital in the context of our study.

In our investigation, a cautious analysis was conducted to compare the quality of speech samples derived from scenarios involving information embedding and those without it. The evaluation process was carried out utilizing PESQ scores, enabling a quantitative assessment of speech quality. Throughout the experimental procedure, our

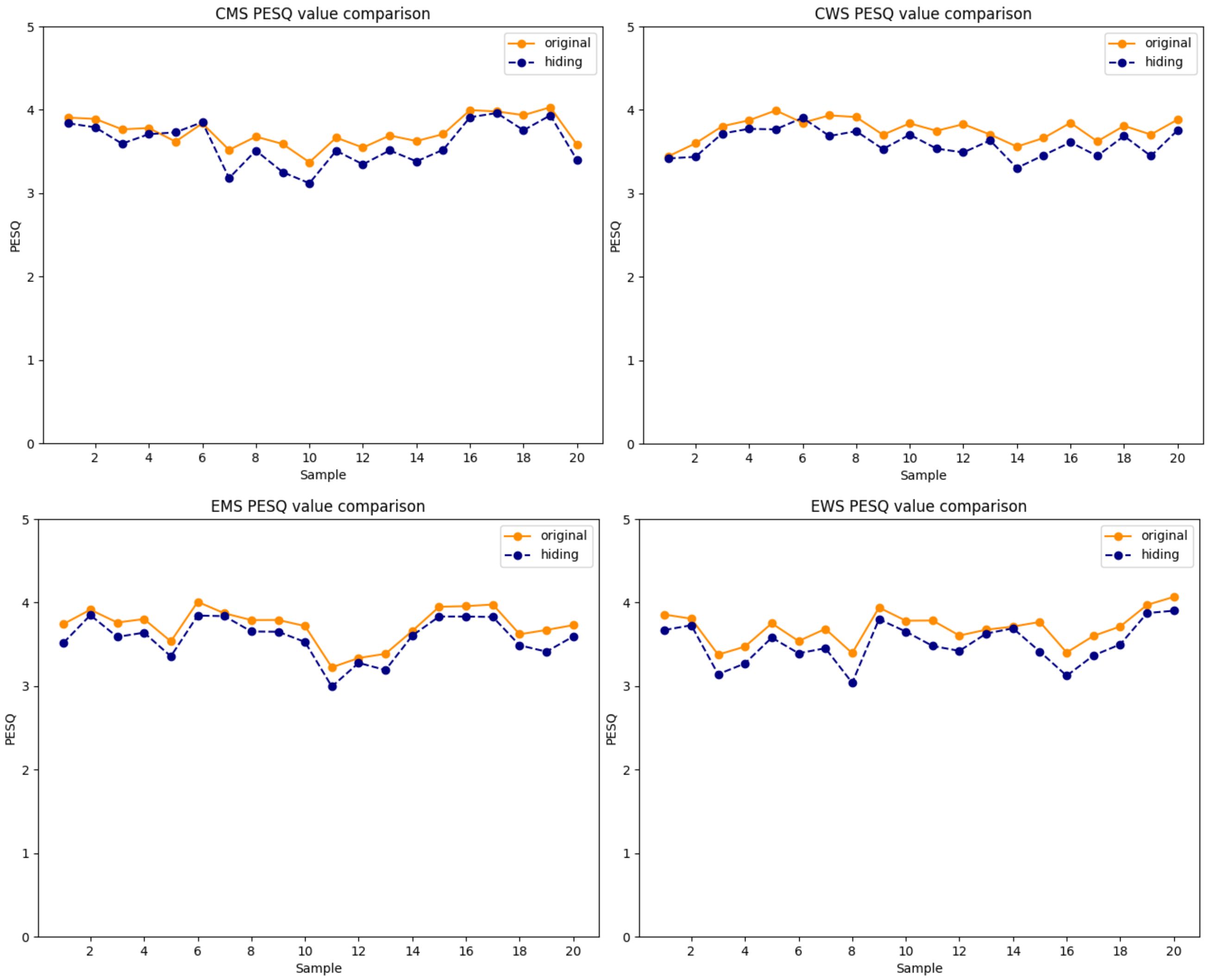


Fig. 4. PESQ values comparison under 30ms frames.

TABLE I. COMPARISON OF PESQ CHANGE, VARIANCE AND CAPACITY AMONG DIFFERENT METHODS.

| **Frame** | **20ms** | | | | | | **30ms** | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Method | ***Start State Residual*** | | | ***Dynamic Codebook [3]*** | | | ***Start State Residual*** | | | ***Dynamic Codebook [3]*** | | |
| Speech | PESQ change | variance | secret bit | PESQ change | variance | secret bit | PESQ change | variance | secret bit | PESQ change | variance | secret bit |
| CMS | 0.2185 | 0.0118 | 57 | 0.1132 | 0.0037 | 9 | 0.146 | 0.0115 | 58 | 0.1265 | 0.0031 | 15 |
| CWS | 0.2258 | 0.0098 | 57 | 0.1585 | 0.0044 | 9 | 0.1634 | 0.0079 | 58 | 0.181 | 0.0026 | 15 |
| EMS | 0.2228 | 0.008 | 57 | 0.077 | 0.0014 | 9 | 0.1463 | 0.0034 | 58 | 0.0965 | 0.0024 | 15 |
| EWS | 0.2558 | 0.0093 | 57 | 0.1653 | 0.0033 | 9 | 0.1887 | 0.008 | 58 | 0.1808 | 0.0056 | 15 |

paramount focus remained on ensuring the accuracy of information extraction.

The embedded information was deliberately randomized to simulate real-world scenarios where the concealed data could vary widely. This deliberate randomness allowed for a robust evaluation of our hiding technique, considering a spectrum of potential hidden data types and patterns.

We tested two frame structures for iLBC encoding to validate the robustness of our concealment technique across both frame lengths. Fig. 3 illustrates the concealment test results under the 20ms frames structure, showing the PESQ values of 20 speech samples with and without hidden information. Similarly, Fig. 4 delves into the concealment test outcomes for the 30ms frames structure, offering a detailed comparison of the corresponding PESQ values.

As it can be clearly observed in Fig. 4, the process of embedding secret messages in the carrier speech does not significantly alter the hidden speech, thereby preserving the quality of the audio. Interestingly, the PESQ values for the 20ms frames length show a greater decrease than those for the 30ms frames length. This indicates that the 20ms frames can accommodate more secret bits within the same timeframe, albeit at the cost of reduced imperceptibility.

Examining different hiding methods, our approach yields slightly lower PESQ values compared to the adaptive codebook method employed in [3]. However, our approach significantly enlarges the concealable capacity. In TABLE I, the four rows CMS, CWS, EMS, EWS stand for "Chinese Man Speech", "Chinese Woman Speech", "English Man Speech", "English Woman Speech", respectively. The average change of PESQ value and variance are also listed in the table. It can be seen that the average variation in PESQ values in our study is only 0.015 greater than that of the method proposed in [3] under 30ms frames. Nevertheless, our method enhances the capacity four-fold for 30ms frames, and six-fold for 20ms frames. This underscores that our method, utilizing start state residuals, achieves a significantly higher hiding capacity with nearly the same PESQ values, surpassing the technique proposed in [3].

Considering the current widespread usage of Secure Real-time Transport Protocol (SRTP), the encryption of information renders it remarkably resistant to easy decryption attempts. Thus, the subtle fluctuations observed in PESQ scores might not raise immediate suspicions. In situations where enhanced imperceptibility is deemed necessary, employing additional encryption methods can be a viable strategy. By encrypting the bitstream of the secret information with pre-negotiated keys, the concealment effect can be substantially supported.

Furthermore, our approach utilizes residual samples in the start state, which possesses distinct advantages. With the ability to conceal multiple bits within these residual samples, our technique offers an expansive and versatile concealment capacity. This remarkable flexibility not only broadens the spectrum of information that can be hidden but also enhances the imperceptibility of the concealed data, as we will elaborate later in this section.

### C. *Impact of Different Secret Information*

The current landscape of information hiding research often lacks specific indications about the nature of the concealed secret messages. However, methods that employ techniques like parity values for hiding, the bitstream hidden within, can significantly impact concealment. In our upcoming analysis, we will explore the influence of the concealed information on PESQ values. Five sets of random bitstreams will be adopted to assess the stability of this information hiding method. These random sets will serve as benchmarks, representing typical usage scenarios. By subjecting our technique to diverse and unpredictable bitstreams, we aim to ensure its adaptability and robustness under different conditions in real-world applications. Additionally, we intend to calculate the worst-case scenario, where all the quantization indices of the residual samples differ from the secret message bitstream. In this situation, 57 samples will be modified in 20ms frames, and 58 samples will be modified in 30ms frames. While this extreme situation is highly unlikely, considering such outliers is essential for evaluating the method's performance boundaries.

This detailed investigation not only showcases the resilience of our information hiding technique but also highlights the challenges posed by worst-case scenarios. Understanding both the strengths and potential vulnerabilities in various contexts is vital. It allows us to adapt and refine our approach, ensuring it remains robust under diverse conditions. This adaptive approach in research is essential as it enables us to address real-world challenges effectively.

The data presented in TABLE II and TABLE III provide valuable insights into the stability of PESQ scores under various conditions. In general, we observe a consistent average change in PESQ scores, with slightly larger variations in the 20ms frames compared to the 30ms

TABLE II. COMPARISON OF PESQ CHANGE, VARIANCE AMONG DIFFERENT SECRET INFORMATION UNDER 20MS FRAMES.

| **Frame** | **20ms** | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Secret<br>Speech | ***Random Set 1*** | | ***Random Set 2*** | | ***Random Set 3*** | | ***Random Set 4*** | | ***Random Set 5*** | | ***Worst Case*** | |
| | PESQ change | variance | PESQ change | variance | PESQ change | variance | PESQ change | variance | PESQ change | variance | PESQ change | variance |
| CMS | 0.2225 | 0.0078 | 0.1832 | 0.008 | 0.2005 | 0.0075 | 0.2229 | 0.0074 | 0.2083 | 0.0089 | 0.4557 | 0.0199 |
| CWS | 0.2197 | 0.0102 | 0.2211 | 0.013 | 0.2198 | 0.0109 | 0.2242 | 0.009 | 0.2336 | 0.0129 | 0.4758 | 0.0191 |
| EMS | 0.2272 | 0.0085 | 0.2116 | 0.0052 | 0.2249 | 0.0087 | 0.2203 | 0.0094 | 0.2168 | 0.0083 | 0.4565 | 0.0134 |
| EWS | 0.2682 | 0.0081 | 0.2569 | 0.0097 | 0.2628 | 0.0098 | 0.2795 | 0.0103 | 0.2782 | 0.0104 | 0.5697 | 0.018 |

TABLE III. COMPARISON OF PESQ MEAN, VARIANCE AMONG DIFFERENT SECRET INFORMATION UNDER 30MS FRAMES.

| **Frame** | **30ms** | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Secret<br>Speech | ***Random Set 1*** | | ***Random Set 2*** | | ***Random Set 3*** | | ***Random Set 4*** | | ***Random Set 5*** | | ***Worst Case*** | |
| | PESQ change | variance | PESQ change | variance | PESQ change | variance | PESQ change | variance | PESQ change | variance | PESQ change | variance |
| CMS | 0.1435 | 0.0092 | 0.1431 | 0.0062 | 0.128 | 0.0068 | 0.1387 | 0.0073 | 0.1254 | 0.0083 | 0.2857 | 0.0129 |
| CWS | 0.1577 | 0.0108 | 0.1621 | 0.0061 | 0.1646 | 0.0083 | 0.1591 | 0.0095 | 0.1643 | 0.0121 | 0.3554 | 0.0213 |
| EMS | 0.1526 | 0.0048 | 0.1472 | 0.0036 | 0.1464 | 0.0034 | 0.145 | 0.0051 | 0.1381 | 0.0082 | 0.3185 | 0.0085 |
| EWS | 0.1931 | 0.0042 | 0.1869 | 0.0052 | 0.1884 | 0.0076 | 0.1964 | 0.007 | 0.1907 | 0.0055 | 0.406 | 0.0157 |

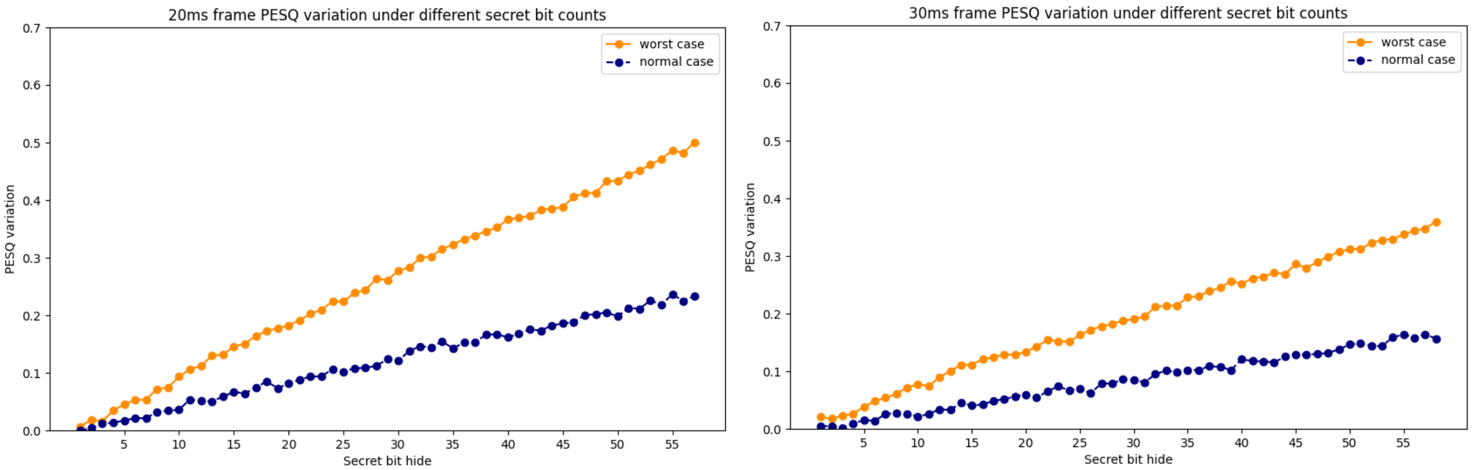


Fig. 5. PESQ values comparison under 30ms frames

frames. This slight compromise in PESQ scores is reasonable, considering the higher transmission rate (bps) achieved. Notably, across different secret messages, subtle variations appear in the four sets of speech samples. However, these variations fall within acceptable ranges, aligning with our initial expectations.

However, when the worst-case scenario expands, the average change in PESQ scores sharply rises. This degree of variation is unacceptable, especially in the case of the 20ms frame, reaching nearly 0.57. Such fluctuations can significantly impact speech quality. It is important to highlight that while these worst-case outcomes are discouraging, they are highly improbable in practical applications.

This analysis prompts an important question about striking the right balance between imperceptibility and capacity. Finding this delicate balance is a challenge that many researchers are struggling with. In the upcoming sub-section, we will delve deeper into this topic, exploring potential strategies to achieve optimal trade-offs between imperceptibility and capacity. This exploration is essential for refining our techniques, ensuring they not only meet theoretical expectations but also demonstrate robustness in practical contexts.

### D. Balance between Capacity and Imperceptibility

Fig. 5 provides a visual representation of the PESQ variations concerning different hidden bit capacities. The secret messages are categorized into worst-case and normal-case scenarios, with PESQ values representing the averages across four sets of speech samples. What stands out is that the PESQ variation in the worst-case scenario is twice as much as that in the normal-case scenario. This pattern aligns with the nature of odd-even information hiding, where half of the secret message has a probability of not modifying the residual sample quantization results.

This visualization serves as a powerful tool, allowing us to strategically fine-tune information hiding mechanisms based on our priorities. If capacity is the main goal, increasing the hidden bits would be the approach. Conversely, if imperceptibility is of utmost importance, reducing the hidden bits can be the strategy. This flexibility equips us with the ability to tailor our information hiding techniques according to specific contexts and needs.

This adaptability is crucial in the realm of information security. It empowers us to make informed decisions, striking the ideal balance between capacity and imperceptibility based on the unique demands of each situation.

## V. Conclusions

In conclusion, our research delved deep into the realm of voice information hiding techniques based on an Internet open source codec iLBC. Utilizing the residual from its start state as the hiding field, we significantly enhanced the hiding capacity four-fold for 30ms frames, while maintaining a high level of PESQ. What sets our study apart is the careful exploration of the worst-case scenarios in odd-even information hiding, a crucial aspect often overlooked by many studies in the field.

One of the pivotal aspects of our study was striking the delicate balance between capacity and imperceptibility. The art of concealing information lies in finding this balance, and our approach demonstrated the ease with which adjustments can be made for varying circumstances. Moreover, our emphasis on utilizing the residual from the start state showcases the potential for innovation within existing codecs like iLBC. By leveraging overlooked aspects, we can often enhance existing techniques as well as paving the way for novel approaches in the realm of information security.

As technology advances, the need for robust information hiding techniques becomes increasingly paramount. Our study not only enriches the existing knowledge base but also charts a course for future exploration. The insights gained and methodologies developed in this research serve as a robust foundation for upcoming advancements, propelling the field of voice information hiding toward new horizons. By embracing adaptability and innovation, we are better equipped to navigate the complexities of information security, ensuring a robust and secure digital landscape for all.

# References

[1] S. Andersen, A. Duric, H. Astrom, R. Hagen, W. Kleijn, and J. Linden, “Internet Low Bit Rate Codec (iLBC),” IETF RFC 3951, 2004.

[2] Z. Wu, J. Guo, C. Zhang, and C. Li, “Steganography and Steganalysis in Voice over IP: A Review,” Sensors, vol. 21, no. 4, p. 1032, 2021.

[3] Y. Huang, W. Yang, and D. Sun, “Steganographic method in self-adaptive codebooks of speech codec,” Computer Engineering and Design, 2013.

[4] B. Chen and G. W. Wornell, "Quantization index modulation: a class of provably good methods for digital watermarking and information embedding," in IEEE Transactions on Information Theory, vol. 47, no. 4, pp. 1423-1443, 2001.

[5] Z. Wu and Y. Sha, "An implementation of speech steganography for iLBC by using fixed codebook," 2016 2nd IEEE International Conference on Computer and Communications (ICCC), 2016.

[6] P. Liu, S. Li, and H. Wang, “Steganography integrated into linear predictive coding for low bit-rate speech codec,” Multimedia Tools and Applications 76, pp. 2837–2859, 2017.

[7] Y. Huang, H. Tao, B. Xiao, and C. Chang, “Steganography in low bit-rate speech streams based on quantization index modulation controlled by keys,” Sci. China Technol. Sci. 60, pp. 1585–1596, 2017.

[8] Y. Ren, H. Wu, and L. Wang, “An AMR adaptive steganography algorithm based on minimizing distortion,” Multimedia Tools and Applications 77, pp. 12095–12110, 2018.

[9] X. Liu, H. Tian, Y. Huang, and J. Lu, “A novel steganographic method for algebraic-code-excited-linear-prediction speech streams based on fractional pitch delay search,” Multimedia Tools and Applications 78, pp. 8447–8461, 2019.

[10] R. Ardila, M. Branson, K. Davis, et al., “Common Voice: A Massively-Multilingual Speech Corpus,” 2019, European Language Resources Association, https://doi.org/10.48550/arXiv.1912.06670