symmetry

MDPI

Article

Secure Speech Content Based on Scrambling and Adaptive Hiding

Dora M. Ballesteros †,‡ **and Diego Renza** *,†,‡

Telecommunications Engineering, Universidad Militar Nueva Granada, Bogotá 11011, Colombia; dora.ballesteros@unimilitar.edu.co

* Correspondence: diego.renza@unimilitar.edu.co; Tel.: +57-1-650-0000

† Carrera 11 No. 101-80, Bogotá, Colombia.

‡ These authors contributed equally to this work.

Received: 26 October 2018; Accepted: 21 November 2018; Published: 3 December 2018

Abstract: This paper presents a method for speech steganography using two levels of security: The first one related to the scrambling process, the second one related to the hiding process. The scrambling block uses a technique based on the ability of adaptation of speech signals to super-Gaussian signals. The security of this block relies on the value of the seed for generating the super-Gaussian signal. Once the speech signal has been scrambled, this is hidden in a non-sensitive speech signal. The hiding process is adaptive and controlled by the value of bits to hold (BH). Several tests were performed in order to quantify the influence of BH in the quality of the stego signal and the recovered message. When BH is equal to six, symmetry was found between the modified bits and unchanged bits, and therefore hiding capacity is 50%. In that case, the quality of the stego signal is 99.2% and of the recovered signal is 97.4%. On the other hand, it is concluded that without knowledge of the seed an intruder cannot reverse the scrambling process because all values of the seed are likely. With the above results, it can be affirmed that the proposed algorithm symmetrically considers both the quality of the signal (stego and recovered) as well as the hiding capacity, with a very large value of the key space.

Keywords: steganography; scrambling; hiding; covert communication

1. Introduction

In the last decades, technological development has highly facilitated communication processes between users, allowing a free flow of information of any kind (i.e., text, image, audio, video, etc.). However, in many cases, confidentiality of information can be a major factor to satisfy. In the case of audio signals, with the purpose of preserving confidentiality, there are three solutions available for covert communication: (i) To encrypt transmitted information, (ii) to manipulate or alter content (scrambling), (iii) to transmit secret information within another signal. In the first two cases, a third unauthorised party can infer that sensitive information is being transmitted, but is ignorant of its content. In the third case, even if the third party intercepts the information being transmitted, the secret content goes unnoticed. Consequently, if the voice signal is transformed through scrambling or encryption, and the result hides in a host signal, the final system will work with at least two levels of security. In the specific case of techniques applied to voice messages, there are works in the area of data hiding (specifically steganography) and in the area of data randomisation (scrambling). In both scenarios, the process can use a secret key to increase system security. Only the authorised user who has the correct secret key will be able to reverse the concealment process and to obtain the original voice message.

Techniques of data concealment generally consist in inserting the information to be transmitted within some other content. To do so, various multimedia signals such as audio [1], text, video [2] or image [3,4] signals, can act as a host that carries secret information. The expression "hiding" can be interpreted as secretly keeping the existence of information (steganography) or making it imperceptible (watermarking). With regard to steganography, its objective is to make the secret information invisible by hiding it, whereas watermarking is based on the necessity to protect the content's author copyright. In steganography, the most important design criteria are transparency and imperceptibility; in watermarking, robustness is the main criterion to satisfy.

The state-of-the-art of steganography techniques include temporal domain [5], frequency [6,7] or time-frequency [6,8] that are based on masking [8–10] or substitution. However, most schemes work with low hiding rates, thus the duration of the secret message is very short. In scrambling, there are techniques based on permutation in the time domain [11,12] or frequency [13,14] domain. The main weaknesses of these kind of methods are related to residual intelligibility and key robustness against attacks. Additionally, many schemes do not guarantee unconditional security, it means that with enough resources (i.e., time, hardware), the key is decipherable.

Covert communication techniques based on Artificial Intelligence (AI) techniques that satisfy the unconditional security encompass, among others, cellular automata (CA) [15,16], genetic algorithms (GA) [17], and imitation [12]. In the case of the cellular automata technique [10], a CA is used to generate a key as long as the secret message, so that mapping is one to one. Every time the system is executed, a different key is generated, and all keys are equally probable; as a result, if an intruder intercepts the hidden message, he/she has no certainty of which key has been selected, and in turn, he/she cannot decipher the secret voice message it contains. The disadvantage of the previous proposal lies in the system not guaranteeing low residual intelligibility; in other words, for certain keys some remnants of the original voice message may remain in the concealed voice message. In the group of genetic algorithms, the GA is used to select the less significant bits (LSB) to hide the encrypted content [17]. The disadvantage is the sensitivity to the cost function. A third type of method that satisfies the principle of unconditional security is the one proposed by Ballesteros and Moreno [8], which is based on the ability of imitation of speech signals. The secret message imitates a voice message of non-sensitive content. The main disadvantage of this method consists in the necessity to have a database of non-sensitive content messages. As an enhancement to this limit, authors of [12] recently proposed an adaptation between a (secret) voice signal and a super-Gaussian signal, which is generated in situ in the randomization system. In this way, it is not necessary to have a database to create the signal to be imitated. As a limitation, this method only contemplates scrambling but not steganography.

According to the comments above, even though there are works in literature framed within hidden communication of speech signals, to date, as far as it is known, there is not an available solution that combines the advantages of the existing methods and overcomes the disadvantages.

The characteristics of our proposal are:

1. The proposed system combines scrambling and steganography.
2. The scrambling block allows a voice signal to imitate a super-Gaussian noise. The output signal (scrambled) is highly similar to the super-Gaussian noise. The key that permits descrambling the signal is a system out.
3. The hiding of the scrambled signal within the host signal is done through an adaptive substitution technique, which allows control of the amount of signal distortion through the parameter bits to hold (BH).
4. The final output signal (stego) is highly similar to the host signal and does not generate suspicion of the existence of a secret message within it. Without knowledge of the seed used to generate the super-Gaussian signal, as well as the value of BIH, a non-authorised user can not reveal the secret content.

2. Imitation Property of Speech Signals

The imitation property of speech signals was presented by first time in 2012 by Ballesteros and Moreno [8], with the hypothesis that any speech signal may seem similar to another speech signal if its wavelet coefficients are sorted. In the second approach with this property (2014), the ability of adaptation was used as a scrambling mechanism for obtaining private messages [11]. The authors analysed that the obtained scrambled signal is unconditionally secure in terms of Shannon's theory. A generalization of the imitation property of speech signals was proposed by Ballesteros, Renza and Camacho in 2016, with the hypothesis that a speech signal with intelligible content can imitate a Gaussian noise signal if the entropy of both signals is similar [12]. Unlike the first hypothesis of imitation, it can be applied in the time domain and, on the other hand, it is feasible that a speech signal imitates not only another speech signal but also a Gaussian noise signal.

3. Proposed Scheme

The proposed scheme for covert communication of a speech signal using scrambling encompasses a hiding module to conceal voice content (confidential) within other content (non-confidential), and a recovery module for the extraction of the secret content. Each of these modules is described in the following.

3.1. Hiding Module

The hiding module carries out two fundamental operations: scrambling of the input samples and hiding data in a host signal. To implement correctly these two processes, the next blocks are used: scrambling, 8-bit conversion, 16-bit conversion, adaptive LSB, Binary (Bin) to Decimal (Dec) conversion (Figure 1). Inputs of this module correspond to the secret signal (voice), host signal (voice), randomization key and *BH* (Bits to Hold). Its output corresponds to the stego signal (voice).

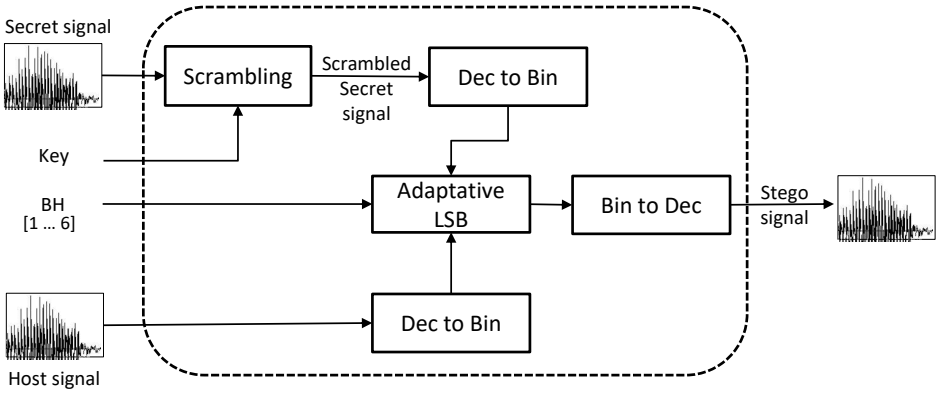

Figure 1. Block diagram of the concealing module. Bits to Hold (BH), decimal (Dec), binary (Bin), less significant bits (LSB).

Next, each block of the transmitter module will be explained.

3.1.1. Scrambling

The objective of this block consists in scrambling the secret content voice signal. To do so, the method proposed in [12] is used, in which a voice signal imitates a super-Gaussian noise signal with similar statistics. The result (scrambled signal), looks like a noise signal. From the adaptive process emerges a key, which allows mapping the scrambled signal positions with the positions of the original noise signal data. This key has the same length as the voice signal.

The first sub-block consists on generating a super Gaussian signal (i.e., a signal with kurtosis higher than three, that looks like a Gaussian signal) with the same (or very similar) statistics and length of the speech signal.

Firstly, a uniform signal is generated, according to:

$$u = rand(L, 1) - 0.5 \tag{1}$$

where u is a uniform signal generated in situ, from a random generator which depends on the seed. The length of the uniform signal is the same of that of the speech signal.

With the above result, a super Gaussian signal, g, is computed, according to (2) and (3).

$$b = \frac{\sigma}{\sqrt{2}} \tag{2}$$

$$g = \begin{cases} \mu - b * \log(1 - 2 * |\mu|) & if \quad u \geq 0 \\ \mu + b * \log(1 - 2 * |\mu|) & if \quad u < 0 \end{cases} \tag{3}$$

Values of μ and σ are 0 and 0.2, respectively, which can be fixed in the system. They are selected according to several tests with the purpose of work with similar statistics of the speech signals. As results, the super-Gaussian signal has similar entropy than speech signals, and therefore is adequate to be imitated. It is worth noting that the super-Gaussian signal is changed every time for every value of the seed.

3.1.2. Binary Representation (Dec to Bin)

The purpose of this block consists in representing in binary format the samples of the scrambled secret signal (for instance 8 or 16 bits). These bits will hide in the host signal through an adaptive LSB scheme. The output of this block is a chain of binary values of length equal to the total samples of the input signal multiplied by its resolution in bits.

3.1.3. Adaptive LSB

This block carries out the hiding process of the binary information derived from the scrambled secret signal within the binary information of the host signal. In contrast to the LSB classic method, a fixed number of the least significant bits from the host signal are not modified, but the total of modified bits depends directly on the sample range and the BH input parameter. The higher the sample value, the higher the number of bits that can be modified. In this case, BH bits numbered from the most significant bit "1" from the input data are preserved intact, and only the remaining less significant bits (LSB) are modified.

Broadly speaking, this block carries out the next steps:

1. Reading the number of bits to hold (BH).
2. Representing the samples of the host signal in 16-bit format. The resulting signal is called i.
3. Determining the minimum quantity of necessary bits to represent the sample of the host signal. This value is assigned to the MSB (most significant bit) variable.
4. Converting to zero the less significant ($MSB - BH$) bits from H (Equation (4)). The resulting signal will be called $\widetilde{H}$.

$$\widetilde{H} = 2^{MSB-BH} * \left\lfloor \frac{H}{2^{MSB-BH}} \right\rfloor \tag{4}$$

where BH is the "Bits to Hold" value, H is the sample of the host signal in 16-bits format, $\widetilde{H}$ corresponds to the modified H sample, MSB is the minimum quantity of bits to represent H, and $\lfloor \ \rfloor$ is the integer part operator.

5. Reading the chain of binary values derived from the Binary Representation block. Selecting a quantity of bits equal to $(MSB - BH)$.
6. Converting the previous binary value to decimal and adding the result to $\widetilde{H}$. The result is called S.
7. Repeating the previous steps until hiding all the bits of the binary chain in the host signal.

In this way, the values of the stego signal are obtained when carrying out an addition operation between $\widetilde{H}$ and the decimal value of the $(MSB - BH)$ bits taken from the chain obtained in the "Binary Representation" (Step 6 of the operation).

3.1.4. Decimal-Binary Conversion

Hereafter, each value of S is normalised so that the peak-to-peak range is the same as the original host signal (i.e., $[-1\ 1]$ V). The result of this normalisation corresponds to the stego signal, in the time domain.

Due to the low modification percentage of the original signal, the stego signal is perceptually equal to the original host signal. It must keep the content of the host signal with the least possible distortion (that is, with a low decrease of the signal quality or with a low SNR (Signal-to-Noise Ratio) value). Distortion is strictly related to the BH parameter. The lower the BH value, the higher the distortion. In the results section, the relation Distortion versus BH will be analysed.

3.2. Recovery Module

The objective of this module is to extract the secret voice message, by means of the following blocks: 16-bit conversion, adaptive LSB extraction, bit-to-sample conversion, and descrambling (Figure 2).

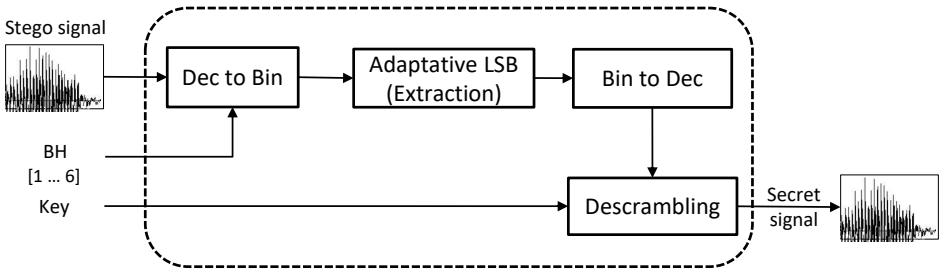

Figure 2. Block diagram of the recovery module.

Below, each block from the receptor module is explained.

3.2.1. Adaptive LSB Extraction

The bits corresponding to the secret message are extracted in this block. Due to the fact that the number of bits that are concealed in the hiding process may vary for each sample of the host signal, it is necessary to extract all the hidden bits first, and then separate and convert them into samples.

The procedure within this block is explained as follows:

1. Reading the stego signal and the BH value.
2. Representing the samples of the stego signal in a 16-bit format. The resulting signal will be referred to as S.
3. Extracting the least significant bits from every sample (Equation (5)).

$$T = S - \left\{ 2^{(MSB-BH)} * \left\lfloor \frac{S}{2^{(MSB-BH)}} \right\rfloor \right\} \tag{5}$$

where T corresponds to the decimal value of extracted bits, which is later converted into a binary value.

Equation (5) is applied on all the samples of the stego signal, creating a bit sequence.

3.2.2. Bit to Sample Conversion

Once all bits are extracted from the stego signal, the following step consists in converting them into samples. Firstly, the number of bits from the secret message in the host signal must be determined. This value is calculated by means of Equation (6).

$$Nbits = ms * N \tag{6}$$

where *Nbits* corresponds to the total number of bits from the secret message hidden in the host signal, ms is the number of samples of the hidden message, and N refers to the number of bits per sample. The value of ms is contained within the key sent by means of an alternative channel.

The process that is carried out in this module is the following one:

1. Calculating the value of *Nbits* by using Equation (6).
2. Taking the first *Nbits* from the binary sequence obtained in the "adaptive LSB extraction" block.
3. Dividing the previous result in groups of non-overlapped *Nbits*.
4. Converting each group of Nbits into decimal values (thus obtaining a sample).
5. Normalisation is applied to each sample, in order for the dynamic range to be the same as the one from the secret message (e.g., [−1 1] V).

The output of this block does not correspond to the recovered secret message yet, since the samples were saved in a scrambled way in the transmission module. In other words, only if the receptor has the scrambling key, he/she will be able to recover the secret message in a proper way. This process is performed in the following step.

3.2.3. Descrambling

The purpose of this block consists in organising the samples obtained from the previous block, according to the positions comprised in the key. Without the information from the key, it is computationally and statistically impossible to determine the secret content in a proper way. By computationally, it is meant that the time to test all options takes several years, given the fact that, thanks to the features of the key, the total number of possible solutions is m factorial, where m is the total number of voice signal samples. This way, if m were equal to 8000 (e.g., a one-second signal with fs = 8 kHz), the total number of possible solutions would be 8000 factorial, a value that greatly exceeds 105,000. Statistically means that if all possible keys that can be used are similarly probable and there is not any preference towards a subset out of them, uncertainty about which of the possible keys is the right one is complete.

In this context, when applying the correct key, a voice signal with (perceptually) intelligible content equal to the original secret message is obtained. Eventual differences between these two signals comply with the normalisation processes of dynamic ranges in the transmission and reception modules.

4. Method Implementation and Validation

This section shows the results of the validation process of the proposed method. The results are analysed in terms of stego signal imperceptibility, recovered message quality, and influence of the *BH* value on the stego signal imperceptibility and quality.

4.1. Testing Protocol

Tests were performed under the following conditions:

- Ten 2-s host signals
- Ten 1-s secret messages

- A sampling frequency of 8 kHz for all signals
- BH is an integer within the range [1 6]
- Each secret message was hidden in each host signal, with different BH values. The total number of tests was 600 (number of host signals × number of secret messages × amount of BH values)
- The secret messages, host signals, stego signals, and recovered signals with correct and incorrect key are published in [18], "Speech steganography based on scrambling and hiding (Spanish audios)", Mendeley Data, v1.

4.2. Preliminary Results

Prior to the analysis of consolidated results, some results of stego signals are shown for different BH values. Figure 3 shows a host signal, a secret signal, and the six stego signals obtained for the six BH values analysed.

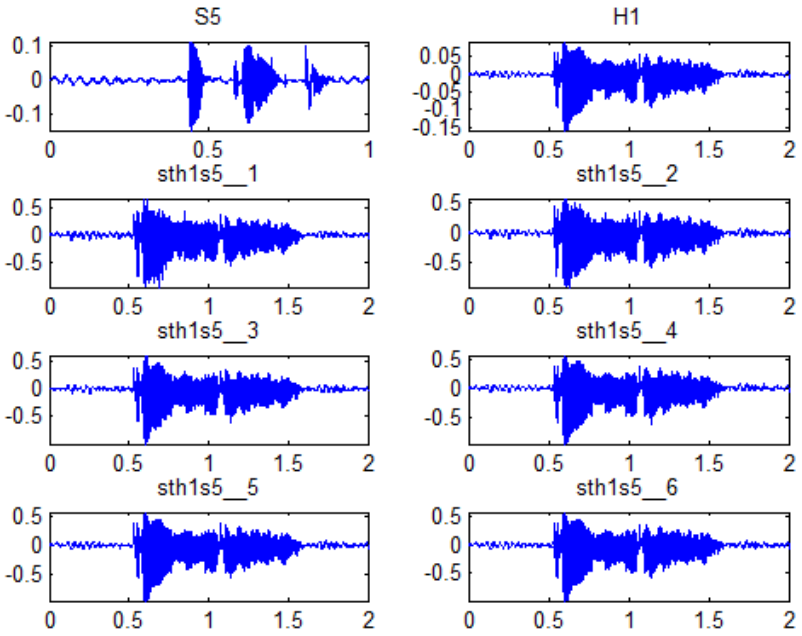

Figure 3. Example signals: Host, secret and stego signals. Source: [18].

When analysing the graphs in Figure 3, it is noted that distortion of stego signals is minimum in relation to the host signal, and the higher the BH value, the lower the localised distortion.

4.3. Imperceptibility

The objective here is to evaluate the similarity between the stego and host signals in a quantitative way. The higher the similarity, the higher imperceptibility is, a desired feature in steganography methodologies [19].

Consolidated results from the 600 imperceptibility tests are shown in Figures 4 and 5. The measurement parameter is PCC (Pearson Correlation Coefficient) between the stego and host signals, where the ideal value is 1 and the null correlation value, 0. The first analysis (Figure 4) is performed by arranging the results from each host signal that was used during the process (10 groups with 60 tests each, i.e. 10 secret messages by 6 BH values). This graph shows confidence intervals, where each box gathers 95% of the results. According to them, it can be inferred that the selection of a host signal has little influence in the secret message imperceptibility. In other words, regardless of the selected host signal, a very high imperceptibility is expected, since the PCC value in the average test was higher than 0.988, and is over 0.985 in at least 95% of all the tests.

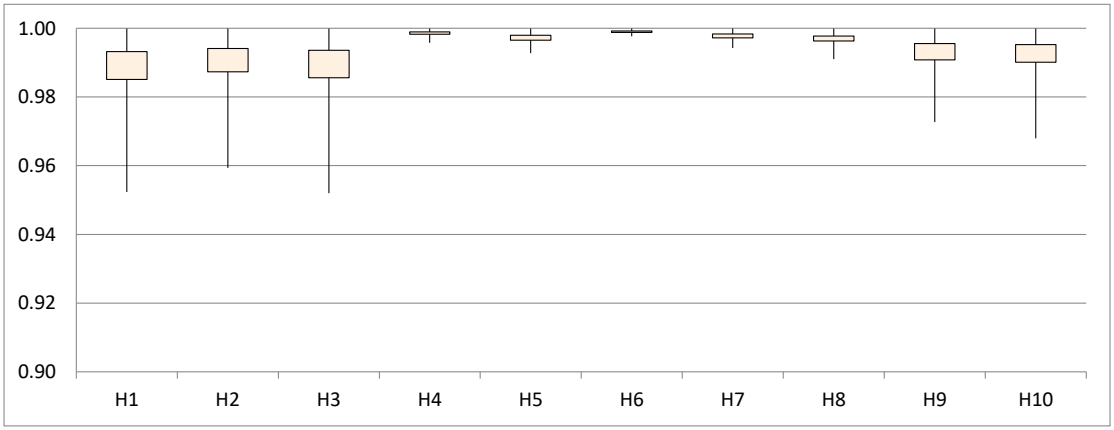

Figure 4. Imperceptibility. Correlation between host and stego signals. Results shown through a confidence interval (95%).

The second analysis involves the influence of the BH parameter on imperceptibility. In order to see this, results are grouped according to secret message and BH value. Thus, 60 groups are set (6 groups by secret message, 10 secret messages in total) as shown in Figure 5. This graph clearly shows how the imperceptibility grows as the BH value increases. The higher the BH value, the more similar the stego and host signals are in an exponential manner. For instance, an improvement in PCC between BH 1 and 2 is much higher than the improvement in PCC between BH 5 and 6.

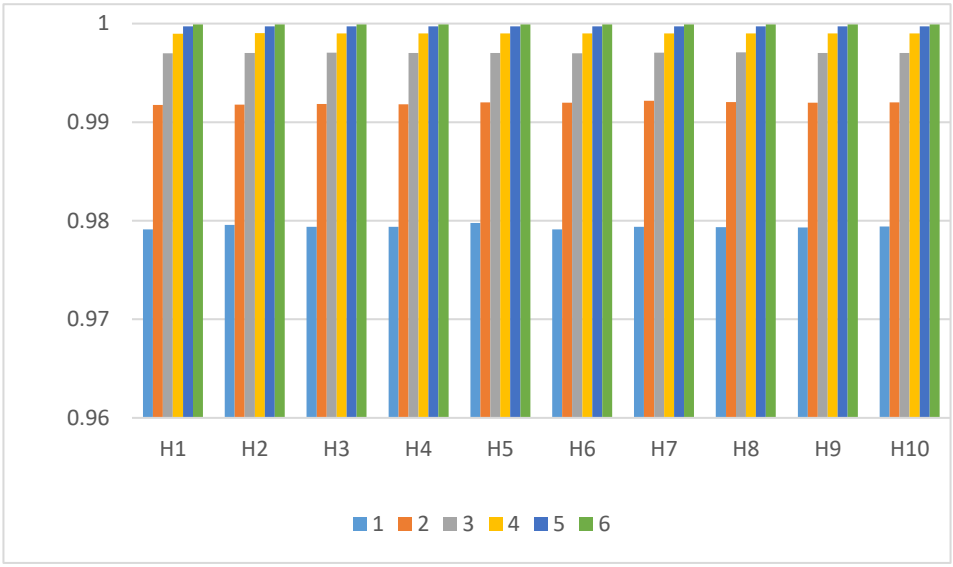

Figure 5. Average results of imperceptibility, arranged by secret message and BH. Correlation between host and stego signals.

In either case, even with a BH of 1, the similarity between both signals is very high, with a minimum PCC value of 0.975. From a BH of 3, the PCC value exceeds 0.955. Another feature that was validated through tests is that imperceptibility results are stable, meaning that the selection of the secret message does not have an influence on the quality of the stego signal. For different secret messages, the approximate same PCC values were obtained.

4.4. Quality of the Recovered Message

The second proposed evaluation criterion consists in measuring the quality of the recovered secret message. An ideal system will recover the message identically to the original file. However, due to the operations of range normalisation in the transmitter and receptor modules, there will be small

differences between both signals. The evaluation of the recovered voice message quality is carried out through the *PCC* parameter between the original secret message and the one recovered in the receptor module.

Similar to the imperceptibility evaluation, the results of the 600 tests for the quality analysis of the recovered secret message are gathered in terms of the selected host signal. Each group contains in turn the result of 60 tests, which consist of varying the secret message that is hidden (ten messages in total) and the *BH* value for each secret message (six different *BH* values). Figure 6 presents the results in confidence intervals of 95%, where it is evident that the recovered signals have soared similarities with the original secret messages, having an *PCC* value above 0.985 in all cases.

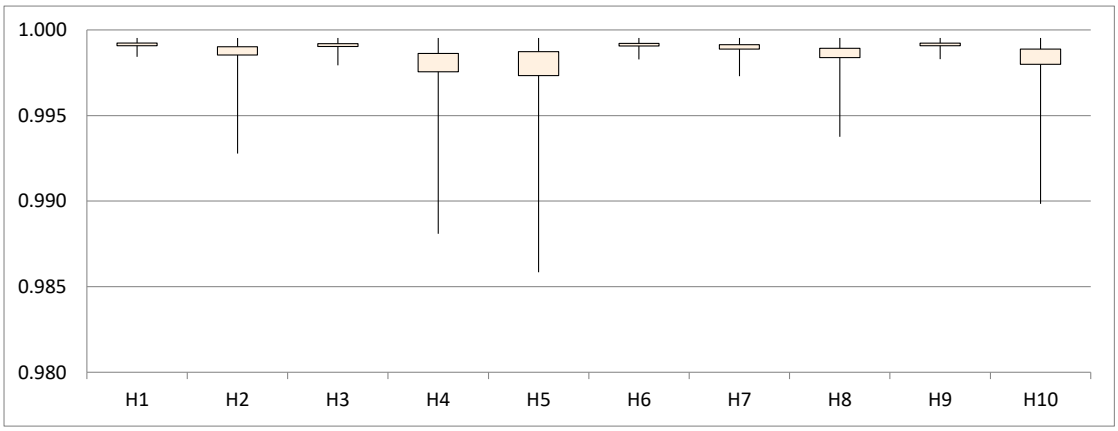

Figure 6. Recovered message quality (confidence range) results, in terms of host signal.

To corroborate the above, the recovered secret messages are shown using 10 host signals for $BH = 4$ (Figure 7).

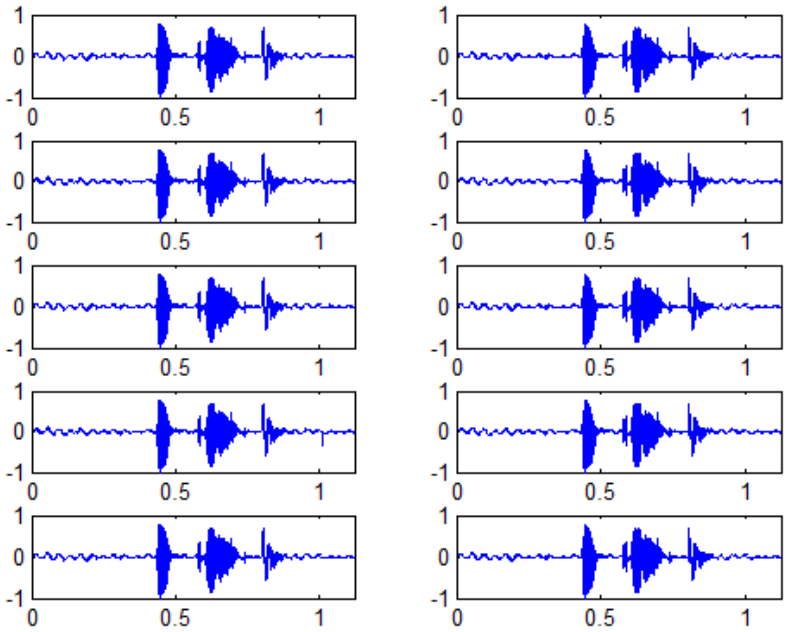

Figure 7. Secret messages recovered from the key/password, varying the host signal and with *BH* equal to 4.

According to the previous results, it is evident that the recovered signals are quite similar to the original ones (perceptually identical in sight and for the human auditory system, HAS), which confirms that the quality of the recovered message does not depend on the selected host signal ($PCC > 0.998$ with a confidence range of 95%).

4.5. Comparison with State-of-the-Art Methods

In this section, we compare the performance of our proposal with some state-of-the-art methods in the field of speech steganography. Table 1 shows the results, h being the host signal, s the stego signal, m the secret message, r the recovered message, $Corr(.,.)$ the correlation coefficient, $PSNR(.,.)$ the Peak Signal-to-Noise Ratio, $BER(.,.)$ the Bit Error Rate, and HC the hiding capacity.

Firstly, every parameter is explained, as follows:

Corr is the square root of *PCC* (Section 4.3). Unlike *PCC*, *Corr* can be positive or negative. If the value is positive it means direct correlation, otherwise, it means opposite correlation. Higher *Corr* means more similarity between the signals.

PSNR measures the difference between the host signal, h, and the stego signal, s. Higher *PSNR* means that the stego signal is more similar to the host signal.

$$PSNR = 10\log_{10}\frac{\sum_{i=1}^{N} h(i)^2}{\sum_{i=1}^{N}[h(i)-s(i)]^2} \tag{7}$$

BER calculates the ratio of the differences (bit to bit) between the message, m, and the recovered message, r. Higher *BER* means that the hiding process is reversible and better.

$$BER = \frac{1}{L}\sum_{i=1}^{L}\begin{cases}1, & m(i)\neq r(i)\\ 0, & m(i)=r(i)\end{cases} \tag{8}$$

Hiding Capacity can be calculated in many ways. For the purpose of this paper, it is the ratio of the size of the secret message by the size of the host signal. Higher *HC* means a better performance of the system in terms of its payload.

Table 1. Performance of some state-of-the-art methods (NR means Not Reported). h being the host signal, s the stego signal, m the secret message, r the recovered message, $Corr(.,.)$ the correlation coefficient, $PSNR(.,.)$ the Peak Signal-to-Noise Ratio, $BER(.,.)$ the Bit Error Rate, and HC the hiding capacity.

Reference	**[20]**	**[21]**	**[22]**	**[23]**	**Ours**
Method	**Hermite Transform (HT) and Threshold**	**RSA + 2-bit LSB**	**LSB Multi-Threshold Based Criterion**	**Variable Low Bit Coding (Up To 3-Bit LSB)**	**Adaptive Scrambling + Adaptive LSB**
$Corr(h,s)$	0.988	NR	NR	NR	0.992
$PSNR(h,s)$	32.7 dB	34.27 dB	NR	47 dB	23.61 dB
$Corr(m,r)$	0.961	NR	NR	NR	0.974
$BER(m,r)$	NR	18%	NR	24%	NR
HC	0.5	<0.125	[0.10.28]	<0.13	>0.5
Key space	NR	NR	280	NR	$L!$

According to Table 1, the greatest strength of our proposal is the simultaneous fulfilment of security requirements, the quality of the signal stego, and quality of the recovered signal. Our system provides very high security through a very large key space; high values of HC, high values of $Corr(h,s)$ as well as $Cor(m,r)$. In terms of *PSNR*, our proposal is the worst, but, it is important to remark that this parameter is not only sensitive to distortion of the signal, but to changes of the amplitude of the signal. It means that the *PSNR* between a signal and its amplified version is not infinite, although both signals have the same content. For that reason, there are more confident in the results of similarity in terms of correlation.

5. Security Analysis

In terms of security, the system must be analysed in the following aspects: exhaustive key search, ciphertext only attack and statistical attack. We assume that the attacker knows the details of the scheme and intercepts some stego signals, but not the value of the key.

5.1. Exhaustive Key Search

A good encryption scheme must overcome the brute-force attack. This implies that the total number of available keys is huge, and then, the number of attempts is great enough. In our proposal, the key is composed of two parts: the first one related to the scrambling process, the second one related to the hiding process. For the first part of the key, its length is equal to the number of samples of the secret message (L), and it corresponds to the numbers $[1\ to\ L]$ in a disordered way. Then, the total number of available keys is $L!$. For example, if the secret message has 8 K samples, an attacker needs to test $L!$ attempts to try to reveal the secret content. For the second part of the key, the total number of keys in the hiding process is very small (there are six available choices). Therefore, the major part of the security in our system relies on the key space of the scrambling process.

For example, if the attacker knows the scheme and accesses stego signals, he can extract the wrong message from the stego signals, if the key is not right (Figure 8). The reader can compare those results with the obtained in Figure 7. In both figures, the recovered messages are from the same stego signals, with BH equal to 4, but for different keys (correct/incorrect). After several tests, we found that the similarity between the original message and the wrong message obtained by the attacker with incorrect keys is lower than 1×10^{-3}. Figure 9 shows the results for the messages obtained from ten different host signals.

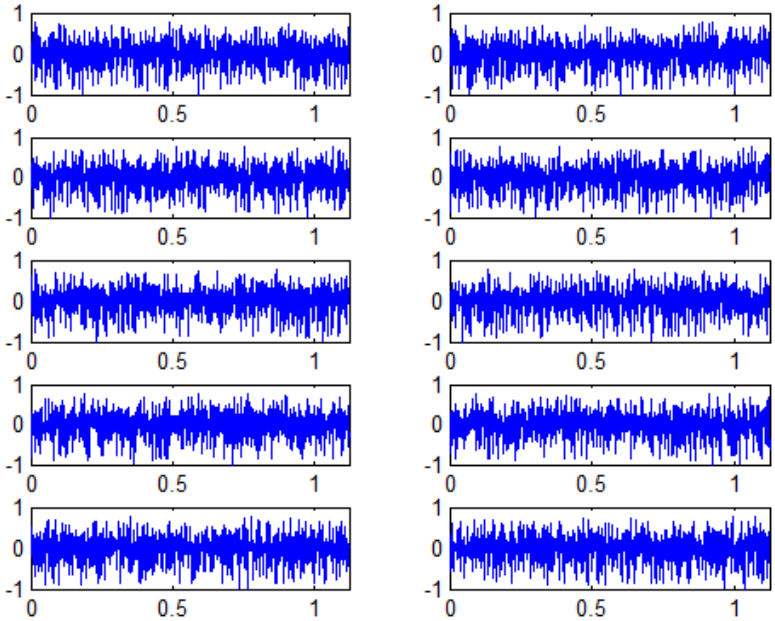

Figure 8. Secret messages recovered without access to the key/password, varying the host signal and with BH equal to 4.

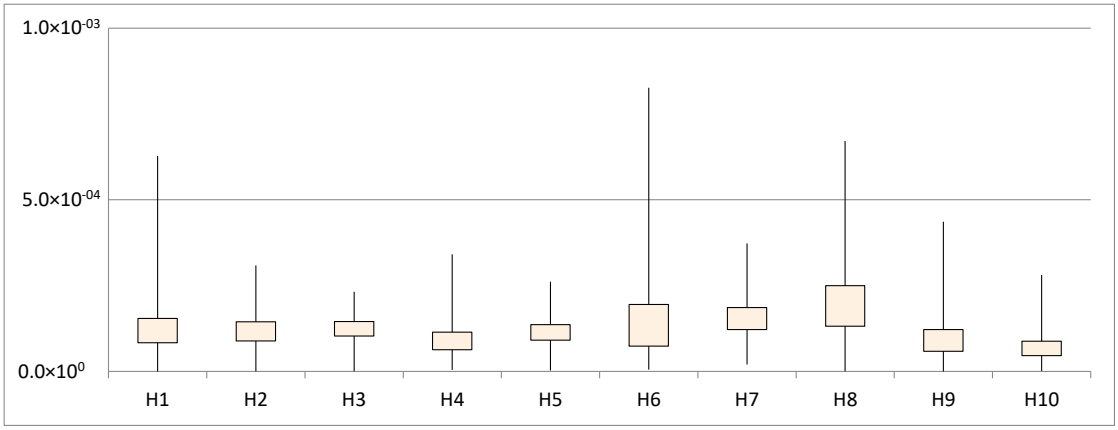

Figure 9. Correlation coefficient trust ranges between the recovered message without having access to the key/password and the original signal, varying the host signal.

5.2. Ciphertext Only Attack

In this case, the attacker accesses some stego signals, but he ignores the value of the key. He tries to reveal the right sequence by analysing, for example, the spectrum of the stego signal. Since all the scrambled signals look like noise, and follow the behaviour of a super-Gaussian signal, their spectrum does not give traces of the original secret message. Then, our system cannot be broken with this kind of attack.

5.3. Statistical Attack and Perfect Secrecy

According to Shannon's theory, a system works with perfect secrecy if the size of the key space is equal to the size of the message space and the size of the scrambled space, and if all the available keys are equally likely. In our case, the size of the possible message with m samples is $L!$, that is the same size of the possible scrambled messages and ordering sequences (keys); and, on the other hand, any key can be used. Then, even if the attacker finds some keys that provide him possible messages, there in no guarantee as to which of them is the right one. Therefore, our system overcomes the statistical attack.

6. Conclusions and Future Work

This document presents a covert communication method, using speech in speech hiding, which combines scrambling and steganography to provide high security in secret content privacy. It highlights the facts that the process is completely reversible and only depends on a key. Without knowledge of the key, users are not authorised, nor able to reveal the secret contents.

In the proposed method, there is a close relation between the hiding capability and stego signal imperceptibility, given by the BH parameter. The higher the BH, the lower the hiding capability of the method. In terms of the recovered signal in the transmitter with knowledge of a key, the method is at least 99.5% reversible. In other words, the recovered signal is perceptually identical to the original secret message. Additionally, the method guarantees that the host signal influence on the recovered signal quality is extremely low. That is, the quality of the recovered message does not depend on the used host signal.

Regarding security, it was tested and verified that if an intruder knows the used steganography method and has access to the transmitted stego signal, but does not know the key, the extracted information from the stego signal has non-legible contents, and differs from the original secret message in a great way.

As future work, the authors of this paper propose the following themes:

- Measuring the resistance of the stego signal against signal manipulation attacks like MP3 (Moving Picture Experts Group Layer-3 Audio) compression, additive noise and filtering.

- Propose alternative ways to obtain the super-Gaussian signal to be imitated by the secret message.
- Explore alternative ways of dynamic LSB substitution.

Author Contributions: Conceptualization, D.M.B.; Formal analysis, D.R.; Investigation, D.R.; Methodology, D.M.B.; Software, D.M.B.; Validation, D.R.; Writing—original draft, D.M.B.; Writing—review & editing, D.R.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Khan, M.F.; Baig, F.; Beg, S. Steganography Between Silence Intervals of Audio in Video Content Using Chaotic Maps. *Circ. Syst. Signal Process.* **2014**, *33*, 3901–3919. [CrossRef]
2. Kapotas, S.K.; Varsaki, E.E.; Skodras, A.N. Data Hiding in H. 264 Encoded Video Sequences. In Proceedings of the 2007 IEEE 9th Workshop on Multimedia Signal Processing, Chania, Greece, 1–3 October 2007. doi:10.1109/mmsp.2007.4412894.
3. Liu, W.L.; Leng, H.S.; Huang, C.K.; Chen, D.C. A Block-Based Division Reversible Data Hiding Method in Encrypted Images. *Symmetry* **2017**, *9*, 308. [CrossRef]
4. Liu, Y.; Chang, C.C.; Huang, P.C.; Hsu, C.Y. Efficient Information Hiding Based on Theory of Numbers. *Symmetry* **2018**, *10*, 19. [CrossRef]
5. Ali, A.H.; George, L.E.; Zaidan, A.A.; Mokhtar, M.R. High capacity, transparent and secure audio steganography model based on fractal coding and chaotic map in temporal domain. *Multimed. Tools Appl.* **2018**, *77*, 31487–31516. [CrossRef]
6. Rekik, S.; Guerchi, D.; Selouani, S.A.; Hamam, H. Speech steganography using wavelet and Fourier transforms. *EURASIP J. Audio Speech Music Process.* **2012**, *2012*. [CrossRef]
7. Gopalan, K. Audio steganography by modification of cepstrum at a pair of frequencies. In Proceedings of the 2008 9th International Conference on Signal Processing, Beijing, China, 26–29 October 2008. doi:10.1109/icosp.2008.4697579.
8. Ballesteros, L.D.M.; Moreno, A.J.M. Highly transparent steganography model of speech signals using Efficient Wavelet Masking. *Expert Syst. Appl.* **2012**, *39*, 9141–9149. [CrossRef]
9. Djebbar, F.; Abed-Meraim, K.; Guerchi, D.; Hamam, H. Dynamic energy based text-in-speech spectrum hiding using speech masking properties. In Proceedings of the 2010 The IEEE 2nd International Conference on Industrial Mechatronics and Automation, Wuhan, China, 30–31 May 2010. doi:10.1109/icindma.2010.5538279.
10. Radhakrishnan, R.; Kharrazi, M.; Memon, N. Data Masking: A New Approach for Steganography? *J. VLSI Signal Process. Syst. Signal Image Video Technol.* **2005**, *41*, 293–303. [CrossRef]
11. Ballesteros, L.D.M.; Moreno, A.J.M. Speech Scrambling Based on Imitation of a Target Speech Signal with Non-confidential Content. *Circ. Syst. Signal Process.* **2014**, *33*, 3475–3498. [CrossRef]
12. Ballesteros, L.D.M.; Renza, D.; Camacho, S. An unconditionally secure speech scrambling scheme based on an imitation process to a gaussian noise signal. *J. Inf. Hiding Multimed. Signal Process.* **2016**, *7*, 233–242.
13. Lim, Y.C.; Lee, J.W.; Foo, S.W. Quality Analog Scramblers Using Frequency-Response Masking Filter Banks. *Circ. Syst. Signal Process.* **2009**, *29*, 135–154. [CrossRef]
14. Elshamy, E.M.; El-Rabaie, E.S.M.; Faragallah, O.S.; Elshakankiry, O.A.; El-Samie, F.E.A.; El-sayed, H.S.; El-Zoghdy, S.F. Efficient audio cryptosystem based on chaotic maps and double random phase encoding. *Int. J. Speech Technol.* **2015**, *18*, 619–631. [CrossRef]
15. Madain, A.; Dalhoum, A.L.A.; Hiary, H.; Ortega, A.; Alfonseca, M. Audio scrambling technique based on cellular automata. *Multimed. Tools Appl.* **2012**, *71*, 1803–1822. [CrossRef]
16. Augustine, N.; George, S.N.; Deepthi, P.P. Sparse representation based audio scrambling using cellular automata. In Proceedings of the 2014 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), Bangalore, India, 6–7 January 2014. doi:10.1109/conecct.2014.6740186.
17. Bhowal, K.; Bhattacharyya, D.; Pal, A.J.; Kim, T.H. A GA based audio steganography with enhanced security. *Telecommun. Syst.* **2011**, *52*, 2197–2204. [CrossRef]

18. Ballesteros, L.D.M.; Renza, D. Speech steganography based on scrambling and hiding (Spanish audios). *Mendeley Data* **2018**. [CrossRef]
19. Lin, Y.; Abdulla, W.H. *Audio Watermark*; Springer International Publishing: Berlin, Germany, 2015.
20. Gomez-Coronel, S.; Escalante-Ramirez, B.; Acevedo-Mosqueda, M.; Mosqueda, M. Steganography in audio files by Hermite Transform. *Appl. Math. Inf. Sci.* **2014**, *8*, 959–966. [CrossRef]
21. Chhabra, A.; Mathur, S. Modified RSA Algorithm: A Secure Approach. In Proceedings of the 2011 International Conference on Computational Intelligence and Communication Networks, Gwalior, India, 7–9 October 2011. doi:10.1109/cicn.2011.117.
22. Kar, D.; Mulkey, C. A multi-threshold based audio steganography scheme. *J. Inf. Secur. Appl.* **2015**, *23*, 54–67. [CrossRef]
23. Xin, G.; Liu, Y.; Yang, T.; Cao, Y. An Adaptive Audio Steganography for Covert Wireless Communication. *Secur. Commun. Netw.* **2018**, *2018*. [CrossRef]