High Scrambling Degree in Audio Through Imitation of an Unintelligible Signal

Dora M. Ballesteros L., Diego Renza(✉), and Steven Camacho

Universidad Militar Nueva Granada, Bogotá, Colombia
{dora.ballesteros,diego.renza,u1400943}@unimilitar.edu.co

Abstract. A reversible scheme of audio scrambling based on the imitation of Gaussian unintelligible signals is presented in this paper. It is supported by the similarities it shares with Gaussian unintelligible signals in terms of the Probability Density Function and the entropy. It is feasible for an audio signal to imitate the behavior of a Gaussian noise signal, and then the residual intelligibility is zero. Our proposed scheme, termed ASGI (Audio Scrambling by Gaussian signal Imitation), is tested with four different music genres, and the experimental tests reveal that the scrambled audio signals look like Gaussian noise signals and have high scrambling degrees. Additionally, our scheme preserves the advantages of imitation-based scrambling schemes.

Keywords: Audio scrambling · Imitation-based scheme · Residual intelligibility · Scrambling degree · Entropy value

1 Introduction

Scrambling and encryption are methodologies aimed at protecting the content of information by altering the original content, which can only be recovered with an appropriate key. In image case, some effective methods consist of histogram modification to obtain uniformly distributed behavior [8,13,16]. However, in audio case, methods are focused on the permutation sequences generation without altering data distribution [1,7]. In both types of methodologies, the main conditions to be satisfied are: residual intelligibility, security level of the key, quality of the recovered message, and computational cost [9].

One way to measure residual intelligibility in audio is through the Scrambling Degree (SD). The lower the residual intelligibility, the higher the SD. In [6], the scrambling process obeys iterative displacement and the obtained SD values are lower than 0.8. With Cellular Automata (CA) as key generator, SD can be high, but its value depends greatly on the number of generations (NOG), the neighborhood types and the boundary condition [12]. When CA is mixed with Compressive Sensing (CS), SD depends on the sub-rate and the content type of the audio signal (e.g. voice, instrumental, or a mix of them) [3]. In terms of secure systems, some approaches include progressive scrambling [14,15], high dimensional matrix transformation [11], mixture with watermarking [10], 2D

J.F. Martínez-Trinidad et al. (Eds.): MCPR 2016, LNCS 9703, pp. 251–259, 2016.
DOI: 10.1007/978-3-319-39393-3_25

Arnold transform [2] or bio-inspired process [4], which have demonstrated that their key-spaces are large enough to resist cryptanalysis [5]. However, a good trade-off among the design conditions is still a challenge. In the proposal of Ballesteros and Moreno [4], an auxiliary signal with intelligible content is used to create the key. The original speech signal imitates an auxiliary signal, and the content is transformed in order to resemble the auxiliary signal content. Imitation is feasible if some conditions are satisfied and the system is unconditionally secure. However, a huge database of auxiliary intelligible signals is required to implement the scrambling scheme. To overcome this problem, in [5] a speech scrambling scheme based on imitation of a Gaussian noise signal was proposed. Unlike [4], a database of auxiliary signals is not necessary because the target signal is created in situ. According to several tests, it was validated that a speech signal can imitate a Gaussian noise signal.

In this paper, the work of [5] is continued and an audio scrambling scheme based on imitation of a Gaussian noise signal (Audio Scrambling by Gaussian signal Imitation, ASGI) is proposed. The aim is to verify if in generic audio case (i.e. speech, instrumental or a mixture of both) the imitation process to a Gaussian noise signal is feasible, and if the audio genre influences the results of SD. Since our proposal has the same characteristics of the scrambling schemes based on imitation, it is expected that the ASGI scheme has low computational cost to create the key, complete reversibility, and high security.

The paper is organized as follows. Section 2 presents the proposed ASGI scheme for the scrambling and recovering modules. In Sect. 3, the ASGI scheme is validated in terms of Scrambling Degree and the results are compared with other proposals. Section 4 presents the conclusions of the work.

2 Proposed ASGI Scheme

The scrambling module of the proposed ASGI scheme is carried out by imitating the behavior of a Gaussian noise signal. If the statistical moments of the audio signal are similar to those of the Gaussian noise signal, entropy and probability density distribution (PDF) are similar too. Therefore, if the audio samples are relocated to resemble the Gaussian noise signal, the result (the scrambled audio signal) looks like the Gaussian noise signal, and the original content of the audio is altered in the process. At the receiver, the scrambling process is completely reversed with the appropriate key.

2.1 Scrambling Module

In this module, the places of the audio signal are relocated through the process of imitating a Gaussian noise signal. In our proposal, the key is not an input of the system because it is generated in situ and corresponds with the mapping positions of the signals involved in the imitation. Figure 1 shows a block diagram of the scrambling module. The steps to obtain the scrambled audio signal are as follows:

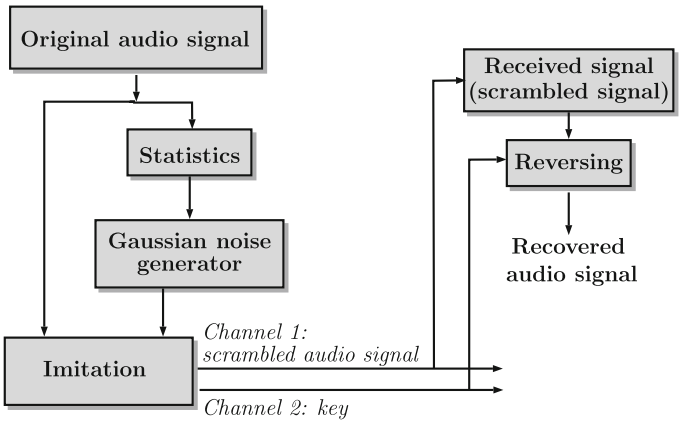


Fig. 1. Block diagram of the proposed ASGI scheme.

1. Calculate the audio signal statistical moments. The first two moments of the audio signal are calculated, namely: mean (Eq. 1) and standard deviation (Eq. 2).

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N-1} \tag{1}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \mu)^2}{N-1}} \tag{2}$$

   Where N is the total number of samples.
2. Generate a Gaussian noise signal from the above statistical moments; the result is an unintelligible signal with similar entropy and PDF values to those of the audio signal.
3. Imitate the Gaussian noise signal; this is the main block of the proposal and it is based on the following hypothesis: "An audio signal can imitate an unintelligible signal if their entropy and PDF values are similar". Suppose that we have two signals (S_a, S_b), S_a with intelligible content and S_b with unintelligible content. Amplitudes of these signals are in the range $x = [x_1, x_2, \ldots, x_n]$, with x_1 as the lowest amplitude and x_n as the highest amplitude. These signals sound different, but they have equal entropy values $(H(S_a), H(S_b))$, and equal PDF. Then, the quantity of samples with amplitude x_1 of the signal S_a is equal to the quantity of samples with amplitude x_1 of the signal S_b, and so on. In other words (Eq. 3),

$$P(S_a = x_i) = P(S_b = x_i)\ for\ i[1, 2, \ldots n] \tag{3}$$

   With $P(\cdot)$ as the data probability. Since the signals have the same entropy and PDF, they can look like each other if one of them imitates the other.

Imitation means that the behavior of one is followed by the other, through a relocation process. Suppose S_a wants to imitate S_b. The first step is to find the position of x_n amplitude both in S_a and S_b. These values are kept in the first place of the vectors I_a and I_b, respectively. The second step is to find the position of x_{n-1} amplitude both in S_a and S_b, keeping the results in the second place of the vectors I_a and I_b, respectively. The above procedure is repeated until the positions of x_1 amplitude in S_a and S_b are found and saved. With I_a and I_b, the mapping process follows the Eq. 4:

$$C(I_b) = S_a(I_a) \tag{4}$$

Where C is the scrambled audio signal. It is worth noting that if entropies and PDF of S_a and S_b are equal, then C is equal to S_b. The key is obtained by means of Eq. 5:

$$key(I_a) = I_b \tag{5}$$

The details of the algorithm are summarized in Algorithm 1.

Algorithm 1. Scrambling module

Inputs: Audio signal *Sa*, Noise signal *Sb*, amplitudes of these signals ordered from highest to lowest x.

Outputs: Scrambled signal C, key K.

```
function SCRAMBLING(Sa,Sb,x)
    audiolength ← length of Sa
    for i = 0 to audiolength do
        Ia_i ← position of x_{n-i} in Sa
        Ib_i ← position of x_{n-i} in Sb
        C_{Ib_i} ← Sa_{Ia_i}
        K_{Ia_i} ← Ib_i
    end for
end function
```

In the case of entropy and PDF values of S_a and S_b being similar but not equal, the scrambled audio signal would be similar (but not equal) to the Gaussian noise signal. However, similarity is enough to alter the original content without leaving a trace of it.

4. Transmit the scrambled audio signal by one channel and the key by another channel.

2.2 Recovering Module

With the scrambled audio signal and the appropriate key, the process is completely reversed and the recovered audio signal is equal to the original one. The recovered speech signal, R, is obtained with the reversing Eq. 6.

$$R = C(key) \tag{6}$$

The details of the algorithm are summarized in Algorithm 2.

Algorithm 2. Recovering module

Inputs: Scrambled signal C, key K.
Outputs: Recovered audio signal R.

```
function RECOVERING(C,K)
    audiolength ← length of C
    for i = 0 to audiolength do
        R_i ← C_{k_i}
    end for
end function
```

3 Experimental Results and Discussion

To evaluate the performance of our ASGI scheme, we select four musical genres (pop, rock, jazz, and classical), and five songs for each one. Six frames of five seconds are extracted from every song in order to apply the scheme. Since the aim is to alter the original content of the audio, performance is measured through the SD between the original audio signal and the scrambled audio signal. Firstly, the difference of the signal, D, is calculated, as follows (Eq. 7):

$$D(i) = \frac{1}{4} \sum_{i=3}^{m-2} \{4 * S(i) - (S(i-1) + S(i-2) + S(i+1) + S(i+2))\} \quad (7)$$

where $S(i)$ is the i^{th} sample of the audio signal, m is the total number of samples and D is a vector with $m - 4$ values.

Next, the sum and the subtraction of the two differences (original signal and scrambled signal) are obtained as follows (Eqs. 8 and 9):

$$B = D_2 - D_1 \quad (8)$$

$$A = D_2 + D_1 \quad (9)$$

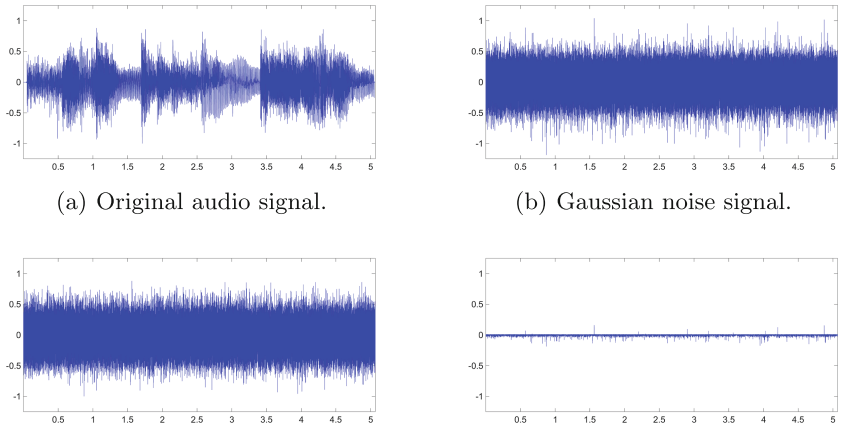
where D_2, is the difference of the scrambled signal and D_1 is the difference of the original signal, calculated by means of Eq. 7.

Finally, the Scrambling Degree, SD, is calculated according to Eq. 10:

$$SD = B/A \quad (10)$$

where B/A solves the system of linear equation $SD * A = B$ for SD. The highest value of SD is 1 and the lowest is 0. If the result is 1, it means that the original content of the audio signal has been completely altered in the scrambling process, or in other words, that the residual intelligibility is zero.

In order to illustrate the performance of the proposed scheme, we show the results of two audio file frames. Then, the results of one hundred and twenty simulations are summarized. Figure 2 shows an example with the rock genre. It contains voices and sounds of musical instruments. In this case, entropy of the Gaussian noise signal is 4.3202 and entropy of the scrambled audio signal is



(a) Original audio signal.

(b) Gaussian noise signal.

(c) Scrambled audio signal.

(d) Subtraction between b and c.

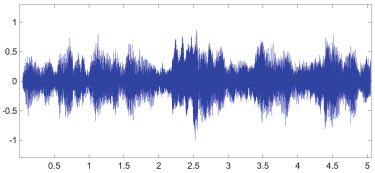
Fig. 2. Example of the proposed ASGI scheme for the rock genre.

4.3186. SD between the original audio signal and the scrambled audio signal is 0.9818.

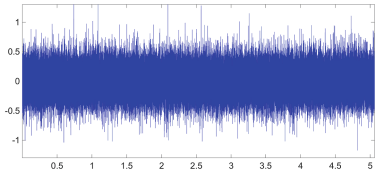
Figure 3 shows an example with the classical genre. It contains only sounds of musical instruments. In this case, entropy of the Gaussian noise signal is 4.3232 and entropy of the scrambled audio signal is 4.6199. SD between the original audio signal and the scrambled audio signal is 0.9999.

Figure 4 shows the summary of the tests. Every genre contains the results of thirty simulations (five songs by six frames by song) with a confidence range of 95 %. These results are higher than 0.87, which is high enough to guarantee very low residual intelligibility. On the other hand, the proposed scheme works with different kinds of audio content, such as speech, instrumental or a mixture of both.

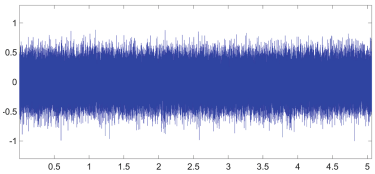
Finally, our proposed scheme is compared with some of the existing techniques (Fig. 5). We use the results reported in the works of Madain et al. [12] and Augustine et al. [3]. In the first one, the results of NOG (number of generations) equal to 1 and 15 were taken into account. In the second one, we use the results of subrate (SR) equal to 0.1 and 0.5. In the work of Madain et al., the value of SD depends very much on the number of generations (NOG). In the proposal of Augustine et al., the subrate (SR) influences the quality of the scrambled signal. In all cases, our proposal has better performance in terms of SD and this value does not depend on adjustable parameters.



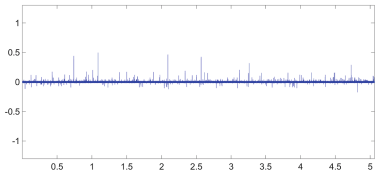
(a) Original audio signal.



(b) Gaussian noise signal.



(c) Scrambled audio signal.



(d) Subtraction between b and c.

Fig. 3. Example of the proposed ASGI scheme for the classic genre.

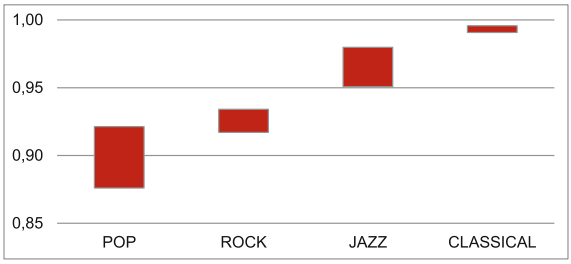


Fig. 4. SD results for the proposed Scheme. Confidence range (95 %) for SD by genre (120 simulations).

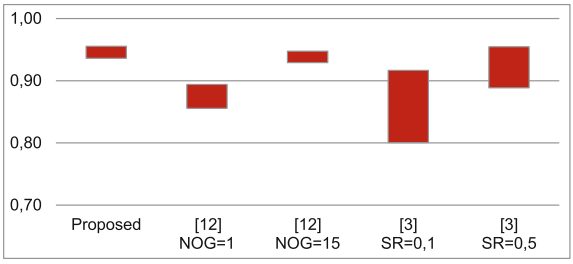


Fig. 5. Comparative results for SD. SD Confidence range (95 %) for the proposed scheme and Madain et al. [12] and Augustine et al. [3] works.

4 Conclusion

In this paper, a scheme of audio scrambling that uses a generalization of the proposal in [5] was presented. Our approach, ASGI, exploits the imitation ability of audio signals and their Gaussian behaviour. The scrambled audio signals look and sound like Gaussian noise signal and there is no trace of the original content. According to our tests, the values of SD are higher than other schemes based on cellular automata [3,12] with the advantage that the behaviour does not depend on initial conditions. On the other hand, we tested the ASGI scheme with different audio genres and it was found that classical music gives a better performance that other genres. However, ASGI works with different audio contents, such as speech, instrumental or a mixture of both.

Acknowledgment. This work is supported by the "Universidad Militar Nueva Granada - Vicerrectoría de Investigaciones" under the grant INV-ING-1910 of 2015.

References

1. Alwahbani, S.M., Bashier, E.: Speech scrambling based on chaotic maps and one time pad. In: 2013 International Conference on Computing, Electrical and Electronics Engineering (ICCEEE), pp. 128–133. IEEE (2013)
2. Augustine, N., George, S.N., Deepthi, P.P.: Compressive sensing based audio scrambling using arnold transform. In: Martínez Pérez, G., Thampi, S.M., Ko, R., Shu, L. (eds.) SNDS 2014. CCIS, vol. 420, pp. 172–183. Springer, Heidelberg (2014)
3. Augustine, N., George, S.N., Deepthi, P.: Sparse representation based audio scrambling using cellular automata. In: 2014 IEEE International Conference on Electronics, Computing and Communication Technologies (IEEE CONECCT), pp. 1–5. IEEE (2014)
4. Ballesteros L, D.M., Moreno A, J.M.: Speech scrambling based on imitation of a target speech signal with non-confidential content. Circ. Syst. Sig. Process. **33**, 3475–3498 (2014)
5. 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 Multimedia Sig. Process. **7**(2), 233–242 (2016)
6. Chen, G., Han, B.: An audio scrambling degree measure based on information criteria. In: 2010 2nd International Conference on Signal Processing Systems (ICSPS), vol. 1, pp. V1–181. IEEE (2010)
7. Da-hui, H., Zhi-guo, D.: An audio watermarking based on logistic map and m-sequence. Int. J. Digital Content Technol. Appl. **6**(1), 1–10 (2012)
8. Ghebleh, M., Kanso, A., Noura, H.: An image encryption scheme based on irregularly decimated chaotic maps. Signal Process. Image Commun. **29**(5), 618–627 (2014)
9. Kulkarni, N.S., Raman, B., Gupta, I.: Multimedia encryption: a brief overview. Recent advances in multimedia signal processing and communications. Springer, Heidelberg (2009)

10. Kwon, G.R., Wang, C., Lian, S., Hwang, S.S.: Advanced partial encryption using watermarking and scrambling in mp3. Multimedia Tools Appl. **59**(3), 885–895 (2012)
11. Li, H., Qin, Z., Shao, L., Zhang, S., Wang, B.: Variable dimension space audio scrambling algorithm against MP3 compression. In: Hua, A., Chang, S.-L. (eds.) ICA3PP 2009. LNCS, vol. 5574, pp. 866–876. Springer, Heidelberg (2009)
12. Madain, A., Dalhoum, A.L.A., Hiary, H., Ortega, A., Alfonseca, M.: Audio scrambling technique based on cellular automata. Multimedia Tools Appl. **71**(3), 1803–1822 (2014)
13. Murillo-Escobar, M., Cruz-Hernández, C., Abundiz-Pérez, F., López-Gutiérrez, R., Del Campo, O.A.: A rgb image encryption algorithm based on total plain image characteristics and chaos. Signal Process. **109**, 119–131 (2015)
14. Oo, T.T., Onoye, T.: Progressive audio scrambling via complete binary tree's traversal and wavelet transform. In: Asia-Pacific Signal and Information Processing Association, 2014 Annual Summit and Conference (APSIPA), pp. 1–7. IEEE (2014)
15. Oo, T.T., Onoye, T.: Progressive audio scrambling via wavelet transform. In: 2014 IEEE Asia Pacific Conference on Circuits and Systems (APCCAS), pp. 97–100. IEEE (2014)
16. Zhou, Y., Bao, L., Chen, C.P.: Image encryption using a new parametric switching chaotic system. Signal Process. **93**(11), 3039–3052 (2013)