Speech Coding based on a Hybrid Approach: DCT, Huffman and Run-length Coding

Assist Lect. Sundos Abdulameer Alazawi
ss.aa.cs@uomustansiriyah.edu.iq
Department of Computer Science - College of Science, Mustansiriyah University, Baghdad / Iraq

Assist. Lect. Esraa Jaffar Baker
es-alshaibany@uomustansiriyah.edu.iq
Department of Computer Science - College of Science, Mustansiriyah University, Baghdad / Iraq

Assist. Lect. Shahbaa Mohammed Abdulmaged
Al-Iraqia University - Law Dept., Baghdad / Iraq
Shahbaa.abdulmaged@aliraqia.edu.iq

ترميز الكلام على أساس نهج هجين: DCT، وهوفمان، والترميز على طول المدى

م. م. سندس عبد الأمير العزاوي
قسم علوم الحاسوب – كلية العلوم، الجامعة المستنصرية، بغداد \ العراق

م. م. اسراء جعفر بكر
قسم علوم الحاسوب – كلية العلوم، الجامعة المستنصرية، بغداد \ العراق

م. م. شهباء محمد عبد المجيد
الجامعة العراقية – قسم الحقوق، بغداد \ العراق

Abstract

The exponential expansion of data in the digital world necessitates the development of effective methods for data transmission and storage. Data compression (DC) strategies are suggested to reduce the quantity of data stored or conveyed due to constrained resources. As a result of DC ideas' ability to efficiently use existing storage space and transmission capacity, different methods have been developed in various areas. Speech coding is a lossy method of coding; therefore, the output signal differs slightly from the input signal. Speech coding is useful for message encryption, communication over long distances and speech quality. In the fields of digital voice processing and telecommunications, speech coding has been a significant problem. In this work, we demonstrate that a DCT with a chaotic system combined with a Hybrid of Huffman and Run-length coding can be utilized to implement very low bit-rate speech coding with high reconstruction quality. The proposed system was conducted on the Lbri-speech dataset and the method is evaluated based on SNR, MSE, and PSNR. The simulation results show good results and the best result was achieved with a compression ratio of about 14%. The results get MSE=0.008, SNR =34.025db, and PSNR=80.1db.

Keywords speech coding, Data compression, transform coding, neural speech coding, speech signal.

المستخلص

يتطلب التوسع الهائل للبيانات في العالم الرقمي تطوير طرق فعالة لنقل البيانات وتخزينها. تُقترح استراتيجيات ضغط البيانات (DC) لتقليل كمية البيانات المخزنة أو المنقولة بسبب الموارد المحدودة. ونتيجة لقدرة أفكار DC على استخدام مساحة التخزين الحالية وقدرة النقل بكفاءة، فقد تم تطوير أساليب مختلفة في مجالات مختلفة. يعد ترميز الكلام طريقة ترميزية ضائعة؛ ولذلك، تختلف إشارة الخرج قليلاً عن إشارة الإدخال. يعد تشفير الكلام مفيدًا لتشفير الرسائل والتواصل عبر المسافات الطويلة وجودة الكلام. في مجالات معالجة الصوت الرقمي والاتصالات السلكية واللاسلكية، كان تشفير الكلام يمثل مشكلة كبيرة. في هذا العمل، نوضح أنه يمكن استخدام DCT مع نظام فوضوي مدمج مع Hybrid of Huffman وترميز Run-length لتنفيذ تشفير الكلام بمعدل بت منخفض للغاية مع جودة إعادة بناء عالية. تم إجراء النظام المقترح على مجموعة بيانات Lbri-speech ويتم تقييم الطريقة بناءً على SNR وMSE وPSNR. أظهرت نتائج المحاكاة نتائج جيدة وتم تحقيق أفضل نتيجة بنسبة ضغط تبلغ حوالي 14%. النتائج تحصل على SNR=34.025db، MSE=0.008، وPSNR=80.1db.

الكلمات المفتاحية تشفير الكلام، ضغط البيانات، تشفير التحويل، تشفير الكلام العصبي، إشارة الكلام.

1. Introduction

The most effective communication method utilized in telephone, mobile communications, and transmissions is speech. One method that aims to use all the communication systems' capabilities and resources is speech compression. By reducing the transmitted voice signal's bit rate or size, compression is achieved.[1][2]. The communication channel's bandwidth is conserved through this technique. Additionally, it reduces the memory needed to store speech files. Because speech signals include a significant amount of redundant information, speech is compressed. A compressed signal is generated by removing non-essential voice information and coding just the important voice information. To reconstruct the original speech with excellent intelligibility, the degree of information to be eliminated must be appropriate. Video teleconferencing systems, voice mail, cellular, and satellite communications are just a few of the current applications for speech compression.[3][4].

Speech coding has been studied extensively for years, leading to various standardized codecs that may be divided into two groups: vocoders and waveform codecs. The decoder synthesizes speech from a set of physiologically essential features that a vocoder, also known as parametric speech coding, distils, such as the spectral envelope (comparable to responses of the vocal tract, including the contribution from tongue position, nasal cavity, and mouth shape), gain (speech-level), and fundamental frequencies. A vocoder often runs with excellent computational efficiency at three kbps or lower, but the quality of the synthesized speech is typically constrained and does not scale to higher bitrates [5][6]. A waveform codec, on the other hand, seeks to precisely recreate the input voice signal and offers up to transparent quality in a high bitrate range. [7][8][9].

The speech coders' objective is to attain bit rate and bandwidth reduction. The memory demanded for speech coders should be reduced, which reduces the bit rate proportionately. Because compressed data may be transferred with fewer bits, the signal's transmission power must be reduced. The coding algorithms offer noise resilience since some saved bits may be employed to protect error-control bits for the speech parameters.[10].

The paper is organized as follows; The basics of speech coding are described in Section 2, the theoretical background is presented in Section 3, the proposed method is illustrated in Section 4, Section 5 shows the experimental results, and Section 6 discusses the conclusion.

2. Speech Coding Basics

Voice coding methods can process massive amounts of data and increase the volume of information transmitted from one point to another. The coding algorithms preserve the original voice quality while representing the data with the least number of bits. The encoder transforms digital data into codes, then broadcasts as frames. After receiving the coded frames, the decoder or receiver conducts synthesis to reconstruct the original signal. The speech coders mainly vary in bit rate, perceptual quality, complexity, and delay of the reproduced speech[10]. Figure 1 illustrates the chart of speech coding methods classification [11].

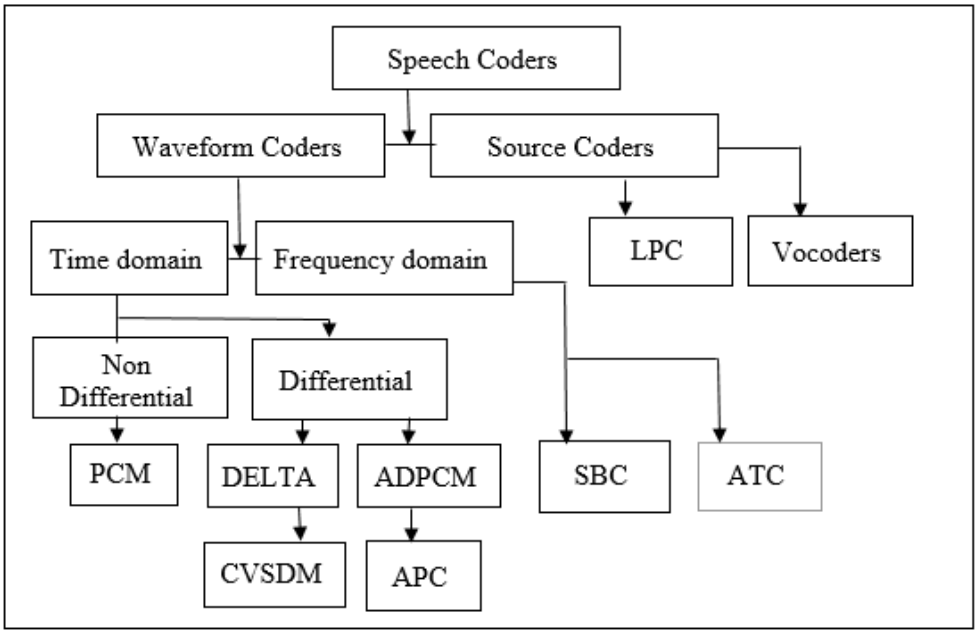

Figure 1: Classification of Speech Coding Methods[11]

The basic objective of parametric and hybrid coding approaches is to improve the quality of a voice signal while reducing the bit rate. The velocity of transmission or storage, voice quality, and computational complexity affect how well a speech signal is delivered. The following characteristics apply to low-bit-rate coding schemes[12]:

- Robustness to various languages and speakers.
- Minimized coding delay. and channel error.
- It must have high speech quality and a low bit rate.
- Less computational complexity and low memory needs.

3. Methodology

the proposed method consists of several steps as follows:

3.1 Discrete cosine transforms (DCT)

The Discrete Cosine Transform (DCT), like other transforms, aims to de-correlate the data. Each transform coefficient may be encoded individually after de-correlation without sacrificing compression performance. This section introduces the DCT and some of its key qualities[13]. The following are the most popular DCT definition of a 1-D sequence of length

N: $$C(u) = a(u)\sum_{i=0}^{N} s(i)\cos(\frac{u\pi(2i+1)}{2N}) \quad (1)$$

$$a(u) = \begin{cases} \sqrt{1/N} & if\ u = 0 \\ \sqrt{2/N} & if\ u \neq 0 \end{cases} \quad (2)$$

Where 0…., N-1, s presents a set of N speech input data values, and C (u) is the uth DCT coefficient.

3.2 Quantization

The process of converting a collection of continuous-valued data to a set of discrete valued data is known as quantization. The goal of quantization is to minimize the amount of information in threshold coefficients. This procedure ensures that errors are kept to a minimum. [14].

3.3 Run Length Coding

Run-length encoding is utilized when symbols do not occur independently but are impacted by their predecessors. Given the occurrence

of a symbol, that symbol is more likely to occur next to others. If this is not the case, using coding runs (rather than symbols) to compress data will not work. Other coding techniques can achieve the same result in a broader sense, but run-length coding consumes extremely minimal cost when the runs are large[15].

3.4 Huffman Coding

David Huffman created this approach as part of a class assignment; in the field of information theory, the class was the first ever, and Robert Fano taught it at MIT. Huffman codes are codes formed utilizing this approach or procedure. These are prefix codes that are optimal for a certain model (set of probabilities). The Huffman technique is based on two considerations about optimal prefix codes.[16].

1. In an optimum code, more often occurring symbols (those with a greater likelihood of recurrence) will have shorter code words than less frequently occurring symbols.
2. In an optimum code the exact length is given to the least frequently occurring symbol.

The first observation can be seen to be true. The average number of bits per symbol would be higher if the code words for symbols that occur more frequently were longer than those that occur less frequently. As a result, an ideal code cannot use longer code words for symbols that appear more frequently.

3.4 Logistic Chaotic Map

Chaos is the pseudorandom behavior that a deterministic, nonlinear dynamical system displays. Chaotic systems have different output values

based on certain beginning conditions and parameter settings. Other parameter values result in oscillations at the system's output with varying periods. A chaotic function or map is what mathematicians refer to as a function that exhibits some chaotic behavior.[17] The logistic map is one of the most straightforward chaotic functions that has lately been researched for cryptography applications. The function of a logistic map is written as [18]

$$X_{n+1} = rX_n(1 - X_n) \qquad (3)$$

The parameter r is a positive constant that accepts values up to 4 in the case when xn takes a value between (0, 1). Its value establishes and investigates the logistic map's behavior. The iterations start to get utterly chaotic at r=3.57 and lend themselves to the goal of encryption.

3.5 Gauss Chaotic Map

A chaotic map is a method that is frequently utilized in encryption because, despite its simplicity, it is challenging to guess. Pseudorandom numbers will be generated via a chaotic map and utilized in the encryption procedure. Chaotic maps provide pseudorandom outputs that rely on the inputted parameters. The Circle map, Tent map, and Gaussian map, among other chaotic maps, may all be employed. Equation (4) can be utilized to produce pseudorandom from a Gaussian map[19][20].

$$X_{N+1} = \exp(-aX_2^N) + \beta \qquad (4)$$

Where β and α are indicate the input parameters that will enormously affect the results of the Gaussian map. A Gaussian map will produce some random values. This sequence will be utilized to randomize other sequences with the exact length. By taking = 4.9 and ß= [-1, +1] values, the graph has been drawn for the Gauss iterated map as illustrated in Figure 2.

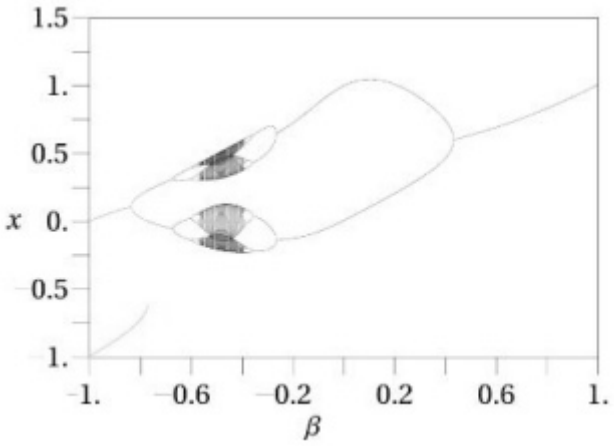

Figure 2. Gaussian map

3.5 Performance evaluation

Several objective tests were performed to assess the overall effectiveness of the suggested speech compression strategy [11][21]. Several criteria are taken into account while evaluating the performance of the reconstructed signal including the compression factor as explain in equation 5.

$$compression\ factor = \frac{length\ of\ orginal\ signal}{length\ of\ compressed\ signal} \quad ...(5)$$

Signal-to-noise ratio that explain in equation 6.

$$SNR = 10log_{10}\left[\frac{{\sigma_x}^2}{{\sigma_e}^2}\right]^2 \quad ... (6)$$

Where σx2 is the speech signal mean square and σe2 is the mean square difference between the reconstructed and original voice signal.

MSE is a useful metric to note that it correlates perfectly with perceived audio quality:

$$MSE = \frac{1}{y}\sum_{n=1}^{y}(I(n) - R(n))^2 \quad \ldots (7)$$

The PSNR, is depend on MSE and it calculated in equation 8:

$$PSNR = 10log_{10}\frac{255^2}{MSE} \quad \ldots (8)$$

4. Proposed Speech Coding Method

The proposed method explains in Figure 3 that The system uses DCT, a Hybrid of Huffman, run length coding, and logistic and Gauss map to encrypt and compress speech signals. The Block diagram of suggested speech encoding and decoding is illustrated in Figure 3. At the first segmentation step, the speech signal is padded and segmented into a fixed-length frame of 5 sec. Then on each frame, DCT and finding the total norm is applied. An absolute for the output of the DCT process is performed to eliminate the negative values and prepare it for the proposed dynamic quantization step.

A Hybrid of Huffman and RLE is proposed for an effective coding process. Then encryption process is implemented on the output of the Hybrid Huffman and RLE algorithm. The encryption stage that is explained in Figure 4 consists of several steps: key generation using the logistic map, permutation, key generation using the Gaussian map, and XOR permuted signal with a key generated from the Gaussian map (Substitution). Fig shows the steps of an encryption stage.

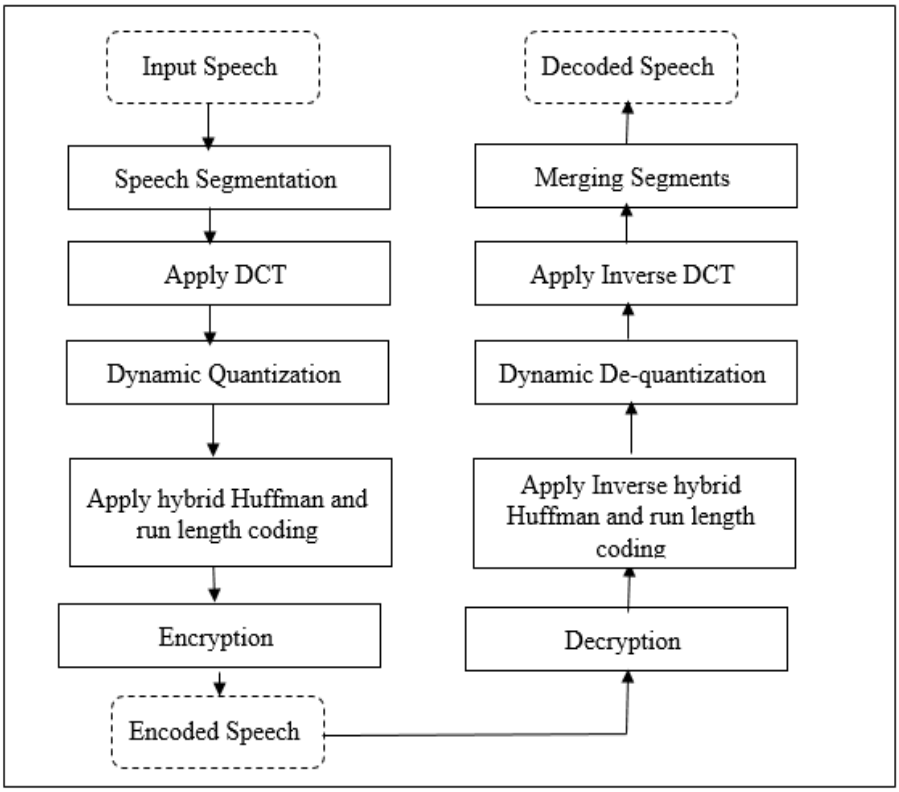

Figure 3: Block diagram of the proposed method

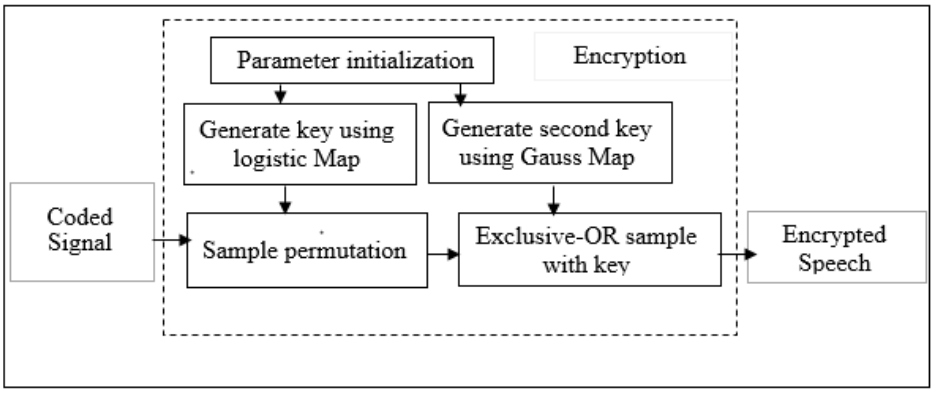

Figure 4: Encryption Stage

The encoding steps of the proposed system algorithm are illustrated in algorithm1.

Algorithm (1): Proposed system method

Input: Speech file

Output: Encoded speech file

Steps:

1: Input speech file

2: set segment as seg ←5*fs

3: cut the segment

4: q←apply DCT using eq(1,2).

5: apply dynamic quantization on q

6: k←apply Huffman algorithm on the quantized speech signal frame

7: apply RLE on the index

8: key1←generate key using logistic map using eq.(3)

9: key2←generate using gauss map using eq.(4)

10: permute the quantized speech signal using key1

11: Convert key2 to hexadecimal.

12: Convert the signal k to hexadecimal.

13: apply Excusive-OR bits of the signal k with the key2.

14: Return Encoded speech

5. Experimental Result and Analysis

The suggested system consists of two phases: encoding and decoding phases. The proposed method has experimented on the Librispeech dataset. Two chaos maps are utilized to generate a key for permutation, as well as DCT, Dynamic quantization, and a Hybrid of Huffman and Run-length encoding is applied to the speech signal.

In this method, the signal is segmented into affixed length frames, each of (5 sec), and determines the quality of the reconstructed signal. The DCT process is implemented on each frame, then calculates the total norm, takes the absolute of DCT coefficients, and sorts DCT coefficients in descending order. The next step is to find the norm for each sample to specify the important part of DCT coefficients. Figure 5 illustrates the Original and constructed signal of the second proposed system.

The hybrid of Huffman and RLE coding is used at the coding step. Dynamic quantization and hybrid Huffman and RLE coding enhanced the system performance. The compression ratio of the samples obtained from the Libri-speech dataset is illustrated in Table 1.

Table 1: Results of the second method for the Libri-speech dataset

the size of wave files in bits =203200*16=3251200					
size in bit	573759	952654	1168730	449601	1397241
compression ratio%	17.64763	29.3	35.95	13.8288	24.97619
MAP	0.001	0.00006	0.00054	0.0008	0.00003
PSNR	46.492	80.52	61.83	80.1	46.15
SNR	5.358	30.2	12.45	34.025	5.45

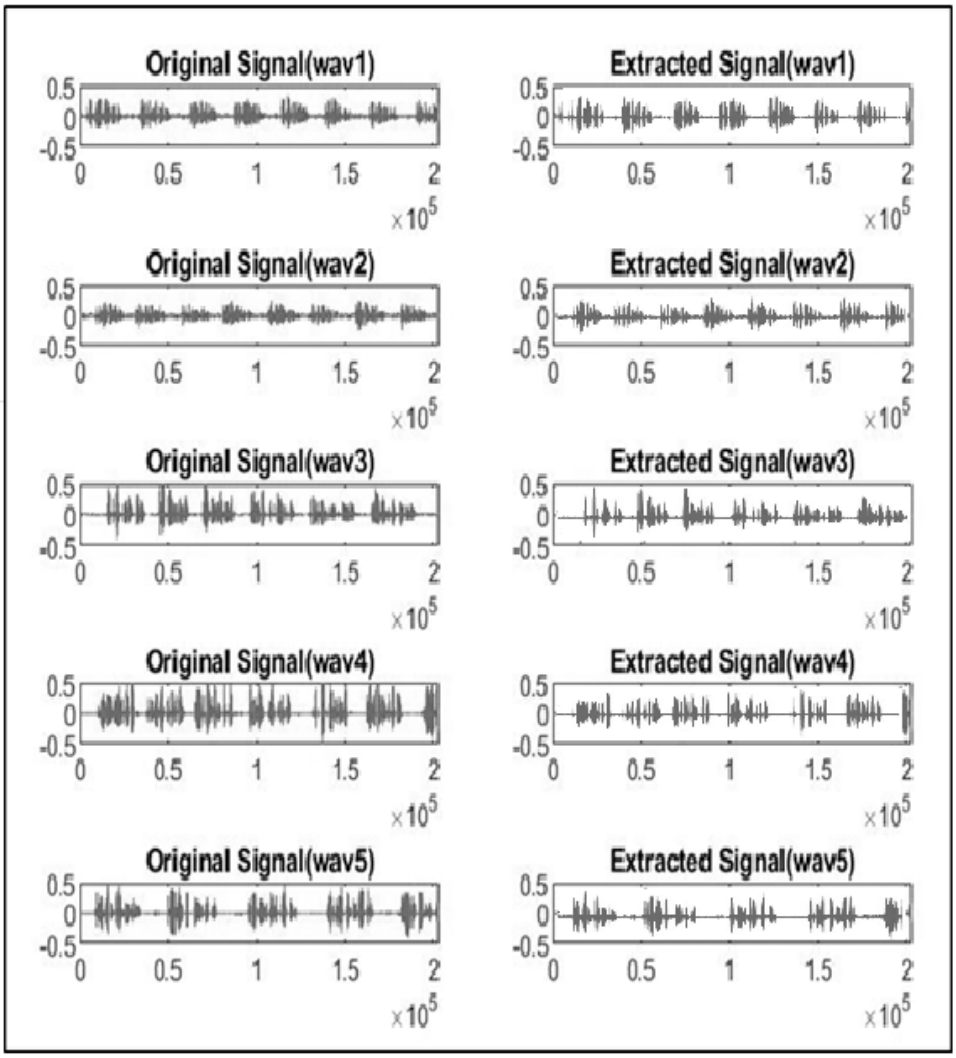

Figure 5: Results of the Second Method tested on Libri-speech samples

6. Conclusion

In this paper, the system proposed encrypts and compresses the voice signal simultaneously. The suggested system consists of two phases: encoding and decoding phases. The proposed method has experimented on the Librispeech dataset. Two chaos maps are utilized to generate a key for permutation, as well as DCT, Dynamic quantization and a Hybrid of Huffman and Run-length encoding is performed to the speech signal. The simulation results achieved a good result and the best result was achieved with a compression ratio of about 14%. The results get MSE=0.008, SNR =34.025db, and PSNR=80.1db.These results reflect the high quality and intelligibility of constructed voice signals.

References

[1] M. Cernak, A. Asaei, and A. Hyafil, "Cognitive Speech Coding: Examining the Impact of Cognitive Speech Processing on Speech Compression," *IEEE Signal Process. Mag.*, vol. 35, no. 3, pp. 97–109, 2018.

[2] S. Dusan, J. L. Flanagan, A. Karve, and M. Balaraman, "Speech Compression by Polynomial Approximation," *IEEE Trans. Audio. Speech. Lang. Processing*, vol. 15, no. 2, pp. 387–395, 2007.

[3] S. M. Joseph and P. B. Anto, "Speech compression using wavelet transform," in *2011 International Conference on Recent Trends in Information Technology (ICRTIT)*, 2011, pp. 754–758.

[4] A. S. Hameed, "Speech compression and encryption based on discrete wavelet transform and chaotic signals," *Multimed. Tools Appl.*, vol. 80, pp. 13663–13676, 2021.

[5] J. Valin and J. Skoglund, "A Real-Time Wideband Neural Vocoder at 1.6 kb/s Using LPCNet," in *Proceedings of the Annual Conference of the International Speech Communication Association (Interspeech)*, 2019.

[6] M. R. Schroeder, "Vocoders: Analysis and synthesis of speech," *Proc. IEEE*, vol. 54, no. 5, pp. 720–734, 1966.

[7] M. Dietz *et al.*, "Overview of the EVS codec architecture," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 5698–5702.

[8] K. Zhen *et al.*, "Scalable and Efficient Neural Speech Coding : A Hybrid Design," *IEEE/ACM Trans. AUDIO, SPEECH, Lang. Process.*, vol. 30, pp. 12–25, 2021.

[9] N. A. Saeed and Z. T. M. Al-Ta'i, "Feature Selection using Hybrid Dragonfly Algorithm in a Heart Disease Predication System," *Int. J. Eng. Adv. Technol.*, vol. 8, no. 6, pp. 2862–2867, 2019.

[10] R. Jage and S. Upadhya, "CELP and MELP Speech Coding Techniques," in *WiSPNET 2016 conference. CELP*, 2016, pp. 1398–1402.

[11] R. Vig and S. S. Chauhan, "Speech Compression using Multi-Resolution Hybrid Wavelet using DCT and Walsh Transforms," in *International Conference on Computational Intelligence and Data Science (ICCIDS 2018). Peer-review under responsibility of the scientific committee of the International Conference on Computational Intelligence and Data Science Procedia Computer Science*, 2018, vol. 132, pp. 1404–1411.

[12] J. D. Gibson, "Speech coding methods, standards, and applications," *IEEE Circuits Syst. Mag.*, vol. 5, no. 4, pp. 30–49, 2005.

[13] Z. J.Ahmed, L. E.George, and R. A. Hadi, "Audio compression using transforms and high order entropy encoding," *Int. J. Electr. Comput. Eng.*, vol. 11, no. 4, pp. 3459–3469, 2021.

[14] P. K. R. Manohar, M. Pratyusha, R. Satheesh, S. Geetanjali, and N. Rajasekhar, "Audio Compression Using Daubechie Wavelet," *IOSR J. Electron. Commun. Eng.*, vol. 10, no. 2, pp. 41–44, 2015.

[15] S. Mishra and S. Singh, "A Survey Paper on Different Data Compression Techniques," *INDIAN J. Appl. Res.*, vol. 6, no. 5, pp. 738–740, 2016.

[16] K. Sayood, *Introduction to data compression*. Morgan Kaufmann, 2017.

[17] A. V Prabu, S. Srinivasarao, T. Apparao, M. J. Rao, and K. B. Rao, "Audio encryption in handsets," *Int. J. Comput. Appl.*, vol. 40, no. 6, pp. 40–45, 2012.

[18] O. S. Faragallah, "An efficient block encryption cipher based on chaotic maps for secure multimedia applications," *Inf. Secur. J. A Glob. Perspect.*, vol. 20, no. 3, pp. 135–147, 2011.

[19] A. Sahay and C. Pradhan, "Multidimensional Comparative Analysis of Image Encryption using Gauss Iterated and Logistic Maps," in *International Conference on Communication and Signal Processing*, 2017, pp. 1347–1351.

[20] W. M. Rahmawati and F. Liantoni, "Image Compression and Encryption Using DCT and Gaussian Map," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 462, no. 1, pp. 12–35, 2019.

[21] N. A. Saeed and Z. T. M. Al-Ta'i, "Heart Disease Prediction System Using Optimization Techniques," in *New Trends in Information and Communications Technology Applications*, 2020, pp. 167–177.