

## Cost-effective long-range secure speech communication system for internet of things-enabled applications

Samer Alabed<sup>1</sup>, Mohammad Al-Rabayah<sup>2</sup>, Bahaa Al-Sheikh<sup>3</sup>, Lama Bou Farah<sup>4</sup>

<sup>1</sup>Department of Biomedical Engineering, School of Applied Medical Sciences, German Jordanian University, Amman, Jordan

<sup>2</sup>College of Engineering and Technology, American University of the Middle East, Egaila, Kuwait

<sup>3</sup>Department of Biomedical Systems and Informatics Engineering, Hijjawi Faculty for Engineering Technology, Yarmouk University, Irbid, Jordan

<sup>4</sup>Department of Biomedical Technologies, Faculty of Public Health, Lebanese German University, Sahel Aalma, Lebanon

### Article Info

#### Article history:

Received Dec 30, 2024

Revised Aug 31, 2025

Accepted Sep 10, 2025

#### Keywords:

Internet of things

Long range technology

Signal processing

Smart city

Sustainability

### ABSTRACT

A new communication framework has been developed that allows voice transmission over long distances for internet of things (IoT) applications such as healthcare, smart cities, and remote monitoring in the least costly way and most secure manner. The system is based on long range (LoRa) technology and takes advantage of its spread spectrum technique, to provide long range transmission without the high-power requirements. The main limitation is LoRa's bandwidth with a maximum throughput of 22 kbps for data. This presents a challenge for voice transmission communications. To address this shortened bandwidth issue, researchers developed an innovative compression solution that compresses voice data to less than 8 kbps to fit into LoRa's capabilities. The compression allows for real practical voice communications and possibly can provide even greater distance than an uncompressed voice transmission update. The voice communications transmissions have cryptographic protection in place to protect the transmitted voice messages from unauthorized access.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



### Corresponding Author:

Samer Alabed

Department of Biomedical Engineering, School of Applied Medical Sciences

German Jordanian University

Amman, 11180 Jordan

Email: [samer.alabed@gju.edu.jo](mailto:samer.alabed@gju.edu.jo)

## 1. INTRODUCTION

There has been an explosion of inexpensive wireless systems able to send large amounts of data long distances over a distance without physical connections [1]–[5]. The applicable communications environments vary widely and include military operations, remote data collection, urban infrastructure, health services, environmental monitoring, agriculture, internet of things (IoT) and telecommunication networks. Wireless long-distance communications can rely on a variety of technologies, including those utilizing satellites, radio frequency (RF) transmission, and microwave communications, each exhibiting distinct frequency use and techniques of operation [6], [7]. Long range (LoRa) technology comes as a specialized communication technology in a specific ecosystem for wireless communication of IoT applications considering its unique long communication range and low power requirements [8], [9]. LoRa manages to maintain its wide range communication and energy efficient operations due to its spread-spectrum modulation techniques. Moreover, economic advantage comes in the form of absence of ongoing payment due to its operations in un-licensed RF bands for a LoRa device's non-recurring RF bandwidth. For IoT applications with wide geographical distribution, the communication range of several kilometers provided by LoRa devices is ideal [9]–[11].

Semtech corporation innovations in the form of LoRa devices makes it an appropriate choice for energy sensitive battery powered IoT devices due to its low power consumption. Moreover, its high-capacity network makes it possible to cover a wide range and densely distribute the devices in many areas such as health care services, asset tracking, environment and agriculture, industrial sectors, home automation, water management, and smart city applications [8]–[10].

Ensuring the security of transmitted data is essential, requiring a communication system that guarantees the privacy and confidentiality of messages [12], [13]. Such a system needs to be made to prevent unauthorized access and ensure data integrity in-transit. End-to-end encryption is a good example of a secure communication system that is both effective and inexpensive, as facilitated by apps like WhatsApp and Signal. End-to-end encryption performs encryption and decryption on the sender and receiver devices so that even third parties that proxy the communications cannot see the content of the communications [14]–[17]. An inherent limitation of LoRa is its low data rate for transmissions, which makes exchange of voice communication difficult. LoRa uses chirp spread spectrum (CSS) which uses chirp pulses that exhibit linear frequency modulation. This enables flexible, energy efficient, low latency data communication; especially valuable for long-range communication where it is felt further value can be gained in radar systems using pulse compression [18]. Use of existing voice encoding techniques can improve the digital transmission and storage of speech [19]–[25]. As brought to light in this research, a speech encoding method based on sinusoidal signal operation was proposed for use in LoRa speech applications enabling voice signal transmission. This method in wired and mobile environments has been omega encoded which allows optimal bandwidth and voice compression while not sacrificing power consumption and transmission distance with acceptable configuration needed for meaningful deployment. Thus, this research solves the limitation of voice communications over LoRa by proposing a new encoding method, which ultimately encodes speech into a few kbps allowing voice transmissions over LoRa systems. The research also provides a new encryption-decryption protocol tailored to LoRa hardware, and emphasizes speed and reliability while keeping cost low, and this was shown through testing.

## 2. RESEARCH METHOD

The sinusoidal model of physical representations of speech first presented in [26], created a voice coding method that capitalized on the properties of sine waves, amplitude, frequency, and phase, to allow for efficient speech processing. The method has been shown to allow for high fidelity speech reconstruction with a relatively low data rate [26]–[30]. In this framework, voice signal segment  $k$  can be expressed mathematically as a finite sum of sinusoidal components, where each component is specified by its amplitude, frequency, and phase:

$$s(n) = \sum_{k=1}^N A_k \cdot \sin(\Omega_k n + \phi_k) \quad (1)$$

In (1),  $A_k$ ,  $\Omega_k$ ,  $\phi_k$ , and  $N$  denote the amplitude, frequency, and phase of the  $k$ -th sinusoidal component respectively, while  $N$  indicates the total number of potential peaks. Another positive aspect of the sinusoidal coding scheme is that it can represent both voiced speech and unvoiced speech. It performs an analysis function by breaking the incoming speech signal into frames and then extracting parameters while analyzing each frame, which is why we call it an analysis function. The parameters extracted during the analysis are then employed at the receiving end to reconstruct the original speech segments.

### 2.1. Encoder stage

The encoding system works to parameter sets, and quantizes those parameters then converts them into binary format for transmission. The paradigm used strives to minimize the data rate necessary to represent the parameter sets while maintaining perceptual quality. The first step is sampling the speech at 8 kHz and segmenting it into many frames. Each frame is further classification into two distinct groups of voiced and unvoiced. Voiced segments will receive more peaks than unvoiced. In addition, the analysis of voiced segments has a further level of analysis which identifies these segments, and sub-classes the segments into  $N$  sub-segments based on energy classification. The sub-segments will also be classified as peeled or not peeled, for example, based on energy levels. The segment with a higher amount of energy will receive more peaks to represent versus the segment with the lower level of energy that will receive the lowest number of peaks. The intention of this level of classification strategy is to identify the most significant parameters with a major extract and returns that based on its highest and most complete representation. The encoding system is made up of two functional blocks that provide details on how the distinction and applications of both methodologies are expressed in the next few sections.

### 2.1.1. Sinusoidal peaks selection strategy

Signal stability is achieved through boundaries in which speech was divided into primary segments of 20-40 milliseconds. These segments were put through fast fourier transform (FFT) processing to expose sinusoidal peaks which are vital to reconstruct speech at the receiving end. One of the main challenges was properly identifying significant peaks and defining their properties, particularly what the optimum window length is to analyze the peaks. It is imperative to weigh the lengths of the component of the sinusoidal waves between two frames of reference. A short window length captures the rapid variations of the signal, and a long one allows for accurate measurements of frequency and separation of closely situated sinusoids. The analytic portion incorporated a windowing function of Hanning. The Hanning window was selected not only for its side lobe reduction capabilities but also for the improvement of speech quality. The sinusoidal components presented were taken from actual signals from short-time fourier transform (STFT) processing, which scanned every 20-40 milliseconds, extracting every peak across many local maxima. The magnitude peaks in STFT, represent sinusoidal components. This methodology of sinusoidal extraction is a standard practice and has been established in many audio compression methods, as long as they protect the bit rate and selected components to create the most efficient data. Threshold based peak detection also optimized system performance where only maxima over a defined threshold would be quantified as legitimate peaks.

An additional organizational layer uses and groups the primary frames, while dividing each frame into 6 smaller parts. Peak identification involves examining transitions in the spectral gradient from high to low values. Each peak is parabolically fitted, a fitting process illustrated in the parabola's vertex determining the exact frequency. If everything goes well, the above description generally produces about eighty peaks, which are then pragmatically reduced to minimize perceivable information loss. The last processing stage simply needs the frequencies of the identified peaks and their critical phase values, to be quantified prior to transmission. A scheme of this type allows for efficient encoding that possesses the essential characteristics necessary for speech reconstruction.

### 2.1.2. Optimization of data rate

The encoding scheme takes a targeted approach and seeks to compress only the most important peaks possible. The approach is to segment the speech into smaller pieces, and the model incorporates a hierarchical classification structure. The overall architecture of the model is described in Figures 1 (a) and (b). For the purposes of grouping and encoding data, speech is first segmented into frames (primary frames) and classified using an energy-based threshold criterion which classifies frames as voiced or unvoiced. Any frames that had energy above the threshold are classified as voiced, and frames that were below the threshold are classified as unvoiced. For voiced segments, the algorithm then processes a voiced frame and segments each frame into N sub-segments and uses another energy-based classification for the sub-segments. A higher energy sub-segment would receive a higher number of peaks to encode. The unvoiced frames use the segmented sub-segments through the same manner; however, there is no way to distinguish the energy levels of the sub-segments - each sub-segment for every unvoiced frame is given the peak allocation from the lowest energy allocation determined for the voiced frames. The segmenting to segmentation and classification enables identification of the most important peaks as possible and encode them while also compressing data while maintain reconstruction of speech.

The model heavily emphasizes parameter reduction as a means to minimize compression-related errors. The system targets a reduction in frame parameters to between 15 and 30 per main frame. Beyond these initial reduction methods, additional compression is achieved during quantization. The encoding process implements three distinct levels of processing following the frame classification and segmentation:

- Peak reduction: focuses on minimizing the number of peaks to reduce redundant data.
- Phase reduction: simplifies phase information without significantly affecting quality.
- Threshold reduction: implements a threshold to eliminate minor peaks, retaining only the most critical ones.

#### a. Number of peaks minimization

The emphasis of this phase is to determine the N most powerful harmonic components within each voiced segment, where N is specified by the target bitrate. These notably include two distinct steps: first, the spectral decomposition of the segments will be performed to determine the dominant frequency components in that segment. Then, if there is more than one amplitude peak in the vicinity, the algorithm will only keep the largest one in order to avoid redundant information and accurately represent the spectral information. This methodical approach allows for high-fidelity reproduction of voice in the decoded output and provides a strong foundation for the next encoding phases.

#### b. Phase minimization

This process intends to enhance phase related parameter extraction, by differentiating sub-frames into voiced and unvoiced sub-frames. Voiced sub-frames display three attributes as evidence of voicedness: i) they

are above a certain amplitude level, exhibit a lower-than-normal frequency of zero-crossings, and they include identifiable pitched characteristics typical of voiced phonemes; ii) the energetic magnitude is designated as the primary metric for classification: we use a simple binary classification process. In the case of voiced sub-frames displaying considerable energy, phase extraction is direct; when the frame is unvoiced, then well-used phase extraction methods may be used, such as methods in [27]-[30]. In this way, it is possible to reduce the number of phase parameters while retaining perceptual quality in speech, since the human ear is essentially insensitive to non-ideal phases.

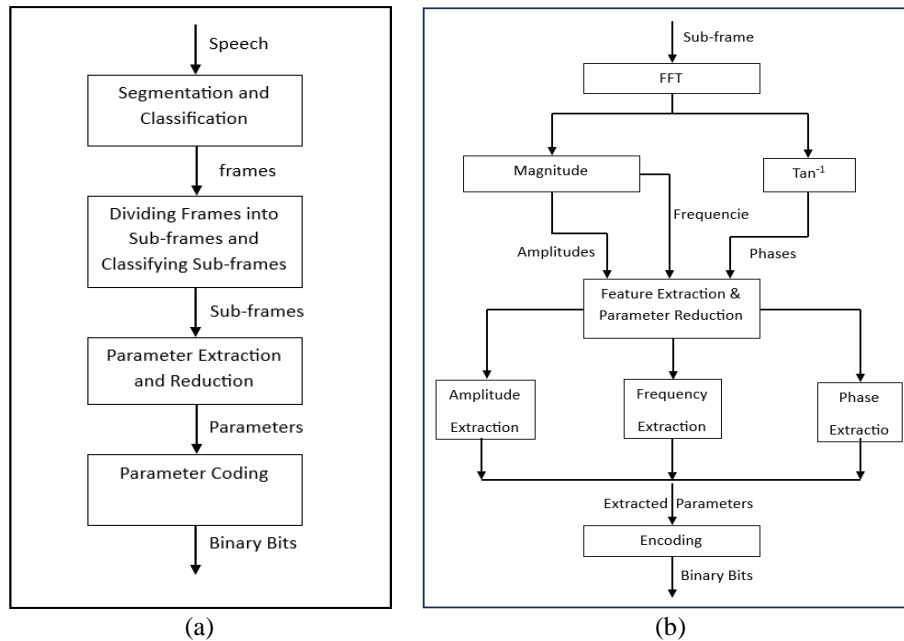


Figure 1. System's architecture: (a) encoder stage and (b) parameter extraction and reduction stage

### c. Threshold optimization

This stage represents the most practical of all previous reduction methods, since it is designed to eliminate the number of sinusoidal components as the fidelity of the voice is maintained. The method employs a threshold based on magnitude, eliminating any peaks below this threshold value. Because each amplitude has a corresponding frequency and phase, this pruning process reduces not only the amplitude values, but also their frequency and phase, thus reducing the magnitude of all the parameters in the model. This process will have two direct benefits. The first benefit is a dramatic reduction in the amount of bandwidth that will be required for transmitting the signal. The second benefit is that the quality of the reconstructed waveform will improve, by removing components that contribute nothing to the voice other than random noise. The filtering will improve the clarity of the signal, but when establishing the threshold value, there needs to be careful consideration to avoid using a threshold that is too high and thus risks eliminating letter- and word-contained spectral components. Because of this, the threshold must be chosen carefully and will most often require a statistical determination of the range of values used to set the threshold.

After the reduction, each primary frame has  $S$  amplitudes,  $S$  frequency components, and  $0.5 S$  phases. This means that there are  $S$  peaks,  $S$  frequencies, and only half as many phases per frame. This implementation uses 6 bits for amplitude and frequency parameter values, while it records 4 bits for each phase parameter. The frame-specific data rate is calculated as  $(6 \times (S + S) + 4 \times (0.5 S)) = 14 S \text{ bits/frame}$ . The system's overall data rate ( $R$ ) is then derived as:  $R = 14 S \text{ bits/frame} \times N \text{ frames/s} = 14 NS \text{ bps}$ . The system may require additional bit allocation for control functions and error management. At this stage, the quantization process becomes crucial, carrying equal weight in ensuring efficient signal encoding and transmission.

### 2.1.3. Modeling and encoding

The quantization process divides the amplitude range of the signal into finite regions. The formula  $W = 2^b$  is used to determine the number of quantization steps, where  $b$  is the number of bits and  $W$  is the number of quantization steps. The system subsequently maps the signal value to the nearest quantization step and then converts to its binary value according to pulse code modulation (PCM). For the system under

discussion, the amplitude, frequency, and phase components of each sinusoid undergo specific quantization procedures, which are elaborated in the following sections.

a. Phase modeling and encoding

The system optimizes phase component bit allocation through entropy reduction of phase values. This optimization utilizes differential prediction techniques, where future phase states are estimated from historical data. Rather than encoding absolute phase values, the system processes phase differentials, which characteristically exhibit reduced entropy compared to unprocessed phase data [27]. The predicted phase follows this mathematical relationship:

$$\hat{\phi}_l^n = \phi_l^{n-1} + \Omega_l^n T \quad l = 1, 2, \dots, L, \quad (2)$$

where  $n$  indicates frame position, represents the  $l$ th sinusoidal component,  $T$  denotes frame duration, and  $L$  signifies the total sinusoidal count. The system then calculates phase differentials, or residues, using:

$$\Delta\phi_l^n = \phi_l^n - \hat{\phi}_l^n \quad l = 1, 2, \dots, L \quad (3)$$

In this framework, initial phase values serve as the basis for computing phase differentials, which are fundamental to residue determination during encoding.

b. Frequency encoding

Following the conversion of speech segments to frequency representation via STFT, frequency components are expressed as integers. Consider a MATLAB implementation: within a 256-sample STFT frame, frequencies map to integer values ( $S_n$ ) from 1 to 256, corresponding to the physical frequency range ( $f_n$ ) of 0 to 4000 Hz. The relationship between these values follows:

$$f_n = \frac{(S_n - 1) \times 4000}{STFT \text{ frame}} \quad (4)$$

In (4),  $S_n$  represents the integer frequency value, with the dimension (such as 256) defining frequency resolution granularity. While conventional systems typically allocate 8 bits per frequency value ( $f_n$ ), this model achieves comparable results with just 6 bits. The system takes the previous refinement of bit allocation further based on the perceptual significance of individual frequency bands. In this system, lower frequency components are more affected by the first frequencies, while higher frequencies are affected by the last frequencies. Because higher frequencies contribute less to the overall perceptual quality of speech, the same bit allocation across all frequencies is not needed. Therefore, the system allocates less bits to the higher frequency components than to the lower frequency. The adaptive bit allocation to achieve a reduction in bandwidth while maintaining a similar quality of speech occurs as follows: First, the system normalizes the frequency components ( $S_n$ ) through division by the STFT frame dimension, yielding a normalized frequency vector ( $z_n$ ). Next, the system transforms these normalized frequencies ( $z_n$ ) into a compressed representation ( $r_n$ ) to minimize encoding bits per frequency value. This compression follows the relationship:

$$r_n = \frac{64 (\log_e(1 + 4z_n))}{1.6} \quad (5)$$

This transformation constrains the values to between 1 and 64, following a structure of a formula similar to the  $\mu$ -Law compression method used in digital speech processing. All traditional processes are followed by the last part of the process: quantization, where the rounded values will be integers and binary for transmission. Thus, the compression of frequency values happens while retaining significant precision characteristics to decrease bandwidth usage, and adhere to the model's design function.

c. Amplitude encoding

Sinusoidal amplitudes are crucial components that are highly sensitive to variations during the quantization process. To address this, we propose an advanced encoding technique that significantly enhances resolution by a factor of 6 to 12 compared to traditional PCM. Given the amplitude vector  $x_n = [x_0 x_1 \dots x_{N-1}]$ , where  $N$  represents the number of peaks, this technique systematically processes amplitudes to optimize their encoding:

1. Logarithmic transformation: compute the base-2 logarithm of each amplitude ( $x_n$ ).
2. Dynamic range adjustment: take the absolute value of the result from the previous step and multiply it by  $\gamma$  to map the values to a dynamic range of (1-64).
3. Amplitude extraction and sorting: extract the amplitudes ( $p_n$ ) using (6) and arrange the ( $p_n$ ) values in ascending order, pairing them with their associated phases and frequencies as a set.

4. Binary conversion of initial amplitude: convert the integer part (floor value) of the first amplitude  $p_0$  into binary, denoted as  $b_0$ .
5. Difference encoding: for each amplitude  $p_n$ , calculate the difference ( $q_n$ ) by subtracting  $p_n$  from the integer part of the preceding amplitude ( $\lfloor p_{n-1} \rfloor$ ).
6. Resolution scaling: choose a scaling factor  $\alpha$  in the range (6-12), and multiply it by the result of the previous step ( $q_n$ ).
7. Quantization and binary conversion: take the floor of the scaled value and convert it into binary, denoted as  $b_{nb\_nbn}$ .
8. Repeat the process: repeat steps (5–7) until all  $p_n$  values have been processed.

The logarithmic conversion of unity-bounded amplitudes generates negative values spanning from  $-1$  to  $-20$ , with  $10^{-6}$  serving as the defined threshold. The system employs a scaling coefficient to modify this range and processes the differential between consecutive amplitude values, exploiting their characteristically small intra-frame variations. The algorithm then sequentially orders these amplitudes from lowest to highest, maintaining their phase and frequency associations to facilitate efficient encoding. This methodical process optimizes amplitude data transmission while preserving resolution and minimizing compression-induced information loss. The amplitude quantization general equations are given by:

$$p_n = \gamma |\log_2(x_n)| \quad (6)$$

$$b_0 = \text{Binary } \lfloor p_0 \rfloor \quad (7)$$

$$b_n = \text{Binary } \lfloor \alpha (p_n - \lfloor p_{n-1} \rfloor) \rfloor \quad (8)$$

## 2.2. Decoder stage

The reconstruction process, depicted in Figure 2, employs a synthesis approach where the decoder regenerates speech segments through the superposition of sinusoidal components. Each sinusoid incorporates its unique combination of amplitude, frequency, and phase parameters, as specified by the encoded data stream from the earlier processing stage.

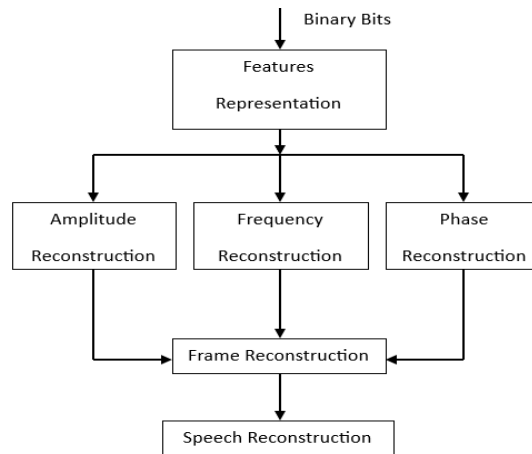


Figure 2. Speech reconstruction process

### 2.2.1. Decoding strategy

The reconstruction process initiates with the conversion of binary-encoded parameters to decimal values. Signal restoration occurs across three distinct levels, addressing amplitude, frequency, and phase components separately. This layered approach minimizes reconstruction error between original and recovered parameters, enabling high-quality signal synthesis.

#### a. Phase component reconstruction

The phase recovery sequence involves:

1. Applying dequantization to the encoded phase differential bits
2. Computing predicted phase values from historical data using (2)
3. Combining the dequantized phase differential with the predicted phase estimate

## b. Frequency component reconstruction

The frequency restoration follows this sequence:

1. Converting frequency-encoded binary data into decimal representation ( $\hat{r}_n$ )
2. Applying the following relationship to recover the normalized frequency vector ( $\hat{z}_n$ ) from ( $\hat{r}_n$ ):

$$\hat{z}_n = \frac{\exp(0.025 \hat{r}_n) - 1}{4} \quad (9)$$

3. Implementing rounding on ( $\hat{z}_n$ )

This process employs (9), which functions as the mathematical inverse of (5) from the encoding stage.

## c. Amplitude decoding

The amplitude reconstruction process begins by transforming binary parameters [ $b_0, b_1, \dots, b_{N-1}$ ] into their decimal equivalents [ $p_0, p_1, \dots, p_{N-1}$ ]. These values undergo further processing through the following relationship:

$$\begin{aligned} d_0 &= p_0 \\ d_1 &= p_0 + \frac{p_1}{\alpha} \\ d_2 &= p_0 + \frac{p_1}{\alpha} + \frac{p_2}{\alpha} \\ d_n &= p_0 + \sum_{i=1}^n \frac{p_i}{\alpha} \end{aligned} \quad (10)$$

where N represents the total peak count. Though this transformation introduces maximum error when  $n = 0$ , this deviation remains insignificant and doesn't materially impact reconstruction fidelity. The final amplitude values ( $y_n$ ) are then recovered using:

$$y_n = -2^{\left(\frac{d_n}{\gamma}\right)} \quad (11)$$

### 2.2.2. Advantages of the proposed Speech coding technique

The analysis of the preceding methodology reveals several key benefits of this speech encoding approach: (i) incorporates highly efficient and effective encoding and decoding processes, (ii) delivers a reconstructed speech signal of exceptional quality at the receiver's end, (iii) achieves a reduced data rate, ranging from 3.6 to 8 kbps, (iv) enhances the quality of the reconstructed signal, even in noisy environments, (v) functions independently of the fundamental pitch of the speech signal, (vi) demonstrates strong resilience to noise interference, (vii) significantly reduces power consumption and the bit rate required for transmission, and (viii) supports the integration of error detection and correction mechanisms for enhanced reliability.

## 3. RESULTS AND DISCUSSION

The implementation utilizes MATLAB for initial signal processing, where speech input undergoes compression and encryption before transfer to an Arduino platform. The Arduino interfaces with a LoRa transceiver, which employs CSS modulation techniques to facilitate extended-range signal propagation. The receiving system includes packet detection and acknowledgment protocols and returns acknowledgment when successful. This two-way acknowledgment protocol improves reconstructed speech quality while keeping the bandwidth at or below 8 kbps. The LoRa protocol is capable of transmission rates up to 22 kbps. In our needs, 22 kbps should be more than sufficient. However, the system uses the principle of distance and tradeoffs by being able to use low transmission speeds that allows for better communication distances. The limitations of data rates maximize performance based on varying receiving and max transmission speeds based on adjustable spreading factors (SF) and bandwidth. The increased SF for example allows for better distances, but gives up transmission data transfer completely. Whereas wider bandwidth can increase transmission data but give up on range. The main goal of the system is to keep speech transmission rate below the 22 kbps rate, so the max distance can be achieved. The system will use a LoRa based SF to increase transmission range, especially when using emergency contact systems where it is important to ensure long distance and reliability.

### 3.1. System block diagram

The block diagrams presented in Figure 3 illustrate the setup where the microcontroller is interfaced with both MATLAB and the LoRa module. The configuration on the transmitter side is identical to that of the receiver. While the transmitter and receiver share similar code structures, their functionalities are distinct: the transmitter as shown in Figure 3(a) is responsible for sending and analyzing the data, whereas the receiver as shown in Figure 3(b) focuses on receiving and synthesizing the communicated information.

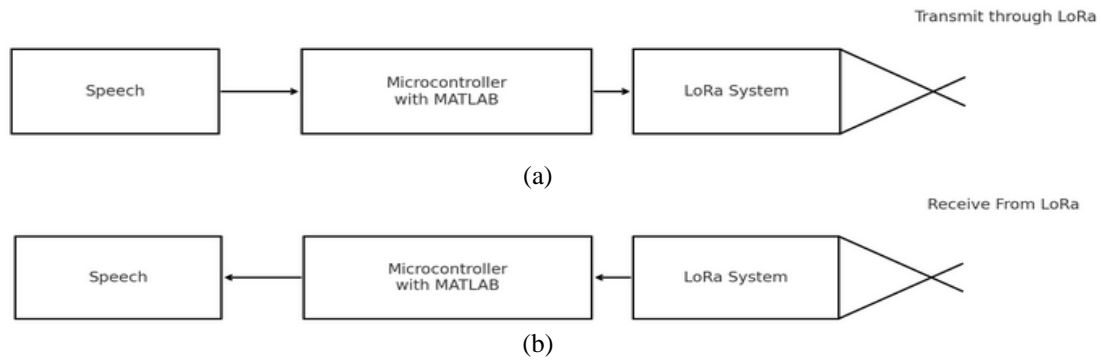


Figure 3. System block diagram: (a) transmitter and (b) receiver

### 3.2. Printed circuit board

Following the prototype phase on breadboard, development progresses to printed circuit board (PCB) fabrication for an individual LoRa transceiver unit, depicted in Figure 4. The implementation of MIMO capabilities would necessitate an alternate board design. The PCB layout illustrated in Figure 4 depicts a single module configuration, featuring integrated ground connectivity and designated connection points for each pin. Each connection point incorporates a copper pathway terminating in a through-hole, enabling versatile connectivity options via male or female connectors. The complete system architecture for MATLAB-Arduino communication is documented in Figures 5 and 6, which detail the transmitter and receiver circuitry respectively.

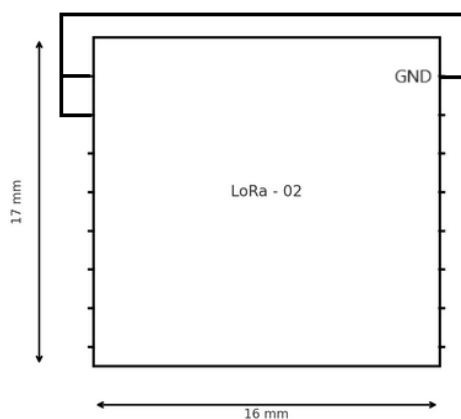


Figure 4. Single LoRa PCB module

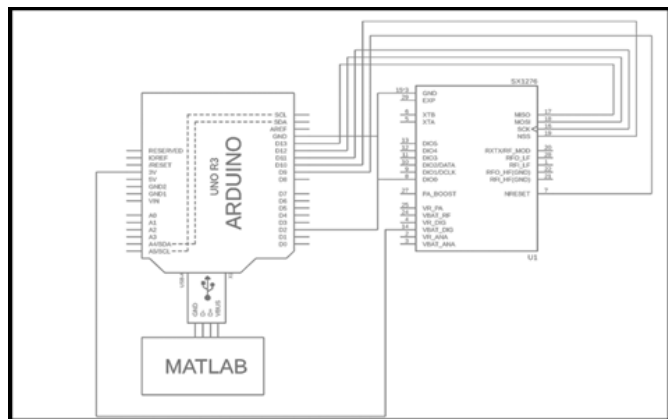


Figure 5. Complete system architecture for transmitter

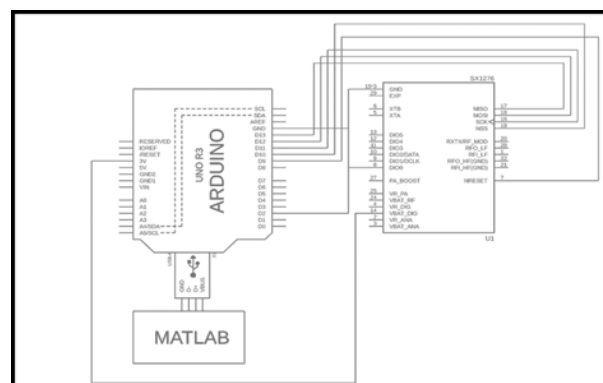


Figure 6. Complete system architecture for receiver



### 3.3. Encryption

The coding methodology outlined in Section 2 achieved efficient data rate reduction, optimizing transmission requirements. The system processes data in 128-bit packet structures, with an example message  $m$ :

$$m = 10100111 \ 00110101 \ 11010010 \ 00111100 \ 10101010 \ 11001100 \ 01011101 \ 11110000 \ 01110011 \ 10011010 \ 11001101 \ 00101011 \ 11100001 \ 00111100 \ 10101011 \ 01011101 \quad (12)$$

Using the encryption key  $k$ :

$$k = 11001100 \ 10101011 \ 00111100 \ 11001101 \ 01011010 \ 00111100 \ 10101011 \ 00110011 \ 10011100 \ 01100111 \ 10101010 \ 11110000 \ 00111100 \ 11001101 \ 01011010 \ 10101100 \quad (13)$$

The encryption process combines  $m$  and  $k$  through XOR operations, producing the encrypted output  $y$ :

$$y = 01101011 \ 10011110 \ 11101110 \ 11110001 \ 11110000 \ 11110000 \ 11110110 \ 11000011 \ 11101111 \ 11111101 \ 01100111 \ 11011011 \ 11011101 \ 11110001 \ 11110001 \ 11110001 \quad (14)$$

At the receiving end, applying the same key  $k$  from (13) to the encrypted message  $y$  from (14) via XOR operations recovers the original message:

$$m = 10100111 \ 00110101 \ 11010010 \ 00111100 \ 10101010 \ 11001100 \ 01011101 \ 11110000 \ 01110011 \ 10011010 \ 11001101 \ 00101011 \ 11100001 \ 00111100 \ 10101011 \ 01011101 \quad (15)$$

In this system,  $m$  represents the initial message,  $k$  denotes the shared cryptographic key, and  $\oplus$  symbolizes the XOR operation. The symmetrical nature of the XOR operation, combined with key reuse for encryption and decryption, ensures secure and accurate message recovery at the receiver.

### 3.4. Implementation

Employing LoRa for voice communication is a novel approach to utilizing the technology beyond sensor data applications. The system mitigates bandwidth usage through advanced coding methods for speech. The system architecture starts with MATLAB for signal processing that includes encryption and compression stages prior to Arduino implementation. The architecture divides the captured processed data into packets using LoRa hardware, which with CSS modulation for data transmission. The receiver architecture includes a method for packet detection and acknowledgment to indicate receipt of the packets, following on with the Arduino implementing processing to reverse the original signal audio waveform. A PCB was developed to demonstrate the optimization of the system, concluding that LoRa viability for voice transmission has merit in voice applications. The transmission architecture discussed in section 2 consists of multiple processing stages that convert acoustic signals into encrypted binary streams. The discussed MATLAB environment provided the connection to acquire and process the signals without hardware adaptation. Initial development steps were to determine the level of bits needed for two-way communication. The system basic voice signals bandwidth requires compression algorithms as shown in Section 2 to ensure signals can be transmitted and reconstructed.

This research overlaps with efforts in LoRa alliance and IoT. The primary aim of the project is to develop a reliable, adaptive speech coding and secure cryptographic protocols for the transmission of voice over LoRa. Implementation was undertaken to promote reliability, security, and access within the IoT environment by decreasing the data rate and encrypting the data. The main goal of a LoRa systems was focused on being cost-effective and utilizing resources effectively, demonstrated through the experiments. These forms of innovations will allow the capabilities of LoRa networks to be employed into a range of applications including municipal infrastructure, environmental sensor systems and health care systems.

## 4. CONCLUSION

The system architecture presents a new method for safe and also low-cost long-distance voice communication and is designed for the IoT ecosystem. It leverages the CSS modulation technique of LoRa technology, which is well known for its standard performance in maximizing distance with power. LoRa is limited in bandwidth and throughput is limited to a few kilobits per second. The system architecture implements speech encoding algorithms as part of the voice encoding process to minimize the bandwidth needed for voice data to accommodate the lower data transmission rates associated with LoRa. The secured voice communications made possible by the system architecture fall under a framework of integrated encryption-decryption security for safeguarding the contents of the transmission.

## ACKNOWLEDGMENTS

The authors would like to express their gratitude to the German Jordanian University, the American University of the Middle East, and Aeon Medical Supplies Company for their generous support of this work.

## FUNDING INFORMATION

This work was funded by Seed Grant number SAMS 03/2023 from the German Jordanian University, under decision number 2023/15/344. Additional support was provided by a grant from Aeon Medical Supplies Company, by the American University of the Middle East, and by the Scientific Research and Innovation Support Fund (SRISF), Jordan, under project number SE-MPH/4/2024. The authors express their gratitude for these generous supports.

## AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Samer Alabed	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Mohammad Al-Rabayah				✓		✓			✓	✓		✓	✓	✓
Bahaa Al-Sheikh				✓		✓	✓			✓	✓	✓		
Lama Bou Farah				✓		✓	✓			✓	✓	✓		

C : Conceptualization

M : Methodology

So : Software

Va : Validation

Fo : Formal analysis

I : Investigation

R : Resources

D : Data Curation

O : Writing - Original Draft

E : Writing - Review & Editing

Vi : Visualization

Su : Supervision

P : Project administration

Fu : Funding acquisition

## CONFLICT OF INTEREST STATEMENT

The authors state no conflict of interest.

## DATA AVAILABILITY

Data availability is not applicable to this paper as no new data were created or analyzed in this study.




## REFERENCES

- [1] G. E. John, "A low cost wireless sensor network for precision agriculture," in *2016 Sixth International Symposium on Embedded Computing and System Design (ISED)*, Dec. 2016, pp. 24–27, doi: 10.1109/ISED.2016.7977048.
- [2] F. E. M. Covenas, R. Palomares, M. A. Milla, J. Verastegui, and J. Cornejo, "Design and development of a low-cost wireless network using IoT technologies for a Mudslides monitoring system," in *2021 IEEE URUCON*, Nov. 2021, pp. 172–176, doi: 10.1109/URUCON53396.2021.9647379.
- [3] D. Taleb, S. Alabed, and M. Pesavento, "Optimal general-rank transmit beamforming technique for single-group multicasting service in modern wireless networks using STTC," in *WSA 2015; 19th International ITG Workshop on Smart Antennas*, 2015, pp. 1–7.
- [4] M. E. I. bin Edi, N. E. Abd Rashid, N. N. Ismail, and K. Cengiz, "Low-cost, long-range unmanned aerial vehicle (UAV) data logger using long range (LoRa) module," in *2022 IEEE Symposium on Wireless Technology & Applications (ISWTA)*, Aug. 2022, pp. 1–7, doi: 10.1109/ISWTA55313.2022.9942751.
- [5] W. San-Um, P. Lekbunyasini, M. Kodyoo, W. Wongsuwan, J. Makfak, and J. Kerd Sri, "A long-range low-power wireless sensor network based on U-LoRa technology for tactical troops tracking systems," in *2017 Third Asian Conference on Defence Technology (ACDT)*, Jan. 2017, pp. 32–35, doi: 10.1109/ACDT.2017.7886152.
- [6] S. Alabed, "Performance analysis of two-way DF relay selection techniques," *ICT Express*, vol. 2, no. 3, pp. 91–95, Sep. 2016, doi: 10.1016/j.ict.2016.08.008.
- [7] S. Alabed, "Performance analysis of bi-directional relay selection strategy for wireless cooperative communications," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, no. 1, Dec. 2019, doi: 10.1186/s13638-019-1417-1.
- [8] P. Lu, "Design and implementation of coal mine wireless sensor Ad Hoc network based on LoRa," in *2022 3rd International Conference on Information Science, Parallel and Distributed Systems (ISPDS)*, Jul. 2022, pp. 54–57, doi: 10.1109/ISPDS56360.2022.9874124.
- [9] S. Mishra, S. Nayak, and R. Yadav, "An energy efficient LoRa-based multi-sensor IoT network for smart sensor agriculture system," in *2023 IEEE Topical Conference on Wireless Sensors and Sensor Networks*, Jan. 2023, pp. 28–31, doi: 10.1109/WISNeT56959.2023.10046242.
- [10] A. I. Ali, S. Z. Partal, S. Kepke, and H. P. Partal, "ZigBee and LoRa based wireless sensors for smart environment and IoT




- applications,” in *2019 1st Global Power, Energy and Communication Conference (GPECOM)*, Jun. 2019, pp. 19–23, doi: 10.1109/GPECOM.2019.8778505.
- [11] Y.-W. Ma and J.-L. Chen, “Toward intelligent agriculture service platform with lora-based wireless sensor network,” in *2018 IEEE International Conference on Applied System Invention (ICASI)*, Apr. 2018, pp. 204–207, doi: 10.1109/ICASI.2018.8394568.
  - [12] S. Yogalakshmi and R. Chakaravathi, “Development of an efficient algorithm in hybrid communication for secure data transmission using LoRa technology,” in *2020 International Conference on Communication and Signal Processing (ICCS)*, Jul. 2020, pp. 1628–1632, doi: 10.1109/ICCS.2020.9182233.
  - [13] Z. AlArnaout, N. Mostafa, S. Alabed, W. H. F. Aly, and A. Shdefat, “RAPT: A robust attack path tracing algorithm to mitigate SYN-flood DDoS cyberattacks,” *Sensors*, vol. 23, no. 1, Dec. 2022, doi: 10.3390/s23010102.
  - [14] W.-T. Sung, S.-J. Hsiao, S.-Y. Wang, and J.-H. Chou, “LoRa-based internet of things secure localization system and application,” in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, Oct. 2019, pp. 1672–1677, doi: 10.1109/SMC.2019.8913875.
  - [15] J. Xing, L. Hou, K. Zhang, and K. Zheng, “An improved secure key management scheme for LoRa system,” in *2019 IEEE 19th International Conference on Communication Technology (ICCT)*, Oct. 2019, pp. 296–301, doi: 10.1109/ICCT46805.2019.8947215.
  - [16] A. Iqbal and T. Iqbal, “Low-cost and secure communication system for remote micro-grids using AES cryptography on ESP32 with LoRa module,” in *2018 IEEE Electrical Power and Energy Conference (EPEC)*, Oct. 2018, pp. 1–5, doi: 10.1109/EPEC.2018.8598380.
  - [17] P. Edward, S. Elzeiny, M. Ashour, and T. Elshabrawy, “On the coexistence of LoRa-and interleaved chirp spreading LoRa-based modulations,” in *2019 International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, Oct. 2019, pp. 1–6, doi: 10.1109/WiMob.2019.8923211.
  - [18] I. B. F. de Almeida, M. Chafii, A. Nimr, and G. Fettweis, “In-phase and quadrature chirp spread spectrum for IoT communications,” in *GLOBECOM 2020 - 2020 IEEE Global Communications Conference*, Dec. 2020, pp. 1–6, doi: 10.1109/GLOBECOM42002.2020.9348094.
  - [19] F. I. Abro, F. Rauf, M. Batool, B. S. C. Dhry, and S. Aslam, “An efficient speech coding technique for secure mobile communications,” in *2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, Nov. 2018, pp. 940–944, doi: 10.1109/IEMCON.2018.8614855.
  - [20] B. Bessette, R. Salami, R. Lefebvre, and M. Jelinek, “Efficient methods for high quality low bit rate wideband speech coding,” in *Speech Coding, 2002, IEEE Workshop Proceedings.*, 2002, pp. 114–116, doi: 10.1109/SCW.2002.1215742.
  - [21] J. Hai and Er Meng Joo, “Improved linear predictive coding method for speech recognition,” in *Fourth International Conference on Information, Communications and Signal Processing, 2003 and the Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint*, 2003, vol. 3, pp. 1614–1618, doi: 10.1109/ICICS.2003.1292740.
  - [22] T. Moriya, Y. Kamamoto, and N. Harada, “Enhanced lossless coding tools of LPC residual for ITU-T G.711.0,” in *2010 Data Compression Conference*, 2010, pp. 546–546, doi: 10.1109/DCC.2010.71.
  - [23] X. Yu, X. You, X. Liu, and C. Li, “An improved algorithm for residual signal excitation based on LPC 10,” in *2022 7th International Conference on Communication, Image and Signal Processing (CCISP)*, Nov. 2022, pp. 318–323, doi: 10.1109/CCISP55629.2022.9974264.
  - [24] F. A. Muin, T. S. Gunawan, E. M. A. Elsheikh, and M. Kartiwi, “Performance analysis of IEEE 1857.2 lossless audio compression linear predictor algorithm,” in *2017 IEEE 4th International Conference on Smart Instrumentation, Measurement and Application (ICSIMA)*, Nov. 2017, pp. 1–6, doi: 10.1109/ICSIMA.2017.8312033.
  - [25] S. Alabed, A. Alsaraira, N. Mostafa, M. Al-Rabayah, Y. Kotb, and O. A. Saraereh, “Implementing and developing secure low-cost long-range system using speech signal processing,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 31, no. 3, pp. 1408–1419, Sep. 2023, doi: 10.11591/ijeecs.v31.i3.pp1408-1419.
  - [26] A. S. Spanias, “Speech coding: a tutorial review,” *Proceedings of the IEEE*, vol. 82, no. 10, pp. 1541–1582, 1994, doi: 10.1109/5.326413.
  - [27] R. McAulay and T. Quatieri, “Speech analysis/Synthesis based on a sinusoidal representation,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 4, pp. 744–754, Aug. 1986, doi: 10.1109/TASSP.1986.1164910.
  - [28] R. McAulay and T. Quatieri, “Processing of acoustic waveforms,” Patent No. Re.36, 478, Assignee: Massachusetts Institute of Technology, Cambridge, Mass, 1999.
  - [29] S. Ahmadi and A. S. Spanias, “New techniques for sinusoidal coding of speech at 2400 bps,” in *Conference Record of The Thirtieth Asilomar Conference on Signals, Systems and Computers*, 1996, vol. 1, pp. 770–774, doi: 10.1109/ACSSC.1996.601158.
  - [30] S. Ahmadi and A. S. Spanias, “Low bit-rate speech coding based on an improved sinusoidal model,” *Speech Communication*, vol. 34, no. 4, pp. 369–390, Jul. 2001, doi: 10.1016/S0167-6393(00)00057-1.

## BIOGRAPHIES OF AUTHORS






**Samer Alabed**    is a Professor of Biomedical and Electrical Engineering and Director of Accreditation and Quality Assurance Center at German Jordanian University (GJU). He served as a Director of E-learning and Academic Performance Improvement Center, Head of Biomedical Engineering Department and Exchange Coordinator at GJU and was an Associate Professor at American University of the Middle East, Kuwait, from 2015 to 2022. He also studied and worked in Darmstadt University of Technology, Darmstadt, Germany, from 2008 to 2015. He can be contacted at email: Samer.Alabed@gu.edu.jo.






**Mohammad Al-Rabayah**    earned his Ph.D. in Electrical Engineering from University of New South Wales. He is serving as an Associate Professor at American University of the Middle East (AUM), where he teaches a variety of courses, advises students, and contributes to departmental excellence. He was an Assistant Professor at Prince Sultan University. During his tenure, he taught courses in the Communication and Networks Department and played a key role in supporting the university's efforts to achieve ABET accreditation. He can be contacted at email: [Mohammad.Alrabayah@aum.edu.kw](mailto:Mohammad.Alrabayah@aum.edu.kw).



**Bahaa Al-Sheikh**    received Ph.D. in Biomedical Engineering from University of Denver, Colorado, USA. From 2009 to 2015, he worked for Yarmouk University as an Assistant Professor and served as the department chairman from 2010 to 2012. He worked as an Assistant and Associate Professor at the American University of the Middle East in Kuwait from 2015 to 2019. He works now as an Associate Professor at Yarmouk University. He can be contacted at email: [alsheikh@yu.edu.jo](mailto:alsheikh@yu.edu.jo).



**Lama Bou Farah**    is currently an Associate Professor, Head of the Biomedical Technologies Department, Acting Head of the Medical Imaging, Department and Academic Quality Coordinator at the Lebanese German University. She received her Ph.D. degree in Biomedical Engineering and Advanced Medicine from Macquarie University, Australia. She has worked as an associate professor, assistant professor, researcher, and lecturer in several international universities and supervised tens of master students and several Ph.D. students. She received several awards and grants. She can be contacted at email: [l.boufarah@lgu.edu.lb](mailto:l.boufarah@lgu.edu.lb).