ChirpMu: Chirp Based Imperceptible Information Broadcasting with Music

Yu Wang*, Xiaojun Zhu*, Hao Han*
*College of Computer Science and Technology,
Nanjing University of Aeronautics and Astronautics,
Nanjing 211106, China
Email: {wongyu, xzhu, hhan}@nuaa.edu.cn

2021 IEEE/ACM 29th International Symposium on Quality of Service (IWQOS) | 978-1-6654-1494-4/21/$31.00 ©2021 IEEE | DOI: 10.1109/IWQOS52092.2021.9521353

***Abstract*—This paper presents ChirpMu, a system that encodes information into chirp symbols and embeds the symbols with music. Users enjoy music without realizing the existence of chirp sounds, while their smartphones can decode the information. ChirpMu can be used to broadcast information such as Wi-Fi secrets or coupons in shopping malls. It features novel chirp symbol design that can combat sound attenuation and environment noise. In addition, ChirpMu properly adjusts the portion of chirp symbols with music, so that the chirp symbols cannot be heard by users but can be decoded by smartphones with low error rate. Various real-world experiments show that ChirpMu can achieve low bit error rate.**

***Index Terms*—chirp, wireless communication, acoustic communication, imperceptible**

Fig. 1. Shopping malls send coupons through ChirpMu.

I. Introduction

Acoustic data transmission technology provides an attractive means to broadcast information to co-located mobile devices without addition communication devices, and hence is widely adopted in many location-based service (LBS) applications. For example, in a shopping mall, store vendors can use data-over-music to broadcast electronic coupons to those customers who are in close vicinity of the store as shown in Fig. 1. Wi-Fi secrets can be broadcast via acoustic signals to reduce free riders [1]. Compared to other wireless communication technologies such as Bluetooth, NFC (near-field communication) and screen-camera communication like QR code, the advantages of the acoustic approach are in terms of transmission distance and coverage.

By investigating existing acoustic data transmission schemes [2]–[4] in the literature, we raise an open question *whether acoustic data transmission can be achieved within certain range of aerial space (saying, 5 meters away) between a speaker and microphones on COTS (commercial off-the-shelf) smartphones without human perception* (i.e., being inaudible or hiding audio data in background music). We are mainly interested in smartphones due to its portability [5]. Since COTS smartphones do not support capturing ultrasound signals (frequency above 20kHz), the information can only be encoded in audible (less than 16kHz) or near-inaudible (16kHz-20kHz) frequencies. To ensure that users do not notice the sounds, information encoded in audible frequencies can only slightly change frequency or amplitude of the background sound [2]. For near-inaudible frequencies, the energy of the encoded sounds shall be small. In both cases, the effective transmission range of existing inaudible acoustic data transmission approaches is usually less than 1m (e.g, 30cm in [3] and 20cm in [4]). The underlying reason is that the coding scheme is not robust to signal fading. Louder sound signals can propagate longer, but they may become audible by humans.

To address this question, this paper proposes ChirpMu, which uses chirp symbols to combat signal fading in near-inaudible frequencies. Chirp modulation has shown success in various scenarios to have robust properties against signal fading and noise, e.g., in LoRa [6]. Because it uses the slope of frequency change to modulate information, it can cope with Doppler effect caused by movements, which traditional frequency modulation cannot handle. We propose to apply chirp modulation to encode information as sounds, with near-inaudible frequencies. In contrast to [4], ChirpMu can achieve a transmission range of 5 meters. We also propose new chirp symbols, which are easier to decode than the traditional chirp symbols, and are suitable for implementation on smartphones.

Our contributions can be enumerated as follows.

1) We propose ChirpMu, which encodes information as chirp symbols in near-inaudible frequencies and mixes them with background music. ChirpMu can achieve longer transmission range than existing coding methods for acoustic communication in smartphones.
2) We propose new chirp symbols, which are easier to

Xiaojun Zhu is the corresponding author (xzhu@nuaa.edu.cn).

978-0-7381-3207-5/21/$31.00 ©2021 IEEE

Authorized licensed use limited to: NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS. Downloaded on January 03,2022 at 12:52:35 UTC from IEEE Xplore. Restrictions apply.

decode than traditional chirp symbols. We also propose a controllable method to adjust the portion of music and chirp symbols, so that the users do not notice the embedded chirp symbols and the bit error rate of chirp signals is low.
3) We conduct various experiments to evaluate ChirpMu.

The rest of the paper is organized as follows. Section II reviews related works. Section III presents the basic idea and challenges. Section IV gives the design of ChirpMu and Section V evaluates it. Section VI concludes the paper.

II. Related Works

Many works employ acoustic channel for device to device communication. There are generally two categories according to whether the frequencies are audible.

A. Acoustic communication in audible frequencies

For audible frequencies, the basic idea is to slightly modify the amplitude and frequencies so that humans cannot distinguish. A major application is to watermarking audio files [2].

Chen et al. [7] use a spread spectrum watermarking technique to deliver information up to 2m with stable performance around 30cm. Dolphin [8] performs orthogonal frequency-division multiplexing (OFDM) on 8kHz-20kHz, with each subchannel alternating between amplitude-shift keying (ASK) and energy difference keying according to whether the original audio has strong signal above 14kHz. In case of ASK, acoustic signals above 14kHz are cleared altogether. Similarly, Eichelberger et. al [3] find subchannels with weak signals in 500Hz to 10kHz, and clear original audio in this subchannel, followed by encoding information in the cleared subchannel.

Zhang et al. [1] propose a system to broadcast Wi-Fi information to customers through sound signal, using method similar to frequency-shift-keying with 3kHz as the base frequency, and the experimental distance is between 10cm and 100cm. The resulting signals are audible.

For these works, the original music is mixed closely with encoded information, which may be annoying to users. It is also not easy to control the ratio of music and encoded information.

B. Acoustic communication in (nearly-)inaudible frequencies

Backdoor [9] enables microphone to record ultrasound (above 20kHz). It observes that special high frequency (e.g., 40kHz) sounds may leave “shadow” in audible frequencies at the microphone, which can be used to encode information. To use Backdoor, the sender should be specially-designed ultrasonic speakers.

PhoneEar [10] uses simple version of frequency-shift keying (FSK) to modulate in the 17kHz-20kHz band, and distances up to 18m. But it is not completely inaudible, and its experiments show that the music produced by this method is somewhat different from the original music.

Hush [4] is a software modem that uses OFDM to send data between commercial smart mobile devices at frequencies 16KHz-20KHz, ideally over a distance of 5-20cm.

Chirp modulation has successful applications in many areas, including distance estimation [11], underwater communication [12] and long distance communication [6]. These successful applications inspire out design of ChirpMu. Lee et al. [13] propose an enhanced modern design for CSS (chirp spread spectrum) -based AAC (aerial acoustic communication) to achieve high reliability and low computational complexity. The proposed quaternary coding unit utilizes the time-shift property to deal with the frequency selectivity of the audio interface of COTS devices, and the frequency range of the signal is between 18.5KHz and 19.5 kHz. They do not mix encoded information with music.

C. Alternative technologies

Instead of using microphones as receivers, VEH-COM [14] propose to use a vibration energy harvesting (VEH) device as receiver to decode sounds from a speaker.

Screen-camera link is an alternative technology for broadcasting. For example, information can be hidden in videos, so that users can watch videos and their smartphone can decode hidden information [15], [16]. When used in our scenario, screen-camera communication require a large screen that is visible to customers at different locations, which need more financial investment. RFID technology [17] can also transfer short messages, but require specialized hardware at the user.

III. Basic Idea and Design Challenges

We consider scenarios where music is played on a daily basis, e.g., in a shopping mall, a restaurant, a gymnasium or a bar. The basic idea is to embed information into music such that humans cannot hear but their smartphones can ignore the ambient noise and successfully decode the information they receive. This approach can be used to broadcast information such as location-based coupons or Wi-Fi secrets [1]. This is a broadcast service, and there is no feedback from receiver to the sender.

In summary, the following requirements should be satisfied.

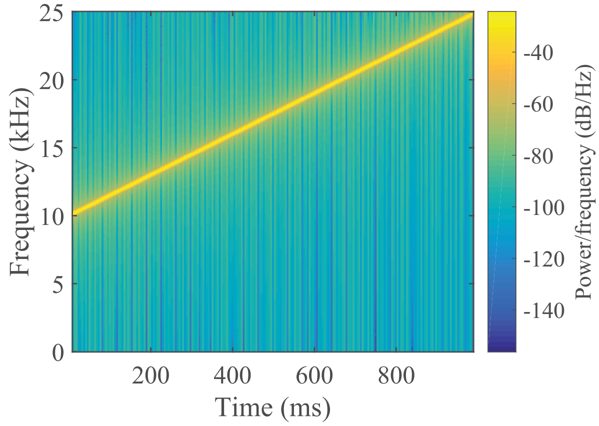
- R1: Humans cannot hear the embedded information, i.e., they cannot tell the difference between modified music and the original music.
- R2: Smartphones can decode the embedded information.
- R3: The distance between sender and receiver can be several meters away, a typical distance is 5 meters.
- R4: There is no feedback from receiver to sender.

For R2, we need to require the smartphone to automatically ignore ambient noise and carrier music during decoding. For R3, we note that a 5-meter distance implies that the system can cover about $10 \times 10\ m^2$ square. This requirement excludes near-field communication technique such as [18].

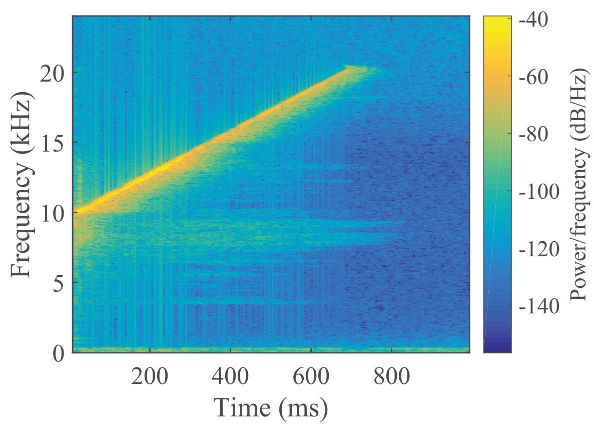
A. Smartphones cannot hear ultrasound

A natural solution to satisfy R1 and R2 is to use ultrasound to encode the broadcast information. Unfortunately, a major challenge we encounter to implement the idea is that commercial smartphones do not have microphones supporting ultrasound signals. By definition, ultrasound is sound that is

Authorized licensed use limited to: NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS. Downloaded on January 03,2022 at 12:52:35 UTC from IEEE Xplore. Restrictions apply.



(a) The original generation frequency increased from 10kHz to 25kHz in 1 second.



(b) Observed frequency of a smartphone when the actual frequency grows from 10kHz to 25kHz in 1 second.

Fig. 2. Original frequency and observed frequency

inaudible by human, and has frequency above 20kHz. Smartphones do not need to support ultrasound because humans do not need this ability. We conduct an experiment where a loudspeaker plays sounds with frequency gradually growing from 10kHz to 25kHz (e.g. 2(a). Fig. 2(b) shows the observed frequency of a smartphone.

We can see that the smartphone is unable to detect sounds above 20kHz. We test other phones and get similar results. In fact, the microphone of most smartphones can only support a sampling rate of 44.1kHz, and by Nyquist-Shannon sampling theorem, this sampling rate can decode signals with frequency at most 22.05kHz. This observation suggests that we should use sounds with frequency no more than 20kHz, or near-inaudible sounds. Backdoor [9] attempts to solve this problem by using specialized ultrasound speakers to generate certain ultrasound that leaves "shadow" in audible frequencies.

B. Near-inaudible frequency may be audible

Sounds with frequency from 16kHz to 20kHz is near-inaudible by humans because the threshold for hearing them increases dramatically after 16kHz. However, an increasing threshold does not necessarily mean that we can use these frequencies arbitrarily. Some people may still hear the sounds if the volume is high [19].

Thus, we encounter a dilemma to use near-inaudible sounds. If the volume is too high, some people may hear it, violating R1. However, if the volume is too low, smartphones may not hear it, violating R2 or R3. This dilemma does not exist for ultrasound since human cannot hear it even for high volume. This dilemma poses a challenge on the coding mechanism, which must tolerate low volume sounds. To the best of our knowledge, existing works using near-inaudible sounds do not restrict the volume to avoid disturbing humans. The reason that no human can hear the sounds is that the conducted experiments has short distance (e.g., [4]), naturally requiring low volume.

Works in audible frequencies either slightly change the frequency or amplitude, having short transmission range [2], or make the resulting mixed sounds noticeable by users sometimes [8], [10]. Essentially, the encoded signals should have low energy in audible frequency, otherwise the original sounds may be corrupted.

In this work, we use near-inaudible frequency as in [4], but we use a robust coding mechanism to tolerate low signal energy and high noise, so that it can support longer distance and has low bit error rate.

C. Requirements to automatically ignore ambient noise and music carriers

A key step to satisfy R2 is to avoid the influence of music carrier and ambient noise as much as possible in the decoding process. Since chirp signals are in a high frequency band, a straightforward approach is to obtain chirp signals by applying a high-pass filter on the mixed sounds. We can use a high-pass filter to directly filter out all the signals below 17kHz and get the chirped signal, so the high-pass filter is suitable for all types of music carriers decoding. However, the use of a filter may lead to distortion of the sound signal, resulting in a certain amount of energy loss, affecting the subsequent decoding.

In this work, we use the method of comparing the energy sum of the up-chirp frequency band and the down-chirp frequency band to decode, so our decoding work is only carried out within the frequency range of the chirp symbol. However, the sound produced in our daily life, whether it is the sound of the carrier music or the environmental noise, is generally below 8kHz, far below the frequency range of the chirped symbols we produce. Therefore, our decoding method can automatically ignore both the music itself and the ambient noise.

IV. ChirpMu Design

To overcome the limitations of existing approaches, we propose ChirpMu, which uses chirp symbols to encode information in near-inaudible frequencies. We first overview ChirpMu, and then present its key components.

Authorized licensed use limited to: NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS. Downloaded on January 03,2022 at 12:52:35 UTC from IEEE Xplore. Restrictions apply.

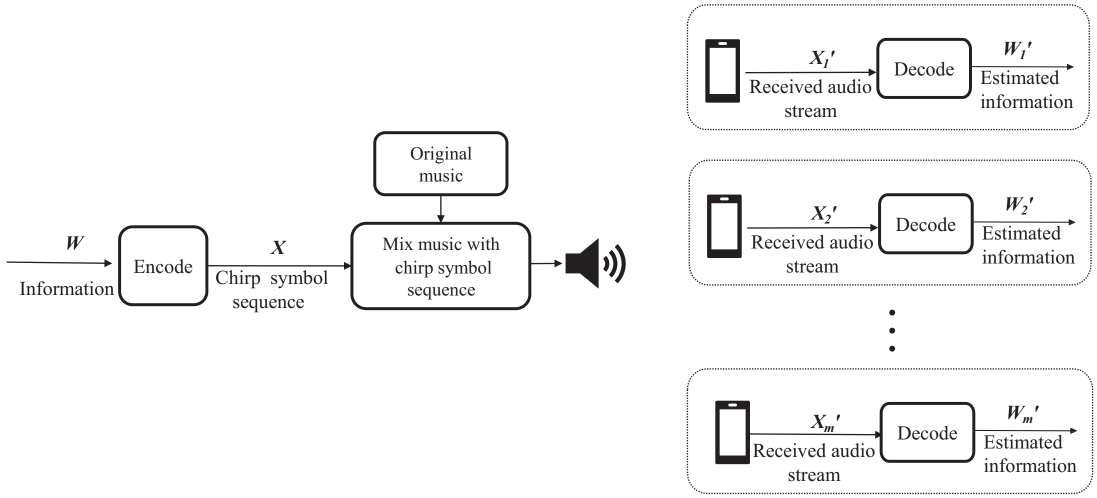


Fig. 3. Overview of ChirpMu

A. Overview

ChirpMu consists of two parts shown in Fig. 3, a sender to create mixed music and m receivers executed on clients' smartphones to extract embedded information in music.

In the sender part, the information W to be broadcast is converted to a sequence of chirp symbols X, and then mixed with the original music stream at a certain proportion to create a mixed audio stream. The mixing operation can be executed either online or offline. In this paper, we implement an offline version, since both the broadcast information and music carriers can be determined beforehand in the considered scenario. The mixed audio streams are transmitted through the air by loudspeakers.

The receiver is executed on a user's smartphone. The smartphone captures mixed music in the air by the built-in microphone, processes the audio stream by fast Fourier transform (FFT) to obtain the amplitude of the frequency, and then calculate the energy of the up-chirp and down-chirp bands and decode the symbols to get the broadcast information.

B. Encoding and decoding chirp symbols

To combat against low energy requirement to avoid making the information audible, we propose to use chirp to encode information, which has shown success in long distance communication in LoRa [6]. It does not necessarily provide high throughput, but can deliver information without disturbing the user.

We encode 0/1 bit to an up-/down-chirp symbol, respectively. Let f_0 and f_1 be the smallest and largest frequency of chirp, and τ be the symbol duration. Let μ be the frequency changing rate defined as $(f_1 - f_0)/\tau$. Then, chirp symbols are obtained by sampling two cosine waves in time for a duration of τ.

Traditionally, the cosine wave for up-chirp is $s(t) = \cos(2\pi f_0 t + \pi\mu t^2)$, and that for a down-chirp is $s(t) = \cos(2\pi f_1 t - \pi\mu t^2)$. Intuitively, an up-chirp symbol is a piece of sound with frequency increasing from f_0 to f_1, and a down-chirp symbol is sound with frequency decreasing from f_1 to f_0. Unfortunately, the decoding method for the above chirp symbols is complex. For example, a classic method is to perform Hough transform on short-time Fourier transform of the sounds [20].

Thus, we propose two new chirp symbols by dividing the frequency range into two equal parts. Let the central frequency be $f_c = (f_0 + f_1)/2$. We set the cosine wave for the up-chirp symbol as

$$s(t) = \cos(2\pi f_c t + \pi\mu/2t^2),$$

and that for the down-chirp symbol as

$$s(t) = \cos(2\pi f_c t - \pi\mu/2t^2).$$

That is, the up-chirp symbol's frequency increases from f_c to f_1 and the down-chirp symbol's frequency decreases from f_c to f_0.

To decode chirp symbols, suppose a piece of sounds corresponding to a chirp symbol is s'. We perform a fast Fourier transform (FFT) on s' to get the magnitude along frequencies. Let $mag(i)$ be the magnitude of frequency i, we compute two energy summations

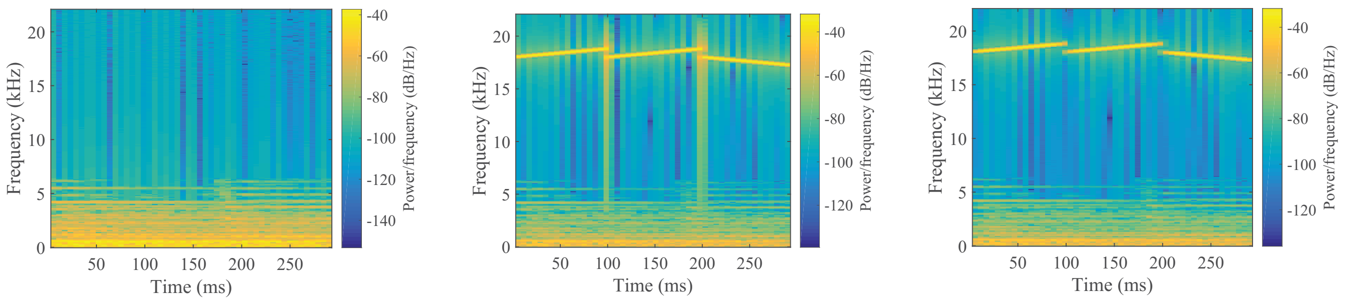
$$e_{up} = \sum_{f_c \le i \le f_1} mag(i)^2$$

and

$$e_{down} = \sum_{f_0 \le i \le f_c} mag(i)^2.$$

If $e_{up} > e_{down}$, then the symbol is a up-chirp, and a down-chirp otherwise. The intuition is that, if the total energy of frequencies in (f_c, f_1) is greater than that in (f_0, f_c), then the symbol is a up-chirp. Our decoding algorithm is much simpler than the traditional Hough transform-based decoder.

Authorized licensed use limited to: NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS. Downloaded on January 03,2022 at 12:52:35 UTC from IEEE Xplore. Restrictions apply.



(a) original music (b) music combined with chirp symbols (c) music combined with smoothed chirp symbols

Fig. 4. Short time Fourier transform on different audio sounds

For the parameters, f_0 and f_1 influence whether the human can hear. We will set them to near-inaudible frequencies mentioned in Section III. The duration of a symbol influences its robustness against fading and noise, and should be configured by experiments.

C. Removing noise due to leaked energy of chirp symbols

After generating chirp symbols, we find that even in the frequency band 18-20kHz, which is insensitive to human ears, the generated chirp audio will still produce an obvious audible click sound. After analyzing the frequency of the generated audio, we find that the connection of the two chirp symbols in the near-inaudible frequency band produces energy in the audible frequency band.

To illustrate the problem, we show the spectrogram of the original music in Fig. 4(a) and the spectrogram of the mixed audio in Fig. 4(b). By comparing the two figures, we observe a high-pitched audible noise (below 16kHz) around the border of two chirp symbols. The main reason is that a chirp symbol can be considered as an infinite sequence multiplied by a finite rectangular window. The window leaks energy at the border.

To address this issue, we propose to preprocess a chirp symbol by smoothing it with a simple triangle window. Specifically, We multiply the first half and the second half of a triangle window at the beginning and the end of each symbol, so that each symbol tends to be continuous. Fig. 4(c) shows the spectrogram of the resulting combined sound. We find that there is no extra energy show in frequency below 16 kHz at the junction of two chirp symbols. We play the combined sound preprocessed by the triangle window and find that the high-pitch noise disappear.

D. Frame synchronization

The client may start listening at any time. To decode correctly, it should determine the start of a frame. To this end, we generate a startup symbol as the startup signal, which aligns the data symbols. We use five fixed chirp symbols at the beginning of a frame. The receiver continuously detect the start flag and align the symbol data using short-time Fourier transform. When we detect energy of frequencies in the range from f_0 to f_1, we enter into a special state, where we attempt to decode the upcoming five chirp symbols. When the five chirp symbols are decoded successfully, the frame is perfectly synchronized. Otherwise, we slide the window by one sample and begin the next attempt.

E. Mixing music with chirp symbols

Suppose the audio stream corresponding to music is a vector $\mathbf{x}$ and that corresponding to chirp symbols is a vector $\mathbf{y}$. The straightforward solution to mix music is to add them directly to produce the combined audio stream $\mathbf{z}$, i.e.,

$$\mathbf{z} = \mathbf{x} + \mathbf{y}.$$

However, we find that the resulting audio would produce bursts and high-pitched noise, which does not occur when playing the two audio streams separately. After examining the mixed audio stream, we find that the reason is due to addition overflow due to each sound sample being represented as a fixed-length binary string.

To address this issue, we normalize music stream so that every sample is a float value in the range from -1 to 1. The chirp audio stream do not need to be normalized since they are sampled from the cosine function, which is already in the range from -1 to 1. Let the the normalized music stream be $\mathbf{x}$. We set the resulting combined stream as

$$\mathbf{z} = ((1-\rho)\mathbf{x}' + \rho\mathbf{y})N_{\max},$$

where ρ is a weighting parameter controlling the portion of chirp in the combined stream, and $N_{\max}$ is the maximum value of a sound sample (e.g., $N_{\max} = 255$ if each audio sample is represented in 8-bit).

Small ρ gives low portion of chirp audios so that they are inaudible, but results in high bit error rate due to low energy. We should set ρ such that the chirp symbols are inaudible and the bit error rate is low. We will set it in experiments.

It is worth mentioning that our decoder still works on the mixed music. There is no need to filter out music carriers and ambient noise, because little music and ambient noise energy follow in the frequency range of chirp symbols, and music carriers and ambient noise is automatically ignored by our decoding method.

Authorized licensed use limited to: NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS. Downloaded on January 03,2022 at 12:52:35 UTC from IEEE Xplore. Restrictions apply.

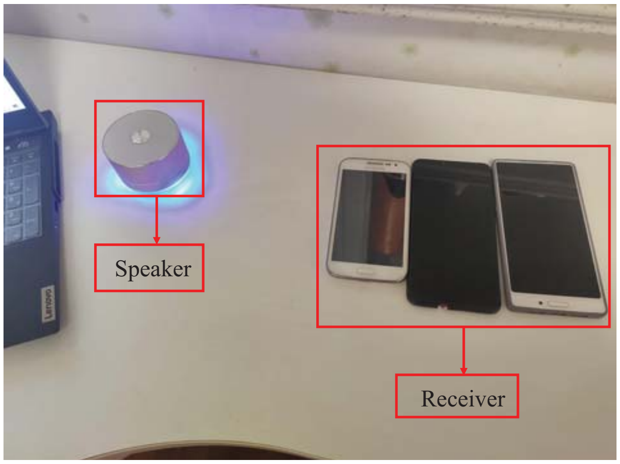


Fig. 5. Equipments in the experiments

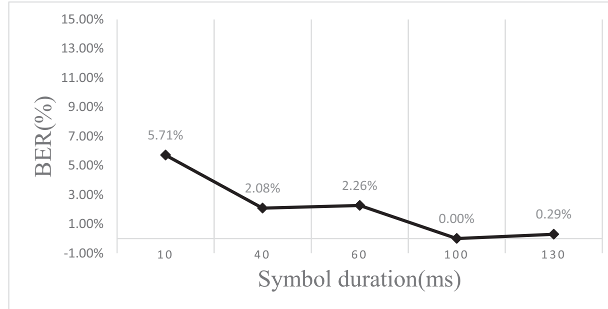


Fig. 6. Impact of duration on BER

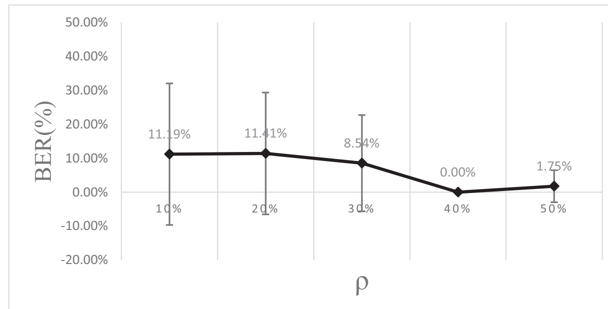


Fig. 7. Impact of chirp proportion on BER($\pm$ standard deviation)

V. Performance Evaluation

In this section, we conduct a large number of scientific and reasonable experiments under different conditions. The experimental equipment is shown in Fig. 5. We study the influence of various factors on the system performance. We first consider how to choose appropriate parameters, including symbol duration, music proportion and center frequency of chirp signal. The influence of the selection of various parameters on the system performance and concrete conclusions are obtained. We then further expanded our research ideas. We study the effects of a number of external factors, including distance, volume, ambient noise, and different music carriers. Experiments on these factors demonstrate the robustness of our system.

A. Parameter selection

In this subsection, we discuss the choice of three key parameters for chirp signal generation. We conduct some comparative experiments to discuss symbol duration, music proportion, and center frequency selection. Only by selecting appropriate parameters can our chirp signal have a better transmission effect in maintaining imperceptible.

1) Symbol duration: We first test the effect of symbol duration on bit error rate. We test the relationship between the duration and the bit error rate (BER) by changing the duration of symbols on the basis of transferring 240 symbols. We conducted the experiment under the premise of chirp audio ratio ρ of 0.4 and center frequency of 18,000Hz. We set the duration respectively to 10ms, 40ms, 60ms, 100ms and 130ms, and the experimental results are shown in Fig. 6.

From the experimental results, it can be concluded that the bit error rate of symbols decreases with the increase of symbol duration. Considering the two aspects of transmission rate and decoding rate, we choose 100ms as the duration of a symbol.

2) Ratio of chirp and music: We then consider the impact of chirp audio ratio on bit error rate. Obviously, when the proportion of music is higher, the similarity between the synthesized audio and the original music is higher, that is, the audio quality is higher. Similarly, when the proportion of chirp audio is higher, the bit error rate of information transmission is lower. To ensure low bit error rate and high audio quality at the same time, it is of great significance to find a proper chirp audio ratio, i.e., ρ. Under the premise that the symbol duration is 100ms and the central frequency is 18000Hz, the world-famous song Canon is selected as the music carrier for the experiment. We change the chirp audio ratio from 0.1 to 0.5 with increments of 0.1, and measure the relationship between the bit error rate and the audio ratio. The results are shown in Fig. 7.

We can see that, when the proportion of chirp audio is 40% or above, the bit error rate is relatively low and tends to a stable state. When the proportion of chirp audio is 30%, it can be seen that the bit error rate increases significantly. When the proportion of chirp audio continues to decrease, the bit error rate continues to rise rapidly. According to the experimental results, selecting a chirp ratio of 0.4 can make the two audio segments reach an appropriate balance point, which will not affect the user's senses and can send data in a better way.

3) Center frequency: In addition, we need to find an appropriate center frequency to generate the chirp signals. Ideally, we prefer high frequency band since it is inaudible to human ears. However, the energy in high frequency band may be low due to hardware constraints on both microphones and load speakers, which may make it difficult for mobile phones to decode. Therefore, we measure the frequency bands detectable by different microphones, and find that the upper limit frequencies vary among smartphones. The upper limit frequency of Redmi phones is about 18.8KHz. Oppo, Smartisan, Samsung

Authorized licensed use limited to: NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS. Downloaded on January 03,2022 at 12:52:35 UTC from IEEE Xplore. Restrictions apply.

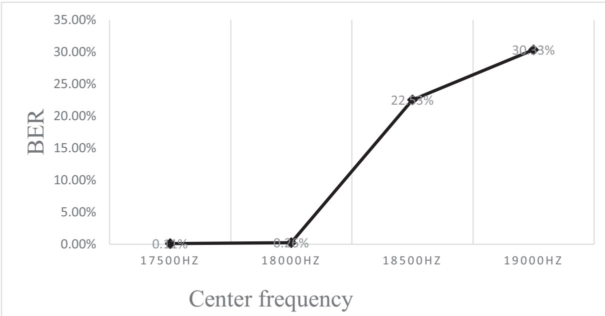


Fig. 8. Impact of center frequency on BER

TABLE I
PARAMETER SETTINGS

Parameter	Default Value
Symbol Duration	100ms
Ratio of Chirp and Music	0.4
Center Frequency	18kHz

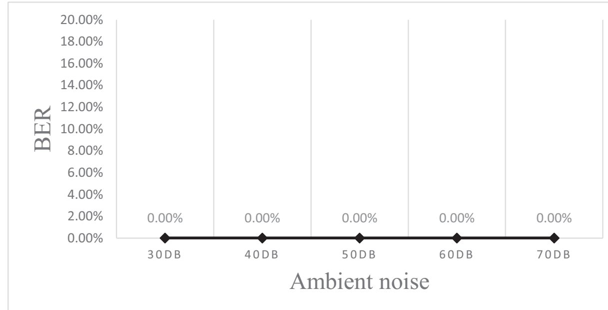


Fig. 9. Impact of ambient noise on BER

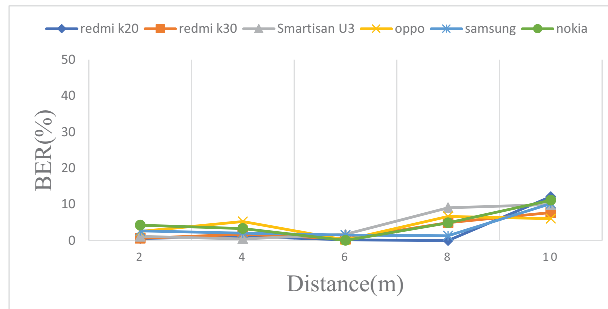


Fig. 10. Impact of distance on BER

and Nokia all have a maximum frequency of nearly 20KHz. We also find that the higher the frequency, the weaker the detection power. Our experiments use Redmi phones for the interest of worst-case evaluation. We set the duration as 100ms and ρ as 0.4. Then we set the chirp center frequency as 17500Hz, 18000Hz, 18500Hz and 19000Hz to measure the bit error rate respectively. The experimental results are shown in Fig. 8.

As can be seen from the experimental results, for Redmi phones, the bit error rate increases rapidly when the chirp center frequency increases to 18500Hz and above. Therefore, to accommodate most phones and for better decoding performance, we set the center frequency to 18000Hz.

We summarize the adopted parameter settings in Table I.

B. *Impact of ambient noise*

Considering that our system will be used in different types of public places, and different types of places will be accompanied by different levels of noise. Therefore, it is necessary to study the impact of environmental noise on the transmission of information. We have five levels of ambient noise and we will test our bit error rate at different levels of ambient noise. Quiet and comfortable ambient noise intensity is 30-40 decibels. Conversation place noise is generally 40-60 decibels. Noise above 60 decibels is considered noisy and 70 decibels is very noisy, interfering with the normal conversation, and making people feel upset, mental concentration, and begin to damage the hearing nerve. We use a decibel meter to measure the ambient noise, and then we measure the bit error rate when the ambient noise is around 30, 40, 50, 60 and 70 decibels respectively. Fig. 9 shows the relation between ambient noise and bit error rate.

The experimental results show that our system still has a good decoding rate under the circumstance of large ambient noise. This is because our decoding work is only carried out within the frequency range of chirp symbols, while the sound produced in our daily life, whether it is the sound of music or ambient noise, is generally below 8000Hz, far below the frequency range of chirp symbols we generate. So our decoding method can automatically ignore both the music itself and the ambient noise. Thus, our system works in almost any place whether it is quiet or noisy.

C. *Impact of distance between the speaker and the receiver*

Fig. 10 shows the effect of the distance to the bit error rate, we use two Redmi devices, a Nokia device, an Oppo device, a Smartisan device and a Samsung device, to conduct the experiment at the distances of 2-10 meters respectively. The sound signals should degenerate with the increase of the distance, thus leading to the increase of the bit error rate. The experimental results show that in a certain distance, the bit error rate fluctuates within a certain error range, and after the certain distance, the bit error rate increases in the fluctuation. It also shows the robustness of our system. Bit error rate is the lowest when a suitable distance is reached. Later, as the distance increases, the sound signal decays and the bit error rate increases.

D. *Impact of volume*

In this subsection, we test the influence of the volume on the bit error rate. As the volume decreases, the sound signal decreases, and the bit error rate should gradually increase. We set the volume as 10%, 30%, 50%, 70%, and 100% of the speaker volume respectively to test the bit error rate. The experimental results are shown in Fig. 11. According to the

Authorized licensed use limited to: NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS. Downloaded on January 03,2022 at 12:52:35 UTC from IEEE Xplore. Restrictions apply.

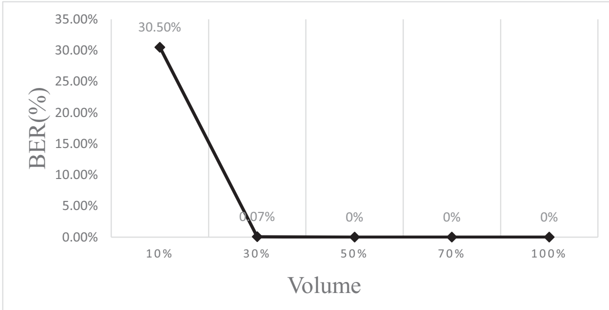


Fig. 11. Impact of volume on BER

TABLE II
BER OF DIFFERENT MUSIC

Songs	$BER[\%]$	$Genre$
Fur ELISE	0	classical
Canon	0	classical
Faded	0	electronic
Closer	0	electronic
Hobo Blues	0	blues
Cross Road Blues	0.18	blues
500 miles	0.65	folk
Yesterday Once More	7.38	folk
Bohemian Rhapsody	0	rock and roll
What Do You Got	0.23	rock and roll

experimental results, the bit error rate is 0 when the volume is 50%, 70% and 100% of the loudspeaker volume. The bit error rate increases slightly to 0.07% when the volume is reduced to 30%. Only when the volume drops to 10% does the bit error rate increase dramatically. The experimental results show that our system is robust, and the volume fluctuation in the normal range will not affect the decoding work, and the bit error rate increases only when the volume is low to a certain extent.

E. Impact of different music carriers

In order to test the applicability of various types of music, we tested different types of songs with a chirp ratio of 0.4. Music can be divided into different types according to its style and characteristics. We select ten world-famous songs of five types, namely electronic music, classical music, blues music, rock and roll music and folk music, to conduct the experiment, and the experimental results are shown in Table II. It can be concluded from the experiment that this system has good adaptability to all kinds of music with low or high spectral centroid. This means that our system is very robust and can be used in different types of places.

F. Discussion

ChirpMu can transmit data to smartphones without humans being able to perceive it. This can be widely used in location-based services (LBS) applications [21]. Our experiments show that ChirpMu can be adapted to a variety of music carriers without being affected by ambient noise, which makes our system suitable for any place where music can be played, such as shopping malls, bars, restaurants, etc. The store vendor can use data-over-music to broadcast electronic coupons to customers near the store. However, if there are many stores in a big shopping mall, we should distribute all kinds of coupons to users in the way of playing music uniformly. However, if each store plays its own music, there will be conflicts. We plan to investigate this problem in the future.

VI. CONCLUSIONS

In this paper, we propose a system for broadcasting information to customers through music. To use our system, the information provider uses loudspeakers to broadcast music mixed with chirped signals. A customer can receive the information using the built-in microphone on the smartphone. The advantage of our system is that it can adapt to a variety of music carrier and can resist ambient noise, so providers can easily send different information to customers according to different scenarios. Our system has a very low bit error rate in both high ambient noise and low ambient noise scenarios. To use our system, customers can be several meters away from the loadspeaker. Our system is characterized by the use of novel chirp symbols to send information, which has a good fault tolerance rate in mobile devices. Customers cannot distinguish between the audio synthesized with chirp signals and the original music, which enables customers to receive information without being disturbed while enjoying music.

ACKNOWLEDGEMENTS

We sincerely thank reviewers for their insightful feedback. This work was supported in part by the National Natural Science Foundation of China (61972199, 61972200).

REFERENCES

[1] L. Zhang, X. Zhu, and X. Wu, "No more free riders: Sharing wifi secrets with acoustic signals," in *Proceedings of the 28th International Conference on Computer Communication and Networks (ICCCN)*, Valencia, Spain, July 2019, pp. 1–8.

[2] D. Kirovski and H. S. Malvar, "Spread-spectrum watermarking of audio signals," *IEEE Transactions on Signal Processing*, vol. 51, no. 4, pp. 1020–1033, 2003.

[3] M. Eichelberger, S. Tanner, G. Voirol, and R. Wattenhofer, "Receiving data hidden in music," in *Proceedings of the 20th International Workshop on Mobile Computing Systems and Applications*, Santa Cruz CA USA, February 2019, pp. 33–38.

[4] E. Novak, Z. Tang, and Q. Li, "Ultrasound proximity networking on smart mobile devices for iot applications," *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 399–409, 2018.

[5] X. Zhu, Q. Li, and G. Chen, "Apt: Accurate outdoor pedestrian tracking with smartphones," in *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*, Turin, Italy, April 2013, pp. 2508–2516.

[6] Wikipedia, "LoRa," 2021, [accessed 24-January-2021]. [Online]. Available: https://en.wikipedia.org/wiki/LoRa

[7] P.-W. Chen, C.-H. Huang, Y.-C. Shen, and J.-L. Wu, "Pushing information over acoustic channels," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal*, Taipei, Taiwan, April 2009, pp. 1421–1424.

[8] Q. Wang, K. Ren, M. Zhou, T. Lei, D. Koutsonikolas, and L. Su, "Messages behind the sound: real-time hidden acoustic signal capture with smartphones," in *Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking*. New York, USA: ACM, July 2016, pp. 29–41.

Authorized licensed use limited to: NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS. Downloaded on January 03,2022 at 12:52:35 UTC from IEEE Xplore. Restrictions apply.

[9] N. Roy, H. Hassanieh, and R. Roy Choudhury, "Backdoor: Making microphones hear inaudible sounds," in *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*. Niagara Falls New York USA: ACM, June 2017, pp. 2–14.

[10] A. S. Nittala, X.-D. Yang, S. Bateman, E. Sharlin, and S. Greenberg, "Phoneear: interactions for mobile devices that hear high-frequency sound-encoded data," in *Proceedings of the 7th ACM SIGCHI Symposium on Engineering Interactive Computing Systems*, Duisburg Germany, June 2015, pp. 174–179.

[11] C. Peng, G. Shen, and Y. Zhang, "Beepbeep: A high-accuracy acoustic-based system for ranging and localization using cots devices," *ACM Transactions on Embedded Computing Systems*, vol. 11, no. 1, pp. 1–29, 2012.

[12] F. Steinmetz, J. Heitmann, and C. Renner, "A practical guide to chirp spread spectrum for acoustic underwater communication in shallow waters," in *Proceedings of the 13th ACM International Conference on Underwater Networks & Systems*, Shenzhen China, December 2018, pp. 1–8.

[13] J. Lee, C. Kwak, S. Kim, and S. Bahk, "Reliable and low-complexity chirp spread spectrum-based aerial acoustic communication," *IEEE Access*, vol. 8, pp. 151 589–151 601, 2020.

[14] G. Lan, W. Xu, S. Khalifa, M. Hassan, and W. Hu, "Veh-com: Demodulating vibration energy harvesting for short range communication," in *Proceedings of the IEEE International Conference on Pervasive Computing and Communications*, Kona, HI, March 2017, pp. 170–179.

[15] A. Wang, Z. Li, C. Peng, G. Shen, G. Fang, and B. Zeng, "Inframe++ achieve simultaneous screen-human viewing and hidden screen-camera communication," in *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*, Florence Italy, May 2015, pp. 181–195.

[16] K. Zhang, Y. Zhao, C. Wu, C. Yang, K. Huang, C. Peng, Y. Liu, and Z. Yang, "Chromacode: A fully imperceptible screen-camera communication system," *IEEE Transactions on Mobile Computing*, vol. 20, no. 3, pp. 861 – 876, 2021.

[17] X. Zhu, L. Xu, X. Wu, and B. Chen, "Minimum time extrema estimation for large-scale radio-frequency identification systems," *Journal of Computer Science and Technology*, vol. 35, pp. 1099–1114, 2020.

[18] R. Nandakumar, K. K. Chintalapudi, V. Padmanabhan, and R. Venkatesan, "Dhwani: secure peer-to-peer acoustic nfc," *ACM SIGCOMM Computer Communication Review*, vol. 43, no. 4, pp. 63–74, 2013.

[19] K. Ashihara, "Hearing thresholds for pure tones above 16 khz," *The Journal of the Acoustical Society of America*, vol. 122, no. 3, pp. 52–57, 2007.

[20] Y. Sun and P. Willett, "Hough transform for long chirp detection," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 38, no. 2, pp. 553–569, 2002.

[21] W. Wei, X. Zhu, and Q. Li, "Lbsnsim: Analyzing and modeling location-based social networks," in *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*, Toronto, Canada, April 2014.

Authorized licensed use limited to: NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS. Downloaded on January 03,2022 at 12:52:35 UTC from IEEE Xplore. Restrictions apply.