A Covert Ultrasonic Phone-to-Phone Communication Scheme

Liming Shi[1], Limin Yu[1], Kaizhu Huang[1], Xu Zhu[2], Zhi Wang[3], Xiaofei Li[4], Wenwu Wang[5], and Xinheng Wang[1]

[1] Xi'an Jiaotong-Liverpool University, Suzhou, China
[2] University of Liverpool, Liverpool, UK
[3] Zhejiang University, Hangzhou, China
[4] Westlake University, Hangzhou, China
[5] University of Surrey, Guilford, UK

Abstract. Smartphone ownership has increased rapidly over the past decade, and the smartphone has become a popular technological product in modern life. The universal wireless communication scheme on smartphones leverages electromagnetic wave transmission, where the spectrum resource becomes scarce in some scenarios. As a supplement to some face-to-face transmission scenarios, we design an aerial ultrasonic communication scheme. The scheme uses chirp-like signal and BPSK modulation, convolutional code encoding with ID-classified interleaving, and pilot method to estimate room impulse response. Through experiments, the error rate of the ultrasonic communication system designed for mobile phones can be within 0.001% in 1-meter range. The limitations of this scheme and further research work are discussed as well.

Keywords: Smartphone · Aerial ultrasonic communication.

1 Introduction

With smartphones being the most popular personal devices and the platforms for supporting smart services, the demands for reliable wireless communications rise rapidly. Communications based on electromagnetic (EM) waves are the main stream of wireless communications. Various protocols have been developed over the years to cover short-range and long-range communications, such as Wi-Fi, Bluetooth, and the latest 5G mobile communications. However, the spectra of electromagnetic waves are becoming scarce and congested. Under this circumstance, exploiting new transmission media and increasing bandwidth efficiency are two mainstream solutions.

Underwater navigation and communications have utilized acoustic wave since the last century [1, 2]. Meanwhile, acoustic wave, especially ultrasonic component, has an abundant untapped bandwidth capacity in the aerial environment, but was not yet fully exploited. This work intends to utilise this resource and design an applicable scheme in face-to-face transmission scenarios. For example,

a customer can get his ticket from the cashier without being in contact in a low-EM-signal-quality environment by using this scheme. This is particularly useful under the emergency situations, such as the outbreak of Covid-19 pandemic. Typically, the transmission rate of this kind of small-data-sized information is low, but security requirements bring challenges to the communication scheme.

Smartphone ownership has increased rapidly in the past decade, and smartphone is expected to become the technological product with the highest ownership rate in modern life. Currently, the most widely used phone-to-phone wireless communication systems are based on radio frequency(RF) and infrared (IR) transmission [3–5]. One significant merit of EM wave is its low decay rate in the air medium so that EM wave can propagate over a long distance. In addition, EM wave is transverse wave that has a propagation velocity of approximately 2.9×10^8m/s, so EM signal under GHz frequency has a long enough wavelength to pass by small-sized solid barriers [6]. In terms of the acoustic wave, the attenuation rate of its propagation in the air is proportional to the wave frequency [7], which causes trouble to the wideband acoustic communications. Therefore, the usable bandwidth of EM wave communication is commonly more extensive than that of acoustic communications, and the data rate of the electromagnetic wave is higher too.

A critical challenge of the ultrasonic communication arises from wave reflection, mechanical energy loss, and Doppler shift: the reflection is mostly caused by multi-path propagation results in significant phase shift; the mechanical energy loss increases distance loss of sound propagation; the Doppler shift causes frequency deviation. Another significant challenge stems from mobile phones: the maximum sampling rate of mainstream smartphone is 48kHz, and the frequency response capabilities vary widely. These conditions limit the coherent bandwidth, data rates, and also the transmission distance.

In the existing research on aerial acoustic communication systems, most work focuses on ultrasonic spectrum for a better signal-to-noise ratio (SNR) [8–12]. A Dual-In-Dual-Out acoustic communication system using frequency shifting keying (FSK) was proposed in [8]. A multi-channel aerial communication scheme using distance-selected binary phase shifting keying (BPSK) and binary amplitude shifting keying (BASK) was proposed in [9]. A mesh network structure based on Multiple Input, Multiple Output (MIMO) acoustic channel communication was proposed in [10]. The above work [8–10] were implemented on the PC or the specific-manufactured speaker&microphone platform, where these platforms all support a higher sampling rate and resolution than the smartphone platform. For research related to the smartphone platform, a low-rate chirp-based BPSK scheme to realize an up-to-25m communication range was proposed in [12]. Researchers in [9] pointed out that the aerial ultrasonic phase characteristics change significantly with distance. Inspired by research work in [9, 12], we have designed a covert aerial ultrasonic communication scheme that leverages chirp-like signal BPSK with ID-classified interleaving.

The main contributions of this work are summarized as follows: (1) We have proposed a short-distance covert aerial ultrasonic communication scheme for

phone-to-phone communications; (2) A convolutional code plus ID-classified interleaving is used; (3) we leverage the phase noise characteristics of aerial ultrasonic wave to limit the effective communication range in the indoor environment.

The remainder of the paper is arranged as following: The system design is described in Section 2, with emphasis on selection of waveform and modulation scheme, estimation of room impulse response and its compensation, channel coding, and receiver design. Extensive experiments and performance evaluation of various aspects of this communication scheme are conducted and investigated in Section 3. Section 4 concludes the paper with discussions of limitations of this scheme and further research work.

2 System Design

Fig. 1 shows a schemetic diagram of the proposed ultrasonic communication system. The transmitter block has two main parts, the encoder and the modulator respectively. The encoder functions to improve the channel gain via applying forward-error-checking code and interleaver; the core challenge in encoder block comes from the trade-off between code rate and channel coding gain. The modulator block takes challenges in the choice of modulation keying, which requires to balance the bit rate and bandwidth usage, communication distance and reliability. On the receiver block, the Room Impulse Response (RIR) estimation block functions to predict the channel state information and then leverage the predicted channel state model to restore the transmitted signal. The blocks in the receiver are paired to that in transmitter respectively: decoder matches the encoder, and demodulator matches the modulator; these blocks in the receiver perform reverse steps of the transmitter blocks to recover the information.

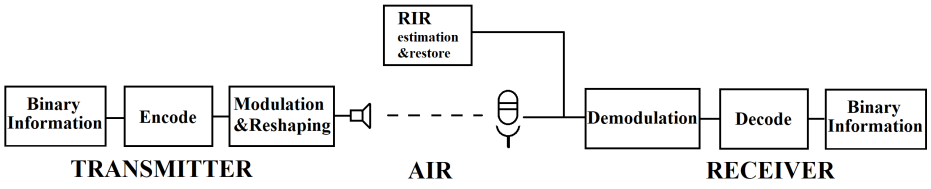

Fig. 1. Overview of the acoustic communication system.

2.1 Waveform and Modulation

The sound collection capabilities of current smartphones are limited to sampling rate of 44.1kHz on iOS phones or 48kHz on Android phones. That is equivalent to maximum signal frequency of 24kHz, which means near ultrasonic. In order to improve the reliability of ultrasonic communications on smartphones, the chirp based BPSK is applied. Compared with frequency shifting keying used in [10, 12, 13], BPSK is less sensitive to the diversity of non-flat frequency response and is more suitable for signal modulation between various types of smartphones. The chirp signal is relatively insensitive to Doppler shift and power-efficiency [13], and the system uses a couple of orthogonal chirp-signals, which are noted as up and down chirps, to spread the frequency band. The expressions of up and down chirps are given in Eqns. (1) and (2), where s_{up} is the up chirp signal, s_{down} is the down chirp signal, f_{low} is the lowest frequency bound of the chirp signal, f_{high} is the highest frequency bound of the chirp signal, t is the time, τ_{chirp} is the duration of the chirp signal.

$$s_{up}(t) = \cos[2\pi(f_{low}t + \frac{f_{high} - f_{low}}{\tau_{chirp}}t^2)] \qquad 0 < t < \tau_{chirp} \tag{1}$$

$$s_{down}(t) = \cos[2\pi(f_{high}t - \frac{f_{high} - f_{low}}{\tau_{chirp}}t^2)] \qquad 0 < t < \tau_{chirp} \tag{2}$$

Fig. 2 demonstrates an example of frame structure to send binary message '110100', which consists of a preamble, guard interval, data payload and trailing bits in sequence from the front to the back. The preamble is a down chirp modulated signal and is partly reshaped by a Blackman window (Fig. 3.(b)), which is used to synchronize the start position of the signal and channel state assessment. The guard interval serves as a transition between the load and the preamble to reduce the influence of inter-symbol interference. The chirp modulation uses up and down chirp to represent the binary '1' and '0' respectively, and the modulated signal of payload data is fully reshaped by the Blackman window (Fig. 3.(c)). A 2-bit trail bit is modulated by up and down chirps sequentially and part reshaped, which indicates the ending of the signal sequence.

In the process of demodulation, we use the matched-filter method to compare the received symbol's correlation coefficient. A Hilbert-transform method is used to detect the envelope of the correlation function. Demonstrated in Eqn. (3), the matched-filter calculates the likelihood of the received symbol corresponding to its correlation to up and down chirps, and the value of likelihood is used in the following maximum likelihood (ML) decoder. In Eqn. (3), the symbol *likelihood* describes the variable that represents the probability of the received bit to become the Boolean value '1', and the symbols $Correlation_{upchirp}$ and $Correlation_{downchirp}$ describe the normalized correlation corresponding to up and down chirps respectively.

$$likelihood = \frac{Correlation_{upchirp}}{Correlation_{downchirp} + Correlation_{upchirp}} \tag{3}$$

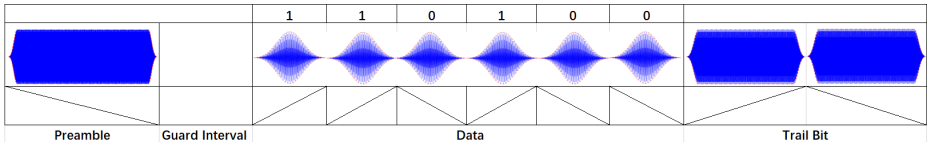

Fig. 2. Signal frame structure

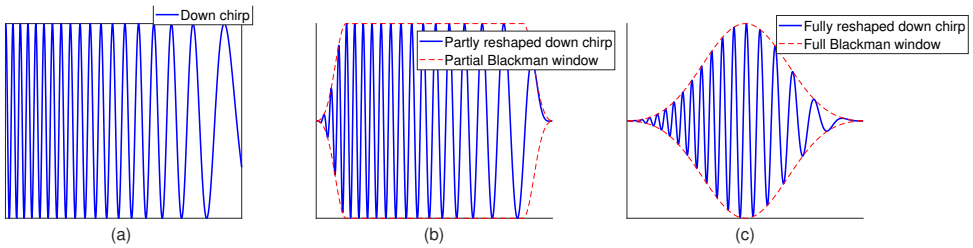

Fig. 3. (a). Down chirp signal; (b). Partly reshaped down chirp signal and partial Blackman window; (c). Fully reshaped down chirp signal and full Blackman window.

2.2 RIR Estimation and Compensation

RIR describes the acoustic channel state between the speaker and receiver. In an indoor scenario, RIR is typically determined by the environmental factors, which vary temporally and spatially. In the situation of face-to-face communications, the relative motion speed between the speaker and receiver is considered to be low enough compared with airborne sound speed, which means the Doppler shift can be ignored. Therefore, the RIR is primarily affected by multipath and shadow fading. In this work, we use a least-square-error method to restore an approximate RIR model from the received preamble.

The simplified channel model is demonstrated in Fig. 4. The received preamble $y[n]$ can be considered as the superposition of the convolution of transmitted preamble $x[n]$ with channel frequency response (the room impulse response) $h[n]$ and additive interference $i[n]$, and the equations of this process can be written as Eqns. (4), (5). The matrix expression of Eqn. (5) can be rewritten as Eqn. (6), and Eqn. (7) simplifies Eqn. (6) by substituting the x-elements matrix by matrix A. The error function E that defines the difference between the received preamble and the transmitted preamble is given in Eqn. (8), and the square error

is obtained in Eqns. (9) and (10), where σ is the variance function of environmental interference and the superscript ^{T} denotes the transposition. To minimize the square error, the partial derivative of the square error function with respect to variable H is shown in Eqn. (11), and the optimal solution can be obtained when Eqn. (11) equals to zero. Finally, we can obtain an approximation of the RIR model only depending on the transmitted and received preamble according to Eqn. (12), where the superscript $^{-1}$ denotes the inverse matrix. Fig. 5 shows the steps of RIR compensation: the algorithm leverages the Cepstrum analysis, using the reciprocal of the fast Fourier transform (FFT), to generate the inverse-RIR filter. The algorithm convolutes the inverse-RIR filter with the received data signal to restore and compensate the data signal.

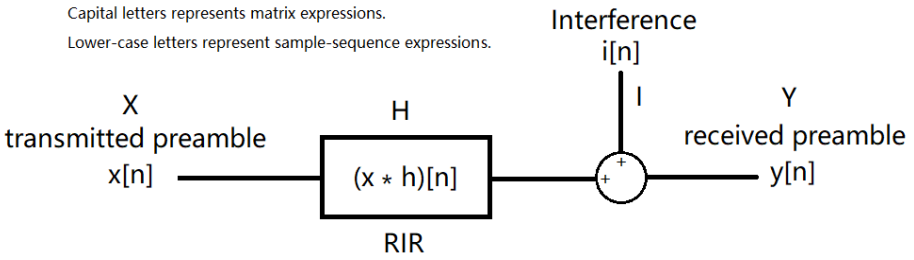

Fig. 4. Simplified channel model.

$$y[n] = x[n] * h[n] + i[n] \tag{4}$$

$$y[n] = \sum_{k=0}^{n} x[n-k]h[k] + i[n] \tag{5}$$

$$\begin{bmatrix} y[0] \\ y[1] \\ \vdots \\ y[n] \end{bmatrix} = \begin{bmatrix} x[0] & x[-1] & \cdots & x[-n] \\ x[1] & x[0] & \cdots & x[1-n] \\ \vdots & \vdots & \ddots & \vdots \\ x[n] & x[n-1] & \cdots & x[0] \end{bmatrix} \begin{bmatrix} h[0] \\ h[1] \\ \vdots \\ h[n] \end{bmatrix} + \begin{bmatrix} i[0] \\ i[1] \\ \vdots \\ i[n] \end{bmatrix} \tag{6}$$

$$Y = AH + I \tag{7}$$

$$E = Y - AH - I \tag{8}$$

$$|e|^2 = (Y - AH)(Y - AH)^T + \sigma \tag{9}$$

$$|e|^2 = Y^T Y - H^T A^T Y - Y^T AH + H^T A^T AH + \sigma \tag{10}$$

$$\frac{\partial}{\partial H}|e|^2 = -2A^T Y + 2A^T AH \tag{11}$$

$$\hat{H} = (A^T A)^{-1} A^T Y \tag{12}$$

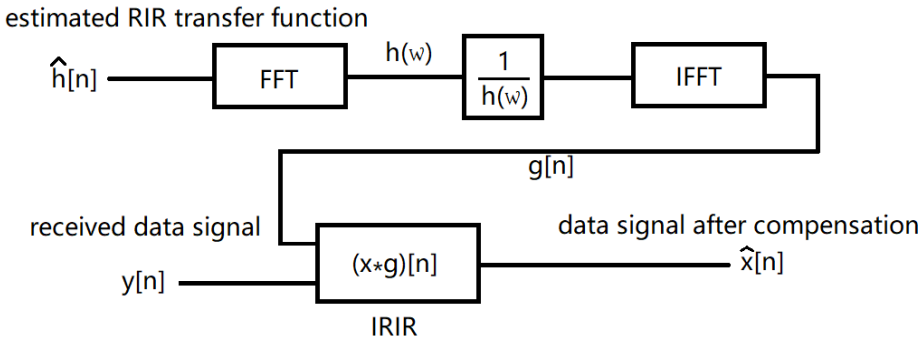

Fig. 5. RIR compensation method.

2.3 Channel Coding Scheme

Inspired by Wi-Fi forward error check coding in IEEE 802.11, the data transmission system applies a 0.5-code-rate convolutional encoding scheme, including 212 convolutional code (Fig. 6.a) and 216 convolutional code (Fig. 6.b). 216 and 212 convolutional codes belong to a single-in-dual-out structure, and 216 convolutional code has more delay units than 212 structure. The single-in-dual-out encoder has two synchronous output ports, generating one information bit and its corresponding redundant bit. The information bits with corresponding redundant bits are reordered into a one-dimensional bitstream by interleaving. We use the Matlab function '*randintrlv*' as the interleaver, the function performs data permutation according to the rand seed. In this coding scheme, the receiver is assigned a unique rand seed as its ID, and the receiver can only use it to recover the information. The decoder uses maximum likelihood (ML) algorithm based on Viterbi [14] algorithm. Fig. 7 shows the theoretical bits error rate performance bound and coding gain of 212 and 216 structure, and the channel coding scheme that leverages 216 convolutional code plus interleaver can implement a zero-error communication under the same condition that the error rate of non-coding communication is around 10^{-2}. Table 1 displays the simulation parameters in Fig. 7.

2.4 Receiver

The diagram of the process of demodulation and decoding is given in Fig. 8. The received signal firstly passes through the frame detector to extract the preamble

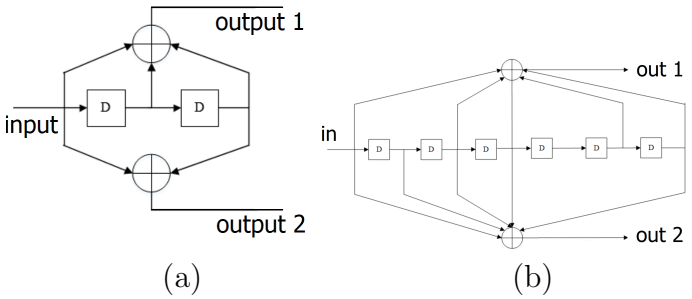

Fig. 6. (a). 212 convolutional code structure, (b). 216 convolutional code structure.

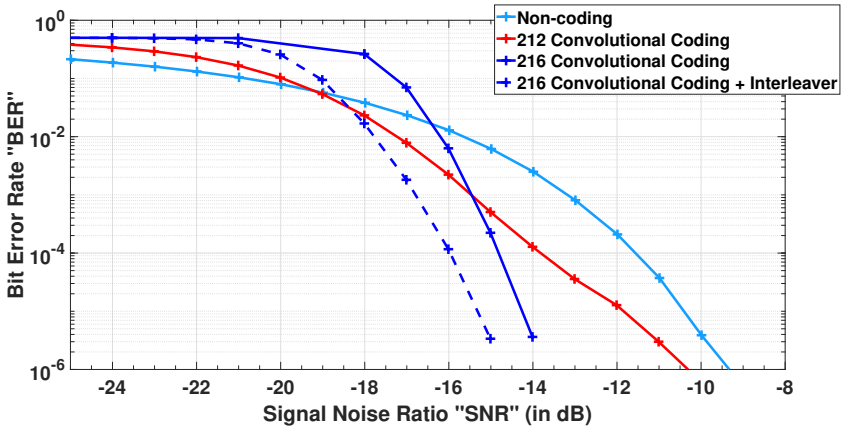

Fig. 7. Comparison of different coding schemes in terms of BER performance

and trail bit frames. The preamble frame is used to estimate the approximate RIR model by using the least square error method, and the RIR model is applied to generate the restore filter to compensate for the data signal. Following the detection of a preamble, the data sequence is divided into segments corresponding to symbols. Then, the symbol segments pass through the matched-filter to calculate the corresponding likelihood, and the receiver uses its unique deinterleaving seed to deinterleave the matched-filtering result. After deinterleaving, symbols are decoded by the ML decoder according to likelihood, and the receiver uses the checksum bits to verify the information.

Table 1. Simulation setup

Bit rate	100 bits/s
Frequency Band	20∼22kHz
Number of Random Bits	10^7
Sampling Rate	48000
Channel	Gaussian noise channel

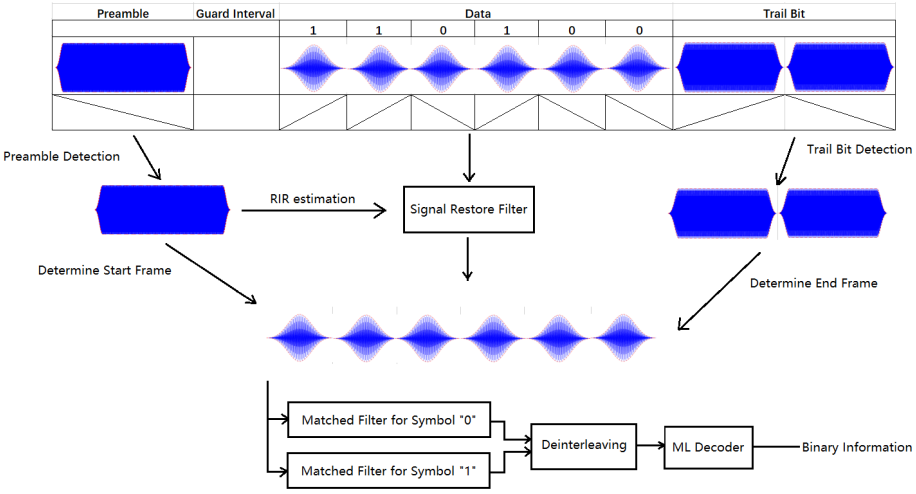

Fig. 8. Demodulation and decoding framework.

3 Experiments

3.1 Experimental Settings for Data Transmission

Table 2. Experiment Parameters

	parameter	**value**
Common	Room size	$5 \times 3 \times 2.7m^3$
	Bit rate	100 bits/s
	Frequency band	20∼22kHz
	Preamble	85.3 ms
	Guard interval	21 ms
	Trail bits	100 ms
	Transmitter Model	Huawei Honor V20
	Receiver Model	Huawei Mate 10
Case I	Distance	1m, 3m
	Data size	1000-bit
Case II	Distance	1m, 1.2m, 1.5m, 3.5m
	Data size	100000-bit

In our experiments, we tested and measured the error rate via simulating a scene (Case I in Table 2): two people holding smartphones exchange messages face-to-face, where the message can be a payment information link, QR code, and hyper-link. The data size of these messages is typically about 1000 bits, so we used a 1000-bits random binary data as the payload of the transitted data. We used Matlab R2019a to encode and modulate these data, and convert the processed data into '.wav' audio format. In the measurement, we used two smartphones to play and record the transmitted signal, and we used Matlab R2019a to demodulate and decode the received signal. In addition, we also tested fixed-point to fixed-point communication as a reference (Case II in Table. 2). The detailed experiment parameters are displayed in Table 2.

3.2 Performance Evaluation

Table 3. Experimental Result

Case	Coding Sheme	Distance(m)	Average Error Rate
I	None	0.5-1.5	0.05%
	212 + Interleaver	0.5-1.5	0
	216 + Interleaver	0.5-1.5	0
	None	2-4	21.5%
	212 + Interleaver	2-4	48.8%
	216 + Interleaver	2-4	50.1%
II	None	1	0.014%
	212 + Interleaver	1	0
	216 + Interleaver	1	0
	None	1.2	0
	212 + Interleaver	1.2	0
	216 + Interleaver	1.2	0
	None	1.5	0.023%
	212 + Interleaver	1.5	0
	216 + Interleaver	1.5	0
	None	3.5	0.978%
	212 + Interleaver	3.5	0
	216 + Interleaver	3.5	0.001%

From the results shown in Table 3, we can find that:

1. In the case I, which is a simulated scenario where two people holding smartphones use the ultrasonic communication, the average error for a none coding communication is 0.05% at a distance around 1 m, and zero error rate for two coding cases. This result can be explained: when two people are holding mobile phones, the speakers and microphones on the two smartphones are in unpredictable motion, which leads to the incoherence of the channel stage. Overall, both schemes using channel channel coding are capable to achieve the short-distance face-to-face communication.
2. In the case I, all schemes fail to implement a zero-error communication when the distance is around 3m. Compared with the same distance condition in the case II, which tests the fixed-point to fixed-point ultrasonic communication, the average error rates of the two channel coding schemes are approximately zero, where the ultrasonic communication quality is much better than that in the case I. This result can be explained: the current RIR estimation and compensation scheme is only applicable to coherent channels, while the channel state in the case I is frequently changing due to the motion between speaker and microphone.

3.3 Distance Loss

To evaluate the distance loss of airborne channel, we play 70s single-symbol modulated signal in a $4.8 \times 8.8 \times 2.7m^3$-sized sitting room. Two testing smartphones are placed on stable supports, which are vertically 0.9m above the ground. We compare the average power spectral density of the received signals tested at 1m, 5m, and 10m distance, and we found the linear distance loss for smartphones is approximately 1dB/m. Based on this estimate, the power density of the signal transmitted by tested smartphones will touch the noise level at 23m, where the signal is undetectable over the distance.

3.4 Phase Noise

In the indoor room-sized environment, multi-path propagation causes significant phase shift that causes confusion in the demodulation of BPSK. Phase degradation characteristics are significant factors that determine the PSK modulation

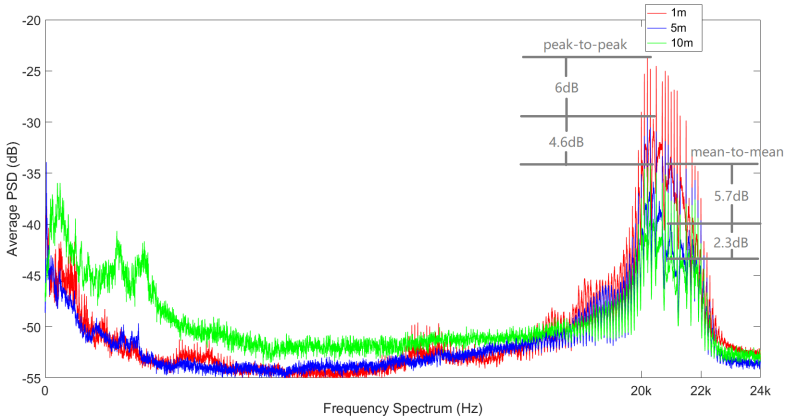

Fig. 9. Average PSD of modulated signal

quality. To evaluate the characteristics in indoor environments, we conducted a chirp signal transmission test. Table 4 shows the experiment parameters, and we used a matched-filter method to evaluate the experimental result.

The convolution of matched-filter and signal outputs the correlation function between the signal and matched-filter-corresponding symbol. Theoretically, the matched-filter synchronizes the phase spectrum of the signal and can reshape the signal into impulse waveform. Under this circumstance, if the phase degradation is significant, the matched-filtering result will not exhibit an impulse waveform.

Fig. 10 demonstrates the matched-filtering result of randomly picked down chirps symbol from results tested in 1m and 5m. In 1m-distance condition, Fig. 10.(a) exhibits a clear impulse in the correlation function of down chirp. However, in the 5m-distance condition, Fig. 10.(c) and Fig. 10.(d) do not exhibit a clear impulse envelope, and two figures both exhibit a low correlation level. This phenomenon shows that the phase noise increases significantly with distance, and the Phase Shifting Keying method is difficult to recover information at a long distance.

Table 4. Experiment Parameters

parameter	**value**
Room size	$5 \times 3 \times 2.7m^3$
Down chirp duration	10ms
Frequency band	20∼22kHz
Data size	1000
Distance	1m, 5m
Transmitter model	Huawei Honor V20
Receiver model	Huawei Mate 10

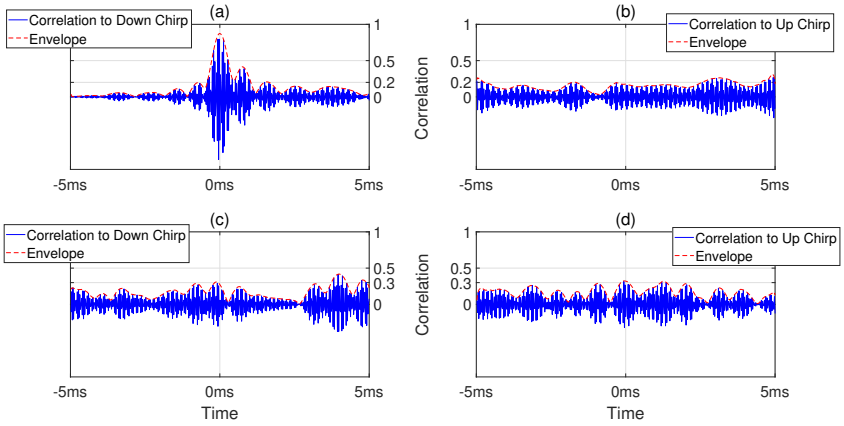

Fig. 10. The correlation functions of the down-chirp signal received at 1m and 5m. (a). Correlation to down-chirp symbol in 1m-distance condition; (b). Correlation to up-chirp symbol in 1m-distance condition; (c). Correlation to down-chirp symbol in 5m-distance condition; (d). Correlation to up-chirp in 5m-distance condition.

4 Conclusion and Future Work

In conclusion, we have developed a chirp-based BPSK modulation communication system on the smartphone platform to realize short-distance indoor communication. The implementation of convolutional coding enables the 1-meter indoor data transmission error rate to be within 0.001%. The experiments on a face to face scenario proved that the work is of practical significance, and the technology can be applied in short-range communication scenarios such as the exchange of payment information. However, the current data transmission efficiency is low, and the system cannot respond to the RIR change in the case of sudden non-line of sight (NLOS). In further work, we plan to design a real-time RIR estimation and compensation scheme to improve data transmission stability over longer distances and take Doppler effect into consideration.

References

1. H. Lichte, "The influence of horizontal temperature layers in sea water and the range of underwater sound signals," *Tracor Sciences & Systems*, 1919.
2. R. F. Estrada and E. A. Starr, "50 years of acoustic signal processing for detection: coping with the digital revolution," *IEEE Annals of the History of Computing*, vol. 27, no. 2, pp. 65–78, 2005.
3. Y. Yu, L. Zheng, J. Zhu, Y. Cao, and B. Hu, "Technology of short-distance wireless communication and its application based on equipment support," in *AIP Conference Proceedings*, vol. 1955. American Institute of Physics Inc., apr 2018, p. 040135. [Online]. Available: http://aip.scitation.org/doi/abs/10.1063/1.5033799
4. R. Kraemer and M. Katz, *Design Rules for Future Short-Range Communication Systems*, 2009, pp. 9–26.
5. H. Du and G. Xu, "Infrared indoor wireless mimo communication system using 1.2ghz ook modulation," *China Communications*, vol. 16, no. 5, pp. 62–69, 2019.
6. A. Nesic, V. Brankovic, G. Oberschmidt, T. Dolle, I. Radnovic, and D. Krupezevic, "Toward new generation of the high data rate in-door communication systems-system and key rf technologies," in *4th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Services. TELSIKS'99 (Cat. No.99EX365)*, vol. 1, 1999, pp. 232–235 vol.1.
7. L. Jakevičius and A. Demčenko, "Ultrasound attenuation dependence on air temperature in closed chambers," *ULTRAGARSAS*, vol. 63, no. 1, pp. 18–22, 2008. [Online]. Available: www.ndt.net/search/docs.php3?MainSource=27
8. A. Sakaushi, M. Okano, K. Kanai, and J. Katto, "Performance evaluations of software-defined acoustic mimo-ofdm transmission," in *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, 2018, pp. 1–6.
9. W. Jiang and W. M. D. Wright, "Multichannel ultrasonic data communications in air using range-dependent modulation schemes," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 63, no. 1, pp. 147–155, 2016.
10. M. Hanspach and M. Goetz, "On Covert Acoustical Mesh Networks in Air," *Communications*, vol. 8, no. 11, nov 2013.
11. S. Holm, O. B. Hovind, S. Rostad, and R. Holm, "Indoors data communications using airborne ultrasound," in *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, vol. 3, 2005, pp. iii/957–iii/960 Vol. 3.
12. H. Lee, T. H. Kim, J. W. Choi, and S. Choi, "Chirp signal-based aerial acoustic communication for smart devices," in *2015 IEEE Conference on Computer Communications (INFOCOM)*, 2015, pp. 2407–2415.
13. A. Nesic, V. Brankovic, G. Oberschmidt, T. Dolle, I. Radnovic, and D. Krupezevic, "Chirp signal-based aerial acoustic communication for smart devices," in *2015 IEEE Conference on Computer Communications*, 2015, pp. 2407–2415.
14. A. Viterbi, "Error bounds for convolutional codes and an asymptotically optimum decoding algorithm," *IEEE Transactions on Information Theory*, vol. 13, no. 2, pp. 260–269, 1967.