A Long-Range Aerial Acoustic Communication Scheme

Liming Shi , Xinheng Wang , Limin Yu , Wenwu Wang , Zhi Wang , Muddesar Iqbal , Charalampos C. Tsimenidis , and Shahid Mumtaz

Abstract—The effective communications range for aerial acoustic communication in indoor environment is a critical challenge for applications on smartphones. An inaudible acoustic communication system on Android phones that breaks the range limit is proposed in this paper, owing to an innovative design of the receiver by considering the time-frequency features of the received signals, using a fractional Fourier transform and 2-D mask filter to denoise the signals, and a hard detection and a soft detection approach to detect the symbols so as to tackle the severe multipath and Doppler effects in large indoor environments. Onsite test results show that an error-free communication distance of 70 m can be achieved at 20 bps, and the estimated maximum communication range could reach at least 85 m.

Index Terms—indoor long-range aerial acoustic communications, smartphone platform, chirp spread spectrum, fractional Fourier transform denoising, multipath delay spread.

I. INTRODUCTION

AERIAL acoustic communication (AAC) systems deliver information by transmitting the modulated acoustic wave from the speaker to the microphone in the aerial medium. AAC systems have been extensively developed in recent years [1]–[10]; however, the communication range is still a challenge for applications in large indoor environments. For example, acoustic indoor localization requires wireless communications to exchange ID information packets (8 payload bits + 3 redundant bits) with updating frequency of 1 Hz between the users and base stations [11], [12], where a throughput of 11 bps is enough for it. Meanwhile, the longer communication range means that the localization system can cover a larger area, which can geometrically reduce the cost of laying out base stations in large indoor scenarios, such as airport lounges. Due to the limitation of effective communication distance, other applications based on correct detection of acoustic signals, such as human motion recognition and object detection [13]–[18] are also affected. There is a strong need to improve the communication distance for the purposes of both communications and also acoustic applications.

As a supplement to electromagnetic (EM) wireless communications, AAC can overcome the weakness of EM shielding in particular scenarios due to the physical nature of the

L. M. Shi, X. H. Wang, and L. M. Yu are with Xi'an Jiaotong-Liverpool University, Suzhou, China
W. Wang is with University of Surrey, Guilford, UK
Z. Wang is with Zhejiang University, Hangzhou, China
M. Iqbal is with Prince Sultan University, Riyadh, Saudi Arabia
Charalampos C. Tsimenidis and S. Mumtaz are with Nottingham Trent University, Nottingham, UK

mechanical wave. Although various acoustic communication methods have been developed for indoor applications, the transmission distance is still a challenge owing to the complexity of the indoor environment. The complexity of the environment and the low speed of sound waves cause multipath reverberation and Doppler shift. Previous research from various researchers demonstrated the difficulty of developing long-distance acoustic communications. The multichannel ultrasonic communications using amplitude shift keying (ASK) achieved a 10 m error-free range in [19]. The application of orthogonal frequency division multiplexing (OFDM) plus ASK [6] or phase shift keying (PSK) [4] reached 10 m and 4 m, respectively. The maximum attainable distance in [20] was 20 m. The work in [3] achieved a 26.4 bps bit rate at 35m with a bit error rate (BER) at around 10%. Application of chirp spread spectrum (CSS) modulation in [8] reached a near error-free communication range of 25 m with a bit rate of 16 bps. An improved robust chirp spread spectrum modulation was proposed in [10], which reached an error-free communication range of 20 m with a bit rate of 125 bps. Currently, the existing studies on AAC are not suitable for our target scenarios.

Smartphones have become popular commercial off-the-shelf (COTS) products for individuals, meanwhile, the embedded micro-electromechanical system (MEMS) microphones and speakers are also the fundamental electronic components on smartphones, where both of them enable the smartphone to play a role as a mature platform for AAC technology. Owing to the variation of MEMS hardware and operating systems among COTS products, the sampling rate-dependent maximum bandwidth is varying from device to device. Currently, two main-stream audio sampling rates for Android and iOS phones are 48 kHz and 44.1 kHz, respectively, which defines the maximum bandwidth of AAC to be less than 22 kHz according to the Nyquist sampling theorem [21]. Based on human intuition, the frequency band under 22 kHz is divided into the inaudible band that is above 18 kHz and the audible band that is below 18 kHz. In audible band, AAC is further divided into "melodic" band which can be noticed by people, and "hidden" band where the used band is hidden by the background "noise" and will not be noticed. The "melodic" communications such as [2] uses frequency modulation to modulate the signal to sound like melody, but this method cannot reach a long transmission range because its frequency band overlaps with the frequency band of natural noise. For another modulation method on the audible band, the modulation on the "hidden" band inserts the modulated signal into the playing audio, where the power spectral density of the

modulated signal is adaptive at a low level which is slightly higher than that of the background audio [3]–[6]. Therefore, the insertion of modulated signal cannot be noticed by people but is detectable by the modulator. In terms of modulation on inaudible frequency band, these AAC systems use the ultrasonic or near-ultrasonic wave that is inaudible to most people [8]–[10].

Although AAC on smartphones is implemented successfully [5], [7], [9], [22], same challenge of short communication distance still remains. In previous research, most researchers realize AAC on smartphone platforms within the distance of 10m, while the maximum attainable communication distance is 25 m in the indoor environment [8].

In order to reduce the negative impact of multipath delay spread, the designed receiver in this paper demodulates the CSS signal based on the time-frequency feature of the signal. The feature is observed to be independent with distance change, and robust to the multipath delay spread and transient Doppler spread interference. In addition, inspired by estimating the chirp rate on radar systems [23], [24], a denoising method using fractional Fourier transform (FrFT) is proposed in this paper. Another innovative approach in the design of the receiver is demodulation. Instead of using correlators, a new method to demodulate the symbols by estimating the chirp rate is proposed. By implementing these approaches, the transmission distance can reach 70 m with a BER less than 10^{-4}, which is significantly longer than the distance reported in previous research. This will enable smartphones to support applications in large indoor environments.

Apart from these three key techniques proposed in this paper, two decoding schemes are proposed to balance the requirements between error correlation and transmission efficiency. All these measures ensure a long-distance and error-free AAC. In summary, the following contributions are made in this paper:

- Time-frequency feature is considered in denoising and demodulating the CSS signals on the receiver. The designed signal is modulated by the chirp rate. The receiver deconstructs and reconstructs CSS signals through fractional Fourier transform, and applies a window-shifted bandpass filter to reduce the non-linear time-frequency featured noise owing to the distortion by multipath delay spread and Doppler spread. The designed receiver demodulates the signal by applying chirp rate estimation, which exhibits better performance than those using conventional correlator demodulators.
- Inspired by the chirp rate estimation in radar systems, an FrFT denoising method is designed and implemented on the receiver. In the tests, the acoustic communication system can achieve 70 m error-free communications.

The rest of the paper is organized as follows: related work is introduced in Section II. The design of the whole system is presented in Section III. The experiments, results analysis, and performance evaluation are presented in Section IV. Finally, the work is concluded in Section V.

II. Related Work

AAC for audio-enabled devices have been studied for over 20 years. To the best of our knowledge, the first related research was presented in [1]. In [1], three AAC protocols were proposed which enabled a maximum data rate of up to 3.4 kbps at a distance of 1 m. The proposed system adopted variants of frequency shift key and amplitude modulation in its two-frequency modulated protocol (audible band), ultrasonic-frequency modulated protocol (inaudible band), and spread-spectrum modulated protocol (wide band covering audible and inaudible band). Because the system in [1] did not consider countermeasures against multipath fading, the given waveform had been already significantly distorted at 1 meter which affected demodulation. In the research of audible band AAC, "hidden" band is a hot topic of AAC that uses audible frequency band but insensible to human beings [3]–[5]. Research in [4] and [5] both adopted OFDM plus PSK or ASK as the modulation scheme. In [4], information was hidden behind music by replacing the high-frequency part of the original audio with the modulated signal. The communication scheme achieved a 4-meter communication range with a data rate of 900 bps and an error rate of approximately 0.1. The Dolphin system was proposed in [5]. Different from research in [4] that cut the original audio, the Dolphin system added ASK-modulated information at the frequency band which has a minimal effect on the original audio. The Dolphin system achieved a 1-meter communication range with a data rate of 500 bps and an error rate of approximately 0.05. In the research of inaudible band AAC for audio-enabled devices, the realization of long-range AAC is the common target for research in [8], [9]. In [8], the chirp signal was first introduced as the base signal in AAC. The work adopted binary orthogonal keying (BOK) and reached a maximum communication range of 25 m with a data rate of 16 bps and an error rate of approximately 0.03. The work in [9] adopted quaternary orthogonal keying (QOK) based on BOK.

Another strategy to extend the communication range is the implementation of relay networks. In [3], the idea of an ad-hoc relay architecture was implemented to improve spatial gain through collaborative acoustic communications among multiple smartphones. A relay-based mesh architecture was adopted in [25], and it demonstrated the benefits of spatial gain through collaborative communication under non-line-of-sight (NLOS) scenarios. However, the multiple-hop architecture increases the costs of deploying more devices and the complexity of the system.

The aforementioned research has a few limitations. The studies of AAC in an audio-band [1], [3]–[5] are not suitable for long-range communication scenarios. In addition to the reason mentioned above that the modulated signal in the audible-band overlaps with the noisy band of ambient noise, the audible sound may become annoying noise, especially in long-range scenarios where high transmission power is required. Research in [8], [9] remains limitations to adapting the multipath fading channel in the further distance. Both BOK [8] and QOK [9] adopted conventional CSS modem that uses correlators to calculate and compare the cross-

correlation between symbols. However, the designs of their equalizer or denoising filter and their symbol decider is not robust enough to handle time-variant phase noise caused by the multipath fading. Phase noise can cause a significant distortion to the result of demodulation based on the correlator, which prominently reduces the peak value in autocorrelation and results in equivocation, especially in low SNR scenarios in the long range communications.

III. SYSTEM DESIGN

Details of the whole AAC system are described in this section, where the key techniques in designing the receiver are described.

A. Transmitter

Fig. 1 displays the schematic design of the transmitter. Firstly, the message is coded by convolutional coding and is rearranged by an interleaver. Secondly, the data bits are modulated depending on the chirp rate. Thirdly, the CSS-modulated signal is multiplied by the carrier signal and passed through a bandpass filter to prevent frequency leakage. Technical details of the waveform and coding are described below:

1) Modulation and Waveform: Indoor acoustic communications are inevitably affected by the multipath effect and mobility that lead to a fast-fading channel. The multipath effect generally arisen from the physical nature of wave propagation where reflection, refraction, and scattering result in the source-to-destination paths that are difficult to enumerate. The number of the maximum reflection times is limited by the energy loss factors, e.g. reflection loss, and atmospheric attenuation, which are dependent on the structure and size of the room. The phenomenon of reverberation is highly associated with the multipath effect, especially the reflection. In addition, audible noise is also a sensitive issue, in order to avoid frequency leakage at the low-frequency band, chirp signal is chosen to be the base spread signal. Therefore, a chirp modulation is adopted in this paper. The function of the adopted chirp waveform $s(t)$ is given as (1):

$$s(t) = \begin{cases} \sin(2\pi f_c t + \mu\pi(-1)^c t^2) & , \forall t \in -\frac{T}{2} \le t \le \frac{T}{2} \\ 0 & , otherwise \end{cases} \quad (1)$$

where f_c denotes the carrier frequency, μ represents the chirp rate that defines how fast the frequency changes with time, T is the symbol duration in time, and c is the control parameter of modulation that defines the chirp rate. A rational frequency band for the modulated signal is between 18 kHz and 22 kHz, where 18 kHz is the limit of hearing for most people [8] and 22 kHz is the upper bound when 44.1 kHz sampling rate is adopted.

2) Convolutional Coding: Convolutional coding is a forward error correction coding (FEC) method that promotes error correction capacity. It is typically applied in non-feedback communication scenarios, e.g., satellite-to-ground communications [26]. Based on the trade-off between channel gain, coding rate, and decoding complexity [27], a code rate of 1/2 with the convolution length of 3 is adopted in the transmitter.

For an input binary bit X, the convolutional coding obtains a dual bit output $Y_{1,2}$.

$$Y_{1,2} = \text{LSB}\lceil X * C_{1,2}\rfloor_2, \quad (2)$$

where LSB means the calculation only considers the least significant bit of the convolution result, symbol $*$ denotes the operator of the convolution, $\lceil \cdot \rfloor_2$ represents the operation of binary conversion, $C_{1,2}$ represents the used two convolutional codes corresponding to the dual bit outputs $Y_{1,2}$ that are $[1,0,1]$ and $[1,1,1]$, respectively. The usage of convolutional coding is assisted with an algorithm-defined random interleaver. After the encoding of convolutional code, the interleaver reorders the sequence with an input of a random seed. The seed is required by the receiver in advance to use it as the authentication in decoding the received signal.

B. Receiver

A brief block diagram is depicted in Fig. 2. The designed receiver consists of five components, which function as signal acquisition, signal denoising, signal synchronization, demodulation, and symbol decision. The function of signal acquisition is achieved by calling the sound recording application in the smartphone's operating system. The sound is acquired by a 48 kHz sampling rate and is encoded and saved as a '.wav' sound file. The details of other components are described in the following subsections:

1) Fractional Fourier denoising and Signal Synchronization: In the time-frequency domain, the power density pattern of a linear chirp exhibits an inclined narrow bar, while the slope of the bar is equal to its chirp rate. With regard to denoising the linear chirp in the time-frequency domain, the ideal purpose is to reserve the narrow bar and deduct the rest. Commonly, a spectrogram in the time-frequency domain is measured by short-term Fourier transform (STFT), where the frequency resolution can be adjusted by changing the length of fast Fourier transform (FFT), and the time resolution can be scaled by setting the overlap rate of the window function. Fig. 3(a) illustrates the least demand of a passband area of a linear chirp filter. Through the adaptation of Fourier transforms length and window overlap rate, the pattern of the filter is depicted in Fig. 3(a) can be rescaled as that in Fig. 3(b). It obverses that rescaling can significantly shrink the passband area. However, rescaling is accompanied by an increase in time-frequency resolution, which means a higher Fourier transform size and more frequency overlapping, prominently increasing the amount of computation. Therefore, the proposed method introduces FrFT [28]. FrFT can rotate the signal to an appropriate angle, as shown in Fig. 3(c) that FrFT rotates the signal vertically. Compared with Fig. 3(a), Fig. 3(c) shrinks the passband without increasing resolution, and it even leaves room for the drop in vertical resolution.

The mathematical expression of continuous fractional Fourier transform is given in (3)

$$\mathcal{F}_\alpha(\omega) = A_\alpha(\omega) \int_{-\infty}^{\infty} e^{-j\csc(\alpha)\omega t + j\frac{\cot(\alpha)t^2}{2}} x(t)\, dt \;, \quad (3)$$

where $j = \sqrt{-1}$ is an imaginary number, ω is the fractional Fourier domain axis whose counter-clockwise angle with time
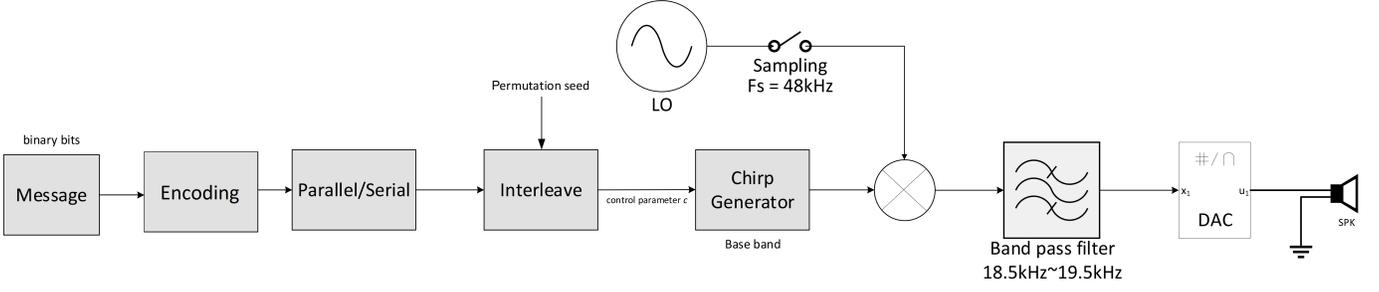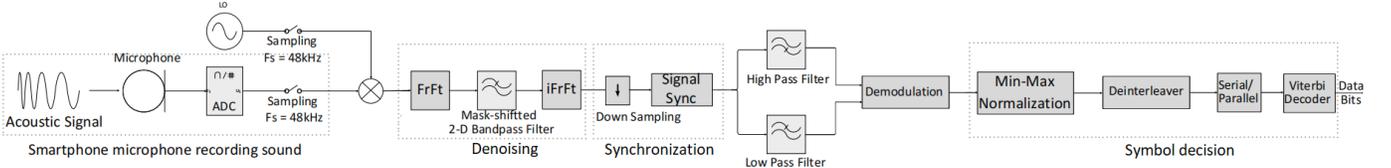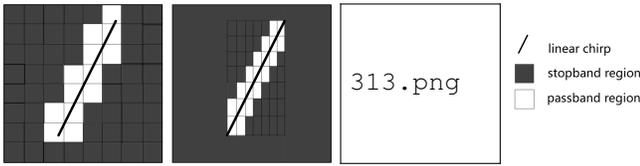
Fig. 1. Block diagram of the transmitter.

Fig. 2. Block diagram of receiver system.

(a) 2-D Filtering in Time Frequency Domain (b) Scaling the Time-Frequency Domain (c) 2-D Filtering in FrFT Domain

Fig. 3. The least filter area for different transform methods.

axis is α in time-frequency complex domain, where α denotes the angle that rotates and

$$A_\alpha(k) = \sqrt{\frac{1 - j\cot(\alpha)}{2\pi}} e^{j\cot(\alpha)\omega^2/2} \qquad (4)$$

Equation (4) describes the transformation on magnitude.

In digital signal processing, a discrete system has a finite bandwidth that is defined by a Nyquist rate [21]. The maximum Android-platform-supported audio sampling rate, F_S, is 48 k samples per second. The transformation from continuous signal $s(t)$ to discrete signal $s[n]$ is presented in (5) where n should be an integer that denotes sample index, and $\delta(t)$ is a Dirac delta function.

$$s[n] = \int_{-\infty}^{\infty} \delta\left(t - \frac{n}{F_S}\right) s(t)dt \qquad (5)$$

In a discrete system, the FrFT for a finite signal sequence $s[n]$ with length M is expressed in (6), where the rightest exponential function in (6) is considered as an expression of linear chirp function $c[n]$ in (8). The linear chirp function $c[n]$ is also the base function of FrFT.

$$\mathcal{F}_\alpha[k] = A_\alpha(k) \sum_{n=0}^{M-1} e^{-j\csc(\alpha)\frac{2\pi kn}{MF_S} + j\frac{\cot(\alpha)n^2}{2F_S^2}} s[n] \ , \qquad (6)$$

where

$$A_\alpha(k) = \sqrt{\frac{1 - j\cot(\alpha)}{2\pi}} e^{\frac{j}{2}\cot(\alpha)(\frac{2\pi k}{M})^2} \qquad (7)$$

$$c[n] = e^{-j\csc(\alpha)\frac{2\pi kn}{MF_S} + j\frac{\cot(\alpha)n^2}{2F_S^2}} \qquad (8)$$

Eq. (6) can be simplified as (9), which simplifies the right component as the convolution between $c[n]$ and time reverse sequence of $s[n]$.

$$\mathcal{F}_\alpha[k] = A[k] \sum_{k=0}^{M-1} c[k]s[k] = A[k]c[n] * s[M - n] \qquad (9)$$

According to the discrete FrFT given in (9), the fractional Fourier dimension, whose axis is labeled ω in Fig. 4(a), rotates from the time axis to the angle of the chirp rate of $c[n]$. The dimension ω analogizes the frequency dimension in the time-frequency domain, and an added axis named t^* that is perpendicular to ω analogizes the time dimension. The linear chirp in FrFT domain in Fig. 4(b) can be filtered by a shadow mask. The width of the mask filter is defined by an equivalent bandwidth in FrFT as (10), where F_H and F_L are the highest and lowest frequency of linear chirp, respectively. T_{symbol} denotes symbol period of linear chirp in the time domain. The period of linear chirp in FrFT domain T_S^* determines the shifting step of mask filter, which is a transformation of symbol period that is expressed in (11).

$$B = \csc(\alpha)(F_H - F_L) + \cot(\alpha)\cos(\alpha)F_S \qquad (10)$$

$$T_S^* = \sin(\alpha)T_{symbol} \qquad (11)$$

Based on discrete FrFT, the denoising method is designed as three steps depicted in Fig. 5. The first step slices the received signal into frames with the size of symbol duration in the time domain, then the algorithm performs FrFT on each frame by

(a) discrete fractional Fourier transform

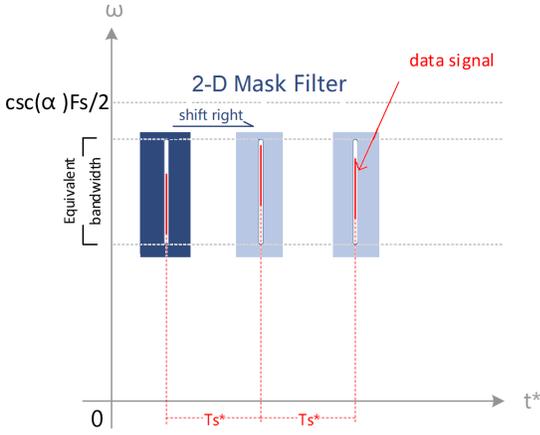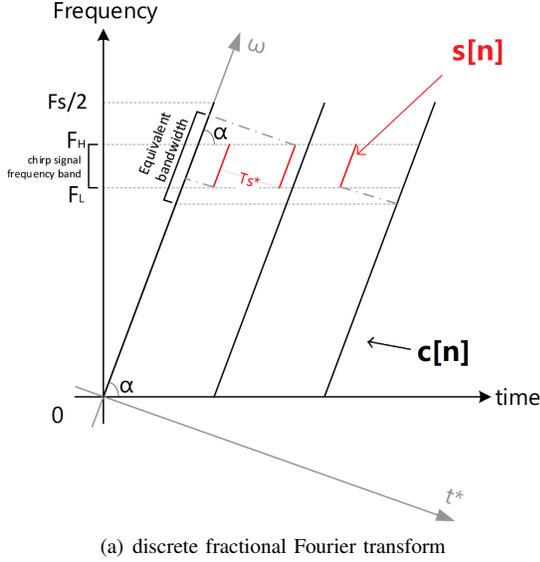(b) data signal presented in fractional Fourier domain

Fig. 4. De-noising method using discrete fractional Fourier transform.

transforming an angle ($\pm\alpha$), corresponding to the chirp rate of two symbols. The second step applies mask filtering on the signal in the FrFT domain. In order to mitigate the effect of inter-symbol interference (ISI), the mask filter is pulsed and shaped by a root raised cosine window with a roll factor of 0.25. In the final step, the algorithm converts the signal back to time domain through FrFT with a reverse angle, and merges the superposition of frames, finally outputs the noise-reduced signal in the time domain.

2) Signal Synchronization: For the proposed denoising algorithm in the previous section, it is demanding to search the position of each symbol during the process of framing. A signal synchronization algorithm is crucial to guarantee the accuracy of framing. In the proposed work, two fixed-up chirp signals are used as preamble bits. Through the knowledge of the initial position of the first preamble bit, the position of the ith symbol p_i can be obtained in (12).

$$p_i = p_0 + (i+1)(T_G + T_S), \forall i \in 1, 2, \cdots, K \quad (12)$$

where p_0 is the position of the first bit in the preamble, K is the number of total payload bits, and T_G and T_S denote the duration of the guard interval and symbol, respectively. Fig. 6 displays the algorithm to search the initial position of the first preamble bit. Firstly, a queue is created to allow the signal in, and the signal in the queue is shaped by a window function, where the window function only allows the passing of two adjacent bit symbols. Then, the algorithm frames all moments during the signal passing through the queue and records the position of the first window. Based on the denoising method mentioned in the previous section, the algorithm performs FrFT on each frame and applies the 2-D mask filtering, then performs inversely FrFT on the filtered result and calculates the root mean square (RMS). The physical meaning of the RMS result is the energy of the frame. Finally, through ranking the energy of each frame, the first found position where the frame energy is the highest can be considered as the most likely p_0.

3) Demodulation: The linear chirp signal is demodulated by estimating its chirp rate. For an approximate estimation of binary chirp modulation, chirp rate estimation can be achieved by observing the sign of the chirp rate. In the proposed algorithm, the frequency band of the chirp symbol is divided into two regions: low-frequency region and high-frequency region, which are separated by a high pass filter $h_H[n]$ and a low pass filter $h_L[n]$, respectively. Preprocessed by denoising and signal synchronization, the obtained output is a noise-reduced signal $\theta[n]$ with the position of each symbol p_i. The demodulation function with input variables $\theta[n]$, p_i is displayed in (13). Table I displays all included parameters. Because the demodulation calculates the relative energy ratio instead of the absolute energy ratio, the demodulation is robust to the frequency selective channel. Considering the symmetry in the demodulating result of binary symbol, the designed high pass and low pass filters require to have the same passband, which is $f_l = B - f_h$. In this condition, the coefficient η is set to $f_h(T_S/B)$, and the coefficient ϵ is an empirical constant with a value of $0.05T_S$. A demodulation example is presented in Fig. 7. The modulated signal $\theta[n]$ is separated into high-frequency and low-frequency parts by filtering, and then the energy is calculated. The output result of $DM[p_i]$ is obtained according to (13) and is forwarded to the step of symbol decision.

$$DM[p_i] = \log \frac{E\{H[p_i : p_i + \epsilon]\}E\{L[p_i - \eta : p_i - \eta + \epsilon]\}}{E\{L[p_i : p_i + \epsilon]\}E\{H[p_i - \eta : p_i - \eta + \epsilon]\}}$$
$$= \log \frac{E\{(\theta * h_H)[p_i : p_i + \epsilon]\}E\{(\theta * h_L)[p_i - \eta : p_i - \eta + \epsilon]\}}{E\{(\theta * h_L)[p_i : p_i + \epsilon]\}E\{(\theta * h_H)[p_i - \eta : p_i - \eta + \epsilon]\}}$$
$$(13)$$

4) Symbol Decision: The proposed symbol decision adopts the sequential steps of deinterleaving, normalization, and decoding. With the pre-ordered permutation seed of interleaver, the deinterleaver first rearranges the demodulation result. The symbol decider first normalizes the demodulation results into a range between 0 and 1 with a criterion of min-max. The min-max applies the sigmoid function to magnify the influence of the ambiguous bits and restrain the influence of the extreme
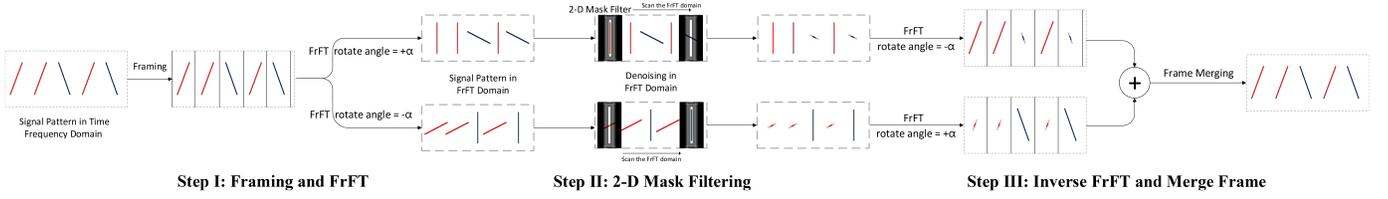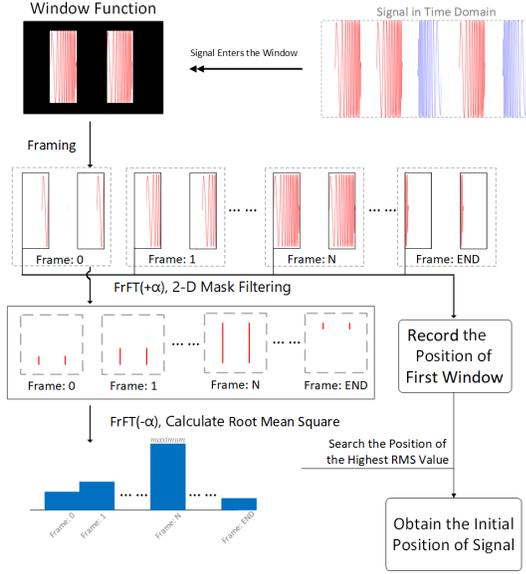
Fig. 5. Three steps of the FrFT denoising method.

Fig. 6. Block diagram of the Signal Synchronization Algorithm.

TABLE I
PARAMETERS IN THE DEMODULATION FUNCTION

Parameters	Description
$DM[p_i]$	Demodulation function.
$E\{x[0:n]\}$	Energy function: $E\{x[0:n]\} = \sum_{k=0}^{n} \|x[k]\|^2$.
$H[n]$	$\theta[n]$'s high-frequency component: $H[n] = (\theta * h_H)[n]$.
f_h	Cut-off frequency of the high-pass filter.
$L[n]$	$\theta[n]$'s low-frequency component: $L[n] = (\theta * h_L)[n]$.
f_l	Cut-off frequency of the low-pass filter.
B	Bandwidth of the chirp waveform.
T_S	Symbol duration.
η	Shift-right coefficient.
ϵ	Signal fragment length in energy calculation.

bits, where the extreme bits are ubiquitously caused by burst error. In binary symbol decision, the process of soft decision assigns the physical meanings of likelihood to the bits. The value closing to 1 means the symbol is highly likely to be binary '1', and vice versa. The symbols with soft decision values around 0.5 are called ambiguous bits. These bits are influenced by the ambient interference that shift themselves from the original value and becomes ambiguous. The decoder adopts Viterbi decoder [29], which calculates the likelihood of all rational combinations of convolutional coding bits. Through searching for the highest likelihood combination, the Viterbi decoder can make the highest likely symbol decision.
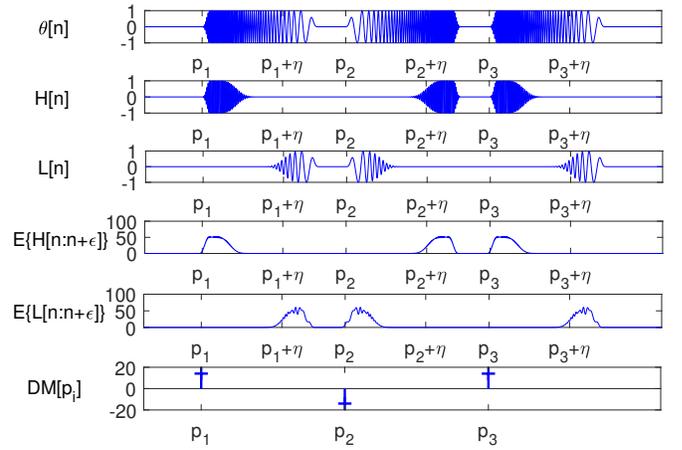

Fig. 7. Demodulation result of binary '010' bit sequence.

IV. EXPERIMENTS, RESULTS ANALYSIS AND PERFORMANCE EVALUATION

The relative experiments are conducted in a long office corridor. A diagram with the precise size of the corridor, which is measured by a Bosch GLM 500 laser meter, is depicted in Fig. 8. The corridor is basically a closed cuboid-like room. In Fig. 8, six points on the left-hand side enumerate the locations where a receiver is placed, while the point at the far right is the fixed location for the transmitter. The distances from the receivers to the transmitter are 70 m, 60 m, 50 m, 40 m, 35 m, and 25 m, respectively. The channel's frequency response and multipath and fading effects are measured and analyzed first to assess the channel quality. Because the hardwares of COTS devices are various, experimental results also vary from the models of smartphones. In this paper, the test on two smartphone samples are evaluated and their performance is evaluated.
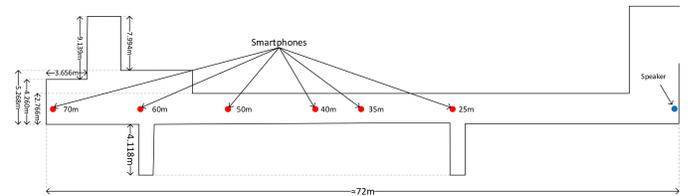

Fig. 8. Diagram of the corridor for testing.

A. Frequency Response Analysis

Due to the fluid motion, unbalanced distribution of heat, humidity, and particles in the air, channel in the uneven air

medium is typically frequency selectively. CSS is a kind of frequency modulation, where the transmission of the CSS signal is sensitive to the frequency response of the channel. Channel frequency response analysis is important to evaluate the potential effect on AAC. In the experiments, a JBL Flip 5 Bluetooth speaker and two smartphones, Honor V30, and Honor 30Lite, are used as transmitters and receivers, respectively. The JBL Flip 5 is directly connected to a laptop via Bluetooth 4.0 with a low complexity subband coding (SBC) codec. The JBL speaker plays white Gaussian noise for 10 seconds, and two smartphones are placed 1 m away from the speaker to record the sounds. Through calculating the average power spectral density (PSD) via FFT, the frequency responses of two smartphones are displayed in Fig. 9. The unit 'dBFS' is the abbreviation of Decibels relative to full scale. It describes the relative power of the digital signal compared to the reference signal with a magnitude of full scale. Generally, both channels exhibit obvious frequency-selective characteristics. The highlighted window denotes the near-ultrasonic frequency band that is inaudible to most people. Within the selected frequency band, the line of the frequency response changes fast and significantly at the frequency band of around 20 kHz for two phone models. It seems that the PSD of two channels at a frequency over 22 kHZ drops to the bottom, which means the weak signal gain is not suitable to be chosen. However, two channels are relatively flat at around 19 kHz, which is suitable as the channel band.

B. Channel Multipath and Fading Analysis

The channel multipath and fading characteristics based on the modeling method in [30] are analyzed. The room impulse response (RIR) and probability distribution function (PDF) of the signal envelope magnitude are measured by experiments and analyzed. RIR represents the channel impulse response in AAC. In experiments, a linear frequency-swept method is used to make the measurement of RIR [31], which is processed by following steps. Firstly, the transmitter sends a chirp signal $s(t)$, where an example of $s(t)$ can be shown as

$$s(t) = \begin{cases} \sin(38000\pi t + 2000\pi t^2) & , -5\,\text{s} < t < 5\,\text{s} \\ 0 & , otherwise \end{cases} \quad (14)$$

Then, the receiver captures the transmitted signal and records it as $y(t)$, while the transmitted signal is transformed by impulse response $h(t)$ and with noise $n(t)$ that is given as (15).

$$y(t) = \int_{-\infty}^{\infty} s(\tau)h(t-\tau)d\tau + n(t) \quad (15)$$

Finally, a convolutional product is calculated by the received signal $y(t)$ with a time-reversal chirp signal $s(-t)$, expressed as

$$y(t) * s(-t) = s(t) * s(-t) * h(t) + n(t) * s(-t), \quad (16)$$

where the function $s(t) * s(-t)$ is the autocorrelation function of $s(t)$, which is an impulse shaped function. The pulse width of the autocorrelation function is inversely proportional to the time-bandwidth product (TBP) of the chirp signal $c(t)$, and the autocorrelation function of the chirp function will be an

ideal impulse function $\delta(t)$ if the TBP tends to infinity. Hence, it is considered that the autocorrelation function of chirp with a large enough TBP is approximately equal to the impulse function, and (16) is equivalent to (17).

$$y(t) * s(-t) \approx \delta(t) * h(t) + n(t) * s(-t) \equiv h(t) + n(t) * s(-t) \quad (17)$$

In a silent test environment, the effect of ambient noise $n(t)$ on RIR calculation is relatively negligible. Based on the method, the RIR measurements in line-of-sight (LOS) scenarios at 25 m, 50 m, and 70m in the corridor are depicted in Fig. 10(a), Fig.10(b), and Fig.10(c), respectively. The effect of multipath to wireless communication is evaluated by calculating the mean excess delay $\bar{\tau}$ in (18), root mean square (RMS) delay spread σ_τ in (19) [32], and maximum excess delay $\tau_{\max}$.

$$\bar{\tau} = \frac{\sum_k a_k^2 \tau_k}{\sum_k a_k^2} \quad (18)$$

$$\sigma_\tau = \sqrt{\bar{\tau^2} - (\bar{\tau})^2}, \quad (19)$$

and

$$\bar{\tau^2} = \frac{\sum_k a_k^2 \tau_k^2}{\sum_k a_k^2}, \quad (20)$$

where a_k represents the amplitude of a single impulse with the index of k, and the τ_k means the relative delay time between kth impulse with the highest impulse. The maximum excess delay is the time difference between the first important impulse and the last important impulse. With the consideration of impulses with normalized magnitude higher than 0.2, the experimental results are displayed in TABLE II. The first arrival impulse is generated by the energy of the acoustic waves propagating along the nearest path, which occupies the dominant percentage of components in three scenarios. Particularly, the most dominant energy component at 25m is generated by the superposition of multipath reverberation other than from the path of first arrival, where the first arrival impulse occupies the second highest energy component; both cases of 50 m and 70 m exhibit the highest impulse at the first arrival component. In addition, the maximum excess delay is inversely proportional to the distance. A rational explanation of this phenomenon is owing to the high atmospheric attenuation of the acoustic wave. Referring to the research on atmospheric attenuation in [33], the atmospheric attenuation level of the tested frequency reaches around 1 dB/m. Therefore, the phenomenon can be explained with consideration of the high atmospheric attenuation: in the case of 25 m, the space of the testing environment is relatively small-scaled, and the distances of the multiple-reflection wave propagation path are relatively shorter compared to those in the other two cases. Since the descent in the distance reduces the energy loss caused by the atmospheric attenuation, the probability of capturing the acoustic wave from multiple reflections is increased, which explains the number of reverberations is the largest among the three in the case of 25 m.

Beta distribution, Nakagami distribution, Rayleigh distribution, Rician distribution, and Weibull distribution are considered to analyze the fading channel model in the corridor. The Kullback-Leibler (KL) divergence [34] is used as a loss
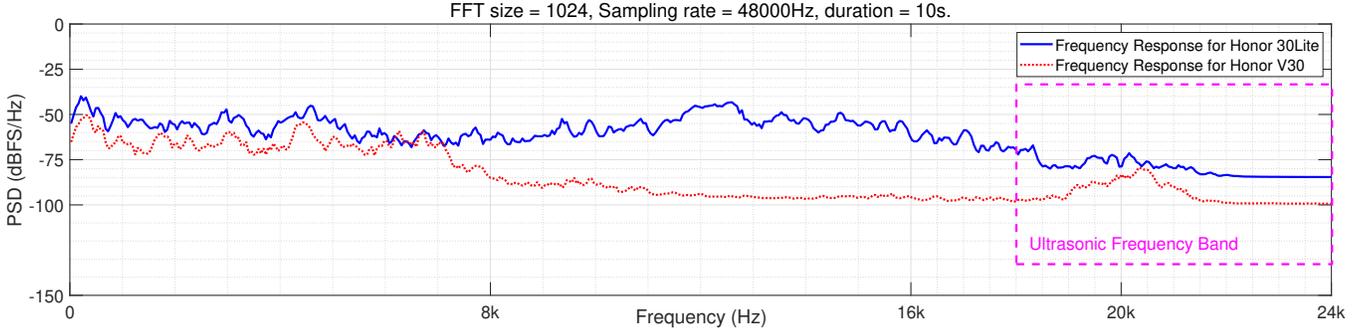
Fig. 9. Frequency response between two smartphones and the JBL Flip 5 Bluetooth Speaker at 1 m.
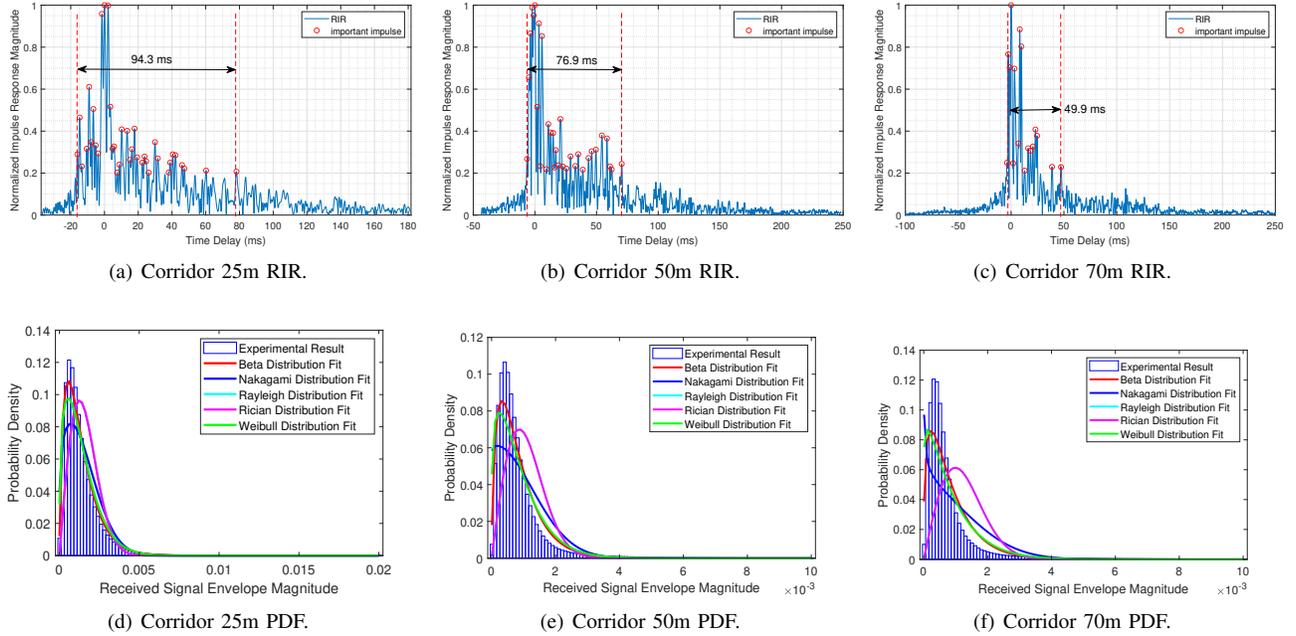

(a) Corridor 25m RIR.

(b) Corridor 50m RIR.

(c) Corridor 70m RIR.

(d) Corridor 25m PDF.

(e) Corridor 50m PDF.

(f) Corridor 70m PDF.

Fig. 10. Experimental results of room impulse response analysis and fading channel modeling.

TABLE II
SPREAD DELAY PARAMETERS

Distance	$\bar{\tau}$	σ_τ	$\tau_{\max}$	Reverberations
25 m	5.4 ms	15.5 ms	94.3 ms	37
50 m	7.9 ms	16.9 ms	76.9 ms	32
70 m	6.2 ms	9.2 ms	49.9 ms	17

function in distribution fitting. KL divergence is a relative entropy that measures the similarity between two probability distribution functions (PDFs). The expression of KL divergence is expressed as

$$D_{\mathrm{KL}}(P \parallel Q) = \sum_{x \in \mathcal{X}} P(x) \log \left(\frac{P(x)}{Q(x)} \right) \qquad (21)$$

where P and Q are two PDFs with variable x that are normalized in the same probability space. Typically, the value of KL divergence is closer to zero when two PDFs are more similar. The fitting results are depicted in Fig. 10(d), Fig.10(e) and Fig.10(f) with KL divergences displayed in TABLE III. Among these distributions, Beta distribution exhibits the

highest similarity with the PDF of received signal envelope magnitude.

TABLE III
KL DIVERGENCE FOR DIFFERENT DISTRIBUTION FIT

Distribution	**Beta**	Nakagami	Rayleigh	Rician	Weibull
Corridor at 25m	**0.02**	0.12	0.29	0.29	0.05
Corridor at 50m	**0.07**	0.25	0.53	0.53	0.11
Corridor at 70m	**0.14**	0.37	0.91	0.91	0.17

C. BER Performance

The performance of the proposed communication method is simulated on an additive Gaussian noise (AWGN) plus Beta distribution fitted fading model. Fig. 11 displays the simulation results of the BER versus E_b/N_0 on different symbol decision strategies, where the symbols E_b/N_0 is defined by energy per bit over noise energy. Three strategies are compared, including hard decision decoder with uncoded modulation, hard decision Viterbi decoder with convolutional coding, and soft decision Viterbi decoder with convolutional coding. The strategy of a

hard decision decoder with uncoded modulation exhibits the baseline of the communication capacity. With the assistance of error correction by applying the convolutional code and Viterbi decoder, both hard-decision and soft-decision strategies achieve prominent coding gains, where the hard-decision strategy achieves 3 dB coding gain when BER is approximately 10^{-3}, meanwhile, soft-decision strategy performs better with 1.5 dB coding gain exceeding that of hard-decision strategy. In addition, it is also observed that interleaving has little effect on this simulated channel.
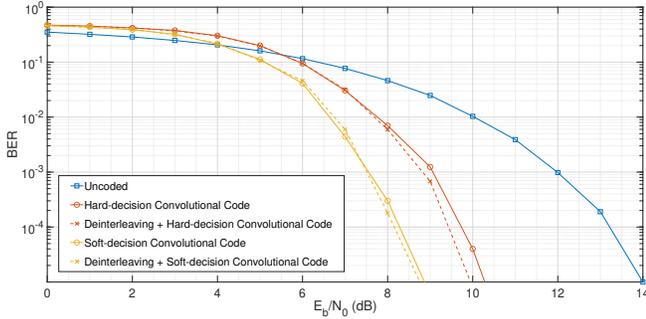

Fig. 11. The simulation of BER versus SNR using the proposed method in the AWGN plus Beta-distribution randomly generated fading channel.

D. Transmission Power Analysis

The output power of the used speaker is analyzed by using ACOEM's 01dB vibration analyzer. In experiments, the signal modulated by the proposed method is played in the corridor, and the vibration analyzer is placed at 1 m, 5 m, 10 m, 30 m, and 50 m far away from the speaker. The selected experimental data is displayed in TABLE IV. Because the modulated frequency is close to 20 kHz, the sound pressure level at 20 kHz is considered to represent the major power of the modulated signal.

TABLE IV
MEASURED SOUND PRESSURE LEVEL IN THE CORRIDOR

Range Frequency	1 m	5 m	10 m	30 m	50 m
	Sound Pressure Level (unit: dBSPL$_{air}$)				
10 Hz	41.9	44	36.6	49.5	39.6
100 Hz	36.8	36.7	43.5	42.5	40.7
1000 Hz	32.3	31.8	32.9	34.8	42.4
10000 Hz	15.5	11.3	24.7	15.5	21.9
16000 Hz	64.8	47	43.8	36.4	22.9
20000 Hz	*82.6*	*66*	*61.9*	*52.6*	*38.5*

E. Burst Noise Experiment

The burst noise experiment is implemented to analyze the resistance of the proposed method to unexpected sudden impulse noise in real life. In the experiment, A random bit sequence is generated on Matlab R2020b in advance, which is used as the test data. The test data is further modulated into a playable audio signal using the proposed algorithm, where the audio is played by the JBL speaker. In addition, an interference

signal with the same duration as the audio is generated, which is composed of a series of discretely distributed impulse noise signals. The impulse signal is generated by the pulse-shaped white Gaussian noise with a pulse width within the range between 2.5 and 200 milliseconds. Each impulse noise signal is randomly distributed in the time domain based on normal distribution, and the quantity of impulse noise is defined as the percentage of the sum of total pulse width length over the test data signal duration. In the experiment, four levels of impulse noise quantity are tested, including 1%, 10%, 20%, and 50%. The impulse noise is played by an additional speaker synchronously with playing the test data, where the energy level of impulse noise is guaranteed to be higher than the data signal by placing the additional speaker closer to the receiver smartphone and setting a higher volume. The audio is recorded and saved on the phone Honor 30Lite, and the audio file is sent to the computer. The proposed algorithm of the receiver is conducted on Matlab R2020b and the results are evaluated. The experimental results for three different impulse noise quantities are listed in Fig. 12. From the results, the mere adoption of soft-decision plus convolutional code strategy performs the worst because the soft-decision decoder of convolutional code is a causal system. Impulse noise affects the soft-decision weight calculation of a segment of bits, and this segment of affected bits will further affect subsequent bit decisions. The usage of interleaver rearranges the noise-affected bits globally, diluting the density of affected bit segments, thereby improving the error correction capacity of convolutional code. Additionally, it is observed that the convolutional code plus interleaving strategy exhibits a lower BER rate when the ratio of burst noise is under 10%. Generally, the scenario that the burst noise ratio is over 10% is rare in most indoor channels, and the decision strategy definitely enhances the resistance to burst noise in most scenarios.
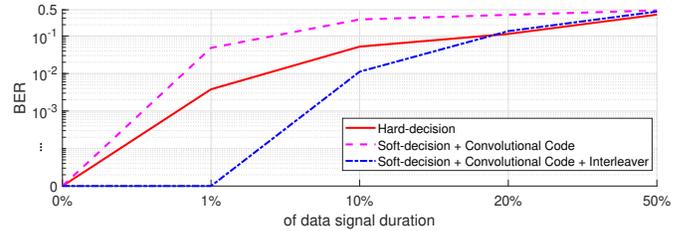

Fig. 12. BER performance of three different decision strategies in the extremely frequent burst-noise channel.

F. Long-range Communication Test

In the long-range communication test in the testing environment mentioned in Fig. 8, the evaluation of bit error rate in different conditions of transmission distance and data rate is focused. The transmission success rate, defined by one minus bit error rate, is evaluated at different distances, and the results are also compared when using different transmission bit rates at the same distance. The preparation of test data is implemented on Matlab R2020b, multiple sets of randomly generated data bits are generated and further modulated into

the playable audio files according to the proposed modulation algorithm. In the experiment, the set of test data is randomly selected to play, and the sound is recorded by smartphones and labeled. After several rounds of experiments, the sound signal is zipped and uploaded to the computer, and demodulation and evaluation are performed on Matlab R2020b. Relevant parameters are presented in TABLE V. The experimental results of transmission success rate are listed in Table VI. Condition "*hard decision no convolutional code*" refers to the direct hard-decision demodulation of the received signal without considering the convolutional code and interleaving. Its transmission success rate is evaluated with the post-interleaved signal, which reflects the baseline of the communication capacity. It is observed that the proposed method achieves the zero-error rate using a bit rate of 10 bps overall testing ranges.

TABLE V
EXPERIMENTAL PARAMETERS IN THE LONG-RANGE
COMMUNICATION TEST

Symbol	Description	Value
Rx	the models of the receivers	Huawei Honor 30Lite, Huawei Honor V30
Tx	the model of the transmitter	JBL Flip 5
V	volume of the speaker	75%
d	distance between the Tx and Rx	25 m, 35 m, 40 m, 50 m, 60 m, 70 m
B	bandwidth of the signal	1 kHz (18.5~19.5 kHz)
F_s	sampling frequency	48 kHz
c	chirp rate of the chirp signal	40 kHz/s
r	bit rate	10 bits/s, 20 bits/s
T_S	symbol duration	25 ms
T_G	guard interval duration	75 ms, 25 ms
Coder	Decoding strategy	hard decision soft decision+Viterbi

TABLE VI
DATA TRANSMISSION SUCCESS RATE AT DIFFERENT
DISTANCES ON TWO SMARTPHONES

Range	Huawei Honor 30Lite			
	hard decision no convolutional code		soft decision + Viterbi decoder	
	10 bits/s	20 bits/s	10 bits/s	20 bits/s
25 m	0.9999	0.9954	1.0000	1.0000
35 m	0.9999	0.9939	1.0000	1.0000
40 m	0.9974	0.9839	1.0000	0.9867
50 m	1.0000	1.0000	1.0000	1.0000
60 m	0.9998	0.9995	1.0000	1.0000
70 m	1.0000	1.0000	1.0000	1.0000
Range	Huawei Honor V30			
	hard decision no convolution code		soft decision + Viterbi decoder	
	10 bits/s	20 bits/s	10 bits/s	20 bits/s
25 m	0.9965	0.9794	1.0000	0.9998
35 m	0.9940	0.9389	1.0000	0.9618
40 m	0.9896	0.9505	1.0000	0.9779
50 m	0.9999	0.9979	1.0000	1.0000
60 m	0.9930	0.9835	1.0000	1.0000
70 m	0.9972	0.9965	1.0000	1.0000

G. Evaluations on Transmission Range Upper Bound

Through calculating the mean power of the received signal measured in Section. IV-F, a plot of communication distance

versus SNR is depicted in Fig. 13. The SNR in Fig. 13 is a rough reference and is calculated in the unit of 'dBFS': the power of the signal is obtained by calculating the average power of the signal frequency band on the spectrogram, and the power of noise is obtained by calculating the average power of the non-signal frequency band but within the ultrasonic frequency band on the spectrogram. Based on the plot, a curve is fitted at a further distance from 70 m to 120 m. According to the simulation results that are depicted in Fig. 11, an SNR threshold of 10^{-5} BER is set. From the cross point, the estimated upper bound of error-free (BER less than 10^{-5}) communication range using the proposed method is approximately 85m.
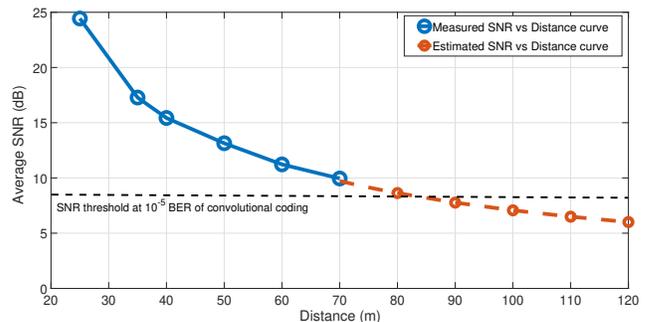

Fig. 13. Evaluate maximum communication range based on the estimated SNR versus distance curve. The average SNR is calculated by the envelope of the signal spectrogram.

H. Communication Test in Non-Line of Sight Scenarios

In a crowed environment, communications under the condition of NLOS frequently occur that has negative impact on the quality of communications. Therefore, this section evaluates the performance of the proposed AAC scheme in a severe NLOS and noisy environment. The experiments were taken place on the same floor as the one mentioned in Section IV-F, and the details of the placement are illustrated in Fig. 14. In Fig. 14, point A denotes the position of the tested smartphones which functions as the receivers; the cross point B is a place where a Bosch GLM 500 laser meter is placed to function as a flag to determine the position of points A and C based on the measured distance of AB and BC; three points of C denote the positions where three JBL Flip 5 Bluetooth speaker were placed. The laboratory where smartphones were placed is connected to the corridor only through a double door, surrounded by cement walls and a large floor-to-ceiling window facing outside. Therefore, the channel path from points Cs to A are all NLOS. In the experiment, two cases of NLOS scenarios were tested: cases I and II represent the case of an absolute NLOS situation, and case III represents the case of a loose NLOS situation. The absolute NLOS defines the situation in that the transmitter has NLOS to the space where the receiver is located. In cases I and II, the speaker was placed next to the wall close to the laboratory, and thus the speaker has no incident angle to the laboratory. In contrast, the loose NLOS represents the situation in that the transmitter has LOS to the space where the receiver is located, but the transmitter

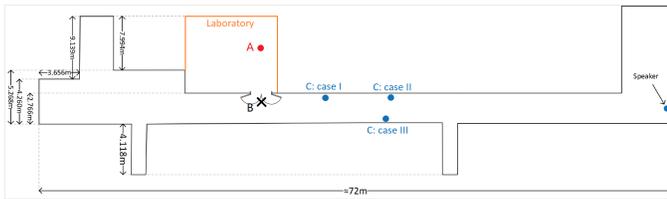is of NLOS to the receiver. The distances and experimental parameters are recorded in TABLE VII.

Fig. 14. Diagram of the environment for NLOS communication testing.

TABLE VII
EXPERIMENT SETUP AND RELEVANT PARAMETERS

Symbol	Description	Value
d_{AB}	distance between point A and B	5 m
d_{BC}	distance between point B and C	
case I		10 m
case II		20 m
case III		20 m
Rx	the models of the receivers	Huawei Honor 30Lite, Huawei Honor V30
Tx	the model of the transmitter	JBL Flip 5
B	bandwidth of the signal	1 kHz (18.5∼19.5 kHz)
c	chirp rate of the chirp signal	40 kHz/s
r	bit rate	20 bits/s
T_S	symbol duration	25 ms
T_G	guard interval duration	25 ms
V	Speaker Volume	75%
Coder	Decoding strategy	hard decision soft decision+Viterbi

In the laboratory, five under-running servers were placed inside the room that made annoying noises during the experiment. A spectrum example of the recorded signal is depicted in Fig. 15. The dashed line-highlighted region is the pattern of the signal spectrum. It is noticed that part of the low-frequency band of the modulated signal overlaps with the noisy band when the receiver is inside the noisy environment. In addition, burst noise generated by human activity and hard disks in servers frequently occurs during the test. The room impulse response in cases II and III was analyzed. As depicted in Fig. 16, Fig. 16(a) and Fig. 16(b) show the RIR of the absolute NLOS condition and the loose NLOS condition, respectively. The experimental results of spread delay are recorded in TABLE VIII. Compared with the results of LOS conditions where the spread delays are recorded in TABLE II, the most significant difference is the increase in the number of reverberations. The distribution of the impulses in Fig. 16 is more dispersed than that in Fig. 10, which means that multipath fading is more severe in NLOS. Meanwhile, the RMS delay spread is relatively larger and the span of the impulses is wider. Because the position of the speaker locates closely in two cases, the global trend of two RIRs depicted in Fig. 16 are similar, and the values of maximum excess delay in the two conditions are approximately the same. The significant difference of the RIRs exhibits in the impulses whose time axis is located before 0 ms. In case II, multiple impulses with overwhelming energy are distributed around 0 ms, but only few impulses with inappreciable energy are distributed near the impulse of the first arrival. Additionally, the boundary of

the first arrival is vague in Fig. 16(a). In terms of the loose NLOS condition in case III, it is noticed that the impulses with the second and third highest energy are distributed near the impulse of the first arrival. The distinction arises from the LOS path from the speaker to the laboratory, where the LOS path derives subpaths to the receiver with short path distance and weak power attenuation.

The experimental results of the data transmission test on two smartphones are recorded in TABLE IX. Generally, the transmission success rate in NLOS condition reduces significantly compared with that in LOS condition shown in TABLE VI, especially in the condition of absolute NLOS. In addition, the task of data transmission performs better in the loose NLOS condition even if the distance of points B and C is shorter in the absolute NLOS condition. It is also noticed that the highest transmission success rate reaches 0.999, which is an acceptable rate for long-range data transmission in target scenarios.
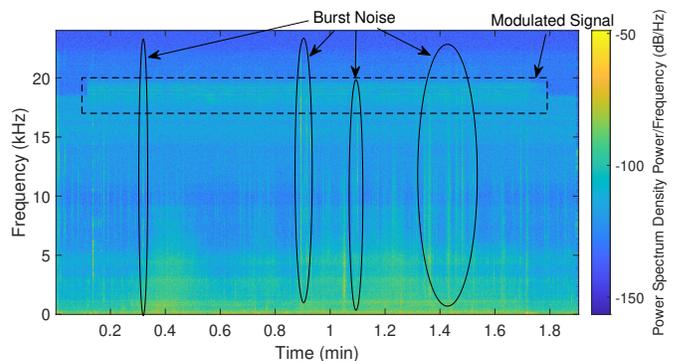

Fig. 15. An example of the recorded signal and noise samples in the laboratory .

TABLE VIII
SPREAD DELAY PARAMETERS

Situation	$\bar{\tau}$	σ_τ	$\tau_{\max}$	Reverberations
Case II	11.2971 ms	19.4161 ms	90.5 ms	63
Case III	16.0669 ms	16.3336 ms	90.6 ms	50

TABLE IX
DATA TRANSMISSION SUCCESS RATE IN DIFFERENT NLOS CONDITIONS

| Situation	hard decision no convolutional code		soft decision + Viterbi decoder	
	30Lite	V30	30Lite	V30
Case I	0.9707	0.9342	0.9978	0.9772
Case II	0.8889	0.8016	0.9038	0.6983
Case III	0.9834	0.9347	0.9990	0.9832

I. BER Comparison with Others Work

A comparative experiment was conducted with the traditional BOK method [8] and the upgraded method attached to a Rake receiver [35] on evaluating the BER of data stream transmission. The experiment was conducted in the described corridor. TABLE X displays the detailed comparison of parameters. Fig. 17(a) displays the comparison of BER

TABLE X
EXPERIMENT PARAMETERS

Parameter	Proposed work		Method I [8]: BOK		Method II: BOK + Rake receiver	
	Value	Samples (@48 kHz)	Value	Samples (@48 kHz)	Value	Samples (@48 kHz)
Frequency band	18.5∼19.5 kHz	-	18.5∼19.5 kHz	-	18.5∼19.5 kHz	-
Symbol duration	25 ms	1,200	20 ms	960	20 ms	960
Guard interval	25 ms	1,200	40 ms	1,920	40 ms	1,920
Preamble	25 ms	1,200	100 ms	4,800	100 ms	4,800
Bit rate	20 bps	-	16.7 bps	-	16.7 bps	-
Modulation method	Chirp Rate Modulation		Binary Orthonogal Keying		Binary Orthonogal Keying	
Demodulation	FrFT denoising + chirp rate estimation		Correlator method		Correlator + Rake receiver	
Decision Decoder	Hard decision decoder, without error-correction code					
Speaker volume	Same volume on JBL flip 5 bluetooth speaker					
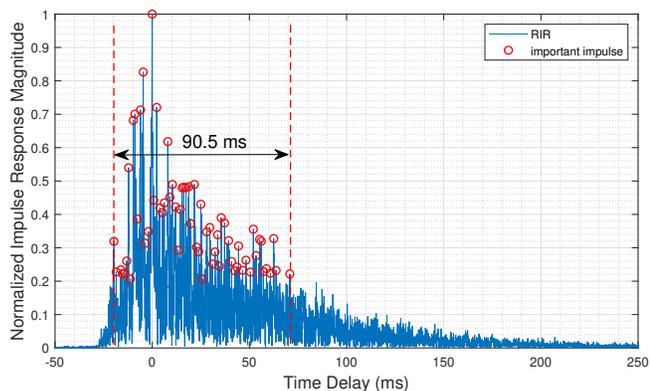
(a) Room impulse response in Case II.
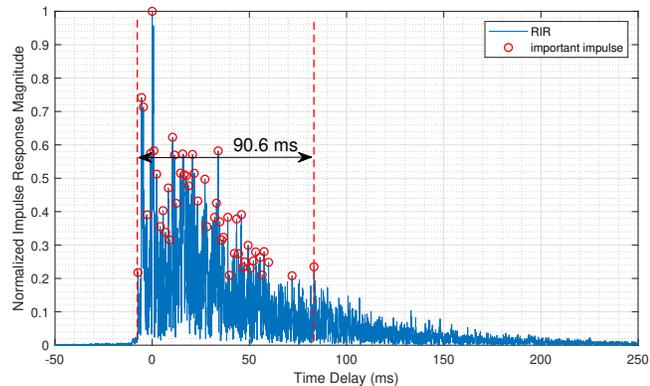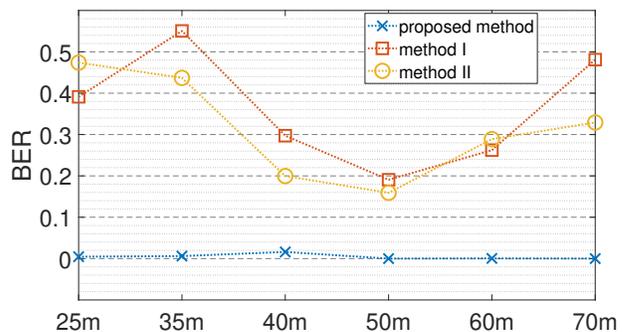

(b) Room impulse response in Case III.

Fig. 16. Experimental results of room impulse response analysis in two NLOS conditions.

(a) Comparison on bit error rate versus distance

(b) Comparison on received power density versus distance

Fig. 17. Comparison between the proposed method and the BOK method.
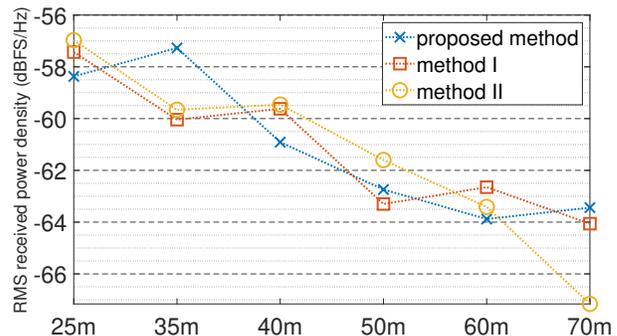
versus communication range, and Fig. 17(b) shows the relative power density based on the calculation of FFT. It is obvious that the proposed work achieves lower BER for long-range AAC in indoor environments where the received energy is approximately the same.

V. CONCLUSION

In this paper, the channel for AAC in the corridor is analyzed and a long-distance indoor acoustic communication scheme associated with the denoising and demodulation method is proposed. The receiver adds the fractional Fourier transform denoising part and applies the chirp rate estimation

method to demodulate the signal based on the adoption of chirp spread spectrum modulation. In symbol decision, the system provides the hard decision strategy and soft-decision-based convolutional decoder strategy to handle the channel state and balance the data rate. The strategy of soft-decision plus convolutional code is associated with interleaver in order to enhance the resistance to burst noise. Finally, the proposed architecture achieves an error-free communication range of 70 m.

References

[1] V. Gerasimov and W. Bender, "Things that talk: Using sound for device-to-device and device-to-human communication," *IBM Systems Journal*, vol. 39, no. 3.4, pp. 530–546, 2000.

[2] A. Madhavapeddy, R. Sharp, D. Scott, and A. Tse, "Audio networking: the forgotten wireless technology," *IEEE Pervasive Computing*, vol. 4, no. 3, pp. 55–60, 2005.

[3] R. Frigg, G. Corbellini, S. Mangold, and T. R. Gross, "Acoustic data transmission to collaborating smartphones? an experimental study," in *2014 11th Annual Conference on Wireless On-demand Network Systems and Services (WONS)*, 2014, pp. 17–24.

[4] Y. Nakashima, H. Matsuoka, and T. Yoshimura, "Evaluation and demonstration of acoustic ofdm," in *2006 Fortieth Asilomar Conference on Signals, Systems and Computers*, 2006, pp. 1747–1751.

[5] 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*, ser. MobiCom '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 29?1. [Online]. Available: https://doi.org/10.1145/2973750.2973765

[6] W. Jiang and W. M. D. Wright, "Indoor wireless communication using airborne ultrasound and ofdm methods," in *2016 IEEE International Ultrasonics Symposium (IUS)*, 2016, pp. 1–4.

[7] M. Zhou, Q. Wang, K. Ren, D. Koutsonikolas, L. Su, and Y. Chen, "Dolphin: Real-time hidden acoustic signal capture with smartphones," *IEEE Transactions on Mobile Computing*, vol. 18, no. 3, pp. 560–573, 2019.

[8] 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.

[9] 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.

[10] C. Cai, Z. Chen, J. Luo, H. Pu, M. Hu, and R. Zheng, "Boosting chirp signal based aerial acoustic communication under dynamic channel conditions," *IEEE Transactions on Mobile Computing*, vol. 21, no. 9, pp. 3110–3121, 2022.

[11] L. Zhang, M. Chen, X. Wang, and Z. Wang, "Toa estimation of chirp signal in dense multipath environment for low-cost acoustic ranging," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 2, pp. 355–367, 2019.

[12] L. Zhang, D. Huang, X. Wang, C. Schindelhauer, and Z. Wang, "Acoustic nlos identification using acoustic channel characteristics for smartphone indoor localization," *Sensors*, vol. 17, no. 4, p. 727, 2017.

[13] J. M. Sabatier and A. E. Ekimov, "Ultrasonic methods for human motion detection," Mississippi Univ University National Center For Physical Acoustic, Tech. Rep., 2006.

[14] P. L. Feintuch, F. A. Reed, M. Cox, and R. A. Borrell, "Method and system for in-air ultrasonic acoustical detection and characterization," Jan. 17 2006, uS Patent 6,987,707.

[15] A. Ekimov and J. M. Sabatier, "Passive ultrasonic method for human footstep detection," in *Unattended Ground, Sea, and Air Sensor Technologies and Applications IX*, vol. 6562. International Society for Optics and Photonics, 2007, p. 656203.

[16] A. M. Zungeru, "Design and development of an ultrasonic motion detector," *arXiv preprint arXiv:1303.1732*, 2013.

[17] Y. Qi, C. B. Soh, E. Gunawan, K.-S. Low, and R. Thomas, "Estimation of spatial-temporal gait parameters using a low-cost ultrasonic motion analysis system," *Sensors*, vol. 14, no. 8, pp. 15 434–15 457, 2014.

[18] G. Arun Francis, M. Arulselvan, P. Elangkumaran, S. Keerthivarman, and J. Vijaya Kumar, "Object detection using ultrasonic sensor," *Int J Inno Technol Explor Eng (IJITEE)*, vol. 8, no. 5, pp. 207–209, 2020.

[19] 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.

[20] S. Holm, O. 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.

[21] H. Nyquist, "Certain topics in telegraph transmission theory," *Transactions of the American Institute of Electrical Engineers*, vol. 47, no. 2, pp. 617–644, 1928.

[22] L. Shi, L. Yu, K. Huang, X. Zhu, Z. Wang, X. Li, W. Wang, and X. Wang, "A covert ultrasonic phone-to-phone communication scheme," in *International Conference on Collaborative Computing: Networking, Applications and Worksharing*. Springer, 2020, pp. 36–48.

[23] R. Tao, B. Deng, W.-Q. Zhang, and Y. Wang, "Sampling and sampling rate conversion of band limited signals in the fractional fourier transform domain," *IEEE Transactions on Signal Processing*, vol. 56, no. 1, pp. 158–171, 2007.

[24] A. Bhandari and P. Marziliano, "Sampling and reconstruction of sparse signals in fractional fourier domain," *IEEE Signal Processing Letters*, vol. 17, no. 3, pp. 221–224, 2009.

[25] M. Hanspach and M. Goetz, "On Covert Acoustical Mesh Networks in Air," *Communications*, vol. 8, no. 11, nov 2013.

[26] J. Hagenauer and E. Lutz, "Forward error correction coding for fading compensation in mobile satellite channels," *IEEE Journal on Selected Areas in Communications*, vol. 5, no. 2, pp. 215–225, 1987.

[27] S. Dhaliwal, N. Singh, and G. Kaur, "Performance analysis of convolutional code over different code rates and constraint length in wireless communication," in *2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, 2017, pp. 464–468.

[28] H. M. Ozaktas, O. Arikan, M. A. Kutay, and G. Bozdagt, "Digital computation of the fractional fourier transform," *IEEE Transactions on signal processing*, vol. 44, no. 9, pp. 2141–2150, 1996.

[29] 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.

[30] H. Kulhandjian and T. Melodia, "Modeling underwater acoustic channels in short-range shallow water environments," in *Proceedings of the International Conference on Underwater Networks & Systems*, 2014, pp. 1–5.

[31] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Audio Engineering Society Convention 108*. Audio Engineering Society, 2000.

[32] T. S. Rappaport *et al.*, *Wireless communications: principles and practice*. prentice hall PTR New Jersey, 1996, vol. 2.

[33] B. D. Lawrence and J. A. Simmons, "Measurements of atmospheric attenuation at ultrasonic frequencies and the significance for echolocation by bats," *The Journal of the Acoustical Society of America*, vol. 71, no. 3, pp. 585–590, 1982.

[34] S. Kullback and R. A. Leibler, "On information and sufficiency," *The annals of mathematical statistics*, vol. 22, no. 1, pp. 79–86, 1951.

[35] G. Bottomley, T. Ottosson, and Y.-P. Wang, "A generalized rake receiver for interference suppression," *IEEE Journal on Selected Areas in Communications*, vol. 18, no. 8, pp. 1536–1545, 2000.

LIMING SHI Liming Shi received the B.E. in electronic science and technology from Xi'an Jiaotong Liverpool University, Suzhou, China, in 2019, and he is currently pursuing the Ph.D. degree with the Department of Electrical and Electronic Engineering. His current research interests include the acoustics, signal processing and cooperative communication.

Xinheng Wang (Senior Member, IEEE) received the B.E. and M.Sc. degrees in electrical engineering from Xian Jiaotong University, Xian, China, in 1991 and 1994, respectively, and the Ph.D. degree in electronics and computer engineering from Brunel University, Uxbridge, U.K., in 2001.

He is currently a Professor with the Department of Mechatronic and Robotics, Xi'an Jiaotong-Liverpool University, Suzhou, China, and an Adjunct Professor with Zhejiang University. He has been an Investigator or Co-Investigator of nearly 30 research projects sponsored by EU, U.K. EPSRC, Innovate U.K., China NSFC, and industry. He has authored or coauthored over 180 referred articles and held 15 granted patents, including 1 U.S., 1 Japan, 4 South Korea, and 9 China patents. His current research interests include Tactile Internet, indoor positioning, Industrial Internet of Things (IIoT), intelligent manufacturing and big data analytics for intelligent services, where he has developed the world's first smart trolley with a company in China. His research has led to a few commercial products in condition monitoring, wireless mesh networks, acoustic localization, and user-centric routing and navigation.

Muddesar Iqbal Dr Iqbal won an EPSRC Doctoral Training Award in 2007 and completed his PhD from Kingston University in 2010 in Wireless Networks. He has Won 17 RD and Capacity building funding Grants from different national and international funding agencies. He has won two Awards of Appreciation for Tutoring the Prize Winner from the Association of Business Executive (ABE) the UK, the first in 2005 for tutoring on Computer Fundamentals and second in 2010 for tutoring on Information System Project. He Also Won Australia Award to complete two months Block-based Business Incubator Management Certificate from Queensland University Australia in 2015. He has Published 100 peer reviewed articles (including IEEE reputed Q1 journals, conference proceedings. Books Collections and book Chapters). He has Co-Invented 5 patented inventions in IoT, Software Defined Networks Technologies and Autonomous Vehicles area. He is member of British Computer Society and also a British Council's Active Citizen for the implementation of British Council Social Action Projects under Youth, Education and Society. He is serving as a Member Board of Governor for United Nation's Independent Monitoring Association for Child Labour (IMAC), which was set up by the International Labour Organisation's International Programme on the Elimination of Child Labour (IPEC) from the Sailkot's soccer ball industry in Pakistan.

Limin Yu (Member, IEEE) received the B.Eng. degree in telecommunications engineering and the M.Sc. degree in radio physics/underwater acoustic communications from Xiamen University, China, and the Ph.D. degree in telecommunications engineering from The University of Adelaide, Australia. She was with ZTE Telecommunication Company Ltd., Shenzhen, China, as a Software Engineer. She was also with the University of South Australia and The University of Adelaide, as a Research Fellow and a Research Associate. She is currently an Associate Professor with the Electrical and Electronic Engineering Department, Xi'an Jiaotong-Liverpool University, Suzhou, China. Her research interests include sonar detection, wavelet analysis, filter bank design, broadband system design, high-mobility wireless local area networks, sensor networks, and medical image analysis.

Zhi Wang (Member, IEEE) received the B.E. degree from Shenyang Jianzhu University, Shenyang, China, in 1991, the M.S. degree from Southeast University, Nanjing, China, in 1997, and the Ph.D. degree from Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, in 2000.

From 2000 to 2002, he was a Postdoctoral Fellow with the Institut National Polytechnique de Lorraine, Vandoeuvre-lès-Nancy, France, and Zhejiang University, Hangzhou, China. He is a Ph.D. Advisor with the College of Control Science and Engineering, Zhejiang University.As a Principal Investigator and Co-Investigator, he has investigated more than 20 international collaboration and national research grants, including nine China–Germany, China–France, China–U.K., and China–Portugal Collaboration Projects, eight National Natural Science Foundation of China Projects, and one China 863 High-Tech Program.He has authored/coauthored over 100 SCI/EI indexed journal papers, including over 20 IEEE transactions and journals, such as the IEEE Transactions on Mobile Computing, the IEEE Transactions on Distributed Systems, the IEEE Transactions on Signal Processing, the IEEE Transactions on Instrumentation and Measurement, the IEEE Internet of Things, the IEEE Transactions on Inudstrial Informatics, and IEEE Access, over 20 top international conference papers, such as IEEE INFOCOM and IEEE SECON. His research areas include acoustic signal and array signal processing, acoustic SLAM and robot positioning and navigation, indoor localization and temporal–spatial data mining, device-free human activity recognition, compressed sensing and deep learning, data fusion of multiple sensing systems, industrial Internet of Things, and privacy protection in crowdsourcing systems.

Wenwu Wang (Senior Member, IEEE) received the B.Sc., M.E., and the Ph.D. degrees from the College of Automation, Harbin Engineering University, Harbin, China, in 1997, 2000, and 2002, respectively.

He then worked in King's College London, London, U.K., from 2002 to 2003, Cardiff University, Cardiff, U.K., from 2004 to 2005, Tao Group Ltd.(now Antix Labs Ltd.) from 2005 to 2006, Creative Labs from 2006 to 2007, before joining the University of Surrey, Surrey, U.K., in May 2007, where he is a Professor in signal processing and machine learning and the Co-Director of the Machine Audition Lab.

Since 2019, he has been a Senior Area Editor, and from 2014 to 2018, was an Associate Editor for the IEEE Transactions on Signal Processing. Since 2020, he has been an Associate Editor for the IEEE/ACM Transactions on Audio Speech and Language Processing, and since 2019, he has been an Associate Editor for the EURASIP Journal on Audio Speech and Music Processing. Since 2020, he has been the Specialty Editor-in-Chief of Frontiers in Signal Processing. He was the Publication Co-Chair for ICASSP 2019, Brighton, U.K. Since 2021, he has been elected as a member of the IEEE Signal Processing Theory and Methods Technical Committee and the IEEE Machine Learning for Signal Processing Technical Committee. Since 2019, he has also been a member of the International Steering Committee of Latent Variable Analysis and Signal Separation.

Charalampos C. Tsimenidis Charalampos C. Tsimenidis (Senior Member, IEEE) received the Ph.D. degree in communications and signal processing from Newcastle University, in 2002. He is currently a Professor of digital innovation at Nottingham Trent University. He has published over 230 conference and journal papers, supervised successfully three M.Phil. and 50 Ph.D. students and made contributions in the area of digital communications to several U.K. and European funded research projects. His main research interests include digital communications and signal processing with specialization in massive multiple-input multiple-output systems, adaptive filter, and demodulation algorithms, error control and network coding for radio frequency, and underwater acoustic channels. He is a member of the IET.

Shahid Mumtaz Shahid Mumtaz has more than 12 years of wireless industry experience and is currently working as Senior Research Scientist and Technical Manager at GS, Portugal. Prior to his current position, he worked as Research Intern at Ericsson and Huawei Research Labs in 2005 at Karlskrona, Sweden. He received his MSc and Ph.D. degrees in Electrical Electronic Engineering from Blekinge Institute of Technology (BTH) Karlskrona, Sweden and University of Aveiro, Portugal in 2006 and 2011, respectively. Dr. Shahid MSc and Ph.D. were funded by the Swedish government and FCT Portugal. He has been involved in several EC RD Projects in the field of green communication and next generation wireless systems. In EC projects, he holds the position of technical manager, where he oversees the project from a scientific and technical side, managing all details of each work packages which gives the maximum impact of the project's results for the further development of commercial solutions. He has been also involved in two Portuguese funded projects (SmartVision Mobilia) in the area of networking coding and development of system level simulator for the 5G wireless system. Dr. Shahid has several years of experience in 3GPP radio systems research with experience in HSPA/LTE/LTE-A and strong track-record in relevant technology field, especially physical layer technologies, LTE cell planning and optimization, protocol stack and system architecture.

Dr Shahid research interests lie in the field of architectural enhancements to 3GPP networks (i.e., LTE-A user plan control plan protocol stack, NAS and EPC), 5G NR related technologies, green communications, cognitive radio, cooperative networking, radio resource management, Network slicing, LAA/LTU, cross-layer design, Backhaul/fronthaul, heterogeneous networks, M2M and D2D communication, and baseband digital signal processing. Dr. Shahid has more than 150 publications in international conferences, journal papers, and book chapters.

Dr. Mumtaz is a senior member of IEEE. He was awarded an "Alain Bensoussan" fellowship by ERCIM to pursue research in communication networks for one year at the VTT Technical Research Centre of Finland in 2012. He was nominated Chair for IEEE new standardization on P1932.1: Standard for Licensed/Unlicensed Spectrum Interoperability in Wireless Mobile Networks. This standardization resulted from Dr. Mumtaz' novel idea on "WiFi in Licensed band". He is also actively involved in 3GPP standardization on LTE release 12 onwards, along with major manufacturers.