IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 13, 2020, accepted August 4, 2020, date of publication August 17, 2020, date of current version August 28, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3017097

Reliable and Low-Complexity Chirp Spread Spectrum-Based Aerial Acoustic Communication

JIHWAN LEE[1,2], (Graduate Student Member, IEEE), CHULYOUNG KWAK[1,2], (Graduate Student Member, IEEE), SEONGWON KIM[3], (Member, IEEE), AND SAEWOONG BAHK[1,2], (Senior Member, IEEE)

[1]Department of Electrical and Computer Engineering (ECE), Seoul National University, Seoul 08826, South Korea
[2]Institute of New Media and Communications (INMC), Seoul National University, Seoul 08826, South Korea
[3]Vision AI Labs, SK Telecom, Seoul 04539, South Korea

Corresponding authors: Seongwon Kim (s1kim@sktbrain.com) and Saewoong Bahk (sbahk@snu.ac.kr)

This work was supported by the National Research Foundation of Korea (NRF) funded by the Korea Government [Ministry of Science and ICT (MIST)] under Grant 2017R1E1A1A01074358.

ABSTRACT Aerial acoustic communication (AAC) has been developed for a variety of Internet-of-things (IoT) applications, thanks to the merit of embedding small amounts of data directly into audio contents and transmitting them through commercial off-the-shelf (COTS) devices without additional infrastructure. However, AAC has clear limitations due to ambient noise, severely frequency-selective acoustic channels, and high power consumption of COTS devices. We adopt chirp spread spectrum (CSS) modulation and design new quaternary symbols to overcome frequency selectivity of audio interfaces of COTS devices. We also develop a computation-efficient method to demodulate the proposed symbols, aiming to reduce the power consumption of COTS devices. We employ a frame combining technique without increasing computational complexity, to mitigate the effects of multipath fading and ambient noise in acoustic channels. To evaluate the proposed methods, we conduct extensive experiments with several smartphones in various environments. The experimental results and evaluation demonstrate that the proposed symbols and frame combining method contribute to improving the frame reception ratio by up to 59.8%p (267.9%) and 14%p (84.3%), respectively. The frame combining increases the possibility of receiving a frame within two attempts at extremely low SNR, by up to 107.9%, resulting in reduced excessive delay. The method to demodulate the proposed symbols lowers power consumption by tens to hundreds of milliwatts, depending on the device. Our proposal significantly helps to achieve high reliability and low power consumption of COTS devices in AAC.

INDEX TERMS Aerial acoustic communication, ambient noise, chirp spread spectrum, correlator, frame combining, software-based digital modem.

I. INTRODUCTION

With the advent of the Internet-of-things (IoT), various wireless communication technologies have been considered to connect a wide range of objects. Unlike traditional views focusing on the data rate maximization, it is more important to focus on each communication technology and to take advantage of its own strengths, considering each IoT scenario.

Aerial acoustic communication (AAC) is a major candidate to consider in the scenarios where small amounts of data are transmitted through commercial off-the-shelf (COTS) devices, such as sharing URLs, user authentication, voucher-delivery services, and TV's second screen services [1]. In general, the frequency range of AAC should be high enough to be inaudible to humans. According to the Nyquist theorem, the frequency range is upper-bounded by approximately 22 kHz among COTS devices, mostly with the sampling rate of 44.1 kHz. Therefore, only narrow bandwidths of a few kHz are available for AAC, resulting in low data rates. However, AAC has attracted commercial and academic attention owing to several advantages of acoustic signals in small data transmission scenarios.

First, unlike traditional wireless communications such as Wi-Fi and Bluetooth, AAC does not require any additional hardware and infrastructure installation. Since most mobile devices already have audio interfaces, i.e., speakers

The associate editor coordinating the review of this manuscript and approving it for publication was Ananya Sen Gupta.

and microphones, they can be reused for data transmission. Second, secondary information about audio content, such as music and video, can be embedded directly into the content itself, which can be of benefit to not only users but also service providers. For example, in the case of TV content, users can get additional information about the content without hassle through acoustic communication applications running in the background of smart devices. Service providers are happy to offer additional advertisements or interactive experiences to users at the right moment. Finally, acoustic signals are drastically attenuated outside a room or building, which creates a geo-fence and enables an acoustic signal to act as a location identifier (ID). For instance, the geo-fence of the acoustic signal can be used to secretly exchange users' security keys between smart devices in close proximity. In cafes and restaurants, the location ID can be used to order items directly through smart devices without going to the counter. Also, in the subway, AAC can help passengers watching video clips on mobile devices not to miss their stop, by sending push-notifications with the location ID of the station.

To exploit these advantages, we target applications that send and receive specific types of short-length context information, e.g., content program ID, user location ID, and security key for smart device pairing, without additional hardware and infrastructure. However, there exist clear limitations on exploiting AAC as follows.

A. LIMITED RELIABILITY

Acoustic channels suffer from severe fading due to multipath propagation and frequency selectivity of reused audio interfaces of COTS devices. Strong ambient noise prevents mobile devices from receiving data frames. To avoid human perception, the transmission power of an acoustic signal should be extremely low. To make matters worse, the transmitter is not sure whether the data was successfully received since in most cases, AAC uses broadcasting without a feedback mechanism.

B. POWER CONSUMPTION

Various AAC systems for COTS devices have been proposed, and they exploit signal processing techniques with high computation complexity to demodulate acoustic signals [2]. This issue is more important because many mobile applications using AAC run in the background and consume excessive power consistently without being perceived by users [3].

The target applications require no high data rate, but data reception should be guaranteed at least once, even in extremely low transmission power. For high reliability, we adopt chirp spread spectrum (CSS) modulation techniques, used in [1], [4], due to the robustness to ambient noise and the frequency selectivity of an acoustic channel, despite its low data rate. To supplement the absence of feedback mechanism, we consider blind retransmission used in LTE-based device-to-device (D2D) communication [5] and AACs [1], [4], i.e., the transmitter repeatedly broadcasts the same bits over time regardless of data reception at the receiver. In this paper, we aim to improve the modem design of CSS-based AAC for high reliability and low complexity. The main contributions of this paper are as follows.

- We verify the defect of the existing symbol designs for CSS modulation and propose a new quaternary chirp symbols, utilizing the cyclic time shift (CTS) property. The proposed symbols have good correlation properties and robustness to frequency selectivity, achieving high reliability.
- We develop a computation-efficient method for demodulating the proposed symbols, through the distributive property of convolution. The proposed method can halve the FFT size of the symbol correlator and reduce power consumption by tens to hundreds of milliwatts, depending on the device.
- We develop a frame combining method for blind retransmission suitable for acoustic channels to mitigate the effects of multipath fading and ambient noise without increasing computational complexity.
- In real environments, the experimental results show that the proposed symbol design reduces symbol error rate (SER) and improves frame reception rate (FRR) by up to 12.6%p (99%) and 59.8%p (267.9%), respectively, compared to the state-of-the-art symbol designs. In addition, the proposed frame combining scheme increases the occurrence of receiving frames within two attempts by up to 107.9%, reducing excessive delay at low SNR.

The rest of the paper is organized as follows. We summarize related work in Section II. We suggest a target scenario and discuss some challenges of AAC in Section III. We present the proposed quaternary chirp symbol design and describe the receiver design with high reliability and low complexity, in Sections V and IV, respectively. We evaluate the proposed designs in Section VI, followed by the conclusion in Section VII.

II. RELATED WORK

AAC for audio-enabled devices has been studied in an audible band and an inaudible band. The systems proposed in [2], [6]–[9] transmit data through an audible band with no consideration for human perception. In [6], multiple tones and a single tone are transmitted through an audible band and an inaudible band, respectively. Digital Voice [7] uses frequency-shift keying (FSK) to transmit information through an audible band under 12 kHz. The authors in [8] propose secure AAC by utilizing 6–7 kHz band, which can be a substitute for near field communication (NFC) with a data rate of up to 800 bps within 20 cm. Similarly, the authors in [9] present a secure acoustic short-range communication system for commercial smartphones, which achieves data rates around 1,000 bps utilizing an audible band over 8 kHz. VEH-COM [2] introduces a vibration energy harvesting (VEH)-based communication system that achieves data rates over 5 bps within 80 cm using an audible band.

To prevent an audible signal from being perceived by humans, some studies propose embedding a signal into

existing sound sources, without damaging the perceived quality of sound sources. In [10], 600 bps data is transmitted based on the phase modification for modulated complex lapped transform (MCLT) coefficients of sound sources. Masking effects are often used to embed data into sound sources without human perception [10]–[14]. In [11], the authors propose acoustic orthogonal frequency division multiplexing (OFDM) to modulate approximately 240 bps of information in existing sound sources in the audible frequencies without discomforting human ears.

Dolphin [12] uses amplitude shift keying (ASK) and energy difference keying (EDK) depending on frequency ranges and achieves average data rates of up to 500 bps using COTS devices. Similarly, the authors in [13] propose a system for transmitting 900 bps data within music through a loudspeaker and a smartphone, where suitable frequency regions for OFDM subcarriers are determined by exploiting tonal harmonics of existing sound sources. As a recent work of [2], the method in [14] enables VEH-based short-range data communication to avoid human perception while achieving up to 14 bps data rate.

Data communication in the high-frequency band is studied to avoid human perception. The authors in [15] use commercial laptops equipped with audio interfaces of flat frequency response to form acoustical mesh networks by utilizing ultrasonic frequency ranges and to achieve a data rate of 20 bps in the 19.7 m range. In [16], the authors present U-wear, a networking framework for commercial wearable devices via near-ultrasonic communication, which achieves data rates up to 2.76 kbps within tens of centimeters. Phoneear [17] transmits data using an acoustic signal in an inaudible band (17–20 kHz) using FSK. The authors in [18] propose an AAC system using an inaudible chirp signal sweeping a frequency range of 19.5–22 kHz. The authors in [1] develop a near-ultrasound chirp-based system for TV's 2nd screen services.

The aforementioned researches, however, have a few limitations. Audible signals may be annoying to AAC users [2], [6]–[9]. Even if audible signals are embedded into existing audible sound sources to avoid human perception [10], [12]–[14], audible sound sources should be played, and thus it is impossible to transmit signals only. Since ambient noise has higher power in the low-frequency band than in the high-frequency band, it significantly interfere with audible signals.

Short-range AAC systems [2], [8], [9], [14], [16] are not suitable for our target scenarios. In [10], [11], [15], the transceiver has an audio interface with a flat frequency response, but in general, most of COTS devices' audio interfaces are highly frequency selective. In [15], [18], a significant portion of an inaudible signal above 20 kHz could be lost, due to small gains of several encoding techniques in the high-frequency band. In [18], the chirp-based acoustic modem requires a high transmission volume that increases the likelihood of audibility. In [1], [17], frequency selectivity of COTS devices is not considered when choosing a modulation scheme.

III. THE PRELIMINARIES

We consider AAC applications that send and receive an acoustic signal embedding a short-length ID between COTS devices equipped with audio interfaces. The service provider either transmits only the acoustic signal itself or mixes it into existing audio sources. To avoid human perception, inaudible acoustic signals use high-frequencies and should be transmitted with very low transmission power. In this section, we discuss several challenges of the target scenario.

A. EXTREMELY LOW SIGNAL POWER

Human hearing spreads widely from 20 Hz to 20 kHz in frequency with different sensitivity, and the hearing threshold increases sharply above 16 kHz [19]. Given the hearing range and threshold, we adopt high-frequencies above 18 kHz as a signal band to avoid human perception. However, if the signal power is high enough, users may hear high-frequency signals. Moreover, unlike the hearing threshold, the threshold of pain is a decreasing function of frequency higher than 2 kHz [19].

In many AAC applications, such as TV's second screen services, when a signal is mixed into an audio content, its power depends on the power of the mixed-signal and the transmission volume of the speaker. Since the service provider can only adjust the mixed-signal power, not the speaker volume, the mixed-signal power should be very low against an unexpectedly large volume of the speaker. In reality, however, if the user selects a small volume, the signal power becomes extremely low due to not only the small volume of the speaker but also the low mixed-signal power. Hence, it is important to receive a signal successfully even at an extremely low signal power. Conventional RF communications have used RAKE receiver and a frame combining technique to improve the signal to noise ratio (SNR) at the receiver. We apply these to AAC to overcome the weaknesses due to low signal power.

B. FREQUENCY SELECTIVITY OF AUDIO INTERFACES OF COTS DEVICES

Audio interfaces in COTS devices, such as smartphones, are not originally intended for AAC. To analyze the characteristics of audio interfaces, we measure the frequency responses of microphones of various smartphones based on the linear frequency sweep (LFS) signal method in [20]. To ignore frequency selectivity of multipath propagation, we conduct all experiments in an anechoic room by using YAMAHA MSP5 as a reference speaker owing to its flat frequency response. The reference speaker transmits sine signals linearly sweeping from 20 to 22,000 Hz for 10 s. We record the signals using a smartphone and convert the recorded data to the frequency response using MATLAB and AUDACITY. To reduce fluctuations in the experiments, we perform ten iterations and average the results.

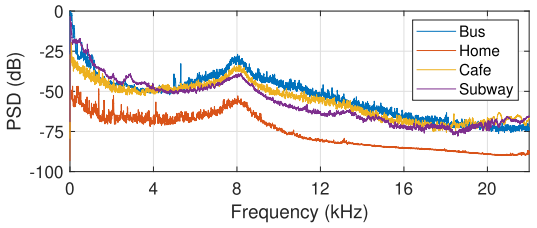
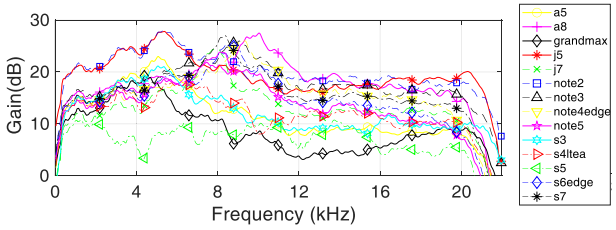
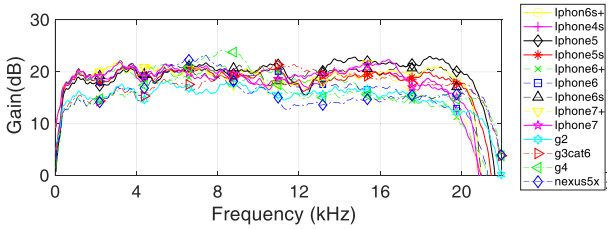


FIGURE 1. PSD of ambient noises in various places.



(a) Samsung devices



(b) LG and Apple devices

FIGURE 2. Frequency response of microphones of smartphones.

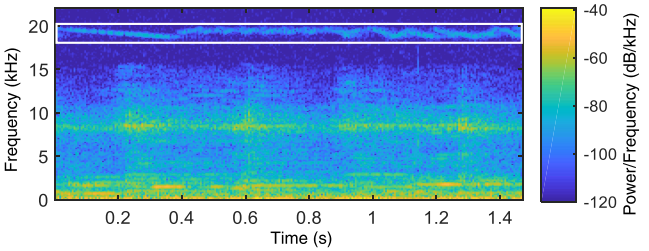
Fig. 2 shows the experimental results of Samsung, LG, and Apple smartphones. We observe that the microphones of all devices suffer severe frequency selectivity in the high-frequency band above 18 kHz. Moreover, audio interfaces of some devices are more frequency selective than the other devices, which leads to a huge difference in performance between devices. To overcome the limitations of audio interfaces of COTS devices and enable reliable communication in the high-frequency band, we introduce the CSS modulation which is robust to frequency selectivity.

C. AMBIENT NOISE

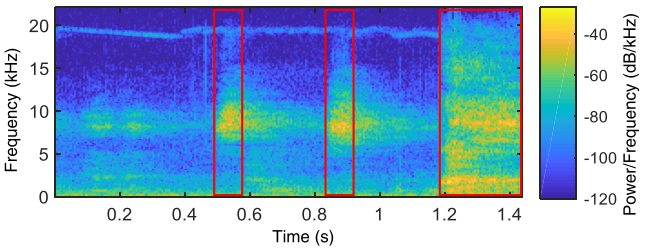
Ambient noise in our daily life distorts acoustic signals and significantly degrades decoding performance for the received data. We record ambient noise in various places using Samsung Galaxy S8 and calculate their power spectral density (PSD). The results in Fig. 1 show that the power level of ambient noise is much higher in a bus, cafe, and subway train compared to in home. However, the gap between home and noisy environments becomes less significant at higher frequencies, which results in quite low PSD at the high frequency regardless of the environments. This observation is supported by the similar results in [12]. Therefore, if we use the high-frequency band to avoid human perception, ambient noise does not seem to be a serious problem.

However, with further experiments, we observe that a certain type of ambient noise, named impulse noise, often has large power even at frequencies above 18 kHz and severely interferes with acoustic signals. Specifically, several subway lines provide notification services that send location information of passengers to their smartphones through acoustic signals. Through collaboration with a subway operator, we have embedded an acoustic signal at frequencies above 18 kHz in existing announcements and advertisement audio clips played on built-in sound systems of a subway train, and recorded it.

Fig. 3 shows the spectrogram of signals (white box) and impulse noises (red box) measured on Samsung Galaxy S8 in a subway train. As shown in Fig. 3(a), the received power of the signals is sufficient to decode embedded information successfully, even at loud ambient noise in the subway train. However, we observe that there are consecutive decoding errors despite the large signal power. This is because intermittent impulse noises in Fig. 3(b) cause severe interference to acoustic signals. Furthermore, we observe the same problem in other noisy environments, such as a cafe and bus. This motivates us to develop a frame combining method suitable for acoustic channels to mitigate the effects of impulse noise for reliable AAC in noisy environments.



(a) Spectrogram of chirp signals without impulse noises



(b) Spectrogram of chirp signals with impulse noises

FIGURE 3. Spectrogram of chirp signals recorded in a subway.

D. POWER CONSUMPTION OF MOBILE DEVICES

Recently, various AAC systems for COTS devices have been proposed, most of which exploit signal processing techniques with high computation complexity to demodulate acoustic signals, e.g., FFT. The experimental results in [2] show that FFT consumes tens or hundreds of times more power of devices than an audio recording process does. Furthermore, the power consumption of mobile devices using AAC is a further critical issue, considering many applications running in the background on mobile devices, e.g., TV's second screen service [1] and the aforementioned notification service. This is because they can consume excessive power consistently without being perceived by the user, which makes users

hesitate to utilize them. To achieve low power consumption in mobile devices, we develop a computation-efficient method for demodulating the proposed symbols.

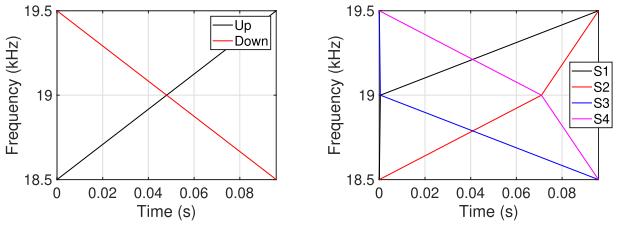
IV. SYMBOL DESIGN

Considering the long propagation delay and low carrier frequency of acoustic signals, only a few modulation techniques can be applied to AAC systems [21]. For example, phase shift keying (PSK) shows significantly degraded performance due to the Doppler shift and imperfect synchronization, and frequency shift keying (FSK) is vulnerable to frequency selectivity of audio interfaces. Furthermore, narrow bandwidths and frequency selectivity of audio interfaces in COTS devices make it difficult to select a modulation method for the target scenario.

To overcome the challenges in the target scenario, we adopt CSS modulation due to its reliability despite its limited data rate. It is worth noting that since the target scenario requires transmission of a short-length ID, data rate is far less important than reliability. There are two well-known CSS modulation methods for AAC: binary orthogonal keying (BOK) [18] and quaternary orthogonal keying (QOK) [1]. In this section, we explain these two techniques and propose a new symbol design for CSS modulation, called CTS-chirp. Here the chirp signal refers to a signal whose carrier frequency varies over time.

A. EXISTING CSS MODULATION SCHEMES FOR AAC

BOK consists of up/down chirp signal whose carrier frequency increases/decreases as shown in Fig. 4(a). Up and down chirp signals are nearly orthogonal to each other, i.e., their cross-correlation is close to zero. Since chirp signals sweep the entire frequency band unlike FSK, they have much narrower auto-correlation peaks and higher resolution in peak detection. Thanks to these correlation properties, chirp signals are detectable under ambient noise and resilient to multipath fading. Furthermore, chirp signals are used to overcome frequency selectivity of audio interfaces of COTS devices. This is because chirp signals mitigate the effects of attenuation at certain frequencies due to frequency selective channels in that they contain all frequency components evenly.



(a) Symbol set of BOK (b) Symbol set of QOK

FIGURE 4. Existing chirp symbol designs.

However, due to its low modulation order, BOK should have a short length of symbol duration to meet requirements on data rates. Since the short length of symbol duration increases cross-correlation and reduces symbol energy at the same transmission power, BOK is not suited for low transmission power. If we increase the transmission power to adopt BOK, the risk of hearing acoustic signals also increases.

The authors in [1] propose QOK to increase the modulation order of BOK. As shown in Fig. 4(b), they designed four nearly orthogonal chirp symbols to minimize cross-correlation between chirp symbols through an exhaustive search in the limited search space. Because of the low cross-correlation, there is no need for guard intervals (GI) between symbols to overcome inter symbol interference (ISI). Therefore, the symbol length and energy per bit increase by raising the modulation order and removing GI, which allows low transmission power without human perception.

However, we verify that QOK has another problem, named asymmetric symbol energy, owing to its use of frequency division in symbol design. As shown in Fig. 4(b), QOK symbols divide the frequency band in half and sweep the upper and lower frequency ranges with different slopes. Each symbol sweeps the frequency range with different duration, thus getting different energy attenuation in frequency selective channels. We discuss this in detail next.

In AAC, sound waves are transmitted from a speaker, propagated through the air, and received by a microphone. The aerial acoustic channel consists of three components: speaker channel, air channel, and microphone channel. We formulate this as

$$Y(f) = H_{speaker}(f)H_{air}(f)H_{mic}(f)X(f) + N(f), \qquad (1)$$

where $X(f)$ is the PSD of the transmitted symbol, $Y(f)$ is the PSD of the received symbol, $H_{speaker}(f)$, $H_{air}(f)$, $H_{mic}(f)$, and $N_{mic}(f)$ are the frequency responses of the speaker channel, air channel, microphone channel, and noise, respectively.

To analyze the effects of frequency selectivity on QOK, we calculate the PSD of each QOK symbol for two cases, as shown in Fig. 5. One is the symbol's PSD ($X(f)$) and the other is the symbol's PSD after propagation through acoustic channel ($H(f)X(f)$). We normalize the energy of each symbol to be one, and replace the acoustic channel with the normalized microphone channel of LG G2 measured in Section III-B, assuming that both air and speaker channels have their own flat frequency response. The results show that

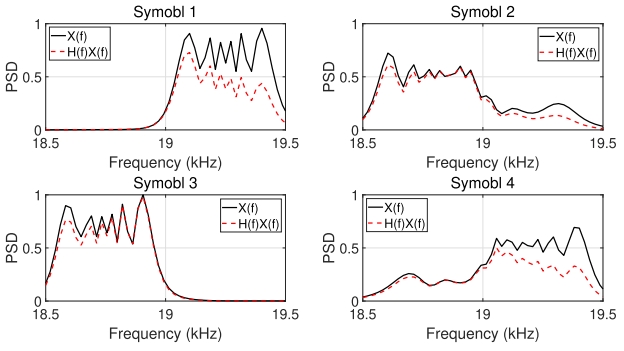


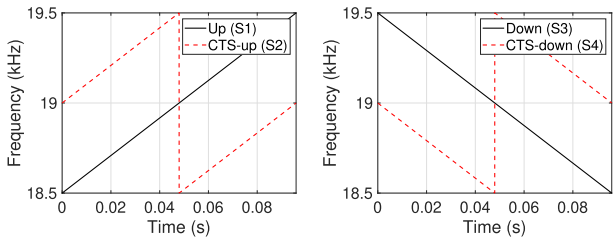
FIGURE 5. PSD of QOK symbols with or without the microphone channel of G2.

Symbols 1 and 4 suffer from more serious fading compared to Symbols 2 and 3, which results in asymmetric symbol energy. This is because QOK symbols use frequency division to raise the modulation order and maintain orthogonality at the same time, resulting in less robust chirp signals against frequency selectivity.

To be specific, as shown in Fig. 4(b), Symbols 1 and 4 sweep 19–19.5 kHz for a long time, and 18.5–19 kHz for a short time. This is opposite to the cases of Symbols 2 and 3. On the other hand, the frequency response of LG G2 has a larger gain at 18.5–19 kHz than 19–19.5 kHz, and thus, Symbols 1 and 4 have lower energy than Symbols 2 and 3. We will verify later that asymmetric symbol energy significantly degrades decoding performance of various smart devices. Since the frequency response of each speaker and microphone varies widely, pre-measuring and pre-tuning for each device can not be a viable solution.

B. PROPOSED CHIRP SYMBOL DESIGN

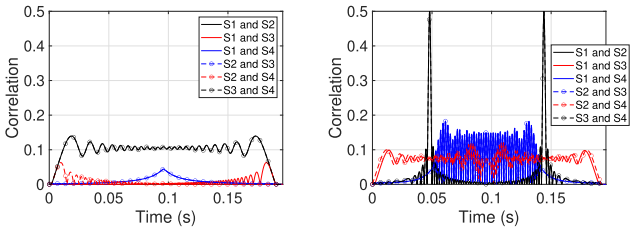
To overcome the low symbol energy of BOK and asymmetric symbol energy of QOK, we propose a new quaternary symbol design, named CTS-chirp signals. We design the proposed symbols to sweep the same frequency band evenly so that they can be resilient to frequency selectivity of acoustic channels. In return, the designed quaternary symbols show higher cross-correlation between symbols, compared to QOK. To mitigate the effects of high cross-correlation, we employ the CTS property. Specifically, we design circularly shifted up and down chirp signals by using half the length of symbol duration and add them to existing up and down chirp signals. Fig. 6 depicts an example of the proposed symbols when the symbol duration is 96 ms, the same as in [1].



(a) Up and CTS-up chirp (b) Down and CTS-down chirp

FIGURE 6. Proposed CTS-chirp symbols.

Since chirp signals have narrow auto-correlation peaks in symbol duration, the symbol duration becomes the sampling timing offset for symbol decoding. Fig. 7 shows the cross-correlation of QOK and CTS when the magnitude of the auto-correlation peak is normalized to one. Thanks to the CTS property, the cross-correlation of CTS is negligible or slightly higher compared to QOK at the sampling timing offset (96 ms), but merges into the two half-size peaks at 48 ms and 144 ms. However, the cross-correlation peaks have little impact on symbol decoding because they are narrow and far away from the sampling offset to detect an auto-correlation peak. Therefore, CTS symbols are more robust to frequency



(a) Cross-correlation of QOK (b) Cross-correlation of CTS

FIGURE 7. Cross-correlation of QOK and CTS-chirp.

selectivity of audio interfaces of COTS devices compared to QOK, keeping good correlation properties. Besides, we use the CTS property not only to design symbols but also to reduce decoding complexity of the proposed symbols.

V. RECEIVER DESIGN

In CSS-based AAC, the procedures for symbol decoding consist of two steps after the recording process. The first step is to get sampling timing offsets from the preamble that has long duration and high energy enough to resist ambient noise. After calculating the correlation between the received signal and preamble and finding the peak of correlations, we set the sampling offsets for each symbol in the frame. The second step is to detect the peak of symbol correlations. At the sampling offsets, the received signal has a higher correlation with the transmitted symbol than the other symbols, due to the narrow and high auto-correlation peak of chirp signals. Hence, we can decide which symbol was transmitted, by calculating the correlation between the received signal and each reference symbol, sampling each correlator output at the timing offsets, and comparing their magnitudes. In this section, we describe how to decode an audio signal with low complexity and high reliability. To reduce power consumption of smart devices, we develop a computation-efficient method to calculate symbol correlations. We employ RAKE receiver to overcome multipath fading and improve SNR at the receiver. Then, we develop a frame combining technique to mitigate the effects of ambient noise.

A. COMPUTATION-EFFICIENT CORRELATOR

The correlation $c(t)$ between symbol $x(t)$ with the symbol duration T and the received signal $y(t)$ is calculated by convolving $y(t)$ with the time-reserved symbol $x(T-t)$.

$$c(t) = y(t) * x(T-t). \tag{2}$$

As shown in Fig. 8, if $x(t)$ is the transmitted symbol, the correlator output has the auto-correlation peak at the sampling offset T. Otherwise, there is only cross-correlation with a small magnitude at time T compared to the auto-correlation peak. After calculating the correlations between each symbol and the received signal, we decide which symbol was transmitted by comparing four correlator outputs sampled at T.

In general, we can reduce the complexity of calculating correlations from $O(\log n^2)$ to $O(n \log n)$ by replacing the

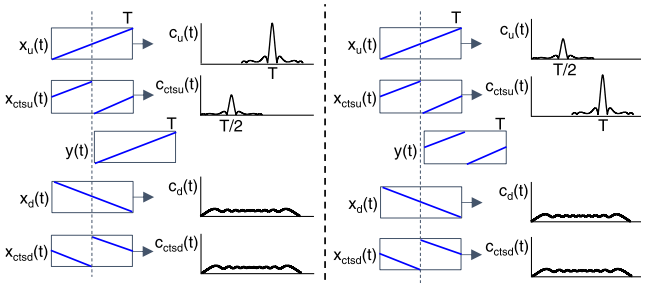


FIGURE 8. Example of an existing correlator.

convolution with FFT and inverse FFT (IFFT), as follows.

$$c(t) = \mathcal{F}^{-1}\Big(\mathcal{F}\big(y(t)\big)\mathcal{F}\big(x(T-t)\big)\Big), \tag{3}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote FFT and IFFT, respectively. The computation complexity of FFT/IFFT depends on the FFT size, selected to be a power of two, which is not smaller than the sum of the received signal length and the symbol length (N_{sym}).

We develop a method to calculate the symbol correlation with low complexity by exploiting the CTS property of the proposed symbols. The proposed symbols consist of two pairs: up-CTS pair and down-CTS pair. Up-CTS pair consists of up chirp signal $x_u(t)$ and CTS-up chirp signal $x_{ctsu}(t)$, and down-CTS pair consists of down chirp signal $x_d(t)$ and CTS-down chirp signal $x_{ctsd}(t)$. Halving the up chirp signal and down chirp signal in the time domain, we get four chirp signal segments $s_1(t)$, $s_2(t)$, $s_3(t)$, and $s_4(t)$. $s_1(t)$ and $s_2(t)$ are the first and second halves of the up chirp signal, respectively, and $s_3(t)$ and $s_4(t)$ are the first and second halves of the down chirp signal, respectively. Thanks to the CTS property, two symbols of the same CTS pair share the same two segments in the reverse order of time.

$$\begin{aligned}
x_u(t) &= s_1(t) + s_2(t - T/2),\\
x_d(t) &= s_3(t) + s_4(t - T/2),\\
x_{ctsu}(t) &= s_1(t - T/2) + s_2(t),\\
x_{ctsd}(t) &= s_3(t - T/2) + s_4(t),\\
s_1(t) &= \begin{cases} x_u(t), & 0 \le t < T/2;\\ 0, & \text{otherwise,}\end{cases}\\
s_2(t) &= \begin{cases} x_u(t + T/2), & 0 \le t < T/2;\\ 0, & \text{otherwise,}\end{cases}\\
s_3(t) &= \begin{cases} x_d(t), & 0 \le t < T/2;\\ 0, & \text{otherwise,}\end{cases}\\
s_4(t) &= \begin{cases} x_d(t + T/2), & 0 \le t < T/2;\\ 0, & \text{otherwise.}\end{cases}
\end{aligned} \tag{4}$$

The above equation indicates that four CTS-chirp symbols are represented as a linear combination of four segments. Hence, we compute the symbol correlations using the linear combinations of correlations between the received signal and each segment. First, we calculate the segment correlation $c_i(t)$ with the segment index $i \in \{1, 2, 3, 4\}$ from (2) as

$$c_i(t) = y(t) * s_i(T/2 - t). \tag{5}$$

Then, we obtain the symbol correlations for up-CTS pair, i.e., up chirp symbol correlation $c_u(t)$ and CTS-up chirp symbol correlation $c_{ctsu}(t)$, using $c_1(t)$ and $c_2(t)$, as

$$\begin{aligned}
c_u(t) &= y(t) * x_u(T - t)\\
&= y(t) * s_1(T - t) + y(t) * s_2(T/2 - t)\\
&= c_1(t - T/2) + c_2(t),\\
c_{ctsu}(t) &= y(t) * x_{ctsu}(T - t)\\
&= y(t) * s_1(T/2 - t) + y(t) * s_2(T - t)\\
&= c_1(t) + c_2(t - T/2).
\end{aligned} \tag{6}$$

Similarly we can compute the symbol correlations for down-CTS pair, i.e. $c_d(t)$ and $c_{ctsd}(t)$, by replacing $c_1(t)$ and $c_2(t)$ with $c_3(t)$ and $c_4(t)$, respectively.

Fig. 9 shows how the transmitted symbol is detected with the proposed correlator. First, the receiver computes the correlation between the received signal and each segment. Second, depending on the symbol, the receiver chooses two of the four segment correlations and samples them at T and $T/2$. Two symbols of CTS pair use the same two segment correlations, and the receiver applies only sampling offsets (T and $T/2$) inversely to the segment correlations. Finally, the receiver computes the correlation output by combining the outputs of two segment correlator linearly, and decides which symbol was transmitted by comparing four combined outputs.

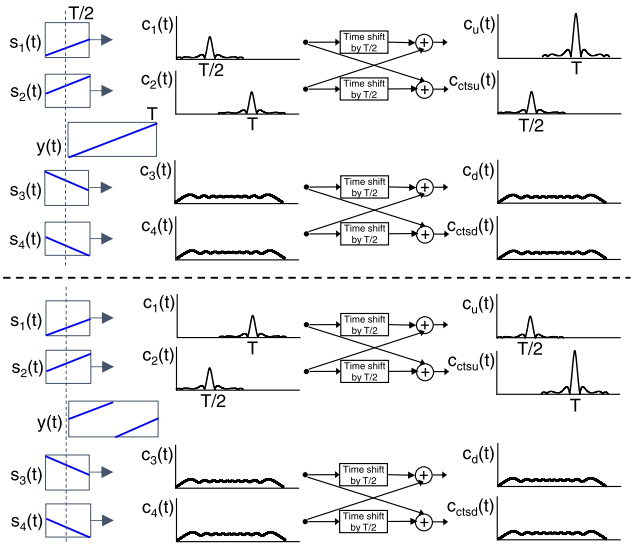


FIGURE 9. Example of the proposed correlator.

To reduce the computation complexity, we compute the segment correlations and symbol correlations using FFT in a similar way to (3).

$$\begin{aligned}
c_i(t) &= \mathcal{F}^{-1}\Big(\mathcal{F}\big(y(t)\big)\mathcal{F}\big(s_i(T/2 - t)\big)\Big),\\
c_u(t) &= \mathcal{F}^{-1}\Big(\mathcal{F}\big(y(t)\big)\mathcal{F}\big(x_u(T - t)\big)\Big)\\
&= c_1(t - T/2) + c_2(t),
\end{aligned}$$

$$c_{ctsu}(t) = \mathcal{F}^{-1}\Big(\mathcal{F}\big(y(t)\big)\mathcal{F}\big(x_{ctsu}(T-t)\big)\Big) = c_1(t) + c_2(t - T/2).$$

Similarly, we compute $c_d(t)$ and $c_{ctsd}(t)$, by replacing $c_1(t)$ and $c_2(t)$ with $c_3(t)$ and $c_4(t)$, respectively.

Unlike (3), the proposed method for calculating the symbol correlation does not require the FFT of time-reversed symbols but the FFT of time-reversed segments. Therefore, in the case of the proposed correlator, the FFT size is chosen to be a power of two, which is not smaller than the sum of the received signal length and the segment length, i.e., $N_{sym}/2$.

If the receiver decodes a signal after recording an entire frame, frame reception delay increases greatly, due to the long duration of the frame and high complexity of the bulk processing. To reduce frame reception delay, the receiver decodes a part of the frame in parallel during recording. We consider symbol-by-symbol recording and decoding, thus the length of the received signal is set to N_{sym}. Accordingly, the sum of the received signal length and the symbol length is $2N_{sym}$ while the sum of the received signal length and the segment length is $(3/2)N_{sym}$. Thus, if a positive integer m satisfying $(3/2)N_{sym} \leq 2^m < 2N_{sym}$ exists, i.e., $2^{m-1} < N_{sym} \leq (2/3)2^m$, the proposed method halves the FFT size of the existing method.

B. RAKE RECEIVER

To mitigate the effects of multipath fading and improve SNR in multipath environments, we apply RAKE receiver to AAC. RAKE receiver, widely used in code division multiple access (CDMA) systems, combines multipath components with different time delays by leveraging high time resolution due to wide bandwidth. Similar to CDMA, CSS modulation exploits wide bandwidth, and thus chirp signals have narrow auto-correlation and high time resolution.

To combine the symbol energy dispersed through multiple paths, RAKE receiver finds the time offset of the major path component and determines which multipath components are combined to improve SNR. To estimate the multipath structure more accurately even under the noise floor, we use the preamble due to its longer duration and higher energy compared to the symbols. Fig. 10 shows the correlation between the original preamble and the received signal after they propagate through a multipath channel. We observe that the auto-correlation peak of the preamble is divided into several peaks due to the characteristics of multipath channel.

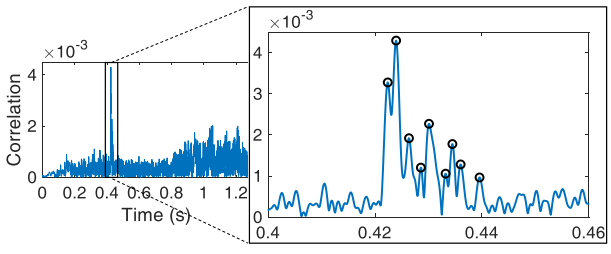


FIGURE 10. Preamble correlation in multipath environments.

We easily get the time offset of the major path by detecting the maximum peak of the preamble correlation.

Then, we select the correlation peaks as the multipath components near the major path if their magnitudes are larger than a threshold. These multipath components are highlighted by black circles in Fig. 10. Otherwise, the peaks are considered noises. This threshold-based method is commonly used to estimate the time delay of multipath components. We empirically set the threshold equal to the magnitude of the main peak multiplied by 0.3. Finally, by sampling the symbol correlation at the timing offsets of multipath components and combining them, we recover the dispersed symbol energy due to multipath fading. To combine the sampled outputs, we employ maximal ratio combining (MRC) that linearly combines multipath components with different weights proportional to the SNR of each path.

C. FRAME COMBINING

In aerial acoustic channels, extremely low transmission power and severe ambient noise may consecutively prevent the receiver from decoding frame successfully. In that case, both the delay in successful frame reception and the resulting power consumption of the receiver increase due to repeated recording and decoding processes. Considering that the transmitter sends the same bits repeatedly to complement the absence of feedback mechanism, we employ a frame combining technique to improve SNR, similar to the hybrid automatic repeat request (HARQ) [22], which results in reduced power consumption of the receiver.

We combine the outputs of RAKE receiver instead of raw signals, between consecutive frames. Due to the long propagation delay of acoustic signals and time-varying channels, combining raw signals that experienced different channels causes the signals to interfere with each other, and thus significantly degrades decoding performance. Moreover, the low sampling frequency of COTS devices causes inaccurate time-synchronization between consecutive frames, which makes it difficult to combine raw signals. However, RAKE receiver estimates the channel that each frame has experienced, and samples the outputs within each frame, without synchronization between frames.

1) IMPULSE NOISE DETECTION

To improve SNR, we employ a frame combining technique using RAKE receiver. However, the resulting SNR may become lower, if any of the frames used for the combination is the frame corrupted by impulse noise that has large power even at high frequencies. Fig. 11 shows the correlation between the chirp symbols and the signal of Fig. 3(b). We observe that, without impulse noise in the frame, the auto-correlation peak is sufficiently large compared to the cross-correlation and the noise floor, and thus the signals may be decoded successfully (black box). However, if there is impulse noise within the same frame, all symbol correlations have a huge noise floor covering up the auto-correlation peak, which causes decoding errors (red box). Once the frame,

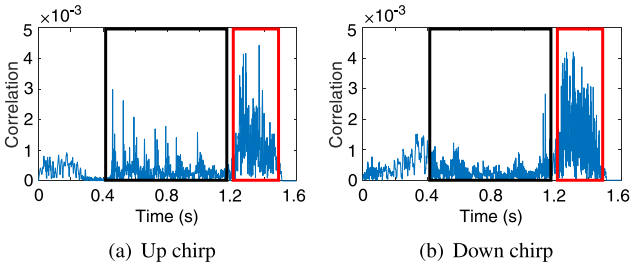


FIGURE 11. Correlations between impulse noise and chirp signal.

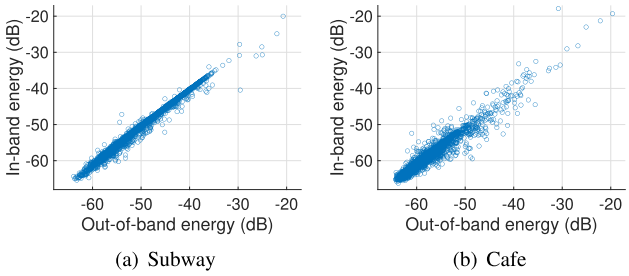


FIGURE 12. Energy of in-band (18.5–19.5 kHz) and out-of-band (17–18 kHz).

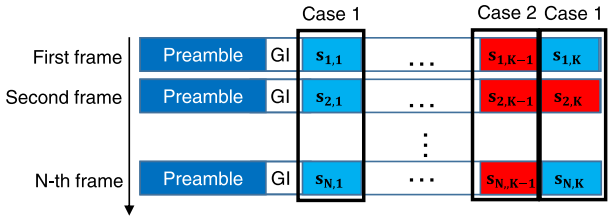


FIGURE 13. Frame combining with impulse noises.

corrupted by impulse noise, is combined, the huge noise floor will remain in the subsequent combining attempts and constantly cause decoding to fail. Hence, the receiver should detect the frames corrupted by impulse noise and exclude the frames from the combination. To detect impulse noise without increasing computational complexity, we take an energy detection (ED) approach. Specifically, we first measure in-band energy of the received signal and determine that the signal is corrupted by impulse noise if the detected energy is larger than a threshold value. To calculate the energy of the received signal, FFT computation to calculate the symbol correlations can be reused without additional computation. However, with only in-band energy, the existence of impulse noise may be misjudged because in-band energy also increases due to other factors, such as speaker volume and transmission range.

Therefore, we exploit not only in-band energy but also out-of-band energy to detect impulse noise, which is motivated by the observation that impulse noise induces large power across the entire audio frequency range [see Fig. 3(b)]. For more observations, we have measured ambient noise without signals in a subway train and a cafe for 30 minutes using Galaxy S8 and calculated in-band and out-of-band energy of 96 ms time samples, with setting in-band to 18.5–19.5 kHz and out-of-band to 17–18 kHz. The experimental results show that when in-band energy of ambient noise is high, out-of-band energy is also high, as shown in Fig. 12. Since our signal does not affect out-of-band energy, the receiver uses this energy to reduce false detection of impulse noise. Similar to In-band energy, out-of-band energy can be also computed by reusing FFT computation for the symbol correlations. In short, the receiver judges that impulse noise exists if in-band energy and out-of-band energy are larger than their respective threshold value. The thresholds are pre-determined from experimental evaluations depending on the device, or adaptively determined based on error statistics. In this paper, we use pre-determined thresholds.

2) THE PROPOSED SCHEME

We propose a frame combining scheme for AAC as follows. First, the receiver determines whether the received frame is correctly decoded, by cyclic redundancy check (CRC). Then, if there occur consecutive frame errors, the receiver attempts to combine frames. We assume that the receiver combines N consecutive frames consisting of a preamble, a preamble GI, and K symbols, as shown in Fig. 13. $s_{n,k}$ denotes the signal with the k-th transmitted symbol in the n-th frame. The receiver apply this combination symbol-by-symbol and use only remaining signals for each symbol after removing the signals corrupted by impulse noise.

To be specific, if there remain one or more signals on the same symbol (Case 1 in Fig. 13), the receiver combines RAKE outputs of remaining signals. On the other hand, if there remains no signal on the same symbol (Case 2 in Fig. 13), the receiver does not have any information about the symbol and declares that the symbol is undecided. Then, in the case of undecided symbols, the receiver conducts exhaustive searching for the symbols with the help of CRC. If the number of undecided symbols is i, there are 4^i candidate sets for the quaternary symbols. Among the 4^i candidates, the receiver looks for a symbol set that satisfies CRC. If more than one candidate meets the CRC, the receiver stops the process and repeats the above processes again for the next frame. We restrict the number of frames used for the combination (N) up to N_{max}. If the number of consecutive frame errors exceeds N_{max}, the receiver discards the oldest frame and combines the rest.

VI. PERFORMANCE EVALUATION

In this section, we present the experimental results and evaluate the performance of the proposed symbols and receiver operation. To test in different levels of ambient noise, we conduct the experiments in two places: a classroom (10×14 m^2) and a noisy cafe (17×12 m^2). A MacBook Pro laptop (Tx) plays wave audio files, and various smartphones (Rx) including Samsung Galaxy S4, S8, LG G2, and G4 record the audio signals. We placed the smartphones on a table at the center of each place, with the transmission range of 8 m in the classroom and 5 m in the cafe, respectively.

We fix the speaker volume of the laptop to 50% of the maximum volume and change the mixed-signal power (dBFS) of the played audio file. We verify that acoustic signals can avoid human perception even with the maximum transmission power. To prevent bias of the results due to time-varying acoustic channels, especially ambient noise, the speaker alternately transmits CTS-chirp frame and QOK frame.

For fair comparison, the frame architecture and parameters are set to be the same as in the comparison scheme [1]. The frame consists of a preamble, a preamble GI, and 11 symbols. The frequency range of the signal lies between 18.5 to 19.5 kHz. The preamble duration, preamble GI duration, and symbol duration are 368, 40, and 96 ms, respectively, making the overall frame duration 1463 ms.

A. CTS-CHIRP

We first investigate symbol error rate (SER). Fig. 14 shows the SER performance of CTS-chirp and QOK in the classroom, averaged over 1,000 repetitions for each power setting. Since the performance of each device varies widely, we summarize the experimental results for different power settings depending on the device. S4 shows much lower performance compared to other devices, thus it is important to achieve high reliability even in low-performance devices, such as S4. It can be observed that CTS-chirp outperforms QOK in all devices, especially in G2 and S4 whose microphones have severe frequency selectivity, reducing SER by up to 12.6%p (99%).

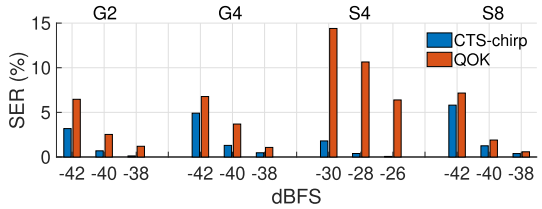
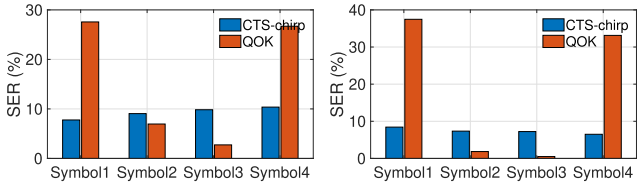


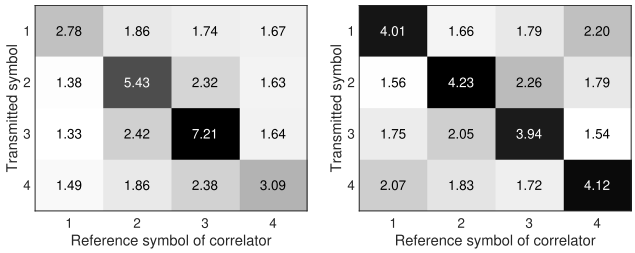
FIGURE 14. Symbol error rate (SER).

We observe the SER depending on the symbol type. Fig. 15 depicts the cases of G2 and S4. The SER of QOK symbols are significantly high in Symbols 1 and 4 while the SER of CTS-chirp symbols is evenly spread regardless of the symbol type. Although Symbols 2 and 3 of QOK have lower SER compared to CTS-chirp symbols, QOK has much higher SER on average, due to Symbols 1 and 4. The asymmetry in the SER of QOK symbols comes from the asymmetric received symbol energy.



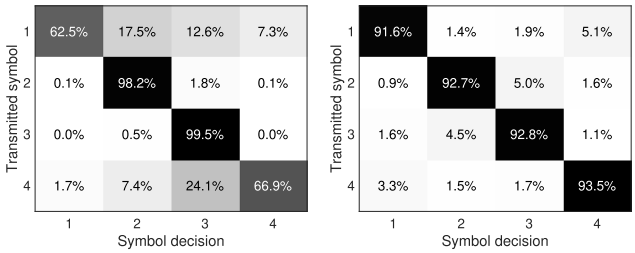
(a) S4 at −30 dBFS (b) G2 at −42 dBFS

FIGURE 15. SER depending on each symbol type.



(a) QOK (b) CTS

FIGURE 16. Correlator output with RAKE receiver (S4 at −30 dBFS).



(a) QOK (b) CTS

FIGURE 17. Symbol decision probability (S4 at −30 dBFS).

Fig. 17 shows the symbol decision probabilities. Specifically, the diagonal elements are the probabilities that each symbol is correctly judged as the transmitted symbol when the symbol was transmitted, and the other elements are the probabilities that each symbol is misjudged as the transmitted symbol when the different symbol was transmitted.

As shown in Fig. 7, Symbol 1 has high cross-correlation with Symbol 2 in QOK, while it has high cross-correlation with Symbol 4 in CTS. Thus, in Fig. 17, we observe high probabilities of misjudging Symbol 1 as Symbol 2 in QOK and Symbol 1 as Symbol 4 in CTS-chirp. Especially, the misjudgment probabilities of QOK are much higher although the cross-correlation of QOK symbols is lower compared to that of CTS-chirp symbols. This is because the frequency selectivity of the COTS devices' audio interfaces cause the asymmetry in received symbol energy.

Fig. 16 shows the average RAKE outputs of the correlation between the received signal and each symbol, depending on the transmitted symbol, with S4 at −28 dBFS. For example, the element in the second row and the first column means the average output of the correlation between the received signal and Symbol 1, when Symbol 2 was transmitted. The diagonal elements and other elements are RAKE outputs of the auto-correlation and the cross-correlation, respectively. Thus, the larger the difference between the diagonal elements and other elements, the better the decoding performance.

Fig. 17 shows the symbol decision probabilities. The diagonal elements are the probabilities that each symbol is correctly judged as the transmitted symbol, and the other elements are the probabilities of misjudgement, thus the sum of each row is one. Symbol 1 has high cross-correlation with Symbol 2 in QOK, while it has high cross-correlation with Symbol 4 in CTS-chirp. Fig. 17 also shows high probabilities

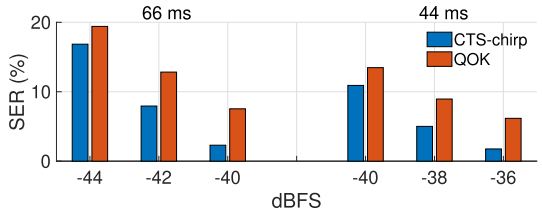


FIGURE 18. Averaged SER depending on symbol length.

of misjudging Symbol 1 as Symbol 2 in QOK and Symbol 1 as Symbol 4 in CTS-chirp. Especially, the misjudgment probabilities in QOK are much higher although the cross-correlation of QOK symbols is lower compared to that of CTS-chirp symbols. This is because frequency selectivity of audio interfaces causes the asymmetric received symbol energy.

After going through the acoustic channel with severe frequency selectivity, Symbol 1 has much lower energy compared to Symbol 2 in QOK. Thus, the auto-correlation of Symbol 1 is small, while the effect of ISI, caused by Symbol 2 in front of Symbol 1, is more serious due to higher energy of Symbol 2, (see (1, 1) and (1, 2) of the left matrix in Fig. 16). On the contrary, Symbol 2 has larger auto-correlation, so it is less vulnerable to ISI compared to Symbol 1. The same issue can be applied to Symbols 3 and 4 in QOK. However, without frequency division, CTS-chirp has resilience to frequency selectivity, thus achieving symmetric received symbol energy and equally low SER regardless of the symbol.

Fig. 19 shows that, unlike CTS-chirp, the asymmetry in the SER of QOK symbols causes significantly high error probabilities at specific bits or symbols in a frame. Thus, when the receiver combines consecutive frames, CTS-chirp has a higher time-diversity gain compared to QOK. To verify the performance of CTS-chirp and QOK for different symbol lengths, we measure SER using G2, G4, and S8 with 500 repetitions, at the same range of transmission power, and then average the measurement values. CTS-chirp shows lower SER with different symbol lengths, i.e., 66 and 44 ms, as shown in Fig. 18.

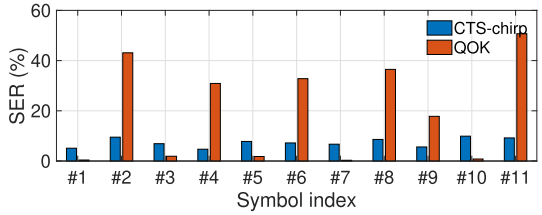


FIGURE 19. SER depending on symbol index (S4 at −30 dBFS).

Due to better SER performance, CTS-chirp has a higher frame reception rate than QOK, especially in S4, improving the FRR by up to 59.8%p (267.9%), as shown in Fig. 20. We evaluate CTS-chirp in a cafe with higher level of ambient noise. Fig. 21 shows that CTS-chirp has a higher FRR than

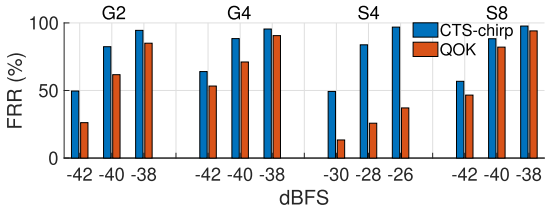


FIGURE 20. Frame reception rate (FRR) in a classroom.

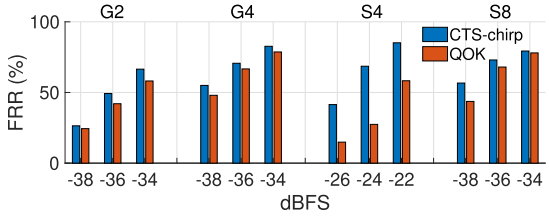


FIGURE 21. Frame reception rate (FRR) in a cafe.

QOK in a noisy cafe, but with less performance improvement in several devices compared to the results in a classroom. This is because, in noisy environments, ambient noise has a larger impact on performance than ISI and equivalently reduces the reliability of CTS-chirp and QOK. However, in the case of a device whose audio interface suffers serious frequency selectivity, such as S4, CTS-chirp has much higher reliability than QOK, even in such noisy place. Therefore, we claim CTS-chirp achieves higher reliability than the existing symbols, under a variety of environments.

B. COMPUTATION-EFFICIENT CORRELATOR

Now we investigate the computational complexity and power consumption of our proposed demodulation method. We consider that the receiver records and decodes a signal symbol-by-symbol, thus the length of the received signal is equal to the symbol length.

Thus, when the receiver calculates symbol correlation with FFT/IFFT computation, the lower bound of FFT size decreases from $2N_{sym}$ to $(3/2)N_{sym}$ in the proposed method. With the symbol duration (96 ms) and sampling frequency at 44.1 kHz, the FFT size of the existing method is selected to be a power of two which is not smaller than 8,468, i.e., 16,384. On the other hand, the FFT size of the proposed method is selected to be a power of two which is not smaller than 6,351, i.e., 8,192.

The power consumption of COTS devices using AAC mainly depends on audio sampling and FFT computation. The authors in [2] conduct experiments to estimate the power consumption of audio sampling and FFT computation. The experimental results show that the power consumption for audio sampling is only several milliwatts, even with the sampling frequency 44.1 kHz. However, the power consumption of FFT is hundreds of milliwatts and increases linearly with the FFT size. Hence, reducing the FFT size is an efficient way to reduce power consumption of COTS devices. According to the results in [2], halving the FFT size reduces power

consumption by tens to hundreds of milliwatts, depending on the device.

C. PROPOSED COMBINING SCHEME

We propose a frame combining technique for AAC to enhance SNR and achieve high reliability. The symbol energies in in-band (18.5–19.5 kHz) and out-of-band (17–18 kHz) are used to mitigate the effects of severe ambient noise. To evaluate the performance of frame combining in blind retransmission, we transmit the same frames consecutively. We evaluate FRR performance without frame combining, with only MRC, and with the proposed combining scheme. We take an average of over 500 repetitions for each power setting, as shown in Tables 1 and 2.

TABLE 1. FRR of G4 in a cafe with frame combining.

dBFS \ Scheme	w/o comb.	MRC	Prop.
−40	28.6%	35%	40.2%
−38	60%	62.8%	69.2%
−36	69%	71.4%	74.8%

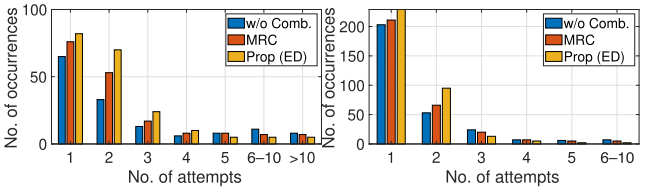
TABLE 2. FRR of S4 in a cafe with frame combining.

dBFS \ Scheme	w/o comb.	MRC	Prop.
−28	16.6%	28.4%	30.6%
−26	31.4%	38.6%	43.4%
−24	63.8%	68.6%	70%

The experimental results present that with G4 and S4 at −40 dBFS, the proposed scheme improves FRR by up to 14%p (84.3%), compared to the case without frame combining. The proposed method has a higher combining gain compared to the method using only MRC, by detecting signals corrupted by impulse noise, through ED. The gain of MRC decreases as the signal power increases, since high FRR reduces the occurrence of consecutive frame errors and the use of frame combining. However, since impulse noise prevents frame reception even at high SNR, the proposed method still outperforms the method with only MRC by a similar amount in all power settings. Thus, we claim that the proposed frame combining method improves the reliability in AAC.

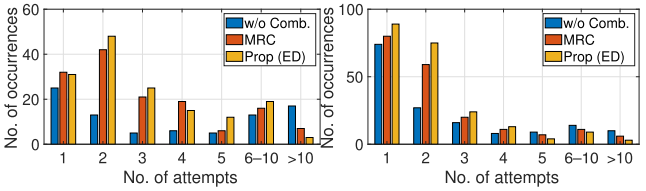
We also evaluate delay performance in frame reception with the proposed combining scheme. The delay is determined by the number of decoding attempts for a successful frame reception with and without frame combining. Figs. 22 and 23 depict the histogram of the number of attempts to receive a frame, i.e., number of consecutive failures in frame reception plus one, in G4 and S4. The proposed combining scheme not only improves SNR but also removes signals corrupted by strong noise, which reduces the possibility of excessive delay, when severe ambient noise continues for a certain period.

The results in G4 show that the proposed combining scheme increases the occurrence of receiving a frame within two attempts at −40 dBFS and −38 dBFS, compared with



(a) −40 dBFS (b) −38 dBFS

FIGURE 22. Delay in frame reception with combining in G4.



(a) −28 dBFS (b) −26 dBFS

FIGURE 23. Delay in frame reception with combining in S4.

the scheme without combining, by 55.1% and 26.6%, respectively. For S4, the proposed scheme increases the occurrences of receiving a frame within two attempts at −28 dBFS and −26 dBFS, by 107.9% and 62.4%, respectively. Therefore, the proposed combining scheme helps a frame to be successfully received within fewer attempts, resulting in high reliability and reduced energy consumption due to the reduced number of recording and decoding attempts.

VII. CONCLUSION

In this paper, we proposed the enhanced modem design of CSS-based AAC, to achieve high reliability and low computational complexity. The proposed quaternary symbols, named CTS-chirp, exploit the time shift property to cope with frequency selectivity of audio interfaces of COTS devices. We developed the computation-efficient method for demodulating the proposed symbols that halves the FFT size of symbol correlator. Considering blind retransmission in broadcasting, we employ RAKE receiver and a frame combining technique, and modify them to suit acoustic channels by mitigating the effects of multipath fading and ambient noise. The extensive experimental study shows that CTS-chirp has lower SER than the existing symbol design, and the proposed combining scheme improves FRR and reduces excessive delay in frame reception, thus achieving high reliability. Our evaluation demonstrates that the proposed method for demodulating symbols reduces power consumption of COTS devices.

REFERENCES

[1] S. Ka, T. H. Kim, J. Y. Ha, S. H. Lim, S. C. Shin, J. W. Choi, C. Kwak, and S. Choi, "Near-ultrasound communication for TV's 2nd screen services," in *Proc. 22nd Annu. Int. Conf. Mobile Comput. Netw.*, Oct. 2016, pp. 42–54.

[2] G. Lan, W. Xu, S. Khalifa, M. Hassan, and W. Hu, "VEH-COM: Demodulating vibration energy harvesting for short range communication," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. (PerCom)*, Mar. 2017, pp. 170–179.

[3] C. Kwak, S. Kim, S. Ka, J. Lee, and S. Choi, "No entry: Anti-noise energy detector for chirp-based acoustic communication," in *Proc. 16th Annu. IEEE Int. Conf. Sens., Commun., Netw. (SECON)*, Jun. 2019, pp. 1–9.

[4] H. Lee, "Aerial acoustic communication using chirp signal," Ph.D. dissertation, Dept. Elect. Comput. Eng., Seoul Nat. Univ., Seoul, South Korea, 2014.
[5] *Proximity-Services (ProSe) Management Objects (MO)*, ver. 13.3.0, document TR 24.333, 3GPP, Jun. 2016.
[6] V. Gerasimov and W. Bender, "Things that talk: Using sound for device-to-device and device-to-human communication," *IBM Syst. J.*, vol. 39, nos. 3–4, pp. 530–546, 2000.
[7] C. V. Lopes and P. M. Q. Aguiar, "Aerial acoustic communications," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, Oct. 2001, pp. 219–222.
[8] R. Nandakumar, K. K. Chintalapudi, V. Padmanabhan, and R. Venkatesan, "Dhwani: Secure peer-to-peer acoustic NFC," in *Proc. ACM SIGCOMM Conf. SIGCOMM (SIGCOMM)*, 2013, pp. 63–74.
[9] B. Zhang, Q. Zhan, S. Chen, M. Li, K. Ren, C. Wang, and D. Ma, "PriWhisper: Enabling keyless secure acoustic communication for smartphones," *IEEE Internet Things J.*, vol. 1, no. 1, pp. 33–45, Feb. 2014.
[10] H. S. Yun, K. Cho, and N. S. Kim, "Acoustic data transmission based on modulated complex lapped transform," *IEEE Signal Process. Lett.*, vol. 17, no. 1, pp. 67–70, Jan. 2010.
[11] H. Matsuoka, Y. Nakashima, and T. Yoshimura, "Acoustic communication system using mobile terminal microphones," *NTT DoCoMo Tech. J*, vol. 8, no. 2, pp. 2–12, 2006.
[12] 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 *Proc. 22nd Annu. Int. Conf. Mobile Comput. Netw.*, Oct. 2016, pp. 29–41.
[13] M. Eichelberger, S. Tanner, G. Voirol, and R. Wattenhofer, "Receiving data hidden in music," in *Proc. 20th Int. Workshop Mobile Comput. Syst. Appl.*, Feb. 2019, pp. 33–38.
[14] G. Lan, D. Ma, M. Hassan, and W. Hu, "HiddenCode: Hidden acoustic signal capture with vibration energy harvesting," in *Proc. IEEE Int. Conf. Pervas. Comput. Commun. (PerCom)*, Mar. 2018, pp. 1–10.
[15] M. Hanspach and M. Goetz, "On covert acoustical mesh networks in air," *J. Commun.*, vol. 8, no. 11, pp. 758–767, 2013.
[16] G. E. Santagati and T. Melodia, "U-Wear: Software-defined ultrasonic networking for wearable devices," in *Proc. 13th Annu. Int. Conf. Mobile Syst., Appl., Services (MobiSys)*, 2015, pp. 241–256.
[17] A. S. Nittala, X.-D. Yang, S. Bateman, E. Sharlin, and S. Greenberg, "PhoneEar: Interactions for mobile devices that hear high-frequency sound-encoded data," in *Proc. 7th ACM SIGCHI Symp. Eng. Interact. Comput. Syst. (EICS)*, 2015, pp. 174–179.
[18] H. Lee, T. H. Kim, J. W. Choi, and S. Choi, "Chirp signal-based aerial acoustic communication for smart devices," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Apr. 2015, pp. 2407–2415.
[19] Y. Lin and W. H. Abdulla, *Audio Watermark*, vol. 146. Heidelberg, Germany: Springer, 2015.
[20] A. Farina, "Advancements in impulse response measurements by sine sweeps," in *Proc. Audio Eng. Soc. Conv.*, Vienna, Austria, no. 123, May 2007, pp. 1–21.
[21] C. Cai, R. Zheng, and M. Hu, "A survey on acoustic sensing," 2019, *arXiv:1901.03450*. [Online]. Available: http://arxiv.org/abs/1901.03450
[22] P. Sindhu, "Retransmission error control with memory," *IEEE Trans. Commun.*, vol. 25, no. 5, pp. 473–479, May 1977.

JIHWAN LEE (Graduate Student Member, IEEE) received the B.S. degree in electrical and computer engineering from Seoul National University (SNU) in 2016, where he is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering. His current research interests include the IoT connectivity and the emerging technologies of next-generation cellular systems.

CHULYOUNG KWAK (Graduate Student Member, IEEE) received the B.S. degree in electrical and computer engineering from Seoul National University (SNU) in 2015, where he is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering. His research interests include wireless communication systems for the IoT and mobile deep learning applications.

SEONGWON KIM (Member, IEEE) received the B.S. degree in electrical engineering from the Pohang University of Science and Technology (POSTECH) in 2011, and the M.S. and Ph.D. degrees in electrical and computer engineering from Seoul National University (SNU), in 2013 and 2017, respectively. Before joining SK Telecom, he was a Postdoctoral Researcher with the SNU, Seoul, from October 2017 to February 2019. He was also a Visiting Research Scholar at the Department of Electrical and Computer Engineering, Rice University, USA, from July 2017 to October 2017. He is a Research Engineer with Vision AI Labs, SK Telecom, Seoul, South Korea. His current research interests include applied AI, the IoT connectivity, and next-generation wireless networks.

SAEWOONG BAHK (Senior Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from Seoul National University (SNU), in 1984 and 1986, respectively, and the Ph.D. degree from the University of Pennsylvania in 1991. He was a Member of Technical Staff with AT&T Bell Laboratories from 1991 to 1994, where he had worked on network management. From 2009 to 2011, he served as the Director of the Institute of New Media and Communications. He is currently a Professor with the SNU. He has been leading many industrial projects on 3G/4G/5G and the IoT connectivity supported by the Korean industry. He has published more than 200 technical articles and holds more than 100 patents. He is a member of the National Academy of Engineering of Korea (NAEK) and Who's Who Professional in Science and Engineering. He is the President of the Korean Institute of Communications and Information Sciences (KICS). He was a recipient of the KICS Haedong Scholar Award in 2012. He was the General Chair of the IEEE Dynamic Spectrum Access and Networks (DySPANs) in 2018 and the Director of the Asia–Pacific Region of the IEEE ComSoc. He is serving as the Chief Information Officer (CIO) of the SNU and the General Chair of the IEEE Wireless Communication and Networking Conference (WCNC). He was the TPC Chair of the IEEE VTC-Spring in 2014 and the General Chair of JCCI in 2015. He was the Co-Editor-in-Chief of the *Journal of Communications and Networks* (JCNs) and served on the Editorial Board of *Computer Networks* (COMNET) journal and the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (TWireless). He is an Editor of *IEEE Network Magazine*.