UltraComm: High-Speed and Inaudible Acoustic Communication

Guoming Zhang, Xiaoyu Ji, Xinyan Zhou, Donglian Qi, and Wenyuan Xu(✉)

Zhejiang University, Hang Zhou, China
{realzgm,xji,xinyanzhou,qidl,wyxu}@zju.edu.cn

Abstract. Acoustic communication has become a research focus without requiring extra hardware on the receiver side and facilitates numerous near-field applications such as mobile payment, data sharing. To communicate, existing researches either use audible frequency band or inaudible one. The former gains a high throughput but endures being audible, which can be annoying to users. The latter, although inaudible, falls short in throughput due to the limited available (near) ultrasonic bandwidth (18–22 kHz). In this paper, we achieve both high speed and inaudibility for acoustic communication by modulating the coded acoustic signal (0–20 kHz) on ultrasonic carrier. By utilizing the nonlinearity effect on microphone, the modulated audible acoustic signal can be demodulated and then decoded. We design and implement UltraComm, an inaudible acoustic communication system with OFDM scheme based on the characteristics of the nonlinear speaker-to-microphone channel. We evaluate UltraComm on different mobile devices and achieve throughput as high as 16.24 kbps, meanwhile, keep inaudibility.

Keywords: Ultrasound · Inaudible acoustic communication · Nonlinearity · Device-to-device communication

1 Introduction

With the widespread use of mobile devices equipped with audio interfaces, acoustic communication has attracted an increasing amount of attention [14,16–19,25,31–34]. Unlike Wi-Fi, Near Field Communication (NFC) or Bluetooth, acoustic communication doesn't need extra hardware on the receiver side and provides a communication channel with only built-in microphones. In addition, it is not necessary to go through pairing handshake before establishing connection. The universality and convenience of acoustic communication makes it a lightweight communication scheme for mobile devices and facilitate numerous applications such as mobile payment, near-field file transferring and even interactive gaming.

W. Xu—Supported by National Key R&D Program of China (2018YFB0904900, 2018YFB0904904).

X. Chu et al. (Eds.): QShine 2019, LNICST 300, pp. 184–204, 2020.
https://doi.org/10.1007/978-3-030-38819-5_12

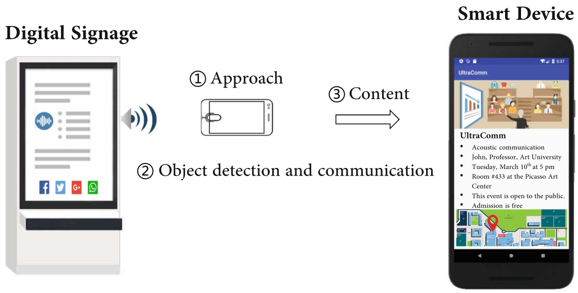


Fig. 1. Inaudible acoustic communication for information sharing. In a school, for example, the detailed information of activities on the screen of a digital signage can be conveniently transmitted to smartphones using the acoustic channel.

Current acoustic communication schemes can be divided into two categories: audible acoustic communication (roughly <18 kHz) [14,16,17,25,29] and inaudible acoustic communication (18–20 kHz) [28,31,33,34]. In general, the throughput for the former is always higher than the latter since it could utilize a wider frequency band, i.e., the audible frequency band. However, audible acoustic communication can be perceived by human ears, which could be annoying. Besides, audible communication also leaks information during transmission. Although information-hiding techniques [17,29] can alleviate the security problem, the sound in which the information is hidden is still audible. On the contrary, inaudible acoustic communication uses near-ultrasonic or ultrasonic sound to achieve inaudibility, while the available frequency band is limited, e.g., less than 4 kHz. We summarize representative literatures in Table 1. We can note that existing schemes fail to accomplish the seemingly conflicting goal of throughput and inaudibility.

We look into the seemingly contradictory goal of being both high-speed and inaudibility. We scrutinize the problem and find that the essence for the contradiction originates from a common belief: people take it for granted that acoustic communication must be within the cutoff frequencies of hardware components, e.g., ADC (analog-to-digital converter). Designed to capture audible sound, the cutoff frequency for an ADC is around 22 kHz, which is only 2 kHz higher than the human-perceivable sound frequency. As a result, the maximum inaudible frequency band is 4 kHz (plus the 18–20 kHz near-ultrasonic frequency band) in an ideal condition, not to mention the losses due to poor frequency responses.

In this paper, we investigate and explore a new approach for acoustic communication, whereby the high-speed and inaudible properties can be simultaneously achieved. We exploit the nonlinearity effect of electronic components (e.g., ADC and operational amplifier) to recover the low-frequency (<20 kHz) signals that are modulated onto a high-frequency (>22 kHz) carrier. That is, the low-frequency signals can be easily "**reproduced**" during the demodulation process caused by the nonlinearity effect of microphones. Utilizing the nonlinearity effect, we first theoretically investigate the throughput of inaudible communication and analyze the maximum throughput. Guided by the formulation, we

Table 1. Comparison with other acoustic communication schemes.

Existing work	Modulation type	Inaudibility	Throughput (bps)
Near-Ultrasound Screen [31]	Chirp QOK	√	15
Chirp [33]	Chirp	√	16
Acoustic OFDM [25]	OFDM	×	40
Dolphin [17]	OFDM	×	500
MCLT [29]	MCLT	×	600
Multi-Tone [28]	MFSK	√	800
PriWhisper [16]	FSK	×	1k
Dhwani [14]	OFDM	×	2.4k
U-wear [34]	GMSK	√	2.76k
BackDoor [27]	FM	√	4k
Ultrasound Proximity [3]	OFDM	√	4.9k
Ultracomm	**OFDM**	√	**16.24k**

develop UltraComm, an inaudible acoustic communication system that modulates audible signals on inaudible frequency. UltraComm is able to: (1) utilize the entire sound frequency band, to satisfy the high-speed requirement, and (2) transmit in high frequency, to maintain inaudibility.

With the increase of communication rate, UltraComm can be applied not only to the payment and authentication but also to some lightweight image and file transfer scenarios. For example, with UltraComm, any type of displaying content, such as activities, exhibitions, and advertising, can be delivered to the user's smartphone when approaching the digital signage. According to the new market research report [2], the digital signage market is expected to grow from USD 20.8 billion in 2019 to USD 29.6 billion by 2024, and it has been widely used in school, museums, transportation systems, and other public spaces, etc. However, the digital signages don't equip with convenient and efficient data transmission channel and cannot send the contents of interest to users to the smartphone. With UltraComm, the user can quickly get the information they want. As shown in Fig. 1, the information of activities can be broadcast to students who put their smartphones close to the digital signage.

The design of UltraComm addresses the following key challenges. First, to approach the theoretical throughput, we specially design the data symbol and use OFDM (Orthogonal frequency-division multiplexing) to increase the frequency efficiency. We exploit 2ASK (2 Amplitude Shift Keying) to modulate each "1" or "0" bit onto a subchannel and carefully choose the duration of symbol and guard interval time. Second, in order to reduce the crosstalk between subchannels caused by the nonlinear distortion, nonlinearity effect should be avoided as much as possible. However, UltraComm needs to demodulate the high-frequency modulated signal by using the nonlinearity effect. So how to utilize the nonlinearity effect while reducing the interference is a challenge. In order to suppress the high bit-error-rate (BER), UltraComm adopt an anti-distortion

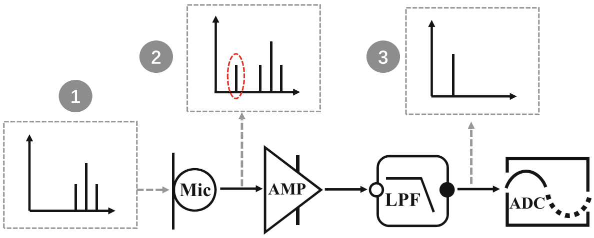


Fig. 2. The hardware structure inside a microphone system. If a modulated signal on high frequency carrier (>20 kHz) is inputted to the microphone ①, a new low-frequency (<20 kHz) signal is generated ②, the high-frequency signals are finally filtered by the LPF with the modulated signals low-frequency signal left ③.

strategy when designing the OFDM symbol. Third, like other electric devices, the speaker-to-microphone channels are high frequency selective due to the non-optimal response on both the speaker and microphone sides. As a result, the demodulated signals at the microphone side are distorted heavily. To maintain equivalent response at each subchannel, UltraComm proposes a differentiated gain control (DGC) mechanism by assigning different power coefficients to each subchannel, i.e., transmitting using various power levels at each subchannel. Fourth, to maximize the nonlinearity effect on the microphone side, we experimentally validate AM (Amplitude Modulation) parameters such as modulation depth and carrier frequency to deduce the proper ones.

In summary, our contributions are summarized as follows:

– We propose UltraComm, an inaudible communication system from a new perspective, which fundamentally addresses the conflict between high-speed and inaudibility for acoustic communication.
– We analyze the characteristic of the nonlinear speaker-to-microphone channel and propose an anti-distortion strategy to find the best subcarriers combination to improve communication performance.
– We implement the UltraComm prototype, evaluate its performance on unmodified mobile devices and achieve a throughput as high as 16.24 kbps.

2 Acoustic Nonlinearity Effect Background

In this section, we first describe the microphone system in mobile devices, and then elaborate the nonlinearity effect.

2.1 Microphone System

A microphone system converts acoustic waves into electrical signals. Microphones on mobile devices can be either Electret Condenser Microphone (ECM) or Micro Electrical Mechanical System (MEMS). Nowadays, MEMS microphones dominate the market of mobile devices such as smartphones, Pads and wearable devices like smart watch due to the miniature package sizes and low power consumption. Thus, we focus mainly on MEMS microphones in this paper yet the analysis suits for both MEMS and ECM microphones.

In order to capture audible sounds, microphones, low-pass filters (LPFs), and analog-to-digital converter (ADC) in the microphone system are used to suppress signals out of the frequency range of audible sounds (i.e., 20 Hz to 20 kHz). The typical structure of signal processing components in a microphone system is shown in Fig. 2. Most sound communication systems are assumed to only accept the audible frequency bands due to suppression of signals above 20 kHz.

2.2 Nonlinearity Effect Principle

Nonlinearity is a phenomenon that demodulated signals can be produced in low-frequency range. It is reported to appear in many electric components, such as operational amplifiers [21–24,26]. For a microphone, nonlinearity means when a modulated high-frequency signal (e.g. >20 kHz) passes through the microphone system, low-frequency signal (e.g. <20 kHz) can be generated and received by the microphone.

Nonlinearity effect has been regarded as distortions of devices and avoided [20]. In UltraComm, however, we attempt to exploit the nonlinearity effect to transmit modulated data on frequency above 20 kHz and recover it in the audible frequency band.

Without loss of generality, let S be a single frequency signal with a frequency of f_S. Theoretically, the output signal under nonlinearity effect can be modeled as:

$$S_{out} = \begin{cases} a_0 S & , f_S \leq f_{rcv} \\ \sum_{i=1}^{\infty} a_i S^i & , f_S > f_{rcv} \end{cases} \tag{1}$$

where a_0 is the gain coefficient for signals of a frequency less than f_{rcv} and f_{rcv} is the maximum receiving frequency of the receiver. In Eq. (1), when $f_S > f_{rcv}$, $S_{out} = a_1 S + a_2 S^2 + \cdots + a_n S^n$. Typically, the values of $a_i, (i \geq 1)$ are related to f_S, and a_i decreases with the increase of i. According to the basic trigonometric theory, the frequency of S^i is higher than f_{rcv} when $i \geq 1$. As a result, the nonlinearity response, i.e., $a_1 S^1$, $a_2 S^2$, $a_3 S^3$, ..., can be eliminated by the LPF in the microphone system. What is more, a_i becomes small when $i \geq 3$, thus we often consider the nonlinearity response as $S_{out} = a_1 S + a_2 S^2$ when $f_S > f_{rcv}$.

Nonlinearity Response for Microphone System. For a microphone system, the nonlinearity effect can be utilized for inaudible communications. Define a signal to microphone as $S_{in} = S_1(1 + S_2)$ where $S_1 = sin\omega_1 t$ is the carrier signal with frequency $f_{S_1} > f_{rcv}$ and $S_2 = sin\omega_2 t$ is the baseband signal with frequency $f_{S_2} < f_{rcv}$. According to Eq. (1), the nonlinearity response through a microphone S_{out} should be:

$$S_{out} = \sum_{i=1}^{\infty} a_i (S_1 + S_1 * S_2)^i \tag{2}$$

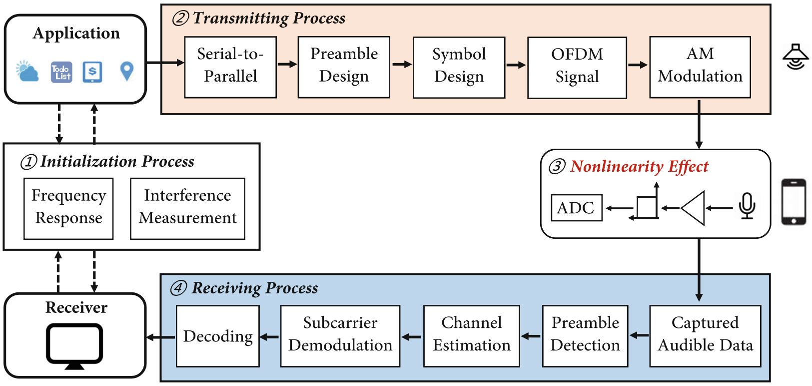


Fig. 3. Overview of the UltraComm system.

After neglecting the high-order harmonics ($i \geq 3$), S_{out} can be represented as:

$$S_{out} = a_1(sin\omega_1 t + sin\omega_1 t * sin\omega_2 t) + a_2(sin\omega_1 t + sin\omega_1 t * sin\omega_2 t)^2 \quad (3)$$

We then expand Eq. (3), and the output of the signal contains the baseband signal frequency f_{S_2}, the carrier signal frequency f_{S_1}, and other harmonics such as $(f_{S_1} - f_{S_2})$, $(f_{S_1} + f_{S_2})$ and other cross frequencies (i.e., $2f_{S_1}, 2f_{S_2}, 2(f_{S_1} + f_{S_2}), 2f_{S_1} + f_{S_2}, 2f_{S_1} - f_{S_2}$).

As mentioned above, the resulting signals from nonlinearity effect with frequency lower than the cut-off frequency of LPF (f_{LPF}, which is often set to 25 kHz) will remain while others are filtered. That is to say, the signal S_2 ($f_{S_2} < f_{rcv}$) modulated on S_1 will be reproduced and output from the microphone. Figure 2 illustrates the whole process where a signal from nonlinearity effect is generated.

3 UltraComm System Design

In this section, we analyze and verify the characteristics of nonlinear speaker-to-microphone channel, and design UltraComm based on them. The design of UltraComm mainly involves three components: initialization process, transmitter, and receiver design. The whole system is shown in Fig. 3 in detail.

The initialization process aims to get the status of the current speaker-to-microphone channel and then select appropriate communication parameters, such as the available bandwidth.

In the transmitter process, the high rate serial data is mapped to lower rate paralleled data by the S/P converter. After the S/P process, the data is modulated onto subcarriers with 2ASK scheme and then adds the preamble. Before the final transmission, we modulate the UltraComm data symbols on an ultrasonic carrier (in ultrasound frequency band, e.g., 40 kHz) with AM modulation. In this way, we transform the originally audible signal into an inaudible signal successfully.

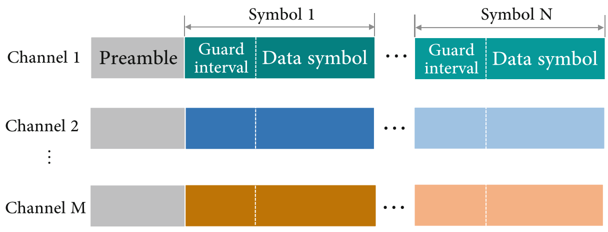


Fig. 4. The frame structure specified in UltraComm.

At the receiver side, demodulation that produces the low-frequency (baseband) signal happens due to the nonlinearity effect on microphones. A preamble in front of a frame is used for synchronization and channel estimation. Once the preamble is matched at the receiver side, the receiver demodulates the OFDM signal and then decodes it to obtain the data.

3.1 Initialization Process

Initialization process is the first step to initialize parameters for transmission, including (1) the frequency response of the speaker-to-microphone channel, (2) the noise level. All the above device-related and environment-dependent parameters are measured before data transmission.

To get the frequency response at each subchannel, we transmit each modulated single tone signal which lasts for 100 ms, and measure the results at the microphone side. The attenuation coefficients can therefore be derived and they are delivered back to the transmitter using existing communication schemes, such as Chirp [33]. According to those measured results, all communication parameters will be elaborately selected at the transmitter side which will be described in the following section and starts the data transmission.

In the following part, we only elaborate the key modules in both transmitter and receiver side.

3.2 Frame Structure

Preamble Design. In UltraComm, data frames are transmitted in the form of OFDM symbols. Figure 4 shows the structure of a frame, which is composed of symbols with a preamble in the front.

In UltraComm, preamble is used for synchronization between the transmitter and the receiver, and channel estimation in frequency–selective environments. As we can see from the Fig. 7(a), the frequency response decreases roughly with the increase of frequency. In order to predict channel state information (CSI) more accurately, we use the combination of two sinusoidal signals as preamble and the frequencies of sinusoidal signals are 5.1 and 15.1 kHz respectively. To estimate the fast-varying channel, preamble is assembled in the front of each frame to facilitate channel estimation [13].

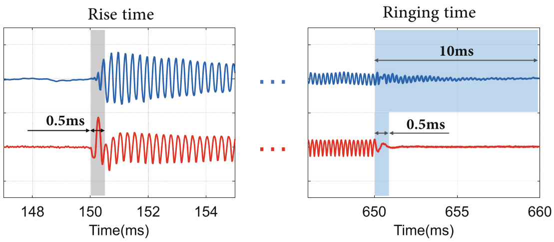


Fig. 5. Rise and ringing time measured at the microphone side using a single-tone 4 kHz signal (the top one) and an AM-modulated (carrier frequency 40 kHz) single-tone signal (the bottom one).

3.3 Symbol Design

As the most important part in transmitter, UltraComm symbol mainly considers subchannel design, subchannel modulation and duration of a symbol as well as the guard interval time.

How Long Should the Duration of Symbol (t_{symbol}) **be?** In UltraComm, we utilize a guard interval between adjacent data symbols to eliminate the inter symbol interference (ISI). Theoretically, the time duration of guard interval (t_{gi}) depends on the multipath effect as well as the rise and ringing time [14] in the nonlinear speaker-to-microphone channel.

Rise and Ringing Time. In order to obtain a proper value for t_{gi}, we investigate and observe the rise and ringing time using an iPhone 7 which has a MEMS microphone. Figure 5 shows the results with different stimulation. To be specific, with the single-tone input (a), the rise and ringing time are about 0.5 ms and 10 ms respectively but only 0.5 ms and 0.5 ms for the AM-modulated signal (b). This proves that modulated signal (AM) can utilize smaller t_{gi}, e.g., $t_{gi} \geq 0.5$ ms is enough. Actually, the larger value of t_{gi} used in existing acoustic communication system is one of the most important reasons for the low throughput.

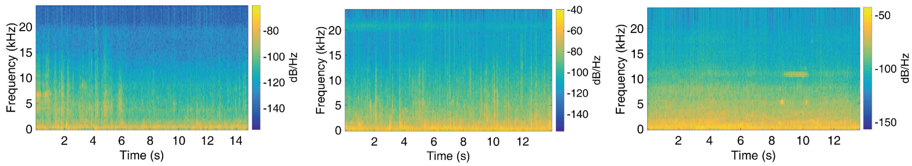
The time duration of data symbol: t_{data} is mainly determined by t_{gi} and the lowest frequency of subcarrier. To avoid excessive transmission rate loss, we require the value of $t_{gi}/t_{data} \leq 0.1$.

$$10 * t_{gi} \leq t_{data} \tag{4}$$

Based on Eq. (4), the minimum value of t_{symbol} can be set to 5.5 ms.

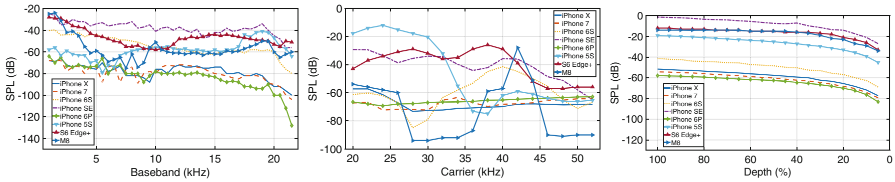
Subchannel Design. UltraComm adopts OFDM to transmit data efficiently. An OFDM signal is the sum of multiple subcarriers that are modulated independently with phase-shift keying (PSK), ASK, or Quadrature Amplitude Modulation (QAM). In UltraComm, each subcarrier is modulated with the 2ASK. Due to the unique characteristics of the nonlinear speaker-to-microphone channel, the design of subcarrier should consider the following questions:

- What is the available frequency for subchannels?
- What is the width of frequency spacing for subcarrier, namely how many subchannels should be divided?



(a) Office (b) Restaurant (c) Street

Fig. 6. Empirical study of ambient noise and its spectrum in three scenarios.



(a) (b) (c)

Fig. 7. Frequency-selectivity of nonlinearity for microphones. (a) Frequency response of modulated signal under nonlinearity differs across devices and the response becomes worse for higher frequency baseband; (b) Frequency response of modulated signal vs. different carrier frequencies at 2 kHz; (c) Frequency response of modulated signal vs. AM depth.

- How to deal with distortion products?
- How to deal with frequency selectivity?

Frequency Range for Subchannels. To achieve a high throughput, the available frequency band Δf should be wide enough. Unfortunately, Δf is often limited in reality.

Minimum Frequency f_{min}. The floor frequency of subcarrier (f_{min}) should be above the ceiling frequency of ambient noise to avoid interference. We conduct an experiment to get a sense of the ambient noise and show the result in Fig. 6. From the figure, the intensity of noises shows the difference for the three scenes (office, restaurant and street). Noise in the street is louder than that of office and restaurant, which is attributed to the crowded people and other sources such as vehicles. Nevertheless, the strength of noise dims when frequency increases, the majority of noise lies below 1 kHz. Thus, f_{min} can be empirically set to 1 kHz to avoid interference. To eliminate the interference of ambient noise, one can also amplify transmitting power to improve the signal-to-noise ratio (SNR) in the lower frequency range, therefore, the f_{min} can also be set lowly.

Maximum Frequency f_{max}. The upper limit frequency of subcarrier f_{max} depends on several factors: (1) Sampling rate of ADC in a microphone. f_{max} should be less than half of the sampling rate of ADC (f_{ADC}) to avoid aliasing. (2) The frequency response of microphone under nonlinearity effect. Microphones may have different frequency-selectivity with the increase of frequency. We will

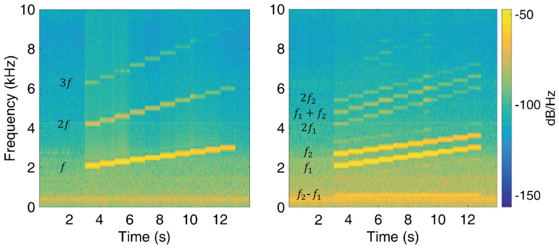


Fig. 8. Harmonic and intermodulation distortion of AM signal over speaker-to-microphone channel. The baseband signals are f and f_1, f_2 respectively.

elaborate the details later. (3) The relation between baseband signal and carrier signal. According to Eq. (3), $f_c - f_{max} > 20\,\text{kHz}$, where f_c is the carrier frequency, should be satisfied. This is to ensure only the baseband signal can be received by microphone under nonlinearity effect. For instance, if we modulate a 14 kHz baseband signal on a 30 kHz carrier, the lower sideband of the AM signal is 16 kHz, which is also audible.

Frequency-Selectivity for Baseband. To investigate the frequency-selectivity of microphones under nonlinearity effect, we conduct an experiment. We use a signal generator to generate a modulated signal and feed it into 8 smartphones respectively to test the frequency responses. As Fig. 7(a) shows, frequency responses vary across the 8 microphones and the responses roughly get worse when frequency increases. Denote the maximum acceptable frequency f_{worst}. In order to guarantee low BER, $f_{max} < f_{worst}$ should be satisfied.

Take all the aforementioned constraints into consideration, the frequency of any subchannel f_{sc} should satisfy the following conditions:

$$\begin{aligned} f_{min} \leq f_{sc} \leq f_{max} \\ f_{min} > f_{Noise} \\ f_{max} < min(f_{ADC}/2,\ f_c - 20k,\ f_{worst}) \end{aligned} \tag{5}$$

Subcarrier Spacing f_{sp}. To ensure the signal can be demodulated accurately, f_{sp} should be an integral multiple of the f_{res}, which is represented as:

$$f_{sc} \mod f_{res} = 0 \tag{6}$$

Besides Eq. (6), f_{sp} is also determined by the f_{min}, i.e., the harmonic distortion and intermodulation distortion (IMD) to be specific.

Dealing with Harmonic Distortion and IMD. Due to the nonlinearity of microphone circuits, the output of the circuits will contain harmonic components. Considering an input signal that contains one single frequency components at f, the output signal will contain higher harmonics with frequencies that are multiple of the frequency of the input signal, which can be expressed as:

$$f_{HAR} = \sum_{k=2}^{\mathbb{N}} k * f \tag{7}$$

In the same time, the output of microphone circuits also contains IMD products which are the result of two or more signals interacting in a nonlinear device to produce additional unwanted signals. For two input signals, the IMD products can be expressed as:

$$f_{IMD} = k_1 * f_1 + k_2 * f_2 \tag{8}$$

Where, k_1, k_2 are integers. The order of the intermodulation product is the sum of the integers $|k_1| + |k_2|$. In UltraComm, the interference of higher order IMD products is generally slight and can be ignored because they have lower amplitudes and are more widely spaced. However, the second-order components, $f_1 \pm f_2$ will interfere with the original signals. For example, if the frequencies of subcarriers are $f_1 = 200\,\text{Hz}$, $f_2 = 400\,\text{Hz}$, the frequencies of second-order products will contain 200 and 600 Hz, which is the same as one of the original frequency.

To verify the distortion in the case of AM signal as input, we modulated two different signals onto carrier respectively. As depicted in Fig. 8, the received signals include not only the original one (f and f_1, f_2), but also the second–order harmonics ($2f$ and $2f_1, 2f_2$), IMD products ($f_2 - f_1, f_1 + f_2$), even third order harmonics ($3f$), which will bring a considerable challenge for accurate decoding. Especially, as the number of baseband signals increases, the received signal is worse and can't be decoded.

In our case, the subcarriers can be represented as:

$$f_{sc} = f_{min} + n * f_{sp},\ n \in \{0, 1, 2, \ldots, \Delta f / f_{sp}\} \tag{9}$$

Based on Eqs. (7), (8) and (9), we find the second–order product may have the same frequency as the subcarrier. To avoid interference caused by IMD and harmonics, we proposed the value of f_{sp} should meet the following condition:

$$f_{sc} \notin n * f_{sp},\ n \in \{0, 1, 2, \ldots, \Delta f / f_{sp}\} \tag{10}$$

Based on Eqs. (6), (9) and (10), we have:

$$\begin{aligned} f_{min} \mod f_{sp} &\neq 0 \\ f_{min} \mod f_{res} &= 0 \\ f_{sp} \mod f_{res} &= 0 \end{aligned} \tag{11}$$

In our implementation, we empirically set $f_{sp} = 200\,Hz$, thus f_{min} can be set to $300 + 200 * i$, where i is non-negative integer.

Subcarrier Modulation. With the selected parameters for subcarriers and symbols, we elaborate subcarrier modulation with 2ASK as follows.

3.4 Modulating Symbols on Inaudible Frequency

In order to achieve inaudibility, the OFDM signal should be modulated on carriers of higher frequency (>20 kHz) before transmission. To fully leverage the

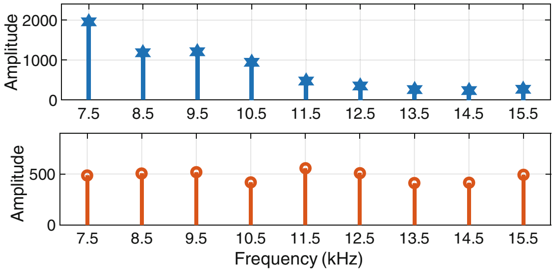


Fig. 9. Performance of DGC in terms of frequency response improvement. The top figure is without DGC while the bottom one utilizes DGC.

nonlinearity of microphone and recover the OFDM signal successfully, UltraComm utilizes AM for baseband (audible sound) modulation.

In the following, we will describe how to select proper AM parameters and cope with the audibility in UltraComm.

Carrier Frequency. Carrier frequency f_c mainly depends on two factors: the available frequency range of ultrasound and the frequency response of the microphone. Note that f_c should be larger than 20 kHz for inaudibility. Besides, f_c has to satisfy the condition of $f_c - f_{max} > 20$ kHz. For instance, given a baseband of 6 kHz, the carrier frequency has to be higher than 26 kHz to ensure the lower sideband of the AM signal is above 20 kHz and inaudible.

Frequency-Selectivity for Carrier. However, microphones are also frequency-selective. For the high-frequency carrier signals, the carrier frequency f_c, therefore, should utilize the carrier with good frequency response.

To discover the frequency-selectivity for the carrier signals, we measure the frequency response of microphones with 8 smartphones. We initially fix the baseband frequency at 2 kHz and then vary the carrier frequency incrementally to get the frequency-selectivity curve. Figure 7(b) reports that the magnitude of the demodulated signal fluctuates with the increase of carrier frequency. In summary, to guarantee the efficiency of communication and inaudibility, the carrier frequency should be carefully chosen. In our implementation, the carrier frequency is set to 40 kHz.

Differentiated Gain Control. Recall that in Fig. 7(a), the frequency response gets worse at higher baseband frequencies under the nonlinearity effect. In order to increase the bandwidth and obtain flat frequency response at the same time, we design a DGC mechanism to balance the frequency response by amplifying the original poor frequency responses. The basic idea is to assign different gain coefficients i.e., using different transmitting power across subcarriers. In this way, the frequency response can be improved.

Denote the finally transmitted signal as:

$$s(t) = \sum_{i=1}^{N_{sc}} A_i g_i(t) cos(2\pi f_i t) \tag{12}$$

where N_{sc} is the number of subchannels and $g_i(t) = 0$ or 1, "0" refers to symbol "0" on the subcarrier and vice versa. A_i is the gain coefficient for the ith subchannel when there is a symbol, i.e., $g_i(t) = 1$. f_i is the frequency for the ith subcarrier. The DGC mechanism assigns different A_i for each subchannel according to its fading coefficient and keeps all $A_i \cdot g_i(t)$ $(i \geq 1)$ roughly the same. The effect of DGC mechanism is shown in Fig. 9. The top figure shows the result with all $A_i = 1$, i.e., without using the DGC mechanism, and the bottom figure demonstrates the case when the DGC mechanism is applied. We can observe a significant difference in terms of frequency response from the two figures. To be specific, with the help of DGC, the frequency response improves across the selected frequency range. This proves the effectiveness of the proposed DGC mechanism in UltraComm.

Modulation Depth. Modulation depth is directly related to the efficiency of the nonlinearity effect of microphones. We investigate the influence of modulation depth upon frequency response with eight microphones. The result in Fig. 7(c) demonstrates that frequency response improves with the increase of modulation depth. In order to maximize the efficiency of the nonlinearity effect, the modulation depth should be set to 100%.

3.5 Receiver Design

Compared to the transmitter, the design of receiver is lightweight. First, UltraComm receiver recovers the OFDM signal by leveraging nonlinearity. UltraComm synchronize the transmitter and the receiver by using the preamble. Then subcarrier demodulation and decoding are performed.

AM Demodulation. As described in Sect. 2, a microphone acts as a demodulator when it receives an AM-modulated signal. Thus, the baseband signals are recovered in a natural way. With our aforementioned design, high-frequency harmonics can be filtered by the LPF and the low-frequency baseband signals can be well received.

Preamble Detection. After recovering the OFDM signal, cross-correlation algorithm is used to detect the preamble. Meanwhile, channel estimation can be performed according to the strength of the preamble.

Subcarrier Demodulation and Symbol Decoding. We use FFT algorithm to demodulate the OFDM signal. After demodulating all the subcarriers, the data on each subchannel can be obtained. The data frame transmitted from the transmitter, is therefore, decoded according to the corresponding encoding schemes.

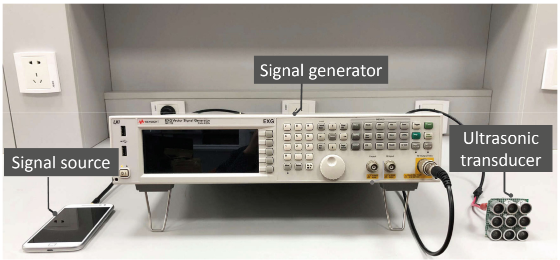


Fig. 10. Experimental setup. A smartphone is used as a signal source, and the signal generator does AM modulation. The modulated signal is transmitted via an array of 9 ultrasonic transducers [15].

4 Evaluation

In order to evaluate the performance of UltraComm, we conduct experiments in an office environment and analyze the influence of multiple factors. The received signal traces are processed and decoded in MATLAB. We summarize the main results as follows:

- A maximum of 16.24 kbps throughput is achieved on an iPhone 5S smartphone with 100 subcarriers ranging from 1.1 kHz to 19.9 kHz.
- The increase of guard interval and symbol duration can lower BER effectively yet undermine throughput.
- The increase of subcarrier number can improve throughput but increase BER.
- Communication distance and ambient noise have fundamental effect upon BER.

4.1 Experimental Setup

In our implementation, we choose different mobile devices as receivers and a set of benchtop equipment as transmitter respectively. Figure 10 is an overview of our experimental setup.

The transmitter is composed of three parts for benchtop-based implementation: (1) a signal source using a smartphone, (2) an AM modulator, which is actually a vector signal generator from Keysight [8], and (3) a narrow-band ultrasonic transducer array [15] or a high-quality full-band ultrasonic speaker Vifa [9]. To elaborate, the smartphone sends baseband signal to the signal generator, by which the baseband signal is modulated onto the high-frequency carrier and the modulated signal is transmitted via the ultrasonic transducer, the driving power is limited below 0.17 W in all the experiments except the user study of inaudibility.

System parameters are listed in Table 2. Some parameters are different among receivers. For example, the available bandwidth is different between iPhone X and iPhone 7, because the nonlinear frequency responses varied from devices.

Table 2. Parameters in UltraComm.

OFDM parameter	Value
Frequency Band	N/A *
Subcarriers	N/A *
Subchannel Bandwidth	200 Hz
Subcarrier Modulation	2ASK
Frame	Preamble + 100 symbols
Preamble	5.1 kHz, 15.1 kHz; 10 ms
Symbol Duration	5 ms, 10 ms
Guard Interval	$\geq$ 0.5 ms
AM Parameter	**Value**
Carrier Frequency	40 kHz
AM Depth	100%

∗ The parameters vary among different receivers.

Table 3. Throughput with different smart devices.

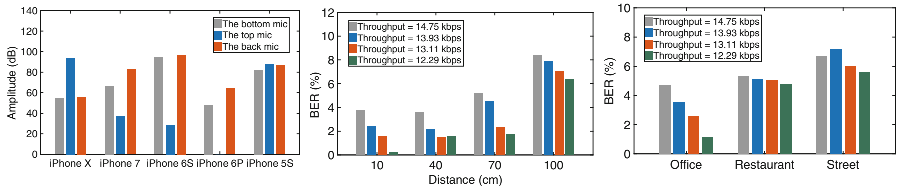
Smart devices	OS	Microphone	Throughput (kbps)	BER (%)
iPhone X	iOS 11.4	Top mic	14.91	1.47
iPhone X *	iOS 11.4	Top mic	13.11	7.6
iPhone 7	iOS 12.1.4	Back mic	15.57	4.61
iPhone 6S	iOS 12.1.4	Back mic	15.24	4.86
iPhone 5S	iOS 10.3.3	Top mic	16.24	1.45
iPhone 5S †	iOS 12.1.4	Top mic	14.75	1.21
iPhone 4S	iOS 9.3.5	Bottom mic	13.11	1.59

∗ The communication distance is 1 m, the others are 20 cm.
† Another iPhone 5S with identical technical spec.

4.2 Macro-benchmark Evaluation

Throughput of UltraComm. To examine the throughput of UltraComm with different receivers, we choose 6 kinds of representative mobile devices. The receiver is placed in front of the transducer array at a distance of 20 cm, with ambient noise of about 45 dB sound pressure level (SPL) measured by a professional decibel meter. Table 3 reports the parameters and results of the 6 different receivers. We can conclude that the throughput, in a manner, depends on the devices and is proportional to the available subcarriers. The diversity is attributed to the difference among audio interfaces, especially the MEMS microphones' nonlinearity frequency response.

Different Microphones. As we know, most of smart devices have several microphones, for example, iPhone X has three microphones, one is placed at the



(a) (b) (c)

Fig. 11. (a) The nonlinear frequency responses of top, bottom and back microphone (mic) with different smart devices; (b) The impact of communication distance on BER; (c) The impact of ambient noise in three scenarios on BER, the range of SPLs are 45–55 dB, 55–65 dB, and 65–80 dB respectively.

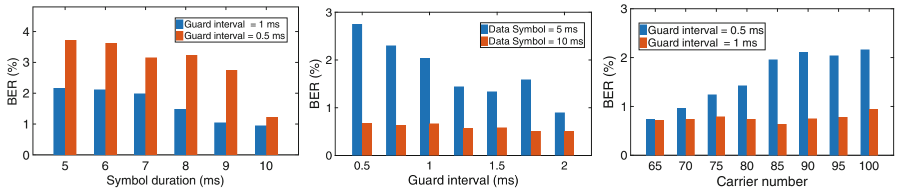
bottom, and two in the top. To achieve a high throughput, we need to choose the microphone with the best nonlinear frequency response. Thus, we tested five different mobile devices, the results are shown in Fig. 11(a). The nonlinearity effect of the top microphone in iPhone 6Plus is almost negligible. For iPhone X and iPhone 5S, the top microphones are preferred, and the back microphones are preferred for iPhone 7 and iPhone 6S in UltraComm.

4.3 Micro-benchmark Evaluation

BER with Different Communication Distance. In this section, we quantify BER at various distances. In general, a shorter communication distance leads to a lower BER, because a shorter distance always means a larger signal-to-noise ratio (SNR) for a given SPL. To explore the impact of communication distance on BER, we use the iPhone X smartphone as a receiver, the ambient noise is measured to be about 45 dB, and the throughput is varied from 12.29 kbps to 14.75 kbps by choosing different number of subcarriers.

Figure 11(b) shows the impact of distance on BER. We can find that BER worsens with the increase of distance. The values of BER are 3.76%, 2.43%, 1.62%, and 0.28% at a distance of 10 cm. When the distance increases to 100 cm, the BER rose to 8.39%, 7.91%, 7.07%, and 6.42% as the SNR worsens. The fundamental influence of distance upon BER, on the other hand, hurts throughput. In order to further increase the distance of UltraComm and reduce BER, we wish to improve distance with more complicated design in the future work.

BER with Different Ambient Noise. In addition to the communication distance, the ambient noise is also a key attribute that influences the performance of UltraComm. Thus, we examine the BERs in three scenarios: office, restaurant, and street by simulating the three scenarios, the SPLs of the corresponding ambient noise are set to 45–55 dB, 55–65 dB, and 65–80 dB respectively. In all experiments, the ultrasonic transducer is positioned 20 cm away from the iPhone X, the transmitter power is set to 0.16 W, and the throughput is also varied from 12.29 kbps to 14.75 kbps by choosing different number of subcarriers. From



(a) (b) (c)

Fig. 12. (a) The impact of symbol duration on BER; (b) The impact of guard interval on BER; (c) The impact of subcarrier number on BER.

Fig. 11(c), we conclude that BER worsens with the increase of ambient noise, which is because the SNR decreases simultaneously for a given transmitter power. Therefore, it's necessary to increase transmitter power to suppress the BER in a noisy scenario.

The Impact of Symbol Duration on UltraComm. The throughput is affected by symbol duration, a short symbol duration may be difficult to decode and cause a high BER in case of poor communication channels that suffer from low SNR. Thus, we evaluate the performance of UltraComm with different symbol duration. We choose iPhone 5S as the receiver, the guard intervals are set to 1 ms and 0.5 ms respectively, the number of subcarriers is 100 with communication distance of 10 cm and ambient noise of about 55 dB in an office.

Figure 12(a) shows the impact of symbol duration on BER. We can find that BER drops significantly after symbol duration is larger than 7 ms, because if the symbol duration increases, the energy per symbol increases. BER drops from 2.17% to the lowest value of 0.95% when the guard interval is 1 ms. Therefore, there exists a tradeoff between BER and symbol duration.

Throughput is affected by symbol duration too, since a shorter symbol duration can transmit more symbols within a certain time, so throughput enhances when symbol duration decreases. And we also noticed that the average BER is lower when the guard interval is 1 ms, because a longer guard interval can improve the signal robustness, we will discuss it later.

The Impact of Guard Interval on UltraComm. Reducing the duration of guard interval can enhance coding rate directly. However, reducing guard interval will increase BER, because the inter-symbol interference (ISI) can't be effectively controlled when the value of guard interval is too short. Choosing a longer guard interval means decreased overhead due to unnecessary idle time. Thus, a proper guard interval is crucial to optimize the relationship between BER and throughput.

As Fig. 5 shows that the rise and ringing time of demodulated signal is about 0.5 ms, then it is possible to eliminate ISI when the value of guard interval is 0.5 ms. To explore the impact of guard interval on BER, we conduct the following experiment. We choose the iPhone 5S as a receiver, 10 cm as the communication distance and alter throughput between 13.38 kbps and 16.96 kbps when symbol duration is set to 5 ms. When symbol duration is 10 ms, throughput is varied from 7.85 kbps to 8.96 kbps. The ambient noise is about 55 dB consistently in this set of experiment.

From Fig. 12(b) shows the effect of guard interval on BER. As expected, BER is higher than the others when guard interval uses a smaller value, say 0.5 ms. With the increase of guard interval, BER remains within a certain range from 2.74% to 0.9% when the data symbol duration is 5 ms, BER keeps at a lower range from 0.68% to 0.51% when the data symbol duration is 10 ms, which means the ISI is effectively suppressed by increasing guard interval and data symbol duration.

Throughput will drop with increases of guard interval, so, Low BER is at the expense of the coding rate (Throughput declines by 12.39% and 21.1% respectively). To make a tradeoff, we choose 0.75 ms as the guard interval for UltraComm.

The Impact of Subcarrier Number on UltraComm. The number of subcarriers is an important factor on the transmission performance. To evaluate the impact of the number of subcarriers on the performance of UltraComm, we conducted the experiment by changing number of subcarriers. The receiver is iPhone 5S, the data symbol durations are 5 ms and 10 ms with guard interval of 1 ms, 10 cm as the communication distance and ambient noise of about 55 dB in an office.

Figure 12(c) shows that reduction of subcarrier number leads to decrease of BER. When the number of subcarriers is 65, BER has the lowest value of 0.74%. Because reduction of subcarrier number can improve the tolerance of UltraComm against the intermodulation distortion which is the reason of crosstalk between subcarriers, then this amount of crosstalk can increase BER. Reduction of subcarrier number also increase the transmission power of each subcarrier. Thus, the SNR increases accordingly. Reduction of the number of subcarriers is a simple method to improve communication performance, but at the expense of reducing throughput, the throughput will drop with the decrease of subcarrier number.

5 Related Work

Acoustic communication can be divided into the following categories, and details can be found in Table 1.

Audible Sound Communication. This is the most common acoustic communication approach and the entire audible frequency band can be utilized. In Dhwani [14], the authors implement a sound-based near field communication (NFC), by transmitting modulated signals in audible frequency band. Zhang et al. propose PriWhisper [16], an audible communication system for secure information exchange in very short range (≤ 0.5 cm). The drawback of this category is that sounds can be heard, resulting in usability and security issues.

Inaudible Sound Communication. To address the poor user-experience issue, several mechanisms have been proposed to achieve inaudible communication by exploiting the information-hiding technique. Lope et al. [30] and Matsuoka et al. [25] design and implement a data transmission system that superposes the transmitting signal in speech or music to avoid discomforting the human ear. To better "hide" sound, more complicated schemes [17,29] have been proposed. Among which, Dolphin [17] implements a dual-channel communication which multiplexes low frequency audible music with high frequency encoded data signals without affecting the existing sound. The limitation of inaudible sound communication is the low throughput, e.g., only 500 bps for the aforementioned Dolphin. The limitation comes from the nature of the information-hiding method.

Near-ultrasonic Sound Communication. Near-ultrasonic sound ranging from 18 kHz to 22 kHz can be inaudible to most people but can still be received by microphones. Utilizing this property, Lee et al. demonstrate a system called Chirp that deliver data in indoor environment using near-ultrasonic audio chirp signal [33]. Santagati et al. [34] design and implement the U-wear communication system with GMSK modulation scheme to acquire 2.76 kbps throughput. Ka et al. [31] enable inaudible acoustic communication with low volume near-ultrasound to ensure low error rate. Ed Novak al. [3] provide a mechanism for ultrasonic proximity networking, it achieves throughput of 4.9 kbps at an ideal distance of 5–20 cm. Although inaudible, the exploitation of near-ultrasonic sound faces the problem of low throughput due to the narrow bandwidth.

Ultrasonic Communication. Real ultrasonic communication has also been studied in depth. In a proposed multi-tone FSK (MFSK) modulation-based communication system [28], ultrasound is used to transmit signals through metal. However, those solutions cannot be transplanted to mobile devices. Roy et al. [27] present a method called BackDoor, a system exploiting two separate speakers operating in ultrasonic frequency band and nonlinearity in microphone to generate audible frequency signal which can be regulated to carray data and finally achieve 4 kbps communication rates. However, FM used in BackDoor has poorer spectral efficiency than some other modulation formats, such as phase modulation and AM. Thus, it's could not achieve higher communication rate with limited bandwidth.

DolphinAttack [7] present a completely inaudible attack on speech recognition system, that modulates voice commands on ultrasonic carriers (e.g., $f > 20$ kHz) to achieve inaudibility. By leveraging the nonlinearity of the microphone circuits, the modulated voice can be successfully demodulated and interpreted by the speech recognition system.

UltraComm stands out from the aforementioned literature, which falls short in either throughput or inaudibility. In essence, UltraComm is able to transmit signals using all available bandwidth of audible sound (e.g, 0–20 kHz), by modulating the OFDM signal onto a higher carrier frequency (>20 kHz) to achieve inaudibility rather than using two separate ultrasonic signals played by two speakers. In this way, UltraComm can simultaneously achieve the two goals.

6 Conclusion

In this paper, we propose UltraComm, a high-speed and inaudible sound communication scheme for mobile devices. UltraComm modulate the coded acoustic signals on ultrasonic carriers before emitting and then the modulated acoustic signal can be recovered by utilizing the nonlinearity effect of microphone circuits and then decoded by UltraComm. We successfully realize and evaluate UltraComm on different mobile devices and achieve the highest throughput of 16.24 kbps at the premise of inaudible as far as we know.

References

1. Dog Park Software Ltd.: iSpectrum - Macintosh Audio Spectrum Analyze. Accessed May 2017
2. MarketsandMarkets. https://www.marketsandmarkets.com/Market-Reports/digital-signage-market-513.html. Accessed May 2019
3. Novak, E., Tang, Z., Li, Q.: Ultrasound proximity networking on smart mobile devices for IoT applications. IEEE Internet Things J. **6**(1), 399–409 (2018)
4. Tseng, W.-K.: A directional audible sound system using ultrasonic transducers. Int. J. Adv. Res. Artif. Intell. **4**(9) (2015)
5. Yan, C., Zhang, G., Ji, X., et al.: The feasibility of injecting inaudible voice commands to voice assistants. IEEE Trans. Dependable Secure Comput. (2019)
6. Nonlinear Acoustics. Academic Press, San Diego (1998)
7. Zhang, G., Yan, C., Ji, X., et al.: DolphinAttack: inaudible voice commands. In: Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, pp. 103–117. ACM (2017)
8. N5172B EXG X-Series RF Vector Signal Generator, 9 kHz to 6 GHz. http://www.keysight.com/en/pdx-x201910-pn-N5172B. Accessed May 2017
9. Ultrasonic Dynamic Speaker Vifa. http://www.avisoft.com/usg/vifa.htm. Accessed 5 May 2017
10. Lazik, P., Rowe, A.: Indoor pseudo-ranging of mobile devices using ultrasonic chirps. In: Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems, pp. 99–112. ACM (2012)
11. Cobham: Intermodulation Distortion. http://aeroflex.com/. Accessed 3 Oct 2018
12. Liang, C.P., Jong, J., Stark, W.E., et al.: Nonlinear amplifier effects in communications systems. IEEE Trans. Microw. Theory Tech. **47**(8), 1461–1466 (1999)
13. Ohno, S., Manasseh, E., Nakamoto, M.: Preamble and pilot symbol design for channel estimation in OFDM systems with null subcarriers. EURASIP J. Wirel. Commun. Netw. **2011**(1), 2 (2011)
14. Nandakumar, R., Chintalapudi, K.K., Padmanabhan, V., et al.: Dhwani: secure peer-to-peer acoustic NFC. In: ACM SIGCOMM Computer Communication Review, vol. 43, no. 4, pp. 63–74. ACM (2013)
15. Jinci Technologies: Open structure product review. http://www.jinci.cn/en/goods/112.html. Accessed 5 May 2017

16. Zhang, B., Zhan, Q., Chen, S., et al.: PriWhisper: enabling keyless secure acoustic communication for smartphones. IEEE Internet Things J. **1**(1), 33–45 (2014)
17. Wang, Q., Ren, K., Zhou, M., et al.: 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, pp. 29–41. ACM (2016)
18. Gerasimov, V., Bender, W.: Things that talk: using sound for device-to-device and device-to-human communication. IBM Syst. J. **39**(3.4), 530–546 (2000)
19. Hanspach, M., Goetz, M.: On covert acoustical mesh networks in air. arXiv preprint arXiv:1406.1213 (2014)
20. Wambacq, P., Sansen, W.M.C.: Distortion Analysis of Analog Integrated Circuits. Springer, Heidelberg (2013)
21. Chen, G.K.C., Whalen, J.J.: Macromodel predictions for EMI in bipolar operational amplifiers. IEEE Trans. Electromagn. Compat. **4**, 262–265 (1980)
22. Fiori, F., Crovetti, P.S.: Nonlinear effects of radio-frequency interference in operational amplifiers. IEEE Trans. Circuits Syst. I: Fundam. Theory Appl. **49**(3), 367–372 (2002)
23. Fiori, F.: A new nonlinear model of EMI-induced distortion phenomena in feedback CMOS operational amplifiers. IEEE Trans. Electromagn. Compat. **44**(4), 495–502 (2002)
24. Graffi, S., Masetti, G., Golzio, D.: New macromodels and measurements for the analysis of EMI effects in 741 op-amp circuits. IEEE Trans. Electromagn. Compat. **33**(1), 25–34 (1991)
25. Matsuoka, H., Nakashima, Y., Yoshimura, T.: Acoustic communication system using mobile terminal microphones. NTT DoCoMo Tech. J **8**(2), 2–12 (2006)
26. Kune, D.F., Backes, J., Clark, S.S., et al.: Ghost talk: mitigating EMI signal injection attacks against analog sensors. In: 2013 IEEE Symposium on Security and Privacy, pp. 145–159. IEEE (2013)
27. Roy, N., Hassanieh, H., Roy Choudhury, R.: Backdoor: making microphones hear inaudible sounds. In: Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services, pp. 2–14. ACM (2017)
28. Hosman, T., Yeary, M., Antonio, J.K., et al.: Multi-tone FSK for ultrasonic communication. In: 2010 IEEE Instrumentation & Measurement Technology Conference Proceedings, pp. 1424–1429. IEEE (2010)
29. Yun, H.S., Cho, K., Kim, N.S.: Acoustic data transmission based on modulated complex lapped transform. IEEE Signal Processing Lett. **17**(1), 67–70 (2009)
30. Lopes, C.V., Aguiar, P.M.Q.: Aerial acoustic communications. In: Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No. 01TH8575), pp. 219–222. IEEE (2001)
31. Ka, S., Kim, T.H., Ha, J.Y., et al.: Near-ultrasound communication for TV's 2nd screen services. In: Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking, pp. 42–54. ACM (2016)
32. Peng, C., Shen, G., Zhang, Y., et al.: BeepBeep: a high accuracy acoustic ranging system using cots mobile devices. In: Proceedings of the 5th International Conference on Embedded Networked Sensor Systems, pp. 1–14. ACM (2007)
33. Lee, H., Kim, T.H., Choi, J.W., et al.: Chirp signal-based aerial acoustic communication for smart devices. In: 2015 IEEE Conference on Computer Communications (INFOCOM), pp. 2407–2415. IEEE (2015)
34. Santagati, G.E., Melodia, T.: A software-defined ultrasonic networking framework for wearable devices. IEEE/ACM Trans. Netw. (TON) **25**(2), 960–973 (2017)