No Entry: Anti-Noise Energy Detector for Chirp-Based Acoustic Communication

Chulyoung Kwak†, Seongwon Kim†, Soonwon Ka‡, Jihwan Lee†, and Sunghyun Choi†
†Department of ECE and INMC, Seoul National University, ‡AI Center, SK Telecom, ◎Soundlly, Inc.
Email: cykwak@mwnl.snu.ac.kr, s1kim@sktair.com, swka@soundl.ly, jhlee16@mwnl.snu.ac.kr, schoi@snu.ac.kr

Abstract—Acoustic communication using microphones and speakers of smart devices is one of the most spotlighted wireless technologies in recent years. In particular, chirp-based acoustic communication is widely adopted for smart device applications because of its robustness to frequency selectivity. Since many chirp-based acoustic applications run in the background on mobile devices, the power consumption of chirp-based acoustic communication is a critically important issue. Using energy detectors (EDs), which determine the existence of a valid acoustic signal based on energy level, applications can reduce power consumption by working only when a valid signal exists. However, conventional ED fails to distinguish between valid signals and high-energy noise in everyday life. In this paper, we propose No Entry, a novel ED for chirp-based acoustic communication systems. No Entry avoids not only high-energy noises but also a different modulation-based acoustic signal by utilizing the frequency sweeping characteristic of chirp signals. We implement prototype Android applications to evaluate the detection accuracy and power consumption. Compared with the state-of-the-art schemes, No Entry reduces energy consumption by 30% while achieving a greater detection performance.

I. INTRODUCTION

Smart mobile devices are no longer special things to modern people. As smart devices become common, they play several roles. For example, they play the role of credit cards, coupon books, and even gaming consoles. Accordingly, various types of wireless communication system have been proposed to deal with several applications. An acoustic communication system using smart mobile devices is one of the most spotlighted wireless technologies.

Most acoustic communication systems using smart devices operate in the near-ultrasound frequency range, i.e., 18–20 kHz. The lower frequency range is avoided to prevent any unwanted audibility by humans, and the higher frequency range is limited by the sampling rate of 44.1 kHz, which is mostly common in off-the-shelf smart devices. A lot of research utilizing the advantages of the sound has been proposed. Relatively slower speed of sound has been flourishing the indoor localization and motion tracking research [1], [2]. In addition, the feature that near-ultrasound acoustic signal can be easily embedded in the music or video contents without user’s perception attracts the attention of the short-range data transmission services [3]–[6].

Acoustic communication exploits various digital modulation schemes, such as frequency-shift keying (FSK), orthogonal frequency-division multiplexing, chirp modulation, etc. In this paper, we focus on chirp signal-based acoustic communication systems. Chirp signals have a clear advantage in that they are robust to frequency selectivity [4]. Thus, chirp signals are suitable for some background applications such as second screen service [5] and motion tracking [1], which requires robust signal transmission as the top priority.

Unlike foreground applications, which a user works on in person, background applications work continuously behind the scene often without being perceived by the user. Thus, background applications pose a risk of excessive power consumption without a user perception. To be specific, in most existing acoustic-based background applications, a receiver does not have any preliminary knowledge about a transmitter, and hence, the receiver has to periodically wake up and try to receive a signal. However, applications consume considerable energy in the receiving process. Applications can reduce energy consumption by detecting a signal preferentially and working only when a valid signal exists.

Energy detector (ED) is a type of such signal detection system. A conventional energy detection method measures in-band energy, i.e., the energy conveyed in the frequency range used by the communication system, during a detection time [7]. However, if we use the conventional ED in acoustic communication, there are two problems. The first problem is caused by some types of noise, called ambient noise, which influences acoustic communication. Fig. 1 shows the power spectral density (PSD) plot of two different types of noise, i.e., a conversation sound and a cough sound. It is shown that the power level of the conversation sound decreases sharply from the frequency of about 10 kHz. On the contrary, the cough sound keeps relatively high power level up to the frequency of 20 kHz. Due to the high power level, an ED can mistake the cough sound as a valid signal. Likewise, ambient
In this paper, we propose No Entry, a novel ED for chirp-based acoustic communication systems that goes beyond existing limitations. The main idea behind No Entry is not merely to sense the in-band energy level, but to verify that the signal has the frequency sweeping characteristic of chirp signals. Since the merit of ED is computational simplicity, we design the detection algorithm not to deteriorate the simplicity. The main contributions of this paper are as follows.

1) We propose No Entry, a novel ED that can avoid not only high-energy noise but also high-energy interference by utilizing the frequency sweeping characteristic of chirp signals.
2) Detection accuracy of No Entry shows that true positive (TP) rate is more than 90% when false positive (FP) rate is 1% even with severe interference.
3) The power consumption of No Entry is measured using a prototype Android application and Monsoon power monitor. No Entry reduces energy consumption by about 30% compared with the state-of-the-art scheme.

The rest of this paper is organized as follows. In Section II and Section III, we describe related work and background, respectively. Section IV presents the overall system of the proposed ED. We discuss several parameters related to implementation and analyze the computation complexity of the proposed method in Section V. In Section VI, we evaluate the performance and conclude the paper in Section VII.

II. RELATED WORK

There have been many studies to deal with the signal detection in acoustic communication. The authors of [4], [8] use peak PSD ratio of in-band, e.g., 19.5–22 kHz frequency range, to out-of-band, e.g., 16–18 kHz frequency range. The authors assume both in-band and out-of-band have similar noise levels when the signal from their acoustic communication system is absent. In this case, peak PSD ratio becomes low because the peak PSD of the in-band and that of the out-of-band are comparable. If the signal exists in the in-band, the peak PSD of the in-band is much greater than that of the out-of-band, so that the peak PSD ratio becomes high. However, this method has limitations. If there exists ambient noise that has higher energy in the in-band than in the out-of-band, peak PSD ratio becomes high thus causing a false positive error. Peak PSD ratio also has a trouble distinguishing different acoustic signals from different communication systems which use the frequency range nearby the in-band.

The authors of [5] consider the influence of ambient noise and propose J-CS algorithm utilizing the shape of a chirp signal correlation. J-CS is able to distinguish the non-chirp signals because it uses chirp signal’s correlation results. However, even though J-CS is able to differentiate interference signals from the chirp signals, it has a limitation in terms of power consumption. To be specific, J-CS hardly reduces power consumption because it determines the existence of the signal at the end of its receiving process.

The authors of [9] propose an energy-efficient acoustic communication system. They utilize an acoustic beacon signal to determine the presence of an acoustic signal before operating the entire process. A receiver wakes up periodically and verifies whether a beacon signal exists or not during short detection time. If the receiver detects the beacon signal, it tries to receive the signal. If the receiver does not detect the beacon signal, on the other hand, it goes to sleep. However, the authors do not specify the behavior of the beacon detection process in the paper, as their main contribution is an accurate spatially-aware interaction.

III. BACKGROUND

A. Chirp Signal

A chirp signal is a signal whose frequency sweeps over time, i.e., the frequency increases or decreases with time. We exploit this frequency sweeping characteristic of chirp signals. Assume that we observe a chirp signal during a time duration shorter than symbol duration ($T_{sym}$), i.e., the detection time, notated by $t_{ED}$, is smaller than $T_{sym}$. Then, the signal sweeps not the entire in-band frequency range but a part of the in-band frequency range. We can divide the in-band frequency range into two groups. We define occupied frequency range ($W_o$) as the frequency range the signal sweeps during $t_{ED}$, and vacant frequency range ($W_v$) as the rest in-band frequency range excluding $W_o$. Fig. 2 shows an example of a chirp signal (a linearly-increasing straight line) of duration $T_{sym}$ sweeping from $f_{start}$ to $f_{end}$, along with $t_{ED}$, $W_o$, and $W_v$ as well as their relationships. Even though the ratio between $W_o$ and $W_v$ would change depending on $t_{ED}$ and chirp’s sweeping rate, each of $W_o$ and $W_v$ is continuous or circularly continuous within the in-band frequency range. Circularly continuous means the end of the in-band frequency ($f_{end}$) is followed by the beginning of the in-band frequency ($f_{start}$), as shown in Fig. 2(b). We utilize two attributes of a chirp to distinguish it from ambient noise, i.e., 1) in-band frequency range consists
of $W_o$ and $W_e$, and 2) each of $W_o$ and $W_e$ is (circularly) continuous. We will present the details in the next section.

B. Noise Analysis

One of the biggest weaknesses of acoustic communication is its vulnerability to ambient noise. Compared to radio frequency wireless technologies, acoustic communication uses relatively low frequency range around kilohertz where noise can be easily generated by human activities. The authors of [5] describe the influence of ambient noise on acoustic communication. However, they do not survey how often ambient noise is generated. If ambient noise rarely appears in everyday life, handling ambient noise in the ED might rather be an unnecessary overhead.

We investigate how often ambient noise is generated in two different environments. We collect data in two offices representing quiet places and two cafes representing loud places. We record one hour per measurement and check how often ambient noise exceeds a certain energy threshold. To set the energy threshold, we select the application in [5], which is a chirp-based background second screen service application, as our target application. In the target application, the signal’s frequency range, i.e., in-band frequency range, is 18.5–19.5 kHz. As shown in the Fig. 3, a frame consists of 368 ms-long up-chirp preamble, 40 ms-long post-preamble guard interval (GI), and 11 symbols with symbol duration ($T_{sym}$) of 96 ms each, making the overall packet duration 1463 ms [5]. We first generate and collect the chirp signals 10,000 samples in a quiet conference room environment using TV as transmitter. To specify the volume of the signal, we measure the sound pressure level and set the sound pressure level of the signal to 32 dBSPL, which is very low volume sound similar to the sound level in a quiet bedroom at night [11]. We then set the threshold as the average energy of the collected data. We measure the ratio of the number of samples exceeding the threshold to the whole samples for a detection time of 50, 100, and 200 ms. The measurement result is shown in Table I.

The result in the offices, which are relatively quiet, is less than 5% because ambient noise is rarely generated in a quiet environment. Due to a fricative sound containing high in-band energy such as a door closing sound, however, ambient noise is sometimes generated as shown in the result in Office 1. On the other hand, the result in the cafes, which are relatively loud, shows that lots of samples pass the energy threshold. In cafes, noise is generated by the coffee machine sound as well as dragging chair or desk sound. These types of noise are common and frequently generated during a busy hour, which generates lots of high-energy ambient noise in the cafe. These results show that ambient noise is prevalent in our daily life.

C. Energy Detector

Energy detectors determine the existence or absence of the signal based on a certain criterion. Generally, EDs use the energy in the in-band frequency range as a criterion. The performance of the EDs depends on how precisely they make decisions. Two terms, namely, TP rate and FP rate, are used to express the ED’s detection accuracy. The TP rate is calculated as the ratio between the number of detection events categorized as signal and the total number of actual signal detection events. The FP rate is calculated as the ratio between the number of detection events wrongly categorized as signal and the total number of actual noise events. In other words, an ideal ED should achieve 0% FP rate and 100% TP rate. The design of the ED depends on the purpose of the system. If it is important for the system to prevent the effect of noise, the optimization problem is to maximize the TP rate for a given FP rate. On the other hand, if the system considers correct detection of the signal as a top priority, the optimization problem is to minimize the FP rate for a given TP rate.

Generally, EDs save the power consumption by ignoring the noise with energy below a configured threshold. However, if the noise has higher energy than the threshold, EDs classify the noise as a valid signal which corresponds to FP. FP errors caused by high-energy noise induce additional operations to decode valid signals, thus causing unnecessary power consumption. To eliminate such unnecessary operations, if we design an ED that can detect not only low-energy noise but also high-energy ambient noise in a short time, we can take a step towards more energy-efficient acoustic communication.

D. FSK modulation

Along with chirp, an FSK modulation is also widely used in acoustic communication due to its simplicity. There are many applications providing services using FSK modulation-based acoustic communication system [12], [13]. As these FSK acoustic communication services are used in common places such as cafes or department stores, we can assume that people could often encounter the FSK signals in everyday life. As mentioned in Section I, due to limited near-ultra sound frequency range, FSK signals are likely to use nearby frequency range to the in-band frequency range. Since FSK signals can cause more fatal interference than noise, FSK signals should be regarded as a major interference in the ED of chirp-based acoustic communication.

![Fig. 3. Chirp frame structure.](image-url)
We distinguish between chirp signal and ambient noise. In case of a chirp signal, only high power, which represents high energy components exist within the in-band. If high energy component is detected in the in-band, the ED assumes the existence of a signal. Ambient noise presence of high energy components in the in-band does not always guarantee the existence of a signal. Ambient noise which checks whether there exists a high-energy component over a certain time. Therefore, it is difficult to distinguish chirp signals from FSK signals only by checking whether the power is concentrated in a specific frequency range within the in-band. We try to discriminate FSK signals from chirp signals based on whether the frequency is fixed or changed for a certain time, which is the greatest difference between FSK and chirp signals. Throughout these processes, No Entry is able to detect a valid chirp signal. The entire process is shown in Fig. 5 and detailed explanation of each process will be described in the next subsections.

B. Low-Energy Noise Filter

The first process of No Entry is low-energy noise filter which checks whether there exists a high-energy component in the in-band frequency range based on the energy level. The performance of low-energy noise filter is closely related with the detection time \( t_{ED} \). Applying short \( t_{ED} \) may result in lack of time samples, thus leading to the degradation in detection accuracy. On the other hand, to utilize the attributes of a chirp, which are mentioned in Section III.A and will be detailed in the next subsection, \( t_{ED} \) should be short enough to divide the in-band into \( W_o \) and \( W_v \). Thus, there are two requirements concerning \( t_{ED} \) in opposition to each other. We have to use a long \( t_{ED} \) to secure the detection accuracy, while the short \( t_{ED} \) is desired to observe \( W_o \) and \( W_v \). Depending on \( t_{sym} \), the required \( t_{ED} \) to meet the second requirement may be too short to securing the detection accuracy. We tackle this challenge by gathering \( N_{Rx} \) segments with \( t_{ED} \) and making them into a large chunk, which is a similar approach to short-time Fourier transform (STFT) without window overlapping. In this way, we can get a short \( t_{ED} \), which can observe \( W_v \) and \( W_o \) for each segment, and at the same time, we can get enough time samples to observe in-band energy by combining several segments.

We empirically set \( t_{ED} \) to 20 ms and set \( N_{Rx} \) to five, which makes a total detection time 100 ms. We denote each segment of the received data in the time domain as \( y_i[t] \), and its fast Fourier transform (FFT) as \( Y_i[f] \), respectively. Each \( Y_i[f] \) has \( N_{freq} \) frequency components within in-band frequency range, where \( N_{freq} \) is determined by FFT size. For example, if we set bandwidth to 1 kHz, FFT size to 1,024, and sampling rate to 44.1 kHz, the frequency resolution becomes 44, 100/1024 \( \approx \) 43 Hz, and hence, \( N_{freq} \) becomes 23 (i.e., \( N_{freq} = \lceil \frac{1000 \text{ Hz}}{43 \text{ Hz}} \rceil = 23 \)). Let \( f_k \) denote the \( k \)-th
frequency component when the nearest frequency component close to $f_{\text{start}}$ is regarded as $f_1$. The frequency components belonging to the in-band frequency range can be expressed as $f_1$ such that $f_1 \leq f_i \leq f_{N_{\text{freq}}}$, thus the in-band energy of the each segment, denoted as $E_i$, is calculated as

$$E_i = \sum_{k=1}^{N_{\text{freq}}} |Y_i[f_k]|^2, \quad i \in \{1, \ldots, N_{\text{Rx}}\}. \quad (1)$$

We compare the sum of the $N_{\text{Rx}}$ segments’ in-band energy ($E_{\text{tot}} = \sum_{i=1}^{N_{\text{Rx}}} E_i$) with a certain threshold ($E_{\text{thresh}}$). If $E_{\text{tot}}$ is greater than $E_{\text{thresh}}$, we go to the next process, assuming that there is a high-energy component in the in-band.

C. Ambient Noise Filter

To verify that the received data passing the low-energy noise filter is not ambient noise, No Entry examines the PSD of the received data. Since there are too many types of ambient noise, it is impossible to find consistent characteristics. Thus, it would be more reasonable to use the attribute of a chirp signal rather than that of ambient noise as a criterion. We focus on $W_o$ and $W_v$ as a special characteristic of the chirp. The frequency component has a different power depending on whether it belongs to either $W_o$ or $W_v$. If the frequency components overlap with $W_o$, then the power of each component is high, otherwise, the power is low. Therefore, the chirp’s frequency components are divided into two sets based on the power and location, while those of noise are not clearly divided like chirp signals. We try to find $W_o$ and $W_v$ by gathering the high power frequency components and the low power frequency components, respectively.

As mentioned in the Section III, $W_v$ is either continuous or circularly continuous depending on the detection timing. In addition, $W_o$ could also be either continuous or circularly continuous depending on the symbol combination. Since the ED performs detection process without synchronizion, some segments can contain parts of two consecutive symbols. The frequency that the signal sweeps during detection varies depending on the symbol combinations at the boundary of two symbols. Fig. 6 shows two different symbol combination examples of the binary chirp. In case of Fig. 6(a), an up-chirp is followed by a down-chirp. In this case, $W_o$ and $W_v$ are made up of continuous frequency ranges during detection time. In case of Fig. 6(b), on the other hand, an up-chirp is followed by another up-chirp. In this case, the first signal’s frequency ends at $f_{\text{end}}$ and the next signal’s frequency starts at $f_{\text{start}}$, which makes $W_o$ circularly continuous. Despite the same detection timing, the different chirp combination makes $W_o$ and $W_v$ different. In other words, we have to consider that both $W_o$ and $W_v$ could be circularly continuous.

We define a metric, named maximum peripheral-to-opposite point peripheral ratio (MOR), to reflect chirp’s $W_o$ and $W_v$. To obtain the MOR, we create two sets of frequency components, $S_{\text{max}}$ and $S_{\text{op}}$, for tracking $W_o$ and $W_v$, respectively. $S_{\text{max}}$ and $S_{\text{op}}$ are created according to the following procedure.

1) Find $f_{i_{\text{max}}}$ having the maximum power among $N_{\text{freq}}$ frequency components and add $f_{i_{\text{max}}}$ to $S_{\text{max}}$.

We set the size of the maximum set ($N_{\text{max}}$) and that of the opposite set ($N_{\text{op}}$) differently to reflect different ranges of $W_o$ and $W_v$. In order to obtain an appropriate MOR through the above procedure, the first selected frequency components, $f_{i_{\text{max}}}$ and $f_{i_{\text{op}}}$, must belong to $W_o$ and $W_v$, respectively. We set a condition that restricts $t_{\text{ED}}$ to be less than a half of $t_{\text{sym}}$. Accordingly, the maximum $W_o$ is not over BW/2 ($\approx |N_{\text{freq}}/2|$) during detection time, which means the range of the absolute value of $W_o$ from point $f_{i_{\text{max}}}$ is always less than BW/2. Therefore, we can ensure the selected $f_{i_{\text{op}}}$ is always contained in $W_v$ unless $f_{i_{\text{max}}}$ is selected out of $W_o$. Moreover, by specifying the location of the opposite frequency components, noise must have similar PSD to that of chirp not only the ratio between high power range and low power range but also the locations of each high power range and low power range. Thus, specifying the opposite frequency component has an effect of strengthening the effect of the ambient noise filter.

We can get $N_{\text{Rx}}$ MORs from the received data and decide whether the received data is a valid signal or noise based on them. However, as shown in the Fig. 3, packet has a post-preamble GI. If a segment contains a GI unfortunately, it is difficult to obtain appropriate MOR that correctly reflects the chirp characteristic. In addition, a GI can also affect one or more segments. To minimize the effect of a GI, we select $N_{\text{comb}}$ segments out of the $N_{\text{Rx}}$ segments based on the in-band energy. Mostly, the in-band energy of the segment including a GI is smaller than that of the segment including a valid signal. We select the top $N_{\text{comb}}$ segments with high in-band energy.

Fig. 6. Different chirp signal combination example.

2) Set the frequency component ($f_{i_{\text{op}}}$) that is $\lfloor N_{\text{freq}}/2 \rfloor$ away from $f_{i_{\text{max}}}$ to the opposite one, i.e., $\lfloor N_{\text{freq}}/2 \rfloor = |i_{\text{max}} - i_{\text{op}}|$, and add $f_{i_{\text{op}}}$ to $S_{\text{op}}$.

3) Compare the power of two frequency components adjacent to $S_{\text{max}}$ and add the higher power frequency component to $S_{\text{max}}$. Repeat this process until the number of frequency components in $S_{\text{max}}$ reaches $N_{\text{max}}$.

4) Compare the power of two frequency components adjacent to $S_{\text{op}}$ and add the lower power frequency component to $S_{\text{op}}$. Repeat this process until the number of frequency components in $S_{\text{op}}$ reaches $N_{\text{op}}$.

We expect that $S_{\text{max}}$ contains a part of $W_o$ and $S_{\text{op}}$ contains a part of $W_v$. MOR can be defined as follows:

$$\text{MOR}_i = \frac{N_{\text{op}} \sum_{k \in S_{\text{max}}} |Y_i[f_k]|^2}{N_{\text{max}} \sum_{k \in S_{\text{op}}} |Y_i[f_k]|^2}, \quad i \in \{1, \ldots, N_{\text{Rx}}\}. \quad (2)$$

We set the size of the maximum set ($N_{\text{max}}$) and that of the opposite set ($N_{\text{op}}$) differently to reflect different ranges of $W_o$ and $W_v$. In order to obtain an appropriate MOR through the above procedure, the first selected frequency components, $f_{i_{\text{max}}}$ and $f_{i_{\text{op}}}$, must belong to $W_o$ and $W_v$, respectively. We set a condition that restricts $t_{\text{ED}}$ to be less than a half of $t_{\text{sym}}$. Accordingly, the maximum $W_o$ is not over BW/2 ($\approx |N_{\text{freq}}/2|$) during detection time, which means the range of the absolute value of $W_o$ from point $f_{i_{\text{max}}}$ is always less than BW/2. Therefore, we can ensure the selected $f_{i_{\text{op}}}$ is always contained in $W_v$ unless $f_{i_{\text{max}}}$ is selected out of $W_o$. Moreover, by specifying the location of the opposite frequency components, noise must have similar PSD to that of chirp not only the ratio between high power range and low power range but also the locations of each high power range and low power range. Thus, specifying the opposite frequency component has an effect of strengthening the effect of the ambient noise filter.

We can get $N_{\text{Rx}}$ MORs from the received data and decide whether the received data is a valid signal or noise based on them. However, as shown in the Fig. 3, packet has a post-preamble GI. If a segment contains a GI unfortunately, it is difficult to obtain appropriate MOR that correctly reflects the chirp characteristic. In addition, a GI can also affect one or more segments. To minimize the effect of a GI, we select $N_{\text{comb}}$ segments out of the $N_{\text{Rx}}$ segments based on the in-band energy. Mostly, the in-band energy of the segment including a GI is smaller than that of the segment including a valid signal. We select the top $N_{\text{comb}}$ segments with high in-band energy.
to exclude the segments including a GI. Therefore, we decide the validity of the received data by comparing $N_{\text{comb}}$ MORs with a threshold ($\text{MOR}_{\text{thres}}$). Unless all $N_{\text{comb}}$ MORs are not larger than $\text{MOR}_{\text{thres}}$, we consider the received data ambient noise.

**D. FSK Signal Filter**

FSK signals change the frequency in a particular time interval. If we look at the power of an FSK signal in the frequency domain, the power is concentrated at the around of the instantaneous frequency. The relationship between concentrated high power frequency range and the other low power frequency range is similar to the relationship between $W_o$ and $W_v$ of chirp signals. For this reason, neither low-energy noise filter nor ambient noise filter properly discriminates FSK signals.

The biggest difference between chirp signals and FSK signals is whether the frequency is fixed or constantly changed for a certain period of time. Therefore, we try to distinguish between chirp signals and FSK signals by observing whether the frequency constantly changes or not. Because the frequency pattern of chirp signals is so varied depending on the symbol combination, it is difficult to find a method of simply confirming that the received segment is a chirp signal. Instead, we used a method that regards the received segments as an FSK signal if they have a fixed frequency during the detection time. We track the frequency of each segment by observing the maximum power frequency component ($f_{\text{max}}$).

We denote the number of pairs of consecutive segments which have the same maximum power frequency component as $Z_{\text{Rx}}$. However, if we simply track the index of the maximum power frequency component, i.e., $i_{\text{max}}$, there is a problem. Due to the limit of the frequency resolution, the maximum power frequency component can not exactly reflect the instantaneous frequency. Accordingly, despite the same instantaneous frequency, the maximum power frequency component is sometimes changed to the nearby frequency component, which fades our approach to distinguishing the signal through $i_{\text{max}}$.

To compensate this problem, we additionally utilize the second highest power frequency component denoted by $f_{2\text{nd}}$. If $i_{\text{max}}$'s of two consecutive segments are not the same, we check the index of the second highest power frequency component, i.e., $i_{2\text{nd}}$, of each segment. If $i_{\text{max}}$ of the former segment equals to $i_{2\text{nd}}$ of the latter segment and also $i_{2\text{nd}}$ of the former segment equals to $i_{\text{max}}$ of the latter segment, this mismatch is considered a transient error and the maximum power frequency components of two segments are considered equal. Finally, we compare $Z_{\text{Rx}}$ with a certain threshold ($Z_{\text{thres}}$). When the received data passes all the processes, then No Entry finally determines that there exists a chirp signal.

**V. LOOK INSIDE NO ENTRY**

This section covers the issues related with implementing No Entry. The important parameters used in No Entry are analyzed, and then the computational complexity is also discussed.

**A. Parameter Analysis and Discussion**

The performance of No Entry depends on various parameters. In this subsection, we will discuss the influence of each parameter. Since the optimization problem of each parameter is a matter of realization, we present a guideline on how to choose the parameters.

1) **Detection time:** Detection time ($t_{\text{ED}}$) is the most important parameter in that it gets involved in the entire process of No Entry directly and indirectly. $t_{\text{ED}}$ determines not only the detection performance but also the operation time. The number of time samples determines FFT size, which is closely related to the frequency resolution and the computational complexity of No Entry. The frequency resolution is associated with how accurately we can track $W_o$ and $W_v$. However, the process of getting MOR is not to find the exact positions of $W_o$ and $W_v$, but to find some frequency components belonging to the groups. Thus, we do not need an excessively fine-grained frequency resolution. We empirically find that No Entry works well with FFT size greater than or equal to 1,024. The detection time will be further discussed below in conjunction with other parameters.

2) **Set size:** Another key parameter is the set size. Two set sizes ($N_{\text{max}}, N_{\text{op}}$) play a key role in getting MOR. Getting MOR is the process of tracking $W_o$ and $W_v$. If we choose inappropriate set sizes, MOR does not reflect the $W_v$ and $W_v$ correctly. In fact, setting the set size is closely related with setting the ratio of $W_o$ and $W_v$ during the detection time, which is closely related with setting $t_{\text{ED}}$. If we shorten $t_{\text{ED}}$, the ratio of $W_v$ within the in-band increases and that of $W_v$ decreases. In this case, we should increase $N_{\text{op}}$ and reduce $N_{\text{max}}$. If we lengthen $t_{\text{ED}}$, on the same principle, we should increase $N_{\text{max}}$ and reduce $N_{\text{op}}$.

If we take a closer look at the process of getting MOR in (2), the noise of the signal is lowered due to a low value belonging to $S_{\text{max}}$ or a high value belonging to $S_{\text{op}}$. Since we divide a large value by a small value, increasing the denominator is much influential than decreasing the numerator. This means setting $N_{\text{op}}$ to a large value is more effective than setting $N_{\text{max}}$ to a small value, which is equivalent to setting a higher ratio of $W_v$ within the in-band. Eventually, getting a effective MOR value comes down to setting $t_{\text{ED}}$ to a small value. From this point of view, we can confirm that the necessary condition for detection time is not a strict condition.

**B. Parameter Selection**

Detection accuracy varies depending on various parameter values. In the target application, the length of the GI after the preamble is 40 ms. Thus, when we collect five 20 ms-long segments, the number of segments that can contain part of a GI does not exceed three. In other words, at least two 20 ms-long segments do not contain a GI at all. Therefore, we choose $N_{\text{comb}} = 2$ to be sure that the selected segments are not affected by a GI, thus minimizing the impact of a GI.

Since $W_o$ and $W_v$ are different for each received data, $N_{\text{max}}$ and $N_{\text{op}}$ should be set so as to be generally included in $W_o$ and $W_v$. Since there are infinite kinds of ambient noise, it is
impossible to get the optimized parameters for all kinds of ambient noise. However, if we set the parameters that work well over many different kinds of ambient noise, we can expect that No Entry copes with the unexpected noise. First of all, we collect noise data set by generating various types of ambient noise that we can meet frequently in our daily life. The types of ambient noise included in the noise data set are clapping, rattling, sneezing, etc. Then we find a set of the parameters that works best on the noise data set.

Fig. 7 shows an example result of Galaxy S7. The graph shows the FP rate by changing $N_{max}$ and $N_{op}$. We set $MOR_{thres}$ to 99% of the ambient noise data passing the low-energy noise filter. The optimal parameters yielding the lowest FP rate are $N_{max} = 2$ and $N_{op} = 8$, which yield 13% FP rate. The performance does not drastically change when the parameters are selected around the optimal set. However, the FP rate considerably increases if the parameters are selected far from the optimal set.

Based on the result, it can be expected that the performance will not be severely degraded unless the parameters are selected significantly far from the optimal set.

C. Computational Complexity Analysis

In this subsection, we analyze the computational complexity of No Entry. One of the reasons for using ED is its low computational complexity [7]. Since the computational complexity is closely related to the amount of the signal processing, it is important to operate with as few calculations as possible. Generally, acoustic communication system uses 16-bit pulse-code modulation (PCM) data, and hence the complexity of multiplication dominates the complexity of addition. Thus, we only consider the number of multiplications in the analysis.

We compare the complexity of No Entry with the in-band ED using the same detection time. In-band ED refers to the conventional ED to reject low-energy noise using a single segment unlike the low-energy filter of No Entry which uses multiple segments. In general, the FFT size is chosen to be a power of two. However, since No Entry uses $N_{Rx}$ segments with $t_{ED}$ so that setting the detection time to $N_{Rx}t_{ED}$ can not guarantee that FFT size of the in-band ED is equal to $N_{Rx}$ times of FFT size in No Entry. Therefore, for a reasonable comparison, we assume that FFT size of the in-band ED is equal to $N_{Rx}$ times of that of No Entry. In other words, if we denote FFT size of No Entry as $N_{FFT}$, FFT size of the in-band ED becomes $N_{Rx}N_{FFT}$. FFT takes part in the largest amount of the computational complexity and the complexity of FFT depends on FFT size. If we denote FFT size as $N_{FFT}$, FFT computation performs $\frac{1}{2}N_{FFT}\log_2 N_{FFT}$ complex multiplications. Secondly, we calculate the complexity of getting the in-band energy. Except for FFT, multiplication and division are performed only in the process of obtaining MOR throughout the process. However, the number of operation is at most $N_{Rx}$, which is negligible. Therefore, we can compare the computational complexity taking into account only the complexity involved in FFT computation. As mentioned above, the number of multiplication operations required for FFT computation is $N_{FFT}(\log_2 N_{FFT})/2$, so that the complexity of the in-band ED is larger by $N_{Rx}N_{FFT}\log_2 N_{Rx}/2$. Therefore, we conclude that the computational complexity of No Entry is much smaller than that of in-band ED, meaning that No Entry can also maintain the advantage of low computational complexity.

VI. PERFORMANCE EVALUATION

We evaluate the performance of No Entry in terms of detection accuracy and power consumption. We use Samsung Galaxy S5, S7, and LG G5, G6 as experiment devices.

A. Detection Accuracy

We compare the detection accuracy of No Entry with three different schemes. We set the in-band ED (Baseline) as a baseline scheme and set J-CS [5] and peak PSD ratio (PPR) [8] as comparison schemes.

1) Noise filtering: We evaluate the detection accuracy in two different environments with the same signal and noise data sets used in Section III. First, we evaluate the detection accuracy with the noise data set of the offices representing quiet places to verify the ability to handle low-energy noise. Second, we evaluate the detection accuracy with the noise data set of the cafes representing loud places to verify the ability to handle ambient noise. The parameters we use in the experiments are described in Table II. No Entry has several thresholds besides $E_{thres}$. Thus, we fix $E_{thres}$ to 100% TP rate and observe the accuracy by changing $MOR_{thres}$. We present the evaluation results using receiver operating characteristic (ROC) curve by changing the threshold defined in each comparison scheme. Fig. 8(a) shows the ROC curves using the noise data set of the offices. As observed in Section III, office environments have little ambient noise. Thus, all of the schemes work properly. Fig. 8(b) shows the ROC curves using the noise data set of the cafes. As expected, the baseline ED shows degraded accuracy due to ambient noise. No Entry and the other comparison schemes, i.e., J-CS and peak PSD ratio, recover the accuracy degradation through their own detection method. As mentioned in Section II, the peak PSD ratio method induces false alarms if ambient noise has a higher energy level in the in-band than that in the out-of-band, which appears in the results of the cafe data. In addition, J-CS, in the case of Galaxy S5, cannot completely distinguish ambient noise and yields degraded performance. On the other hand, No Entry restores the accuracy almost completely as it is designed to withstand ambient noise.
2) **FSK signal filtering**: We next evaluate the detection accuracy of No Entry and comparison schemes with an FSK signal. We collect the FSK signal data that is actually being used for order services in cafes [13]. Fig. 9 shows the spectrogram of the collected FSK signal. The FSK signal uses a frequency range close to the in-band of the target application, thus possibly causing the false alarms on the ED.

First of all, we verify that No Entry and comparison schemes have difficulty filtering the FSK signal. Fig. 10(a) shows the ROC curves of three comparison schemes and No Entry without FSK signal filter. It is shown that the baseline ED, the peak PSD ratio, and No Entry, which are energy based methods, hardly distinguish the FSK signal. On the other hand, since J-CS detects the chirp signal based on the correlation, it is possible to distinguish the FSK signal to some extent. However, its detection accuracy with the FSK signal is degraded compared with the result with ambient noise, i.e., Fig. 8

Fig. 10(b) shows the ROC curves of No Entry with the FSK signal filter. $Z_{thres}$ can range from zero to $N_{Rx}$ − 1, so we examine the detection accuracy by changing $Z_{thres}$ from zero to four. Even a valid chirp signal does not always have zero $Z_{thres}$, because chirp signal can have a consecutive frequency at the boundary of two symbols. Therefore, When $Z_{thres}$ is set to zero or one, the 100% TP rate is not satisfied. On the other hand, if $Z_{thres}$ is set three or four, the effect of preventing interference is reduced. When we set $Z_{thres}$ as two, No Entry shows the best performance, which outperforms J-CS.

**B. Power Consumption**

EDs can save energy by stopping the application after quickly detecting the absence of target signal. We measure power consumption to evaluate how much No Entry can save energy. The power consumption of the experiment device is measured using Monsoon power monitor [14]. We use Galaxy S5 for the measurement because the other experiment devices are equipped with built-in batteries, which make the measurement difficult. We implement three types of Android

---

**TABLE II**

**EXPERIMENT PARAMETER SETTING**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Galaxy S5</th>
<th>Galaxy S7</th>
<th>LG G5</th>
<th>LG G6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_{ED}$</td>
<td>20 ms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N_{Rx}$</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N_{FFT}$</td>
<td>1024</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N_{max}$</td>
<td>2 2 1 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N_{op}$</td>
<td>7 8 8 9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N_{comb}$</td>
<td>2 2 2 2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
applications, which adopt 1) Baseline ED 2) J-CS, and 3) No Entry, respectively. The application periodically wakes up and tries to receive the signal. Depending on the detection result, the application performs the decoding process if it determines that there exists a signal. If the application determines that there is no signal, it immediately goes to sleep and waits for the next wake-up period.

There are various components consuming power in smart devices. For example, display and wireless technologies such as Wi-Fi and Bluetooth consume a considerable amount of power. To the extent possible, thus, we shut down the other processes to measure the power consumption of our target application. Furthermore, the authors of [10] find that applications consume substantial power when applications wake up from the idle state. Therefore, we set the wake-up period 10 s and perform the detection process twice at each wake-up to reduce the overhead of the wake-up process.

Fig. 11 shows an example of measured power consumption. The results with and without No Entry are similar when the signal exists. On the other hand, if the signal is absent, No Entry detects the absence of the signal and quickly stops working. In this way, No Entry can reduce power consumption of the application. The result using the baseline ED shows similar patterns to this.

We also measure the energy consumption of three types of applications for five minutes in two different places. We perform experiments in three different cases in each place: 1) without any signal, 2) with chirp signal, and 3) with FSK signal. At each place, we place the mobile phone on the table and transmit each signal using a laptop speaker. Fig. 12 shows measurement results. For the application with J-CS, it shows consistent energy consumption regardless of the cases because it performs signal detection at the end of the system. The baseline ED reduces energy consumption by filtering low-energy noise. The reduced amount of energy consumption in the cafe without signal is less than that in the office, which shows the limited performance of the baseline ED in the presence of ambient noise. In addition, if the FSK signal exists, the baseline ED hardly saves energy. No Entry, on the other hand, shows outstanding performance in all cases. No Entry works properly in the presence of ambient noise as well as the FSK signal. Compared with J-CS, No Entry reduces the energy consumption by about 30%. Considering the energy consumption caused by wake-up process, the performance of No Entry with the target application itself would be more significant than it appears now.

VII. CONCLUDING REMARKS

In this paper, we present No Entry, a novel ED for chirp-based acoustic communication. No Entry avoids both high-energy ambient noise and high-energy interference by utilizing the sweeping characteristic of chirp signals. In particular, we show that ambient noise is prevalent in the everyday life and show that the vulnerability of the conventional ED to ambient noise. We then propose No Entry consisting of three filters, i.e., low-energy noise filter, ambient noise filter, and FSK signal filter. Our evaluation shows No Entry outperforms the comparison schemes from the perspectives of detection accuracy and power consumption. Our future work includes designing an ED, which automatically changes detection period by estimating the severity of the interference based on the detection result.

ACKNOWLEDGMENT

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) [No.2017-0-01507], Developing ultrasound communication for an extensible 2nd-screen service to make set-top’s VODs play by mobiles.

REFERENCES