SilentInformer: A Framework for Information Dispersion using Inaudible Acoustic Signals

Utkalika Satapathy*, Soumyajit Chatterjee†, Sandip Chakraborty†
*Department of CSE, IIIT Bhubaneswar, India
†Department of CSE, IIT Kharagpur, India

***Abstract*—Disbursing information in real-life noisy environments is challenging. The problem gets further compounded when the users of the system are aged or have sensory impairements. We, in this paper, develop a system called *SilentInformer*, for advanced information sharing ,over smartphones, by exploiting inaudible acoustic signals. The results depict the potential of the system by achieving a minimum bit error rate (BER) $\leq 10\%$ with message length ≤ 4 symbols and an average BER $\leq 30\%$ with a message length ≤ 8 symbols, from a distance of 27ft in realistic outdoor conditions.**

***Index Terms*—Acoustic Communications, Mobile Computing**

I. Introduction

Information sharing in noisy environments have always been challenging. Be it locating a display board in a busy airport or an audio announcement for arrival and departure of trains in a crowded railway station. These situations get further aggravated when the persons concerned are aged or have sensory impairments related to hearing and sight. We, in this paper, try to solve this problem by exploiting the capabilities of modern-day mobiles, which are equipped with more than one microphone and can record acoustic signals in the range of 50–20000Hz. This makes the mobiles capable of listening to sounds that are well beyond the range of human audibility. Leveraging on these inaudible frequencies, which are less contaminated in comparison to audible sounds [1], one can embed information that can be received by these smart devices and can be reported after processing them accordingly. This processed information can then be displayed on their smartphones or notified in the form of alerts to assist the users.

Although the design of such a system is challenging, we identify the following requirements that can allow us to embed information in inaudible sounds. Firstly, the message should be interpolated with the original audio, also known as the *host signal*(HS), in real-time with low or no processing overhead. Secondly, the message should be retrievable on any receptor that can record the audio, including mobile devices. Thirdly, the transfer of messages should be resilient to the background noise. Finally, the message, transmitted in the form of a *message signal* (MS), should not degrade the quality of the HS. Recently, attempts have been made to use inaudible frequencies [2]–[4] to communicate over COTS smartphones. But most of these either not noise resilient [5] or require additional computation for embedding.

The objective of this poster is to develop *SilentInformer*, a system to broadcast information in the form of embedded messages over audio signals in real-time. To perform this real-time embedding of messages in the HS, we adopt the technique of *sonic watermarking* [6], which uses the concept of mixing the HS with the MS in-the-air without employing any sophisticated computations (Section II). In this context, we mitigate the problem of delay between HS and MS, by encapsulating the message as a payload in a packet format with a preamble. Once transmitted it can be decoded and retrieved at the receiver end [7]. We further apply techniques multi-layer error correction techniques along with suitable modulations to ensure the noise-resilient transmission of the MS. Thorough testings of *SilentInformer* over different realistic environments indicate the potential of the system (Section III).

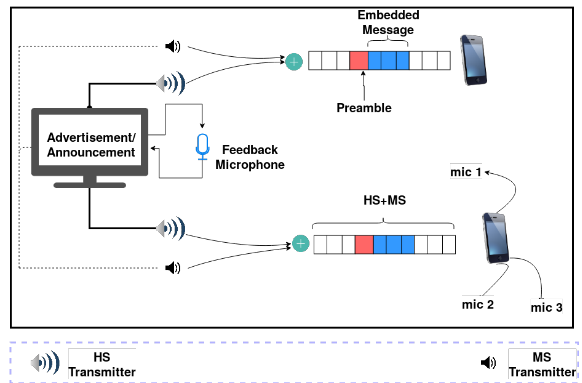
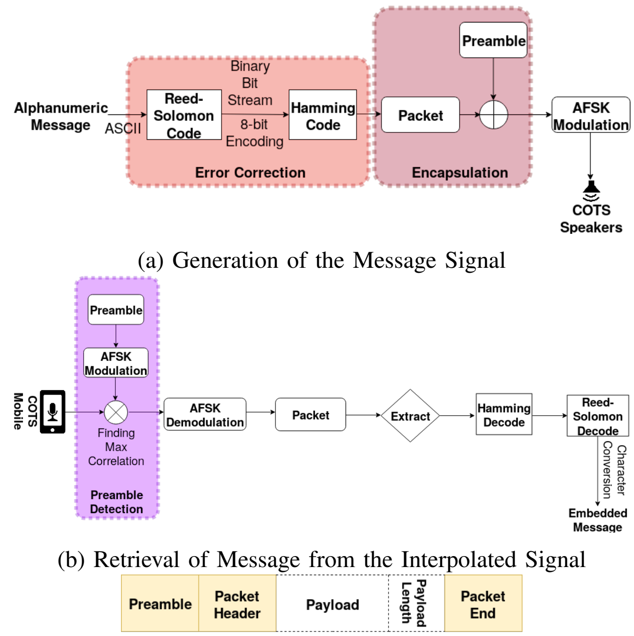


Fig. 1: Typical Setup of SilentInformer: Mic1, 2 and 3 are the possible positions of microphones in different COTS mobiles.

II. SilentInformer: System Design

A typical setup of *SilentInformer* is shown in Fig. 1, where one or more secondary speakers are placed along with the primary speakers (the speakers that broadcast the HS) to generate and broadcast the message signal (MS). The HS and the MS generated from different speakers are mixed in-the-air for producing the combined audio. The secondary speakers for broadcasting the MS can be any COTS speakers with signal-to-noise ratio (SNR) $\geq$ 80dB.

SilentInformer has two primary modules – (a) generation of the message signal using the inaudible range of audio signals and interpolating it with the HS from the live audio, and (b) retrieval of the message from the received signal (HS+MS) recorded using a mobile. Fig. 2 shows various functional components for the above two modules. The details of the two modules follow next.



(a) Generation of the Message Signal

(b) Retrieval of Message from the Interpolated Signal

(c) Packet format

Fig. 2: Functional Components of SilentInformer

A. *Message Signal Generation*

The message signal can be any predefined alphanumeric digital content as the message. This message is embedded with the original audio by mixing the MS with the HS in the air. To do this, our first task is to generate the MS from the predefined message such that the signal can be decoded from the live audio recorded by a COTS mobile phone. Let us assume that the message contains k symbols (alphanumeric digits). We need to ensure two things – (a) these k symbols need to be transmitted to the receiver mobile over an error-prone acoustic channel in such a way so that any error during the transmission can be detected and mitigated efficiently and (b) the message needs to be detectable over a mixed signal. To ensure the first requirement, we use a two-level error correcting code to mitigate the symbol-level errors as well as bit-level errors. To detect the message symbols from a mixed signal, we use the traditional communication technique by using a message boundary through preambles and apply suitable modulation technique to generate the MS. The details follow next.

1. Error Correction The MS needs to be transmitted over an error-prone acoustic channel, in the presence of background noise and the device level processing noise due to A/D and D/A conversions. To make the scheme noise-resilient, we employ a two-level error correction scheme. In the first level, we mitigate the symbol level errors by using *Reed-Solomon* (RS) codes. This is further followed by another layer of *4-bit Hamming code* to correct the granular bit level errors.

At the symbol level, say we want to make the system noise-resilient for errors up to s symbols. Then, we apply RS(n,k) code to convert k symbols of the alphanumeric message into a codeword of n symbols, where $2s = n - k$. This introduces symbol-level redundancies. At the next level, we encode each of the n symbols (both message symbols and redundant symbols from RS code) into binary using $m = 8$ bits. This binary string, of length $n \times m$, is then passed to the *4-bit Hamming encoder*, which helps in correcting a large number of errors by introducing parity bits for every 4-bit window. This final codeword is then passed to the next level for the generation of the MS after introducing the packet boundaries.

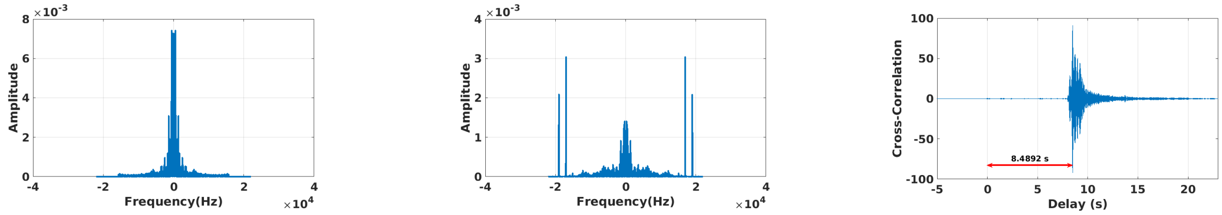
2. Encapsulation In *SilentInformer*, the MS is mixed with the HS in-the-air; therefore, we need to develop a mechanism for detecting the message text inside the MS, which requires explicit knowledge of the beginning of the message in the mixed signal. This is challenging because the HS and the MS are generated from different speakers; hence, there would be a time delay between the two signals. To mitigate the problems associated with this delay and detection of the embedded message in the received signal, we encapsulate the message bits into a suitable packet format. As shown in Fig. 2c, the packet starts with a preamble header of 6-bits, followed by a 3-bit packet header. The binary data obtained from the error correction block then forms the actual payload followed by the its length encoded in 8-bit binary format. Finally, enclosed by a 3-bit packet end sequence.

3. Modulation Finally, we use the modulation technique to generate the MS from the predefined message text. Modulation is the process of carrying a message signal inside another carrier signal that can be physically transmitted. For this application of transmitting messages over error-prone acoustic channel, we choose *Audio Frequency Shift Keying* (AFSK) modulation, in which the data is represented by the variations in the frequency of the audio.The main reason behind choosing AFSK is its high resilience to noises generated by human speech and interference [2]. In the proposed scheme, the packetized payload is encoded using the standard AFSK modulation, with a frequency spectrum of 17-19kHz and a sampling rate of 44.1kHz. We have used 16-bit encoding for AFSK to support recording using most Android apps. This finally converts the alphanumeric message to a digital signal ready for transmission through a COTS speaker.

B. *Retrieval of the Message*

The mixed signal (HS+MS) is recorded through the microphones of a mobile phone; the information can be extracted from the HS through the detection of the message. The frequency analysis of a typical HS before and after interpolation (in-the-air) with the message signal is shown in Fig. 3. From the figure, we can observe that the message is embedded at the higher frequencies (two peaks appear at the higher-end of the frequencies) without disturbing the original HS. The detailed methodology of extracting the message is given as follows.

1. Preamble Detection in the Interpolated Signal One of the prime concerns that arise because of the dependence on the technique of *sonic watermarking* is the problem of intermittent delay between the MS and HS. The presence of this delay makes the detection of the message in the finally mixed signal more challenging. In case of the classical communication-based approaches, envelop detection [3] is one



(a) At the Source (b) After Interpolation with MS (c) Detection of the preamble

Fig. 3: Frequency Analysis of the HS and MS

of the most preferred mechanisms, albeit the technique fails in the presence of noise and is highly susceptible to distortions[1].

To mitigate this concern, we apply a signal-correlation based technique to find out the probable location of the preamble in the mixed signal [7]. For this at the receiver side, we first apply AFSK modulation over the preamble text to generate a signal component of the preamble text. Then, we perform a cross-correlation between the modulated preamble and the received recording of the mixed signal [8], [9]. The point of maximum correlation denotes the most probable location of the message in the received signal. In the example shown in Fig. 3c, the preamble is detected within the combined signal with the maximum similarity at a delay of 8.4982s.

2. *Obtaining the Embedded Message* Once the preamble is detected, the actual MS is obtained by demodulating the recorded signal from the start of the preamble. Once this is done, the payload is pointed out from the demodulated signal using the packet header and ending sequences. The payload is then passed to the Hamming decoder for the first level of error correction. Once this is done, we obtain the ASCII-coded symbols by clubbing 8 binary bits into integers. This forms the n symbol string, which then serves as the input to the RS decoder which then removes the symbol level errors and generates the output of k symbols in ASCII format. Finally, these ASCII coded integers are converted to characters for obtaining the actual alphanumeric message.

III. Evaluation

We test the feasibility and performance of the proposed system as follows.

A. Implementation

For the transmission of the MS, we choose JBL Micro II (max volume = 80 dB at 1ft). As the receiving device, we use COTS mobile phone models – Redmi Note 4, Samsung Galaxy Note 5, Moto G2 and G2s, running Android OS. Finally, for the entire *Generation and Detection of the Message*, we have implemented the prototype of *SilentInformer* in Matlab R2017a, running on a system with 2.9 GHz CPU and 16 GB.

B. Performance Evaluation

To report the performance of the system, we use the metric *Bit Error Rate* (BER) represented in percentages. For this, we calculate the total number of bits as $t \times m \times k$, where t is the number of samples for each experimental instance, m is the number of bits used for encoding a symbol in binary, and k is the number of symbols. Throughout the experiment, we choose the value of $t = 5$ and $m = 8$. Also, to emulate real-life conditions, all the experiments are done in the presence of an HS, played using JBL Flip3 speakers (max volume 90 dB at 1ft). We have performed the experiments in diverse environments like a lab (sound level: 36dB), seminar hall (sound level: 35dB) and an outdoor arena (sound level: 40dB).

1. *Impact of Distance from the MS* As sound intensity decreases by the square of the distance[2], it becomes more prone to noise with increasing distance. We test the resilience of the scheme by varying the distance from 1-5ft in different environments. As shown in Fig. 4a, *SilentInformer* performs well with an average BER $\leq 20\%$. On further investigation in the outdoor environment, we see that the message signal can be transmitted to more than 27ft ($\approx$ 8m).

2. *Impact of the Length of the Embedded Message* To test the resilience of the system; we vary the length of the message from 1-10 symbols in both indoor and outdoor environments. As shown in Fig. 4b the proposed system achieves a significantly low BER ($\leq 20\%$) even with $k = 8$ and distance 27ft. Besides, this is very comforting for us to know that the complete embedded message is retrieved in at least 1 out of the 5 samples even for high lengths as well, irrespective of the environmental condition. To achieve such a low BER

k	n	MS (in secs)	k	n	MS (in secs)
4	12	11	8	24	22
5	15	14	9	27	24
6	18	17	10	30	27
7	21	19	-	-	-

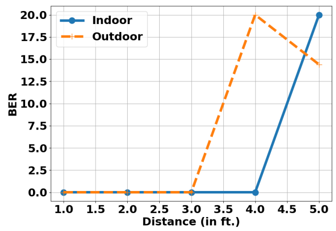
TABLE I: Combinations of **n** and **k** used for RS code

for lengthy messages, we typically increase the redundancies and parity bits for error correction as shown in Table I. This increase in the overall payload size and triggers the increase in expected generation and detection time as shown in Fig. 4c.

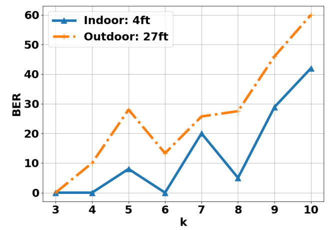
3. *Impact of Compression and Multihop Recording* To evaluate the impact of compression, we introduce the scenario of multihop recording. In this scenario, each user records the audio from the mobile of the user in the previous hop. This experiment helps us in investigating three important characteristics about the system developed - (a) whether the

[1]https://en.wikipedia.org/wiki/Envelope_detector Last Accessed: November 4, 2019

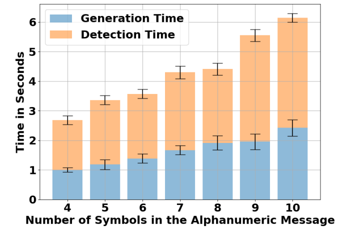
[2]https://en.wikipedia.org/wiki/Sound_intensity. Last Accessed: October 3, 2018



(a) Impact of distance

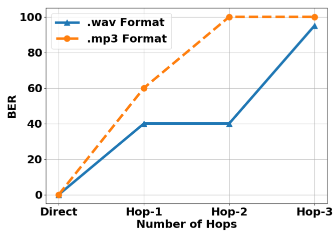


(b) Impact of message length

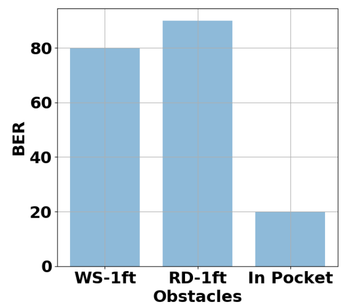


(c) Impact of message length on generation and detection time

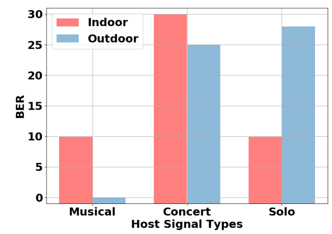
Fig. 4: Evaluation of the system



(a) Impact of multihop recording



(b) With obstacles. RD: Recording Device.



(c) Impact of type of host signal

Fig. 5: Evaluation of the system with respect to various other factors

system is resilient to multiple A/D and D/A conversions, (b) whether the compression of the audio files have an impact on the systems performance and (c) if the information can be exchanged even from a pre-recorded source.

From the results, shown in Fig. 5a, it is clear that for direct recording, there is no adverse impact on the detection of the embedded message irrespective of the formats. This proves that *SilentInformer* is MP3 compression resilient, albeit as MP3 format employs a lossy compression technique, the quality of the original audio and the embedded information degrades rapidly in case of multihop recording unlike WAV.

4. Impact of the Obstacles We choose some of the practical obstacles like human and clothing material that may appear between the MS and RD. For this experiment, we positioned the obstacles in such a way that the line of sight (LoS) between the source of the MS and the recording mobile is blocked entirely. As shown in Fig. 5b, *SilentInformer* performs well even with obstructions; albeit with human as an obstacle, the BER increases rapidly because of low chances of diffraction with an obstacle in the close vicinity.

5. Impact of the Environment The type of HS has a significant contribution to the kind of background noise generated. We experiment the system for this by choosing three different kinds of HS signals - (a) musical advertisements, (b) cacophonic concert audios and (c) solo speeches. As shown in Fig 5c, the *SilentInformer* performs well, with BER $\leq 30\%$ irrespective of the environment.

IV. Conclusion

In this paper, we develop the system *SilentInformer* for broadcasting information leveraging on the inaudible range of audio signals from 17-19 KHz. The usage of this frequency range does not obstruct the normal hearing of the humans but are still audible to microphones embedded in COTS smartphones. The system uses this as an advantage to share information through these higher frequencies especially in cases where the external noise or some physical inability is barring the user from receiving the information in the actual form. Evaluations show the potential of *SilentInformer* for sharing information in challenging real-life conditions.

References

[1] H. Watanabe and T. Terada, "Improving ultrasound-based gesture recognition using a partially shielded single microphone," in *ISWC '18*. ACM, 2018, pp. 9–16.

[2] A. Nittala, X. Yang, S. Bateman, E. Sharlin, and S. Greenberg, "Phoneear: interactions for mobile devices that hear high-frequency sound-encoded data," in *EICS*. ACM, 2015, pp. 174–179.

[3] Q. Wang, K. Ren, M. Zhou, T. Lei, D. Koutsonikolas, and L. Su, "Messages behind the sound: real-time hidden acoustic signal capture with smartphones," in *MobiCom*. ACM, 2016, pp. 29–41.

[4] M. Eichelberger, S. Tanner, G. Voirol, and R. Wattenhofer, "Receiving data hidden in music," in *HotMobile '19*. ACM, 2019, pp. 33–38.

[5] N. Roy, H. Hassanieh, and R. Roy Choudhury, "Backdoor: Making microphones hear inaudible sounds," in *MobiSys*. ACM, 2017, pp. 2–14.

[6] R. Tachibana, "Audio watermarking for live performance," in *Security and Watermarking of Multimedia Contents V*. SPIE, 2003, pp. 32–44.

[7] M. Nabeel, B. Bloessl, and F. Dressler, "Selective signal sample forwarding for receive diversity in energy-constrained sensor networks," in *2017 IEEE International Conference on Communications (ICC)*, 2017, pp. 1–6.

[8] (2012) A High-Performance Sound-Card AX.25 Modem-Sivan Toledo. [Online]. Available: https://sivantoledotech.wordpress.com/2012/07/18/ax25-modem/

[9] (2017) AFSK - EECS: www-inst.eecs.berkeley.edu. [Online]. Available: https://inst.eecs.berkeley.edu/\%7Eee123/sp17/lab/lab5/Lab5_Part_2-Audio_Frequency_Shift_Keying(AFSK).html