Delivery of QR Codes to Cellular Phones through Data Embedding in Audio

Itai Dagan, Gal Binyamin, Alon Eilam
Signal and Image Processing Laboratory (SIPL)
Andrew and Erna Viterbi Faculty of Electrical Engineering, Technion – Israel Institute of Technology
Technion City, Haifa, Israel, http://sipl.technion.ac.il/

***Abstract* – We describe a system that delivers a website address to a cellular phone by encoding inaudible binary data in an analogue audio signal, which is received by the microphone of the cellular phone. This is an alternative to encoding a web site address in a QR code label, which is scanned by the cellular phones camera.**

Data embedding in the audio signal is done by modifying the phase of the signal's modulated complex lapped transform (MCLT) coefficients, while the perceived quality of the embedded audio signal remains the same as that of the original audio signal.

A whole system was implemented and tested both in simulation and in reality. The data rate achieved in reality at a distance of 1.5 meters was 48 bits per second. This rate is sufficient to deliver the same amount of information contained in a QR code label faster and easier than by camera scan.

Keywords – Acoustic communication, acoustic data transmission, data hiding, modulated complex lapped transform, MCLT.

I. Introduction

QR (Quick Response) codes were invented in 1994 by Denso Wave, a Toyota subsidiary, as a way to track vehicles in assembly lines and to scan components at high speed. In cellular applications, QR codes have become widely used since they provide a way to access websites more quickly than by manually entering a URL. Typically, a smartphone is used as a QR code scanner, bringing the user to a website once the scan is completed. Scanning a QR code label requires aiming the smartphone with a steady hand squarely at the code and not at an angle. It is advised to minimize any glare from stray light sources, and scan the code in an evenly lit environment. These constraints imply that only a single user may scan a QR code label at a time.

We describe here a method to transmit data to a cellular phone over an acoustic channel, using a loudspeaker in the transmitter, and a cellular phone microphone in the receiver. While using this method, which is referred herein as acoustic QR, there is no need to aim a camera and the usage is as simple as launching an application. Additionally, several users may receive an acoustic QR code in parallel.

With acoustic QR, data is embedded in an audio signal, yet it is inaudible to a human listener. To provide for media streaming of the audio, it should survive MPEG Layer III audio encoding (MP3) standard, so that that data may be successfully extracted from the received audio.

In earlier works, data embedding methods based on ASK and FSK modulations of audio signals were proposed [1]. The downside of these methods is that the embedded audio signal contains components, which are annoying to a human listener. Other methods that transform consecutive segments of the audio signal to the frequency domain, perform data embedding and transform back to the time domain [2][3] introduce, in addition to annoying audio components, blocking artifacts at the audio segments boundaries.

We are applying Modulated Complex Lapped Transform (MCLT) [4] as the basis for data embedding. Data encoding is done by phase modulation of selected frequency bins in the transformed audio. Since the human hearing system is less sensitive to the phase of an audio signal, moderate phase modulations of several frequency bins is hardly noticeable. Blocking artifacts are mitigated by the segments overlap property of the MCLT transform.

A practical advantage of using MCLT is that fast algorithms exist to implement it [5], since it can be applied using Discrete Fourier Transform (DFT). This shall provide efficient real-time implementation of the receiver in a cellular phone.

An earlier work used MCLT phase modulation to deliver embedded data in audio over an acoustic channel [6]. In this paper we demonstrate a comprehensive system solution, suitable for cellular phones implementation. Performance measurements are provided as well, both in simulation and in reality.

II. Data Embedding

An MCLT complex frame of size M is generated from transforming input signal frames of size 2M samples each. Let the vector $\mathbf{x}_i$ in (1) be the i^{th} input frame:

$$\mathbf{x}_i = \left[x(iM), x(iM+1), \ldots, x(iM+2M-1) \right]^T \quad (1)$$

As provided in [4], the transformed vector $\mathbf{X}_i$ is computed by

$$\mathbf{X}_i = (\mathbf{C} - j\mathbf{S})\,\mathbf{W}\,\mathbf{x}_i \quad (2)$$

Where $\mathbf{C}, \mathbf{S}$ are $M \times 2M$ cosine and sine transform matrices and $\mathbf{W}$ is a $2M \times 2M$ window matrix that avoids the blocking artefacts:

$$(\mathbf{W})_{nn} = -\sin\left[\left(n+\frac{1}{2}\right)\frac{\pi}{2M}\right] \quad (3)$$

Embedding of binary data frames is done by modulating the phase of a subset of the components of the vector $\mathbf{X}_i$ indexed by k, as depicted in equation 4.

$$\hat{X}_i(k) = \left|X_i(k)\right| \cdot e^{jd_i(k)} \quad (4)$$

The values of $d_i(k) \in \{0, \pi\}$. Zero phase encodes a logic '1' and π phase encodes a logic '0'. An illustration of a frame with length M=512 is shown in Fig. 1.

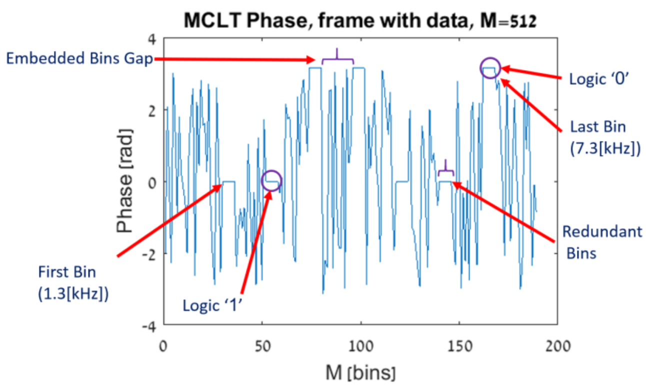

Figure 1: Phase of data embedded frame

The embedding parameters, such as the phase modulated bins locations and the repetition value, which is defined as the number of consecutive bins modified in the same manner, are configurable. These parameters determine the data rate vs. immunity to ambient noise. It is also apparent that spreading the modulated bins over a wider frequency range improves noise immunity.

Following embedding of data in the frequency domain, an inverse transform [4] is performed in order to reconstruct a data embedded audio signal in the time domain. At this stage the signal is ready for transmission by a loudspeaker to a cellular phone over an acoustic channel.

III. System Overview

Data is embedded in an audio signal by the encoder as depicted in Fig. 2. Prior to embedding the data, Reed-Solomon error correction code as well as checksum are appended to provide the receiver means to verify data integrity.

Additionally, a start byte and an end byte are added before the first data frame (start delimiter) and after the last data frame (end delimiter) for synchronization purposes.

At the receiver the audio signal is picked by the cellular phone's microphone, as shown in Fig. 3.

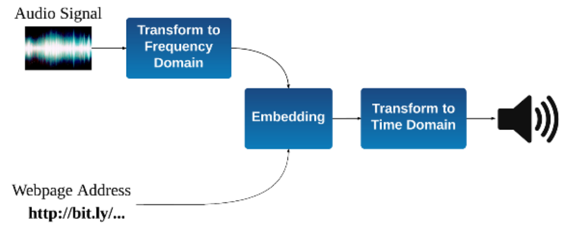

Figure 2: Encoder block diagram

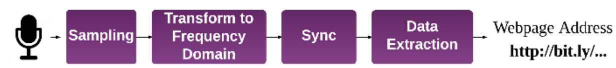

Figure 3: Decoder block diagram

The received signal is sampled and MCLT transformed to the frequency domain. In order to synchronize the decoder with the received data stream, consecutive iterations are performed until matching is achieved with the start delimiter.

Once the receiver is synchronized, data is extracted, Reed-Solomon error correction is performed and the message is delivered to the user.

IV. Performance Evaluation

Evaluation of system performance was done by embedding 21 ASCII characters in various audio signals. The effect of MPEG Layer III audio encoding (MP3) standard was evaluated by compressing the audio signal at 192 Kbits per second, and decompressing to reconstruct the audio signal. We have applied a sampling rate of 44.1 kHz and an MCLT block size of M = 512.

Several tests were performed using a room model simulation, based on [7]. The simulated room dimensions were 5m length, 6m width and 7m height. Reverberation time was set to 0.35 seconds.

Performance was evaluated with two signal types: white Gaussian noise and music, and for various distance from loudspeaker to microphone.

Testing with embedded data in white Gaussian noise signal provides an estimation of the upper limit of the receiver's performance under various conditions, since there is high probability that the transmitted audio signal will contain significant energy levels in all frequency bins. This shall improve the immunity to additive echo and ambient noise in the acoustic path from the loudspeaker to the microphone.

Fig. 4 depicts the message reception success rate for data embedded in white Gaussian noise. The success rate was near 100% up to a distance of 1.5 meters. It is apparent that using Reed-Solomon error correction codes greatly improved reception success rate.

Music tracks were used as another type of signal in the room model test. Twenty tracks of various genres were embedded with data that included Reed-Solomon error correction codes. The test was performed with several redundancy values, where redundancy value is defined as the number of repetitions of the entire data sequence. Success rates were high up to a distance of 1.5 meters as depicted in Fig. 5.

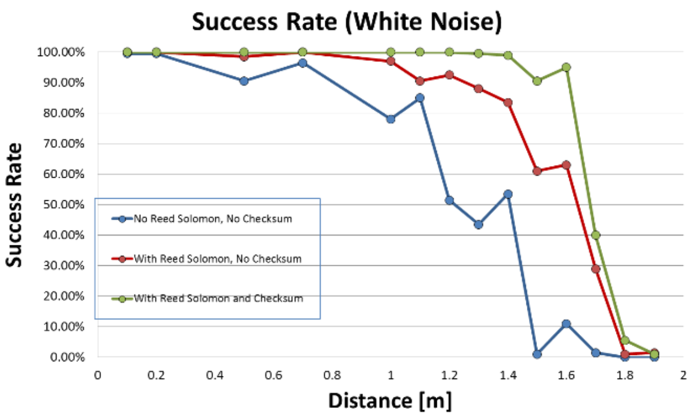

Figure 4: Message reception success rate for embedded data in white Gaussian noise

.

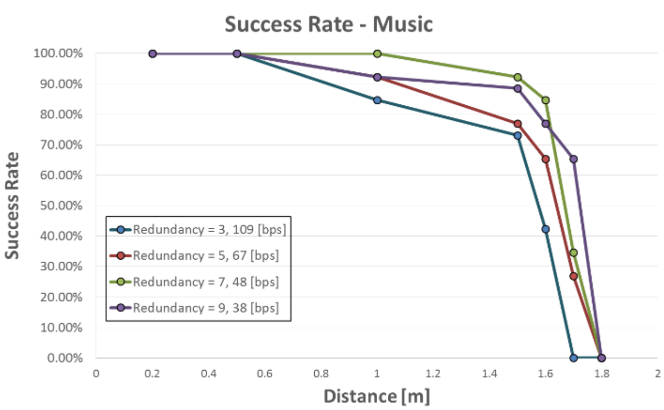

Figure 5: Message reception success rate for embedded data in music

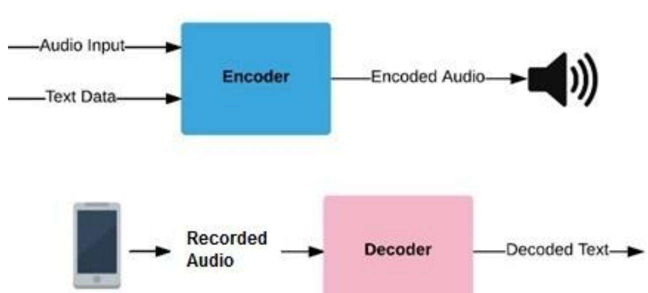

Figure 6: Real environment performance test setup

Following the tests with a room model, performance in a real environment was evaluated by using a computer loudspeaker as an audio source and a cellular phone to pick and record the received audio, as depicted in Fig. 6. Room dimensions were 4m length, 5m width and 3m height. The distance from the loudspeaker to the cellular phone in this test was 1.5m.

At data rate of 48 bits per second, message reception success rate was above 90%, as predicted by the room model simulation.

V. CONCLUSIONS

The motivation for this work was to develop an alternative to optical QR, in order to deliver a web site address to a cellular phone application over an acoustic channel. It is about to provide ease of use by eliminating the need to aim the cellular phone camera at an object, usage in any lighting condition and support of multiple users in parallel. The system has met the design goals, demonstrating good performance in real environment at distances up to 1.5 meters, delivering 48 bits per second and surviving MPEG Layer III audio encoding (MP3) standard. Adding Reed-Solomon error correction codes to the data sequence have improved the performance significantly.

The system was developed and tested with the aid of a room simulation, which has proven to provide a reliable model for real environment operation.

The algorithm implementation was done in MATLAB on a computer; to complete the development, the decoder should be implemented in a cellular phone application running in real time.

ACKNOWLEDGMENTS

The authors are grateful to Benny Saban and to Prontoly Ltd for their kind support of this work, and for exploring various use cases of its implementation. The authors would also like to thank Prof. David Malah, head of SIPL, and Nimrod Peleg, chief engineer of SIPL, for their support, advice and helpful comments.

REFERENCES

[1] Lopes, Cristina Videira, and Pedro MQ Aguiar. "Aerial acoustic communications." Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the. IEEE, 2001.

[2] Lazic, Nevena, and Parham Aarabi. "Communication over an acoustic channel using data hiding techniques." IEEE transactions on multimedia 8.5 (2006): 918-924.

[3] W. Bender, W. Butera, D. Gruhl, R. Hwang, F. J. Paiz, S. Pogreb, "Techniques for data hiding", IBM Systems Journal, Volume 39 , Issue 3-4, July 2000, pp. 547 – 568.

[4] Malvar, Henrique S. "A modulated complex lapped transform and its applications to audio processing." Acoustics, Speech, and Signal Processing, 1999. Proceedings, 1999 IEEE International Conference on. Vol. 3. IEEE, 1999.

[5] Malvar, Henrique S. "Fast algorithm for the modulated complex lapped transform." IEEE Signal Processing Letters 10.1 (2003): 8-10.

[6] Yun, Hwan Sik, Kiho Cho, and Nam Soo Kim. "Acoustic data transmission based on modulated complex lapped transform." IEEE Signal Processing Letters 17.1 (2010): 67-70.

[7] Habets, Emanuel AP. "Room impulse response generator." Technische Universiteit Eindhoven, Tech. Rep 2.2.4 (2006): 1.