# PUSHING INFORMATION OVER ACOUSTIC CHANNELS

*Po-Wei Chen* *Chun-Hsiang Huang* *Yun-Chung Shen* *Ja-Ling Wu*

Department of Computer Science and Information Engineering,
National Taiwan University, Taipei, Taiwan
E-mail: {doublech, bh, cazindo, wjl}@cmlab.csie.ntu.edu.tw

## ABSTRACT

We propose a unidirectional communication system capable of pushing information from audio speakers to nearby audience via acoustic channels. Though constrained by short transmission distances and low data throughput of sound waves, information push services based on acoustic channels can be regarded as a cost-free message delivery scheme since music playback systems and handheld devices equipped with recording capabilities have become ubiquitous in our life. Furthermore, the proposed approach can provide better non-intrusiveness and more convenient user access than other auxiliary information delivery approaches like 2-D barcodes or watermarking schemes based on visual contents. To the best of our knowledge, our system is the first one that can effectively push information to receivers at feasible transmission throughput/robustness/distance within common noisy environments. Various interesting applications and new business models can be facilitated with the proposed scheme.

***Index Terms***— Push technologies, acoustic channels, data hiding, error control codes.

## 1. INTRODUCTION

In this paper, we demonstrate feasible implementations for a novel push methodology, in which information is pushed from audio source equipped with speakers to nearby devices equipped with microphones. This approach is firstly denoted as communications over acoustic channels in [1]. In fact, pushed information is imperceptibly carried by audio sounds. Due to the characteristics of sound waves, maximal transmission distance of this push service is limited to meters. Furthermore, this approach can only provide one-way channels unless other communication channels are involved. However, even with these constraints, this technology is still quite suitable for pushing location-aware/real-time information to nearby users in a cost-effective manner without the need to interrupt normal music playback or deploy additional communication-specific devices.

Currently, several applications for information delivery over acoustic channels have been proposed for sonic/audio watermarking schemes. For example, delivering flight schedules in airports, providing discount information in shopping malls, as well as broadcasting data for positioning within local areas, are potential usages suggested by [1]. Obviously, the performance is infeasible since users must wait for a long period to receive a short and erroneous message. In [2], copyright information is embedded into audio sequence of live performance to deter bootleg copies. In [3], sounds broadcasted from multiple speakers carry respective localization information so that the audience who illegally records the live performance can be traced back as long as ticketing information is available. Since the purpose of audio data embedding systems proposed in [4, 5], denoted as sonic watermarking, is copyright protection, data capacity is not regarded as a very important criterion. In [4], information delivery systems over acoustic models are proposed to provide real-time information to car passengers without networking capabilities. Although the achieved data capacity is high, however, the reported bit error rate is too high (about 20%) to be feasible for real applications. In addition to these schemes, conventional audio watermarking schemes, such as [5, 6] possess high complexities in system design and synchronization mechanisms and thus eliminate the possibility to directly apply them to acoustic channel-based communication systems.

In this paper, we illustrate feasible implementations to facilitate short-distance push of location-aware and real-time information like advertisement or news. Fig. 1 illustrates the typical application scenario. Within the range that broadcasted audio can be heard, users who carry an adequate receiver, e.g. a widely available mobile phone with common recording and computing capabilities, can correctly receive pushed information. From the perspective of mobile-device users, received information can add value to mobile devices without the need to pay for additional short-distance communication device like Bluetooth/infra-red ports or even expensive communication fees. From the viewpoint of auxiliary information provider, free acoustic channels can provide new business opportunities without the need to interrupt normal music playback or purchase additional information broadcast devices.

The remaining parts of this paper are organized as follows. Section 2 illustrates the requirements and overall architecture of the proposed system. Section 3 discusses the implementation details of all important function blocks. Performance of the proposed scheme is proved in Section 4 via extensive experimental results. Section 5 discusses several interesting applications and the current limitations of our scheme. Conclusions and future research directions are provided in Section 6.

## 2. REQUIREMENTS AND THE SYSTEM ARCHITECTURE

In order to effectively push information to users via music delivered over acoustic models, the proposed system must address the following significant issues:

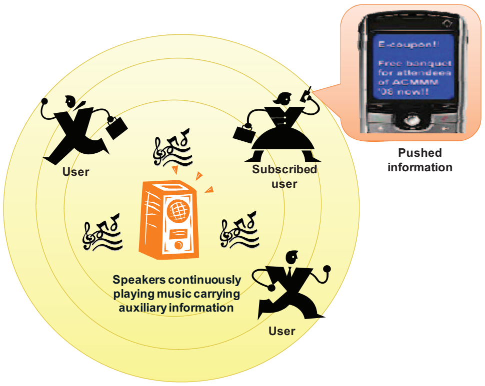


**Figure 1. A scenario illustrating the proposed information push technology.**

(1) Fidelity – the audience shall not be able to tell the differences between marked and original music broadcasted out of the speakers. In other words, the quality of played audio shall be at least as good as that of the music we heard in public places every day.

(2) Robustness – the error rate of the extracted messages must be suppressed as low as possible. Users frequently receiving erroneous messages will no more trust the information delivery service.

(3) Throughput – data rate of the pushed information must be as large as possible. At least, users shall be able to interpret the pushed information at the same speed as they are reading the same type of information via general media or channels

(4) Time/memory requirement of decoding – although real-time extraction is not always necessary for most applications, the decoding process cannot be too slow to catch up the speed of information feed. Furthermore, the required buffer size needs to be reduced for the cost concerns.

To address these requirements, a system architecture based on error control coding and spread-spectrum data hiding is proposed, as depicted in Fig. 2. Detailed descriptions of major function blocks are provided in the next section.

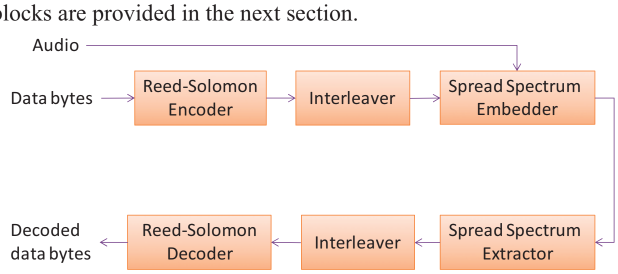


**Figure 2. Architecture of the proposed system.**

## 3. IMPLEMENTATION DETAILS

The proposed system consists of three pairs of processes: forward error correction, interleaving/de-interleaving and spread-spectrum data hiding. In addition, two-level synchronization mechanisms are integrated to resist serious out-of-synchronization errors introduced by acoustic channels.

### 3.1. Spread-Spectrum Data Hiding

#### *3.1.1. DCT- Domain Spread-Spectrum Embedding*

Fig.3 illustrates the details of the DCT-domain embedding process.

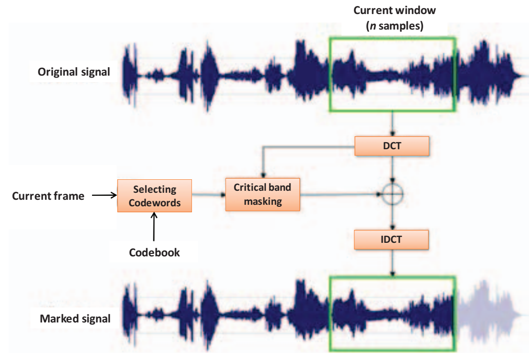


**Figure 3. The adopted DCT-domain spread-spectrum embedding.**

In our implementation, DCT coefficients of the host audio signal are divided into consecutive n-sample windows. We denote the kth coefficient of the ith window as Xi[k], k=1,…,n. Messages are represented by codewords in an m-element codebook, and each codeword Cj in a codebook is a pseudo-randomly generated binary sequence consisting of l bits. The length of codeword l affects the robustness of the data hiding system since the auto-correlation characteristic of pseudo-random sequences is better when l is larger. But the fidelity performance will be reduced while l is too long.

#### *3.1.2. Message Extraction*

Fig. 4 depicts the correlation-based extraction process. As in common spread-spectrum watermarking schemes, the correlation values between the DCT frame Yi'[k] and all the codewords are calculated. The codeword corresponds to the highest correlation value is regarded as the embedded message. Note that the de-scaling operation that is expected to eliminate the masking-required scaling operation is not performed during extraction. The reasons are as follows. In addition to speed-up the extraction process, we found that, according to our empirical test shown in Section 4, performing the de-scaling operation may not necessarily improve the correct extraction ratio. This is a reasonable phenomenon since acoustic channels introduce serious non-linear distortions to the transmitted audio sequences.

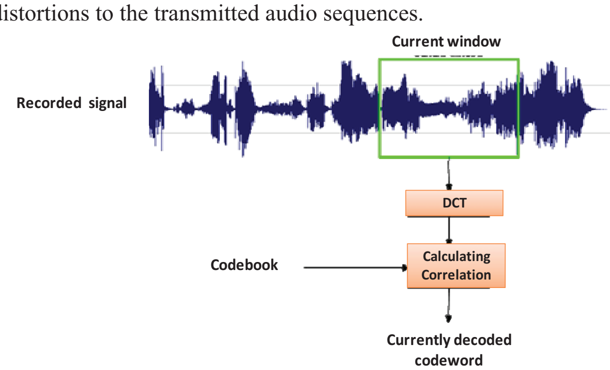


**Figure 4. The corresponding spread-spectrum extraction.**

### 3.2. Error-Control Mechanisms

Although the adopted spread-spectrum data-hiding scheme and the pseudo-random codebook design have provided certain degree of error-correcting capability, we also apply the (14, 4) Reed-

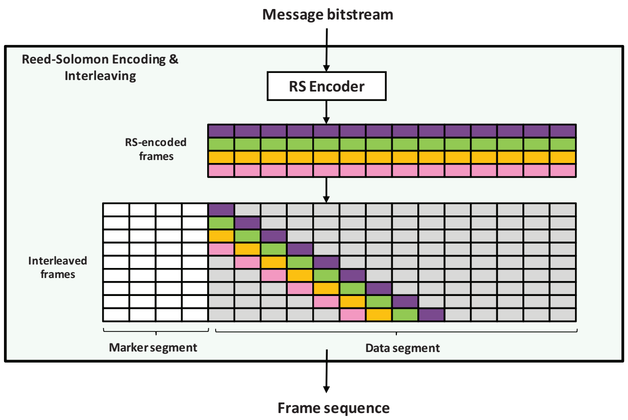


**Figure 5. The interleaver and the RS encoder. Note that colored frames denotes data frames, while white frames stand for marker frames.**

Solomon code to protect the message symbols from burst errors that are frequently encountered in the proposed applications. And to spread the errors further, an interleaving process is adopted just after the RS encoder.Fig.5 depicts the details of RS encoding and interleaving. RS-encoded messages are divided into fixed-length frames. Note that the length of each data frame is determined by the prescribed codebook size m. The corresponding de-interleaving and Reed-Solomon decoding are performed in the decoding process to correctly extract the embedded data bytes.

## 4. EXPERIMENTAL RESULTS

We investigated the performance with twelve 20-second mono audio clips sampled at 44.1 kHz with the quantization steps of 16 bits/sample. The genre of tested audio includes pop music (English and Chinese), rock & roll, heavy metal, country music, Latin, Jazz and classical music (violin, piano and orchestra plays), as well as two speech sequences (respectively from male and female radio DJ). Since the proposed scheme is designed for music sequence, the performances for the two speech sequences in the following experiments are calculated with the aid of simple silence detection mechanisms that eliminate the synchronization problem caused by silence between speeches.

Throughout all the experimental results, the window size of DCT coefficients, denoted as n, is set as 512. The codebook size m is 17, with one specific codeword used to form the marker frame. The length of the pseudo-randomly generated codeword is chosen to be 128 bits in order to provide good self-correlation characteristics. $\Delta$, the number of unaltered low-frequency coefficients, is 128 in order to preserve the fidelity of marked audio signal. Furthermore, 4 marker frames are added after encoding every 14 data frames. In addition, the message data is protected using the (4, 14) Reed-Solomon code. Finally, the value of the weighting factor p is empirically set as 0.4. Therefore, the achieved data rate is equal to:

$$44{,}100*(4/512)*(14/18)*(4/14) = 76.5625 \text{ (bps)},$$

i.e., the users can receive text messages at the rate of 10 ASCII-encoded English characters or 5 characters in the format of Unicode per second. As for the waveform level synchronization, the number of buffered DCT windows, denoted as s, is set as 10 to reduce the required amount of memory in receiving devices. And the interval between two waveform synchronizations, denoted as w, is set as 40 to strike a balance between the robustness against out-of-synchronization error and the efficiency of decoding process.

Fig. 6 illustrates the relationship between correctly extracted ratio versus the distance between speaker and microphone, given different audio volumes. Different scales of audio volumes are adjusted using the interface of Windows XP built-in Volume Control module and all the tested volumes are subjectively non-annoying for a user located one meter away from the speaker. Due to the uncertainty of environmental noises, every test based on a specific configuration is repeatedly performed five times to obtain the average value of correctly extraction ratio.

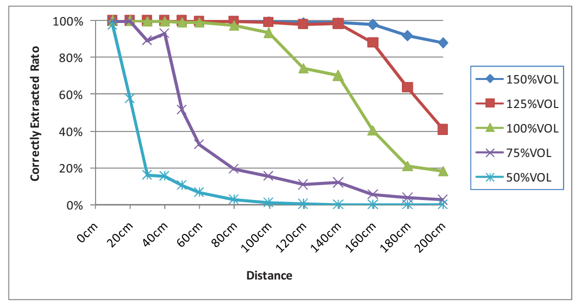


**Figure 6. The relationship between the correctly extracted ratios and distances between the speaker and receiver, given different audio volumes and the data rate of 76.6 bps.**

Fig. 7 shows the SNR values of marked audio clips. All the marked audio sequence show acceptable quality in objective measures and in subjective tests.

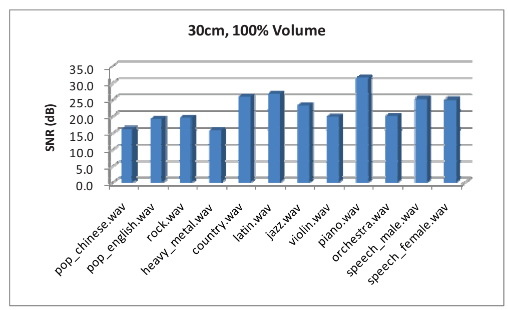


**Fig. 7. The SNR value of recorded audio clips that suffer from severe distortions of acoustic channels**

Fig. 8 illustrates the reason why the de-scaling operation is not performed in the decoding process. According to Fig. 8, performing de-scaling does not significantly improve the extraction ratio, but nevertheless reduce the extraction ratio in several cases. As indicated before, this is mainly due to the fact that the acoustic channel inevitably introduces strong non-linear distortions to the received audio sounds, thus eliminating the need of de-scaling operation. Furthermore, when performing the de-scaling operation during the decoding process, the CPU usage will increase from 20% (the CPU usage for the decoding process that de-scaling is not performed) to 23%.

Fig. 9 shows the CPU usages of respective function blocks during the embedding (shown as red bars) and decoding processes (depicted as blue bars). As we expected, the most computation-

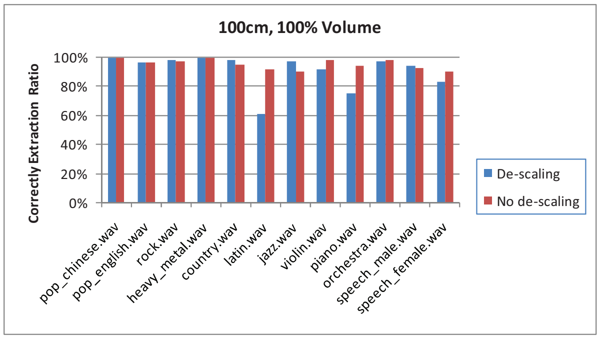


**Figure 8. Comparing extraction results to determine whether the de-scaling operation is required.**

intensive operation is the two-level synchronization that requires repetitive comparing and correlation computations. Fortunately, overall decoding process requires about 20% CPU usage, and encoding needs only 4% CPU usage. In other words, the proposed scheme is lightweight and is quite suitable to be ported to mobile devices with limited computation power.

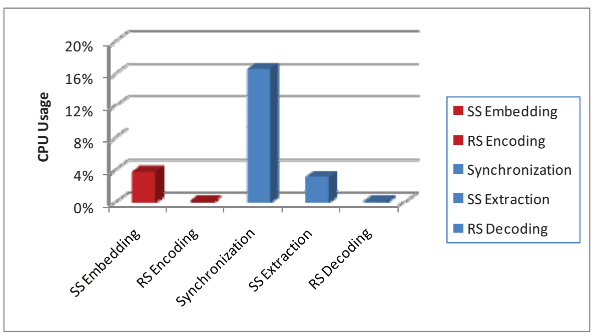


**Figure. 9. CPU usages of major encoding/decoding functions**

## 5. POTENTIAL APPLICATIONS

### 5.1. Linkage between Digital and Real World

Systems based on invisible watermarking schemes, e.g. MediaBridge[10], have been proposed to connect users to digital websites or contents. As long as recording printed images on magazines or newspapers with camcorders equipped by laptops, users will be automatically directed to the websites of certain products. When compared with this type of existing systems, information push over acoustic channels possesses the advantages of better convenience and less user intervention because audio sounds inherently propagate in all directions. In other words, the users of the proposed system only have to enable the information receiving function of their devices to be linked to web-based services or advertisements, without the need for adjusting their devices to a certain direction so that the information audio can be recorded.

### 5.2. Location-Aware Information Delivery

The proposed scheme can be readily used to deliver location-aware information like distributing discount information to privileged customers in shopping malls, sharing detailed flight information at boarding gates or illustrating exhibition contents in museums or zoos. For this application, the most important advantage lies in the savings of equipment purchasing since music playback systems and mobile devices with recording capabilities are widely available. Furthermore, the users can always enjoy musical audio and actively determine whether he or she would like to receive the pushed information, rather than passively and uniformly accepting annoying advertisement cliché or unhandy information.

## 6. CONCLUSIONS AND FUTURE WORKS

In this paper, we propose a feasible implementation to facilitate information push service over acoustic channels for short-distance users. Implementation details, experimental results, as well as its important applications and limitations are all addressed. Currently, we are working on devising schemes with better robustness against environmental noises and achieving higher information throughput. We hope that more creative applications of the proposed scheme could be invented in the near future.

## 7. REFERENCES

[1] Lazic, N., and Aarabi, P. 2006. Communication over an Acoustic Channel Using Data Hiding Techniques. IEEE Trans. Multimedia. 8, 5 (Oct. 2006).

[2] Tachibana, R. 2004. Sonic Watermarking. EURASIP Journal on Applied Signal Processing. 2004, 13 (2004).

[3] Nakashima, Y., Tachibana, R., Nishimura, M., and Babaguchi, N., 2006. Estimation of Recording Location Using Audio Watermarking. In Proceedings of the ACM Multimedia and Security Workshop 2006 (Sep. 2006, Geneva, Switzerland).

[4] Huang, C. H., Chen, P. W., Hsieh P. Y., and Wu, J. L., 2007. Information Delivery systems for Car Passengers without Networking Capabilities. In Proceedings of the Pacific-Rim Conference on Multimedia (December 11-14, 2007 in Hong Kong, China).

[5] Bassia, P., Pitas, I., and Nikolaidis, N. 2001. Robust Audio Watermarking in the Time Domain. IEEE Trans. Multimedia. 3, 2. (Jun. 2001).

[6] Wang, X., Qi, W., and Niu, P.2007. A New Adaptive Digital Audio Watermarking Based on Support Vector Regression. IEEE Trans. Audio, Speech and Language Processing. 15, 8. (Nov. 2007)

[7] Wu, S. Q., Huang, J. W., Huang, D., and Shi, Y. Q. 2005. Efficiently Self-Synchronized Audio Watermarking for Assured Audio Data Transmission. IEEE Transactions on Broadcasting. 51, 1 ( Mar. 2005)

[8] Cox, I. J., Kilian,J., Leighton, T. and Shamoon, T. 1997. Secure Spread Spectrum Watermarking for Multimedia. IEEE Trans. Image Processing, 6, 12 (Dec. 1997).

[9] Xi, J., and Chicharo, J. F. Computing Running DCTs and DSTs Based on Their Second-Order Shift Properties. 2000 IEEE Trans. Circuits and Systems - I: Fundamental Theory and Applications, 47, 5. (May 2000).

[10] Perry, B., MacIntosh, B., and Cushman, D. Digimarc MediaBridge: the Birth of a Consumer Product from Concept to Commercial Application. 2002. In Proceedings of the SPIE Conference on Security and Watermarking of Multimedia Contents IV, 2002, pp. 118-123.