Sonic Quick Response Codes (SQRC) for embedding inaudible metadata in sound files

Sheppard, M., Toulson, R. and Lopez, M.

This paper was presented at the *141st Audio Engineering Society Convention,* Los Angeles, 29 Sep 2016 to 02 Oct 2016, as paper number 9662. The full published version can be found at:

https://secure.aes.org/forum/pubs/conventions/?elib=18466

Audio Engineering Society

Convention Paper

Presented at the 141st Convention
2016 September 29–October 2 Los Angeles, USA

Sonic Quick Response Codes (SQRC) for embedding inaudible metadata in sound files

Mark Sheppard[1], Rob Toulson[2], Mariana Lopez[1]

[1] *Anglia Ruskin University, Cambridge, UK*

[2] *University of Westminster, London, UK*

Correspondence should be addressed to Mark Sheppard (mark.sheppard2@pgr.anglia.ac.uk)

ABSTRACT

With the advent of high definition recording and playback systems, a proportion of the ultrasonic frequency spectrum can potentially be used as a container for unperceivable data and used to trigger events or to hold metadata in the form of text, ISRC (International Standard Recording Code) or a website URL. The Sonic Quick Response Code (SQRC) algorithm is proposed as a method for embedding inaudible acoustic metadata within a 96 kHz audio file in the 30-35 kHz bandwidth range. Thus any receiver that has sufficient bandwidth and decode software installed can immediately find metadata on the audio being played. SQRC data was mixed at random periods into 96 kHz music audio files and listening subjects were asked to identify if they perceived the introduction of the high frequency content. Results show that none of the subjects in this pilot study could perceive the 30-35 kHz material. As a result, it is shown that it is possible to conduct high-resolution audio testing without significant or perceptible artifacts caused by intermodulation distortion.

1 Introduction

A Sonic Quick Response Code (SQRC) algorithm is proposed as a potential method for embedding inaudible metadata within a 96 kHz audio file. As a comparison with visual Quick Response (QR) codes, which display binary image data representing an internet web-link (ISO/IEC 18004:2000), the proposed SQRC holds organised acoustic energy in the 30-35 kHz bandwidth range in order to perform the same function.

QR codes are two-dimensional matrix barcodes which are read by smart phone and tablet based applications, along with dedicated QR reading devices. The encoded information contained within the QR code can consist of any alphanumeric combination and represent a variety of information such as website addresses, email links and catalogue information. Visual QR codes can be read by any digital camera system that has sufficient resolution to capture the image. With SQRC, the encoded data can be read by any digital audio system with a microphone of sufficient resolution. The proposed benefit of embedding an SQRC within 96 kHz audio and music files is that any receiver with sufficient bandwidth and decode software installed can immediately find metadata on the audio being played, without the need for complex audio fingerprinting algorithms, such as those used by Shazam [1], which rely on the network transmission of audio data and large databases of catalogue fingerprints to identify

an audio source. Psychoacoustic watermarking is another current method for embedding metadata within an audio waveform; however watermarks, to date, have been applied to frequencies within the human hearing range (20 – 20,000 Hz), bringing the potential to add distortions and audible artefacts to the carrier audio signal. Common sub-20 kHz watermarking techniques include those described by Bender *et al* [2] and Sinha *et al* [3].

Digital music presented as 96 kHz pulse code modulation (PCM) audio is envisaged to become the future standard audio format for both industry professionals and the consumer. As a front-runner in this development, the Apple Mastered for iTunes programme has already implemented the 96 kHz standard for professional delivery of files to the iTunes Music Store [4]. Additionally many online music stores specialising in high-resolution audio also exist for delivering 96 kHz music to consumers, for example HD Tracks [5], Qobuz [6] and Pro Studio Masters [7].

One unique application of SQRC is in embedding ISRC (International Standard Recording Code) data within an audio waveform, so that broadcast reporting and music cataloguing processes can be more easily automated, which is of significant value to the music industry as discussed previously by Toulson *et al* [8].

This paper gives a description of the SQRC method, presents the results of testing conducted to date, and proposes a number of improvements that will be evaluated in future experiments. It is acknowledged that the SQRC method proposed here is not intended to be optimised for data efficiency, of which many efficient processing methods already exist, including those used for watermarking [2] [3]. Moreover this research is intended to identify if the 30-35 kHz SQRC energy is perceived by listeners, and whether data can be easily encoded and decoded over file transfer protocol (FTP) exchange and also through acoustic (loudspeaker) transmission using a 40 kHz rated loudspeaker and a 40 kHz microphone setup.

2 The SQRC Method

Short (100 ms) sine wave bursts between 30 and 35 kHz are encoded into a 96 kHz audio file to represent alphanumeric characters. Characters are encoded at 50 Hz intervals, for example a 100 ms burst at 30,000 Hz represents the character ‘A’ and 30,050 Hz represents the character ‘B’. A frequency of 32,300 Hz subsequently represents the character ‘Z’ with numerics and symbols at higher intervals up to 33,100 kHz. A spectrogram of all the alphanumeric signals being played sequentially through an Adam A7X loudspeaker, and recorded with a Earthworks SR40 microphone are shown in Figure 1. Both the loudspeaker and microphone are rated with a flat frequency response up to and above 40 kHz.

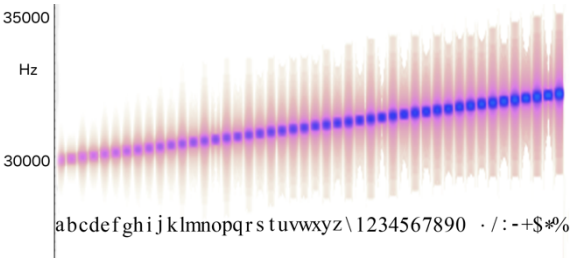

Figure 1. Spectrogram of 30 – 33.1 kHz frequencies representing alphanumeric characters.

The alphanumeric sequence of characters encoded in Figure 1 are “abcdefghijklmnopqrstuvwxyz\1234567890Δ./:-+$*%” (note Δ = Space).

Characters can be combined sequentially to form a website URL, ISRC data or descriptive text, as shown in Figure 2, which contains the sequence “www.anglia.ac.uk/code”.

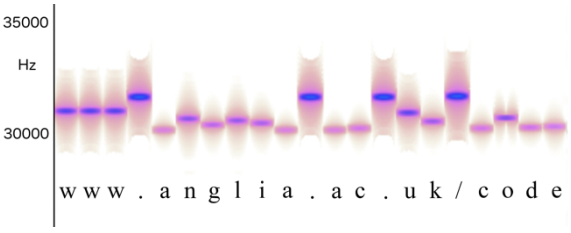

Figure 2. Spectrogram of website URL encoded as SQRC data.

3 Listening Test Design

Reiss gives an overview and meta-analysis of published psychoacoustic testing for evaluating audio

signals sampled at higher than the compact disc standard of 44.1 kHz [9]. Reiss reports that, to date, there is no conclusive or credible proof that listeners can or cannot distinguish between standard resolution (e.g. 44.1 kHz) and high-resolution (e.g. 96 kHz) audio. This particular field of research is complex and immature, given that the methods of audio recording, production, playback, and the design of listening tests are all sensitive and critical for attaining valid and conclusive results. The research presented in this paper seeks to identify whether listeners can perceive 30-35 kHz SQRC data in a 96 kHz audio file, which is a subtly different investigation to those associated with identifying perception between low-resolution and high-resolution audio. Nevertheless, this study provides new and additional research data to the field of high-resolution audio.

SQRC encoded audio files have been evaluated in perceptual listening tests. A novel listening test procedure was developed in tandem, as an alternative to standard ABX testing, building on the work described by Clark [10] and Nousaine [11]. This listening test method allows the perceptual testing to occur in a continuum rather than by comparing separate audio samples (as with the ABX method), reducing potential listener fatigue that was identified by Hicks & Tharpe [12]. Here, a 60 second SQRC was multiplied with a randomly chosen volume map (as shown in Figure 3) which, in turn, produced defined regions where the SQRC was present at zero or full volume, (normalised to -16 dB LUFS). In Figure 3, dark regions indicate periods where the SQRC is at 100% volume, light regions represent 0% volume.

Each volume map is produced with 5-10 second periods of zero or full volume and applied to the test audio sample. Ten, 60-second randomly derived volume maps were created and the applied volume map was chosen randomly from this pool when conducting a listening test.

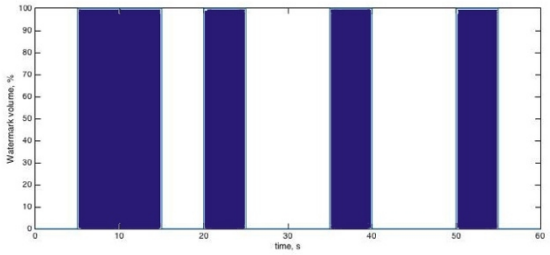

Figure 3. Example volume map for listening tests.

The volume-mapped SQRC data was subsequently mixed with a 96 kHz music file, and played to a listening subject, who was invited to mouse-click a graphical interface button when or if they perceived an artefact of noise in the signal. Different SQRC and control signals were used to identify if embedded SQRC audio data is perceived. The full list of test signals embedded with the 96 kHz carrier music were:

A1 = 5 kHz saw tooth wave

A2 = 5-10 kHz band limited white noise

U1 = 30 kHz saw tooth wave

U2 = Band limited 30-35 kHz white noise

U3 = Stepped SQRC data in 30-35 kHz region

C = Carrier signal alone (no noise added)

(A = audible, U = ultrasonic, C = control)

Test signals A1 and A2 were expected to be audible to the listener, whereas U1, U2 and U3 contained energy purely above the 20 kHz threshold of human hearing. Figure 4 shows, for example, the spectrogram of test signal U2; a 96 kHz music recording with band limited 30-35 kHz white noise embedded within.

The carrier signal was also normalised to -16 dB LUFS and embedded at 100% volume for the full duration of the listening tests. Playback was on Adam A7X loudspeakers calibrated to 85 dB SPL A-weighted at the listener position.

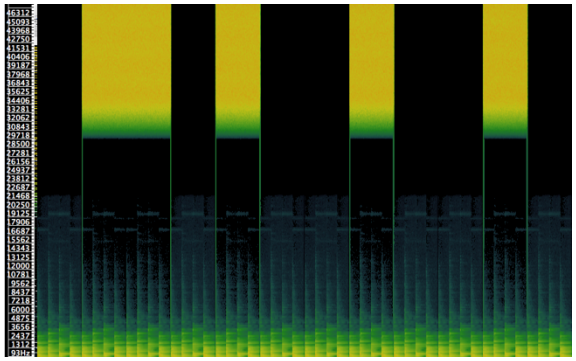

Figure 4. Spectrogram of 30-35 kHz band-limited white noise audio with volume map applied, embedded into 96 kHz music carrier signal.

For the purposes of this pilot study, five test subjects was adequate; as previous studies have shown that only a small cohort of subjects is required to achieve significant results [13].

4 Listening Test Results

The listening test results conclusively showed that high frequency sound in the range of 30-35 kHz was not perceived. Listening test results are shown in Figure 5.

In Figure 5, the y-axis represents the number of listening subject detection responses to the embeded volume mapped audio. Each of the available volume maps contianed four periods of embedded test signal, so the maximum score attainable was four for each trial. The mean number of detection responses are shown in the lower right hand panel. It can be seen that listening test subjects all identified some of the periods where the 5-10 kHz signals (A1 and A2) were embedded audio within the test carrier signal. The 30-35 kHz ultrasonic audio (U1, U2 and U3) was not perceived by any subjects. Figure 5 also shows that no false-positive responses were observed with the control audio file C.

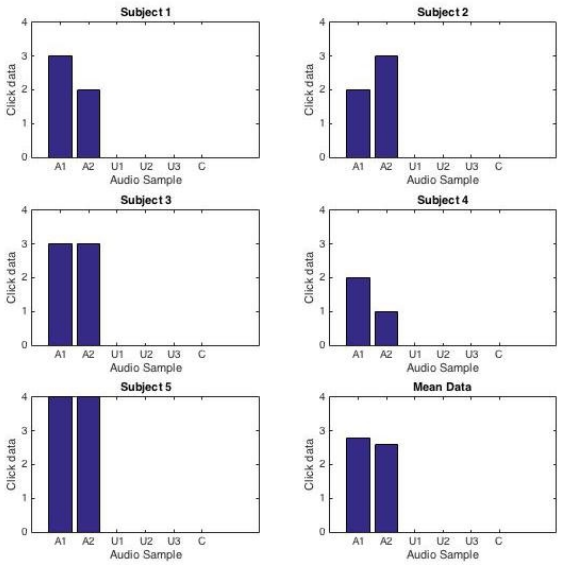

Figure 5. Pilot study listening test results.

5 Encode and Decode Testing

A number of website URLs were encoded as SQRC into an 96kHz 24-bit wave audio (.wav) files. The encode/decode process was coded using MatLab R2015b and produced a 100% decoding efficacy when using FTP as the transmission vector. The MatLab decode program identifies the maximum FFT peaks seen for 100ms in the 30-35 kHz range and translates each one to its corresponding SQRC alphanumeric character. The encode/decode method works efficiently for larger text samples, ISRC data and other alphanumeric string data also.

Acoustic transmission using a 40 kHz rated loudspeaker and a 40 kHz microphone setup has also been verified, though further testing to evaluate performance with respect to hardware choices, background noise and distance between loudspeaker and microphone need to be conducted in future research.

6 Discussion and Conclusions

During this experiemnt, the 30-35 kHz SQRC data was not perceived by the Human Auditory System. It has also been shown that SQRC data can be

effectively encoded and decoded over file transfer protocol exchange.

An interesting discussion arises around the topic of intermodulation distortion, which has been evaluated in detail previously by, for example, Toulson *et al* [14]. A number of discussions, such as that by Albano [15], refer to the conjecture that the reason for some listening tests showing a positive result in subjects identifying ultrasonic audio components is owing to intermodulation distortion that is reflected back into the sub-20 kHz range by the ultrasonic material. This experiment showed that no subjects could hear the 30-35 kHz noise, so it proves that audible intermodulation distortion was not present in the playback chain. Equally, intermodulation distortion in the sub-20 kHz was not detected on spectrogram data when playing back purely 30-35 kHz material. Clearly more investigation into the affects of intermodulation distortion on high resolution audio is needed, but this research has shown that, if correctly rated playback equipment is used, it is possible to conduct high resolution audio listening tests without spurious results being caused through intermodulation distortion.

7 Future Work

Investigations into the performance of SQRC in an acoustic transmission and reception scenario will be continued. Performance in increasingly noisy environments will be evaluated to define the resilience and robustness of the proposed algorithm. Furthermore, standard (22 kHz rated) audio systems will be tested with the SQRC in order to deduce the performance of SQRC when using existing consumer audio devices, such as mobile phones, tablet devices and commercially available recording equipment. This will incorporate methodology to counteract the effects of intermodulation distortion while maintaining an optimum output level of SQRC Other investigations will ascertain the effectiveness of SQRC over distance and the development of multi-user based SQRC information systems and video media [16]. Research into watermark encryption methodology, such as that discussed by Liao & Lee [17], as a secure metadata delivery method will also be explored. Additional listening tests will be conducted to further verify that SQRC encoding is inaudible to humans and EEG analysis will also be conducted to verify that no subconscious perception is encountered, building on the past work by Oohashi et al. [18] [19]

8 References

[1] A. Wang. "The Shazam music recognition service." *Communications of the ACM 49.8*, pp. 44-48 (2006).

[2] W. Bender, N. Morimoto,, & A. Lu. “Techniques for data hiding.” *IBM Systems Journal*, *35*, 313–336 (1996).

[3] M. Sinha, R.K. Rai,., & P.G. Kumar. “Study of Different Digital Watermarking Schemes and Its Applications.” *International Journal of Scientific Progress and Research*, *3*(2), 6–15 (2014).

[4] B. Katz. *“iTunes Music: Mastering High Resolution Audio Delivery: Produce Great Sounding Music with Mastered for iTunes”*, Focal Press (2013).

[5] www.hdtracks.com

[6] www.qobuz.com

[7] www.prostudiomasters.com

[8] R. Toulson, B. Grint and R. Staff. “Embedding ISRC Identifiers in Broadcast Wave Audio Files”, *Innovation In Music 2013* (2014).

[9] J. D. Reiss, 'A meta-analysis of high resolution audio perceptual evaluation,' *Journal of the Audio Engineering Society*, 2016.

[10] D. Clark. “Ten years of a/b/x testing.” *In Engineering Society Convention 91* (1991).

[11] T. Nousaine. “Can You Trust Your Ears?” *Engineering Society Convention 91*, 3 (1991).

[12] C.B. Hicks and A.M. Tharpe. "Listening effort and fatigue in school-age children with

and without hearing loss." *Journal of Speech, Language, and Hearing Research* 45.3 : pp.573-584 (2002).

[13] S. Werner, J. Liebetrau. and T. Sporer. "Audio-visual discrepancy and the influence on vertical sound source localization." Fourth International Workshop on Quality of Multimedia Experience, pp.133–139 (2012).

[14] R. Toulson. W. Campbell and J. Paterson. "Evaluating harmonic and intermodulation distortion of mixed signals processed with dynamic range compression." *KES Transactions on Innovation in Music*, 1(1), pp. 224–246 (2013).

[15] J. Albano. "How High Is High Enough For Hi-Resolution Audio?" www.askaudio.com (2016). Available from https://ask.audio/articles/how-high-is-high-enough-for-hiresolution-audio

[16] Rao, T Srinivasa, Kurra, R. "A Smart Intelligent Way of Video Authentication Using." *International Journal of Computer Trends and Technology*," 10(3), pp.136–142 (2014).

[17] K.C. Liao and W.H. Lee. "A Novel User Authentication Scheme Based on QR-Code. "*Journal of Networks*, 5(8), pp. 937–941 (2010).

[18] T. Oohashi,, E. Nishina, M. Honda, Y. Yonekura, Y. Fuwamoto, N. Kawai,, H. Shibasaki "Inaudible high-frequency sounds affect brain activity: hypersonic effect." *Journal of Neurophysiology*, 83(6), pp.3548–3558 (2000).

[19] Oohashi, T., Nishina, E., & Honda, M. "Multidisciplinary study on the hypersonic effect." *International Congress Series*, 1226, pp. 27–42 (2002).