Evaluating white noise degradation on Sonic Quick Response Code (SQRC) decode efficacy

Toulson, R. and Sheppard, M.

This paper was presented at the 146th AES Convention of the Audio Engineering Society, as paper number 10197. The full published version can be found at http://www.aes.org/e-lib/browse.cfm?elib=20330

Audio Engineering Society

Convention Paper

Presented at the 146th Convention
2019 March 20–23, Dublin, Ireland

Evaluating white noise degradation on Sonic Quick Response Code (SQRC) decode efficacy

Mark Sheppard[1] and Rob Toulson[2]

[1] *Anglia Ruskin University, Cambridge Campus, East Road, Cambridge, CB1 1PT*

[2] *University of Westminster, Harrow Campus, Northwick Park, London, HA1 3TP*

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

ABSTRACT

With the advent of high-resolution recording and playback systems, a proportion of the ultrasonic frequency spectrum can potentially be utilised as a carrier for imperceptible data, which can be used to trigger events or to hold metadata in the form of, for example, an ISRC (International Standard Recording Code), a website address or audio track liner notes. The Sonic Quick Response Code (SQRC) algorithm was previously proposed as a method for encoding inaudible acoustic metadata within a 96 kHz audio file in the 30-35 kHz range. This paper demonstrates the effectiveness of the SQRC decode algorithm when acoustically transmitted over distance, whilst evaluating the degradation effect of adding ultrasonic banded white noise to the pre and post transmission SQRC signal.

1 Introduction

Sonic Quick Response Code (SQRC) algorithms incorporate a methodology for introducing inaudible metadata within a high definition 24-bit 96 kHz sampled audio file [1]. This metadata insertion concept is analogous with visual Quick Response (QR) codes which display binary image data representing an internet web-link (ISO/IEC 18004:2000). QR codes are two-dimensional matrix barcodes that are read by smart phone and tablet based applications, along with dedicated QR reading devices. The encoded information contained within the visual QR code can consist of any alphanumeric combination and represent a variety of information such as website addresses, email links and catalogue information, and these can be read by any digital camera system that has sufficient resolution to capture the image. In order to perform the same function as a visual QR code, SQRC holds organised acoustic energy in the 30-35 kHz bandwidth range. Audio embedded metadata can therefore be transmitted and decoded efficiently and inaudibly using FTP (File Transfer Protocol) or via acoustic transmission over distance using a 48 kHz rated transmitting loudspeaker and a high-resolution microphone. The SQRC frequency range is defined as ultra-audible, as it is represented both above the limit of the human auditory system (HAS) and below the represented upper frequency range defined by the Nyquist theorem, which states that the highest frequency that can be represented is half the sampling rate.

Methodologies such as smartphone positioning systems have been developed with sub-20 kHz audio

[2], but It is possible to utilise the higher frequency range of 22–48 kHz in 96 kHz sampled recordings to increase the amount of bandwidth available for embedding data and algorithms (see Figure 1).

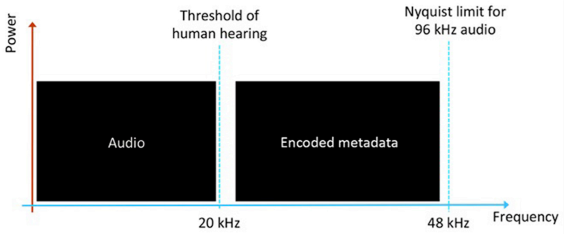

Figure 1. A schematic showing the frequency range available for insertion of high frequency metadata when using 96 kHz sampled audio. The highest frequency that can be represented is defined as being half that of the sample rate. Thus the Nyquist limit for a 96 kHz sampled track is 48 kHz.

This utilization of the high frequency bandwidth can be demonstrated by inserting a series of sequentially encoded alphanumeric characters as discretely coded acoustic energy frequencies. Figure 2 shows, a 24-bit 96 kHz source audio WAV file encoded with the SQRC frequencies that represent the 26 characters of the English alphabet and alphanumeric characters in the range of 30–33.1 kHz, with each character frequency interspaced with 100 ms of silence. Characters are encoded at 50 Hz intervals, so, for example a 100 ms burst at 30,000 Hz represents the character 'A' and 30,050 Hz represents the character 'B', with other alphanumeric characters following.

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 [3], which rely on the network transmission of audio data and large databases of catalogue fingerprints to identify an audio source.

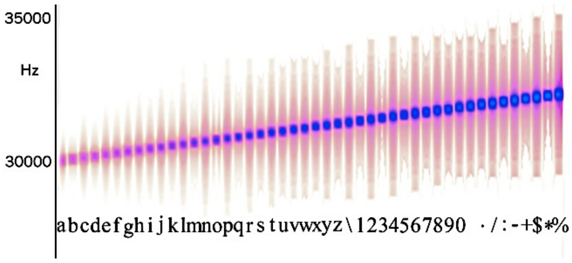

Figure 2. Spectrogram of 30–33.1 kHz frequencies representing alphanumeric characters. The sequence of characters encoded in this figure are "abcdefghijklmnopqrstuvwxyz\1234567890Δ./:-+$*%" (note Δ = Space).

2 Full band ultra-audible white noise degradation effect on SQRC efficacy

In this investigation, a variable power 16-48 kHz white noise degradation experiment was carried out on SQRC recorded over a distance of one metre from the transmission monitor, using high-definition audio equipment (i.e a loudspeaker and microphone each with a flat frequency response extending to above 40 kHz), see Figure 3.

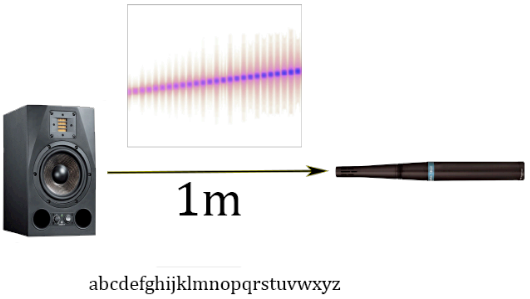

Figure 3. SQRC are transmitted over a distance of 1m at 80 dB LUFS (A), by an Adam A7X speaker and recorded with an Earthworks SR40 microphone.

The 24-bit source audio was sampled at 96 kHz, represented as a 26 character A-Z English alphabet, and acoustically transmitted at 80 dB LUFS. The 16–48 kHz white noise was algorithmically added to the recorded SQRC at variable power magnitudes, expressed in decibels (dB). Both FTP and acoustically transmitted audio samples were treated by digitally mixing high frequency white noise to the SQRC embedded samples. This strategy was employed to provide a standardised degradation protocol on the acoustically transmitted SQRC, and was applied both to the recorded signal and at source in the microphone input feed. As white noise is essentially formed of randomised uniform intensity sound over a pre-defined frequency range, placing it algorithmically on the input or acoustically transmitted output should produce a similar effect to acoustically adding it in an acoustically unidirectional manner. The decode efficiency was calculated as a percentage of correctly translated characters.

3 Results

The experimental results for both FTP (Scenario A) and Acoustic Transmission (Scenario B) after algorithmic addition of white noise are tabulated below in Tables 1-3.

Table 1 gives results for the scenario where white noise of varying levels is added to a computer generated SQRC based WAV file, and is subsequently decoded after file transfer. It can be seen from Table 1 that for all levels of white noise added, the algorithm fails to effectively decode the SQRC data at all noise levels.

Table 2 gives results for the scenario where white noise is combined with the SQRC and acoustically transmitted, and decoded at a distance of one meter. Here it can be seen that the SQRC decoder has successful results for many noise levels, maintaining a near 100% performance at white noise levels up to -50 dB.

Table 3 shows results for the scenario where white noise is combined with the SQRC after it is acoustically transmitted over a distance of one meter. The decode algorithm is then applied. As with Table 2, it can be seen that the SQRC works with 100% efficacy up to noise levels of -50dB, but in this case, the degradation is more rapid thereafter, with 0% efficacy recorded for all tests with higher noise levels.

Noise level (dB)	Decode Efficiency (%)		
	FTP Trial 1	FTP Trial 2	FTP Trial 3
-10	0	0	0
-20	0	0	0
-30	0	0	0
-40	0	0	0
-50	0	0	0
-60	0	0	0
-70	0	0	0
-80	0	0	0
-90	0	0	0
-100	0	0	0
-110	0	0	0
-120	0	0	0
-130	100	100	100

Table 1. Decode efficiency of FTP transmitted SQRC after exposure to varying dB levels of 16-48 kHz white noise.

Noise level (dB)	Decode Efficiency (%)		
	AT (Input) Trial 1	AT (Input) Trial 2	AT (Input) Trial 3
-10	69.2	100	88.5
-20	80.7	100	100
-30	92.3	92.3	96.2
-40	80.7	80.7	100
-50	100	100	100
-60	100	100	100
-70	100	100	96.2
-80	100	100	100
-90	100	100	100
-100	100	100	100
-110	100	100	100
-120	100	100	100
-130	100	100	100

Table 2. Decode efficiency of acoustically transmitted SQRC after exposure to varying dB levels of 16-48 kHz white noise. AT (Input) = White noise added to the SQRC pre-acoustic transmission.

Noise level (db.)	Decode Efficiency (%)		
	AT (Output) Trial 1	AT (Output) Trial 2	AT (Output) Trial 3
-10	0	0	0
-20	0	0	0
-30	0	0	0
-40	0	0	0
-50	100	100	100
-60	100	100	100
-70	100	100	100
-80	100	100	100
-90	100	100	100
-100	100	100	100
-110	100	100	100
-120	100	100	100
-130	100	100	100

Table 3. Decode efficiency of acoustically transmitted SQRC after exposure to varying dB levels of 16-48 kHz white noise. AT (Output) = White noise added to the SQRC post-acoustic transmission.

Note that for both test scenarios, i.e. FTP transfer and acoustic transmission, the SQRC decode method is previously shown and confirmed to operate with 100% efficacy when there is no white noise present or added [1].

4 Conclusions

It can be seen that the SQRC decode efficiency over FTP (Table 1) is less robust from degradation by 16-48 kHz white noise when compared with acoustically transmitted SQRC (Tables 2 and 3). This may be explained by the fact that acoustically transmitted sound has an additional energy component produced by surface reflection as the ultra-audible SQRC interacts with its environment. This additional ultra-audible energy allows an acoustically transmitted SQRC to have a greater degree of robustness when degraded by direct summation of high frequency white noise. Adding ultra-audible white noise to the SQRC post transmission creates a scenario where a greater collective energy of white noise is added to the SQRC signal, effectively causing a greater degree of masking, which in turn reduces decode efficacy.

As a result of this investigation, further research and development work will be conducted to increase the robustness of the SQRC algorithm against noise in the same frequency range. This may be achieved by modulating the SQRC at different frequency ranges in order to produce a combined transmitted signal with a greater overall power. The decode algorithm could also incorporate a threshold filter to eliminate the relatively lower energy white noise. Additional testing may also be conducted to simulate real-world noise environments, and so evaluate the efficacy of the SQRC algorithm in scenarios where it may be realistically utilised, such as public spaces and home listening environments.

References

[1] Sheppard, M., Toulson, R. and Lopez, M. “Sonic Quick Response Codes (SQRC) for embedding inaudible metadata in sound files” Audio Engineering Society Convention 141. Audio Engineering Society, pp. 1–7 (2016).

[2] Lopes, S.I., Vieira, J.M.N. and Albuquerque, D.F. “ Analysis of the Perceptual Impact of High Frequency Audio Pulses in Smartphone-based Positioning Systems” Industrial Technology (ICIT), 2015 IEEE International Conference on, pp. 3398 – 3403 (2015).

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