Acoustic transmission of metadata in audio files using Sonic Quick Response Codes (SQRC)

Sheppard, M., Toulson, R. and Fachner, J.

This is an accepted manuscript of a book chapter published by Routledge in Innovation in Music: Performance, Production, Technology, and Business on 23 April 2018, available online:

http://www.routledge.com/9781138498198

Acoustic transmission of metadata in audio files using Sonic Quick Response Codes (SQRC)

Mark Sheppard[1], Rob Toulson[2] & Jörg Fachner[3]

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

mark.sheppard2@pgr.anglia.ac.uk

[2]University of Westminster, Harrow Campus, Northwick Park, HA1

R.Toulson@westminster.ac.uk

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

jorg.fachner@anglia.ac.uk

Abstract

With the advent of high-resolution recording and playback systems, a proportion of the ultrasonic frequency spectrum can potentially be utilized 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 URL 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 algorithm when acoustically transmitted over distance, whilst evaluating combinations of high and lower resolution audio equipment.

1. Introduction

Sonic Quick Response Code (SQRC) algorithms offer a methodology for introducing inaudible metadata within a high definition 96 kHz sampled audio file (Sheppard et al, 2016). 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 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. The proposed SQRC holds organized acoustic energy in the 30-35 kHz bandwidth range in order to perform a similar function. This audio embedded metadata can be transmitted and decoded efficiently using FTP (File Transfer Protocol) or via acoustic transmission over distance using a 48 kHz rated loudspeaker, and decoded after capture with a high-resolution (40 kHz bandwidth) microphone. In the research presented here, other loudspeaker/ microphone combinations are also trailed and their SQRC decoding efficiency is quantified.

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 (Wang, 2006), 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, with the caveat that the data is imperceptible to the human auditory system (Cvejic & Seppanen, 2001). However psychoacoustic watermarks, to date, have been applied to frequencies within the limits of the human hearing range (20 – 20,000 Hz), bringing the potential to add distortions and audible artifacts to the carrier audio signal. These sub-20 kHz watermarking techniques include those described by Bender et al (1996) and Sinha et al (2014), and are particularly effective despite having limited space for metadata insertion without compromising the integrity of the 44.1 kHz or 48 kHz sampled carrier audio. Adding metadata to frequency ranges above 20 kHz negates this space limitation issue, but it also has the potential to introduce intermodulation distortion artefacts into the sub-20 kHz audio range (Toulson et al, 2014).

Sampling the carrier audio at more than twice the rate of the commercial compact disc (CD) standard (i.e. sampling at 96 kHz) theoretically gives greater resolution and accessibility of higher frequency ranges as defined by the Nyquist criterion, which states that the upper limit of the sampling rate is twice the highest frequency within the signal. Thus, a sampling rate of 96 kHz allows a maximum metadata frequency of 48 kHz to be embedded. Currently, the SQRC encoding strategy is deployed in a stenographic manner of hiding data in plain sight, but can easily be expanded by using a multitude of encryption strategies, such as embedding a spread spectrum watermark (Wojtuń, 2011) or using a more linear pulsed approach (Lopes et al, 2015). Embedding high frequency sounds at a greater bit depth to increase signal clarity could in turn expand these strategies.

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 (Albano, 2017). As a front-runner in this development, the Apple Mastered for iTunes program has already implemented the 96 kHz standard for professional delivery of files to the iTunes Music Store (Katz, 2013). Additionally many online music stores specializing in high-resolution audio also exist for delivering 96 kHz music to consumers, for example HD Tracks (www.hdtracks.com), Qobuz (www.qobuz.com) and Pro Studio Masters (www.prostudiomasters.com). 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 (2014). This paper gives a description of the SQRC method, presents the results of audio encode/decode experiments conducted to date, and proposes a number of improvements that will be evaluated in future experiments.

2. Encoding Sonic Quick Response Codes (SQRCs)

2.1. The SQRC algorithm

Previously, perceptual audio investigations have been carried out on pulsed frequencies up to 22 kHz (Lopes *et al,* 2015), but it is suggested in this research that it is potentially possible to utilize the inaudible frequency range of 22–48 kHz in 96 kHz sampled recordings for pulsed metadata insertion (Figure 1). This can be achieved by inserting a series of sequentially encoded alphanumeric characters as discrete signal data, using Matlab R2015a

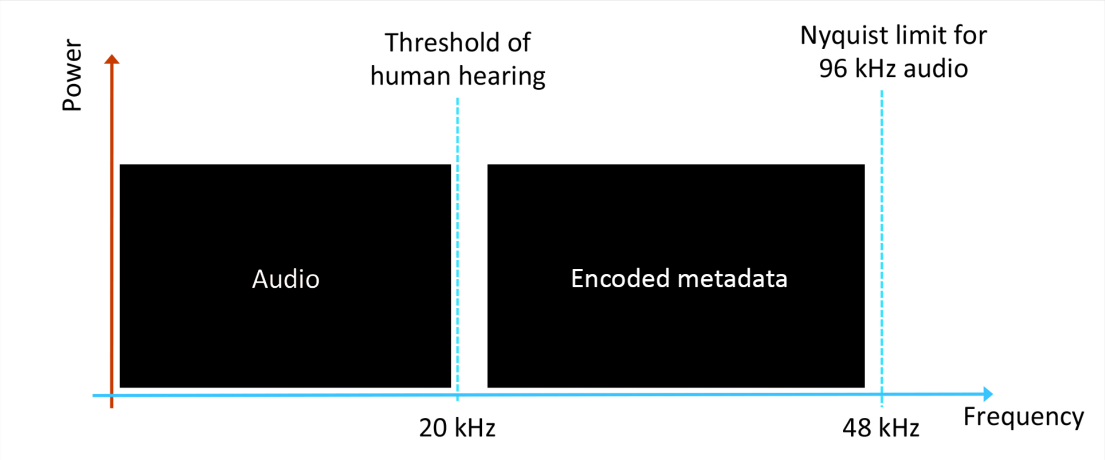

Figure 1. A schematic showing the frequency range available for insertion of high frequency metadata when using 96 kHz sampled audio.

As shown previously by Sheppard et al (2016), a 96 kHz 24 bit source audio wave (WAV) file is encoded with the SQRC frequencies that refer to the 26 characters of the

English alphabet. Alphanumeric characters are encoded as 100 ms sinusoid bursts with a unique inaudible frequency representing each character. Characters are encoded upwards from 30 kHz at 50 Hz intervals; for example a 100 ms burst at 30,000 Hz represents the character ‘A’ and 30,050 Hz therefore represents the character ‘B’. A frequency of 31,250 Hz subsequently represents the character ‘Z’ with numerics and symbols at higher intervals up to 32,250 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 2.

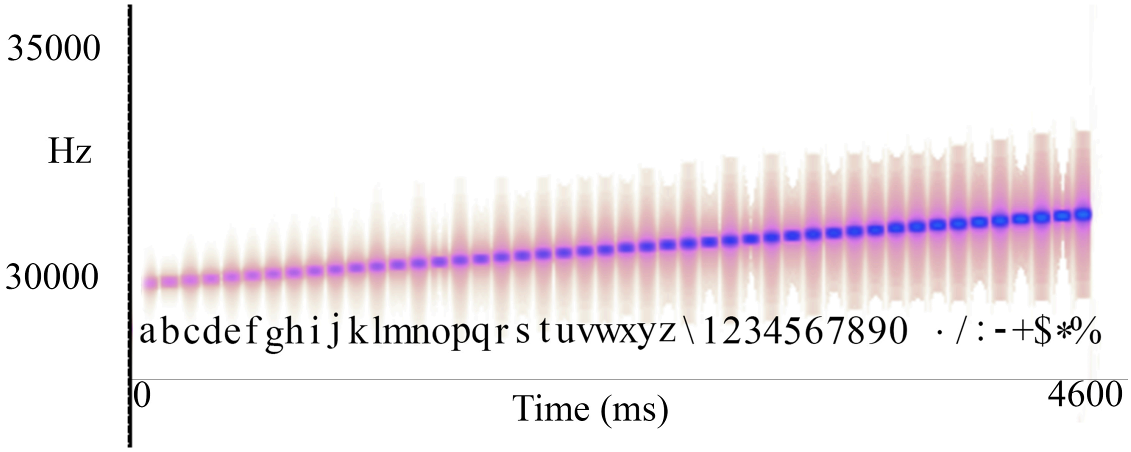

Figure 2. Spectrogram of 30 – 32.25 kHz frequencies representing alphanumeric characters. The alphanumeric sequence of characters encoded in this figure are “abcdefghijklmnopqrstuvwxyz\1234567890Δ./:-+$*%” (note Δ = Space).

Characters can be combined sequentially to form a website URL, ISRC data or descriptive text, as shown in Figure 3, which contains the sequence “http://www.anglia.ac.uk/code”.

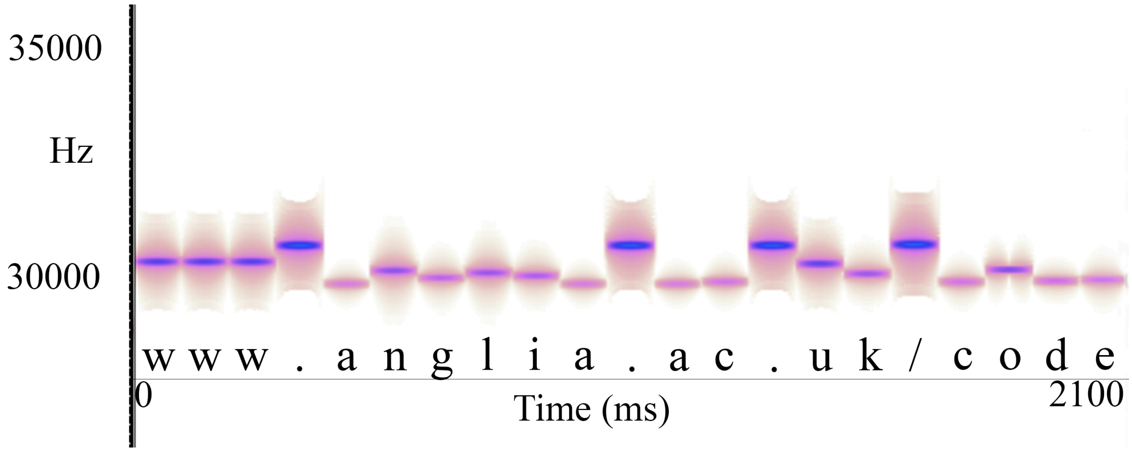

Figure 3. Spectrogram of website URL encoded as SQRC data.

3. Acoustic Transmission Results using High Resolution Audio Devices

Three sets of high-frequency encoded SQRC were transmitted acoustically over a range of distances. The SQRC were transmitted at a calibrated 85dB LUFS from the primary source monitor (ADAM A7X loudspeaker) and samples were recorded at each distance interval using an Earthworks SR40 high definition microphone, and decoded using a Matlab SQRC algorithm. Background noise levels throughout the investigation were in the range of 30-35 dB (A weighting). The three samples acoustically transmitted over a distance of 0.1 – 5 meters in this experiment were:

- 26 sequential characters of the English alphabet (A to Z).
- URL: www.bbc.co.uk
- ISRC: gbpaj1500001

Distance (m)	Decode Efficiency (%)
0.1	100
1	100
2	100
3	100
4	100
5	100

Table 1. Three different audio samples are acoustically transmitted over distance and include a 26 character English alphabet (A-Z), an ISRC and a URL.

4 Acoustic Transmission of SQRC with Standard-Resolution Devices

The viability of using lower resolution (22 kHz rated) equipment for audio transmission and reception, with effective coding and decoding of the SQRC metadata, was investigated. For capture, Samson C01 and iPhone 5S standard-resolution microphones were trialed against the high-resolution Earthworks SR40. For transmission, generic 6 cm and 20 cm loudspeakers (Sony model 1-826-115-11 and Peavey Blazer 10W respectively) were compared against the ADAM A7X 48 kHz monitor. The acoustic transmission testing was carried out over a distance of 5m, with different microphone/monitor speaker combinations (Figure 4).

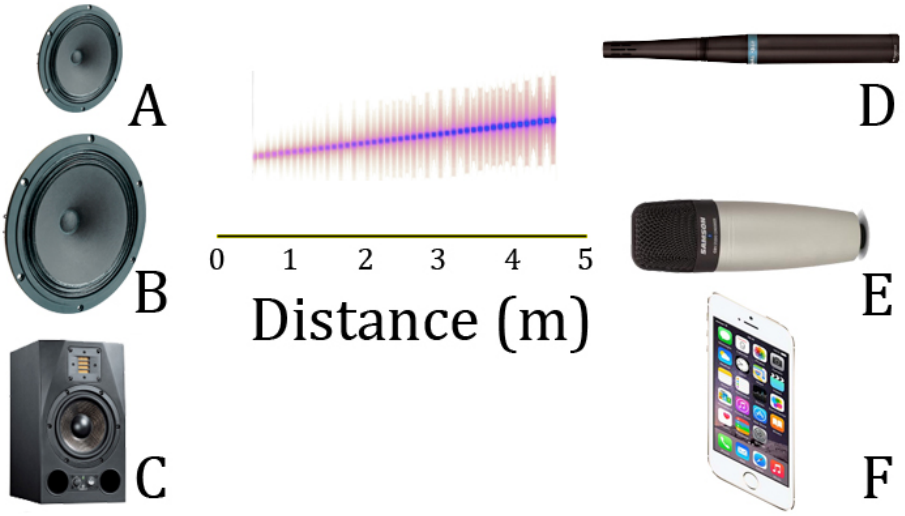

Figure 4. Acoustic transmission of SQRC from various speaker sources over 5m. Speaker/ Microphone combinations: A = 6cm speaker, B = 20cm speaker, C = ADAM A7X monitor, D = Earthworks SR40, E = Samson C01, F = iPhone 5s.

The SQRC used in the following acoustic transmission experiments consisted of a website URL, an ISRC and a 26 character English alphabet set. These SQRC were transmitted at a loudness of 85dB LUFS from the primary source speaker, with background noise levels in the range of 30-35 dB (A-weighting). In each transmission experiment, the microphone under test captured the monitor output as a 24 bit/ 96 kHz WAVE audio file, which was then translated using a Matlab derived SQRC decode algorithm. In the following experiments decode efficacy is scored as a percentage of correctly translated characters.

The decode efficacy of acoustically transmitted SQRC over a distance of 5m is detailed in figure 5. In this scenario a 6cm (Sony model 1-826-115-11) speaker is used as the transmission device, and three recording microphones are compared: A high definition Earthworks SR40 microphone against two standard definition microphones (Samson C01 and the iPhone 5S). The decode efficiency at each distance is calculated from the number of correctly translated SQRC, and displayed as a percentage. Note that the closest measurement to the transmission speaker cone is taken at a distance of 0.1 meters. The high definition Earthworks SR40 microphone has the highest decode efficiency, though in this scenario is only 100% effective at a distance of 0.1 meters from the transmission speaker. Interestingly, the standard definition Samson C01 is able to partially decode SQRC at this distance. The iPhone 5S microphone is unable to translate any SQRC.

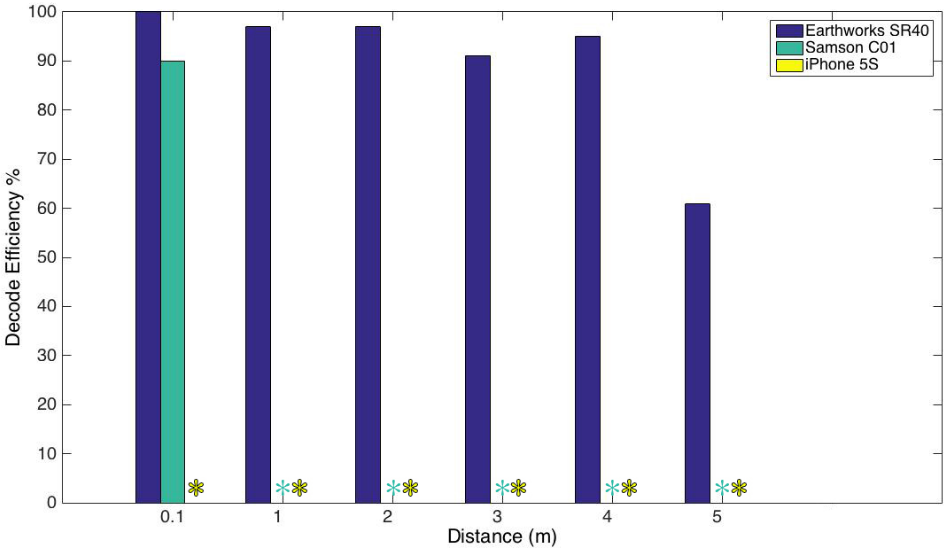

Figure 5. Decode efficiency of SQRC after acoustic transmission over 5m using a 6cm speaker with Earthworks SR40/Samson C01/iPhone 5S microphones. * Denotes zero decode.

Figure 6 shows the acoustic transmission decode results from the scenario where a 20cm (Peavey Blazer 10W) speaker is used as the transmission device, and the same three recording microphones are compared (Earthworks SR40, Samson C01 and the iPhone 5S). The decode efficiency at each distance is significantly reduced with the Earthworks SR40 and the Samson C01. The iPhone 5S still displays zero decode efficacy at all the sampled distances.

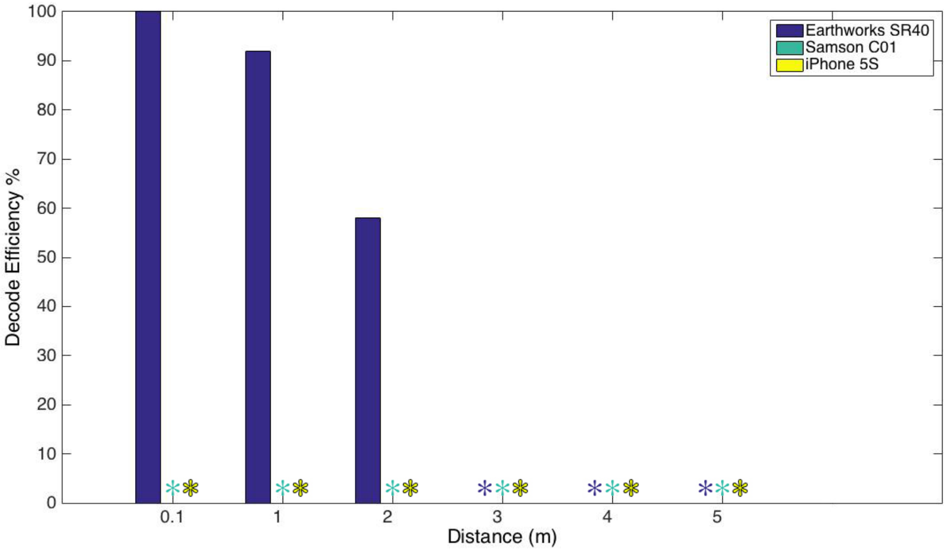

Figure 6. Decode efficiency of SQRC after acoustic transmission over 5m using a 20cm speaker with Earthworks SR40/Samson C01/iPhone 5S microphones. * Denotes zero decode.

Figure 7 shows the effect of increasing decode efficacy when two high definition components are combined. i.e. When the ADAM A7X monitor is utilized for acoustic transmission of SQRC and the Earthworks SR40 microphone is used to record them. Decode efficiency is also increased in the Samson C01 recording scenario, though the iPhone 5S microphone still shows no improvement in decode efficacy at any distance even when the transmission monitor is high definition.

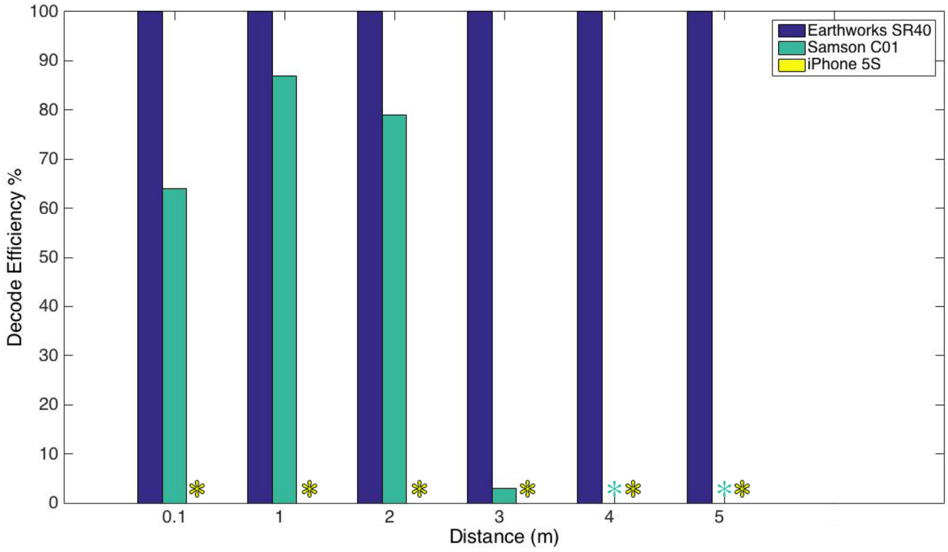

Figure 7. Decode efficiency of SQRC after acoustic transmission over 5m using a ADAM A7X monitor with Earthworks SR40/Samson C01/iPhone 5S microphones. * Denotes zero decode.

5 Full-Band Ultrasonic White Noise Degradation on SQRC Efficacy

A variable power 16-48kHz white noise degradation experiment was carried out on an SQRC recorded by an Earthworks SR40 microphone at a distance of 1m from the transmission loudspeaker (ADAM A7X monitor). The source audio was sampled at 96 kHz and represented a 26 character A-Z English alphabet. The 16–48 kHz white noise was algorithmically added to the recorded SQRC at variable power magnitudes, expressed in dB, and the decode efficiency calculated as a percentage (Figure 6). For both the acoustic transmission and FTP wave file analysis, 16–48 kHz white noise was combined algorithmically using a Matlab script.

Noise level (dB)	FTP Efficacy (%)	Acoustic Efficacy (%)
0	100	100
-10	0	0
-20	0	0
- 30	0	0
- 40	0	0
- 50	0	100
- 60	0	100
- 70	0	100
- 80	0	100
-90	0	100
- 100	0	100
- 110	0	100
- 120	0	100
- 130	100	100

Table 2. Decode efficiency of SQRC after exposure to varying dB levels of 16-48 kHz white noise. Both on SQRC embedded audio transmitted over FTP and acoustic transmission

In order to do a direct comparison of exposure to high frequency white noise on acoustically and FTP transmitted metadata, a Matlab generated white noise algorithm was embedded into both audio file types. The threshold limit at which the white noise had an impact was elucidated, and it was found that relatively low dB white noise insertions reduced decode efficiency markedly.

6 Discussion and Conclusions

SQRC data can be both effectively encoded and decoded over file transfer protocol exchange and acoustic transmission, using high-resolution transmission and recording apparatus. This can be used in conjunction with lower resolution equipment to attain a less optimal level of decode. The decode process is robust and resilient in noisy environments as the majority of environmental noise is below the threshold of the SQRC. Thus with this rationale, any receiver that has sufficient bandwidth and decode software installed can immediately find SQRC derived metadata embedded in the audio being transmitted, without requiring fingerprint analysis or an active Internet connection. The research shows that the SQRC method is a viable acoustic protocol that could feasibly be utilized in a variety of practical applications.

The SQRC are very susceptible to white noise attack in the same transmission frequency range, with even low levels of white noise disrupting the decoding process. SQRC decode efficiency over FTP is less robust from 16-48 kHz white noise attack than SQRC transmitted acoustically. This may be explained by the fact that acoustically transmitted sound has additional energy from reflected surfaces, which in turn allows it to be less affected by direct summation via a digital algorithm.

The high definition Earthworks SR40 microphone was able to achieve greater decode efficacy over 5m across all three speaker types (ADAM A7X monitor, 6cm Sony model 1-826-115-11 speaker and the 10 W Peavey Blazer 20cm speaker), than the Samson C01 and iPhone 5S microphones. Translation efficacy was at its most optimal with the ADAM A7X speaker. Conversely the Samson C01 was only able to produce partial decode of SQRC metadata even when the high definition ADAM A7X monitor was utilized as the transmitter source. Though decode is sub-optimal with standard speakers at present, there is potential to develop the decode algorithm further to increase translation efficacy and facilitate the use of standard definition speakers for SQRC transmission.

The SQRC decode efficiency in the iPhone was shown to be poor, due to the maximum frequency range of the microphone to be 22-24kHz (Aguilera et al, 2013). Reducing the upper frequency limit of the SQRC to 22-24kHz could optimize the iPhone for use as an SQRC receiver. Conversely, if the manufacturer were to extend the microphone frequency range, then the iPhone could be used to receive SQRC metadata.

7 Future Work

Investigations into the performance of SQRC in an acoustic transmission and reception scenario will be continued, and performance in increasingly high frequency noisy environments will be evaluated to define the resilience and robustness of the proposed algorithm as a method of embedding metadata into suitably high definition digital media. The paper itself does not concern itself with high-resolution audio perception (Reiss, 2016), but will investigate this with aspect with EEG analysis of SQRC. This will be conducted to verify that no subconscious perception of SQRC is encountered, building on the past work by Oohashi et al. (2000; 2002).

8 References

Aguilera, T., Peredes, J.A., Alvarez, F.J., Suarez, J.I., and Hernandez, A. (2013). Acoustic local positioning system using an iOS device. *International Conference on Indoor Positioning and Indoor Navigation.*

Albano, J. (2017). How High Is High Enough For Hi-Resolution Audio? Accessed October 2017 from https://ask.audio/articles/how-high-is-highenough-for-hiresolution-audio

Bender, W., Morimoto,N. and Lu, A. (1996). Techniques for data hiding. *IBM Systems. Journal*, 35, pp. 313–336.

Brann, Noel L. *The Abbot Trithemius (1462-1516): the renaissance of monastic humanism*. Vol. 24. Brill, 1981.

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

Lopes, S. I., Vieira, J. M. N., & Albuquerque, D. F. (2015). 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.

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

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

Reiss, J. D. (2016). A meta-analysis of high resolution audio perceptual evaluation, *Journal of the Audio Engineering Society*.

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, (2016), pp. 1–7.

Sinha, M., Rai, R.K., and Kumar, P.G. (2014). Study of Different Digital Watermarking Schemes and its Applications. *International Journal of Scientific Progress and Rese*arch, 3(2), pp. 6–15.

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

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

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

Wojtuń, J., Piotrowski, Z., Gajewski, P. (2011). Implementation of the DSSS method in watermarking digital audio objects. Annales UMCS, *Informatica*, 11(3), pp. 57-70.