IS THERE AN AUDIBLE DIFFERENCE BETWEEN STANDARD AND HIGH-DEFINITION SAMPLE RATES?

Joseph Fossard University of Central Lancashire

1 INTRODUCTION

The aim of the tests described in this report was to determine whether an audible difference could be detected between a standard sampling-frequency of 44.1 thousand samples per second (44.1ks/sec) and the higher sampling-frequency of 96ks/sec, and if so, under what conditions a difference could be perceived. The tests were carried out with just one stereo analogue-to-digital conversion on each recording system, and one digital-to-analogue conversion on each reproduction system, before sending the sound to the different monitor loudspeaker systems for the listeners to assess.

There has been a continuing debate since the introduction of higher sampling-frequencies (often referred to as *High-Definition* audio) as to whether there is really is any improvement of sound quality with the higher rates when applied to a two-channel stereo recording. Most tests published to date appear to have used either a single source-file with a conversion process (where a duplicate file of a different sample rate was created), or the comparison between sample rates has not used identical equipment in either the recording or the playback process. However, to compare a converted file to an original file is not the same as comparing two original files recorded at different frequencies – such that the test has simultaneous recordings at both frequencies on identical equipment. (For example, see Pras and Guastavino, 2010.)

The most obvious advantage of using a higher sampling-frequency would appear at first glance to be the extended bandwidth of the recording. The Nyquist-Shannon sampling theorem explains that the bandwidth is dependent on number of samples per second, and that for any frequency to be recorded there must be a sample-rate of at least double that frequency. (Shannon and Weaver 1999) This limits the standard 44.1ks/sec recording to 22,050Hz, and 'high definition' (HD) 96ks/sec recording to 48,000Hz, but the most obvious critical argument against this issue is that most human hearing does not even reach as far as 20kHz, let alone exceed it. Loudspeakers, also, do not normally respond much higher than 20kHz – at least not without a considerable roll off.

Critics of lower sampling-rates argue that a low-pass filter above 20kHz needs to be very steep if it is to filter out everything above 22.05kHz (half of the sampling frequency – to avoid aliasing). Although it is true that the different filters can sound different, we must also take into account the characteristics of the equipment that we use to record the sounds. Most microphones do not respond flat up to 20kHz due to their mechanical and physical properties, and neither does much of the equipment in many recording chains. Furthermore, only a very few instruments produce high frequencies above 20kHz, and even those which do will produce them at much lower relative levels than the frequencies that we actually hear

Taking all this into account, it would seem that the difference between sample rates can only be assumed to be very subtle, as it must in fact be for there to be so much contrasting opinion regarding the subject, and with this in mind, consideration had to be given to the ways in which these tests could be representatively monitored. Even though a monitor chain itself may add its own complexities (due to DSP artefacts, phase irregularities etc.), it must at least be capable of resolving any differences which may exist, or they will obviously not be heard. Indeed, at the 1999 Reproduced Sound conference, John Watkinson presented a paper on the need to consider the loudspeakers as information channels, with a 'resolution' of its own'. (Watkinson, J. and Salter, R., 1999)

[It should be noted that the tests described in this report were made more difficult than originally envisaged due to the pandemic restrictions, not only on travelling, but also on the meeting of groups of people for the listening tests. Consequently, these tests were not extensive as had originally been intended, but their results still pose interesting questions for future work.]

2 RECORDING PROCEDURES

2.1 The recording equipment

Two brand-new DPA 4091 microphones were used to capture the signals for the recordings, along with two Audient EVO4 interfaces. Two Midas XL4 pre-amplifiers were fed a split signal into the two interfaces which recorded simultaneously into Logic Pro X, on two identical MacBook Pros. The Midas pre-amps have a frequency bandwidth of up to 40kHz and the DPA microphones respond uniformly up to 20kHz (before tailing-off at about 24dB per octave). The EVO4 interfaces were specifically chosen for the fact that the gain can be set on the pre-amp remotely, with no physical gain pot, and because they also have a bandwidth of up to 40kHz. The two recording systems will be referred to as recording systems “A” and “B”, and although they were nominally identical, a record was nonetheless kept of which system was used at which frequency on each recording. Figure 1, in the Appendix, shows the set-up in graphic form

2.2 Source Material

Six files were selected for the tests, with three signal sources of very different natures. The first file was a recording of water falling into an outside pool, and this also contained some ambient background noise. For this recording, system “A” was set at 96ks/sec. and system B at 44.1ks/sec. The second and third files were both of drum kits, recorded at *Loudness Studios*, in Lisbon. ‘Drums1’ was recorded at 96ks/sec on system A, while system B recorded at the lower frequency, with the frequencies being swapped for the ‘Drums 2’ recording. A short piano piece was recorded in *Tcha Tcha Tcha Studios*, also in Lisbon, with system A recording at the lower frequency and system B at the higher frequency. The two last files for the test both consisted of recordings of line-level signals from a Minimoog, being recorded at 96ks/sec on system A, and at 44.1ks/sec on system B, with the opposite being the case for the TB-303 file. The recorded sounds were therefore very different in nature, although speech was not considered a suitable option for these tests due to the fact that, as a system of communication, the listeners could be listening more to the message within the words rather than the quality of sound. Random words, on the other hand, would have no realistic context.

3 THE LISTENING TESTS

Four different rooms were chosen for the tests, with a range of different acoustic treatments and monitor loudspeakers, but all were representative of the environments in which much popular music is currently produced, mixed and mastered.

All the listeners were played two stereo signals simultaneously, via a two-way switch, one being from laptop (computer) A and the other from laptop B, with their respective Audient EVO4 interfaces. The participants could listen to either signal, and change between them virtually instantaneously. The resulting signal was then fed into a volume-control which could be set at any desired level, and from where the signal then passed directly into the studio monitors in each listening environment. Figure 2, in the Appendix, shows a diagram of the listening set-up.

Test 1

Test 1 was carried out in a newly-built ‘non-environment’ control room with highly-damped acoustics, having an approximately 200 ms decay time. The two identical laptops, both with Audient EVO4 interfaces, fed audio into a “master-mix” rotary potentiometer 4-channel mixer, and from there was sent directly into a pair of Barefoot loudspeakers (the first-generation MicroMain 27 model) which use a DSP sampling rate of 48ks/second. It was initially uncertain if the processing, itself, could affect the detection of any noticeable differences compared with all-analogue monitor systems, so the results were keenly awaited.

Both laptops had identical copies of all the recorded material, and either version of the recording could be played from either laptop. All Logic sessions were played back as one stereo track, with the fader set at unity gain in this test, and without plug-ins or processing of any kind.

All the participants in Test 1 noticed a difference in the first three files, but none could distinguish between the last three files. Also, two of the participants preferred the higher sampling frequency, and one preferred the lower.

Test 2

The second test took place in a small mastering-studio in Liverpool, UK. The studio had *Kii Three* monitor loudspeakers, and the room was considerably more lively than the room used for Test 1. It was noted that the *Kii Three* monitors use a sampling rate of approximately 93.75ks/second, which was significantly different from that used in the Barefoot loudspeakers of Test 1. Also, an ‘A/B’ switch and a volume control replaced the rotary-pot mixer use in Test 1, to give the listeners both personal control of the volume and an instant switch-over from laptop A to laptop B. In addition, the studio provided a pair of open-backed Audeze ‘reference’ headphones (frequently reported to be used during mastering), through which the test was repeated by the mastering engineer.

In Test 2, somewhat surprisingly, no difference was distinguished between any of the files, yet that was still the conclusion after several hours of listening. The files were undistinguishable by either the in-house engineer or the author, and nor were any differences noticed when the test was repeated by the mastering engineer using the Audeze headphones.

Test 3

A third test was organized at *Spirit Studios*, formally known as *School of Sound Recording*, in Manchester, and a suitable room was chosen to perform an identical test to that of Test 2. This control room was equipped with a pair of Genetic 8331A loudspeakers, free standing, and a pair of Quested VH3208 loudspeakers built into the monitor wall, but with only a limited degree of acoustic treatment in the room. The fact that the two loudspeaker systems were very different in their design concepts provided a useful additional comparison.

Test 4

After analyzing the data from the first three tests, a modest set-up in the author’s home provided a further way of assessing to the files. This system had both laptops running through an A/B switch and into a passive volume control, connected directly to a pair of Yamaha *HS5* self-powered loudspeakers. As there were no time constraints in this room, more time was spent on these tests compared to the previous ones, and each test was carried out with only one listener present at a time.

The results of all the tests are discussed in Section 5, and a complete table of the tests is shown in the Appendix (Section 9.1), which also includes information about each listener’s occupation and age.

4 MONITOR LOUDSPEAKERS AND ROOM ACOUSTICS

One major factor to consider is that of the characteristics of each of the monitoring systems used in these tests. For all the tests, an analogue input was sent directly to the monitors. However, this is not to say that no sample-rate conversion had subsequently taken place in the amplifiers of the loudspeakers systems, or the on-board DSP, because each system used its own dedicated means for delivering the signal to the loudspeaker voice coils. Some of the monitors which were used were supplied with clear information about how that signal is finally delivered, but other manufacturers were not so forthcoming. It must also be taken in to consideration that some of the monitoring systems had been equalised to compensate for some acoustical characteristics of the rooms.

“Any result of the analysis of an acoustic space that is represented as a simple number or graphical line must be taken with a huge amount of caution, and claims that “This is the result”, without further qualification of that result, are at best for entertainment purposes only, as they serve no function and would most certainly not withstand solid scientific scrutiny.” (Newell, J., 2021) Consequently, whether this equalisation has been conducive to improved resolution, or not, is an open question, but it does represent the current *status quo*.

A major factor in this listening process was also to consider the acoustic spaces in which the tests were conducted. When listening to monitor loudspeaker systems, no matter how good or bad we may consider them to be, the acoustic space must be taken into consideration due to the fact that the sound reaching the listeners ears is a combination of direct sound from the monitors and reflexions in the room.

Moreover, *“It is important to note the difference between simply listening for pleasure and for monitoring purposes here. It is quite possible to listen to a sound in a coloured environment, whether it is a familiar voice or* *a symphony, and recognize the elements that characterize it. This is possible as the brain deconvolves the sound, ….'Monitoring', however, is not consistently possible in many such environments as the artefacts may conceal aspects of the programme, and modify them in ways that make clear objectivity impossible."* (Newell, P., 2003)

Indeed, when we are listening specifically for details, we may not hear in the same way as when we listen for pleasure, in a relaxed manner. The brain can function differently if it is concentrating, and the perception of sound can change. (Newell, P., 2015, also Newell. P. and Holland, K., 2007)

Any or all of these aspects may complicate the question of the 'standardization' of the monitoring, yet this is the reality of how things may currently be judged, even in professional environments. (Section 5.2 will also discuss the likely variability in the listeners, themselves.)

5 DISCUSSION OF RESULTS

5.1 Findings and repeatability

After an initial trial run, repeated tests were made to determine whether any variability in the results could be due to the test-process itself, but after several repetitions, it was concluded that the results were highly repeatable. Table 1, below, shows each listener's response in Test 1 to the questions "Can you hear a difference?", and 'if so, which one do you preferer?'. In each case, there were four responses to the first question for each of the samples. The listeners were played the four following examples in no particular order:

a) computer A playing 44.1ks/sec, computer B playing 96ks/sec

b) computer A playing 44.1ks/sec, computer B playing 44.1ks/sec

c) computer A playing 96ks/sec, computer B playing 44.1ks/sec

d) computer A playing 96ks/sec, computer B playing 96ks/sec

It can be seen from the first three lines of Table 1 that the results were quite consistent between the listeners regarding the audibility (or otherwise) of any difference between the two sample-rates with the different sounds, but when asked about which file was *preferred*, one participant clearly preferred the lower sampling-rate.

Listeners	**Water**	**Drums 1**	**Drums 2**	**Piano**	**Moog**	**TB303**
A	YNYN	YNYN	YNYN	NNNN	NNNN	NNNN
E	YNYN	YNYN	YNYN	NNNN	NNNN	NNNN
J	YNYN	YNYN	YNYN	NNNN	NNNN	NNNN
Prefered file						
A	96ks/sec	96ks/sec	96ks/sec	N/A	N/A	N/A
E	96ks/sec	96ks/sec	96ks/sec	N/A	N/A	N/A
J	44.1ks/sec	44.1ks/sec	44.1ks/sec	N/A	N/A	N/A

Table 1 The results from Test 1

After analysis, the results showing the detection of a difference was not entirely unexpected because a difference was noticed by the author during the tests, even from his position to the left of the left monitor. It was also evident that the lower sampling-rate of the loudspeakers had not inhibited the differentiation between the sampling rates of the files: despite this being a frequently discussed contention in audio circles.

In Test 2, no differences were detected, and hence no preferences existed. Again, however, it was initially suspected that the sampling-rate of less than 96kHz in the monitor system could preclude the detection of any benefit of the higher sampling-rate in the tests, but the results from Test 1 seem to preclude this from being generally relevant.

Listening Test 3 started out very slowly, and after hearing no difference between types of files from the first three sounds, a non-conclusive overall result was expected, just as had occurred in Test 2. Nevertheless, a difference *was* noticed by all listeners when listening to the piano file and the Moog file. Even a change of monitors (from the Genelecs to the Questeds) resulted in the same observation, so the remaining time was spent trying to distinguish between the ambient and drum files, albeit unsuccessfully. A thorough check of equipment was then made, and the tests were repeated, but the results remained the same, as shown in Table 2, below.

Listeners	**Water**	**Drums 1**	**Drums 2**	**Piano**	**Moog**	**TB303**
P	NNNN	NNNN	NNNN	YNYN	YNYN	NNNN
D	NNNN	NNNN	NNNN	YNYN	YNYN	NNNN
Prefered file						
P	N/A	N/A	N/A	44.1ks/sec ‘more natural’	N/A	N/A
D	N/A	N/A	N/A	96ks/sec ‘Brighter’	N/A	N/A

Table 2 The results from Test 3

Test 4 was conducted in the home of the author (himself an experienced recording engineer and producer.) Four further participants took part, and the results are shown in Table 3, below.

Listeners	**Water**	**Drums 1**	**Drums 2**	**Piano**	**Moog**	**TB303**
PD	NNNN	YNYN	NNYN	YNYN	YNNN	NNNN
L	YNYN	YNYN	NNNN	YNYN	YNYN	NNNN
M	YNYN	YNYN	NNNN	YNYN	YNYN	NNNN
JW	YNYN	NNNN	NNNN	YNYN	YNYY	NNNN
Prefered file						
PD	N/A	44.1ks/sec ‘fuller’	96ks/sec. ‘Brighter’	96ks/sec. ‘Brighter’	44.1ks/sec ‘fuller’	N/A
L	96ks/sec. ‘Brighter’	96ks/sec. ‘Brighter’	N/A	96ks/sec. ‘Brighter’	44.1ks/sec ‘more edge’	N/A
M	44.1ks/sec ‘clearer’	96ks/sec. ‘Clearer’	N/A	44.1ks/sec ‘clearer’	96ks/sec ‘Brighter’	N/A
JW	44.1ks/sec ‘fuller’	N/A	N/A	96ks/sec. ‘More realist-ic’		N/A

Table 3 The results from Test 4

Other than the TB303 file appearing to show no detectable difference, there is little else in the four tests that shows any degree of consistency in either the detectability of differences or the preferences of the different sample rates with the different sounds. This fact is made even more apparent if viewed in the overall table in Appendix 9.1.

5.2 Possible criticisms of the tests

One possible criticism of this test is that perhaps a more expensive audio interface could have been used, which may have resulted in different findings. However, such a test would say little about the more general detectability of the differences. In the event, differences *were* heard, but making similar tests with only 'high-end' converters is certainly something for the future.

Another possible criticism is in regard to the ages of the participants and their expected hearing acuity. Despite the fact that the listeners were not screened for hearing loss, the general level of detectability, or otherwise, was largely similar.

Indeed, had it been the case that the differences were only detectable by people of 'normal' hearing via very high-quality converters, the results may not have been as relevant to the wider population.

6 CONCLUSIONS

Even though this particular project is only in its preliminary stages, and with a statistically insufficient number of tests completed, it is important to remember that the primary question asked was simply 'Can a difference be heard between these two formats?' In response to that question, from the results presented here (and putting the reasons aside), it can be said that the sounds which were recorded in this manner, and with the system described, *did* sound recognizably different in some situations, and quite noticeably so.

Further planned tests, using the same recorded files and a superior digital-to-analogue conversion system, will go some way towards eliminating the question of the converter quality in the interfaces used. Nevertheless, the differences of the sound followed the files when swapped across computers, and so this is probably not the issue. Also, modern converters for 'serious' music-recording equipment are almost all of a very respectable performance compared with those of 20 years ago.

The next most obvious stage would be to try the recordings using a different system entirely, and producing a new set of files before repeating the testing process. This argument brings forth the question that if a higher-quality conversion of analogue-to-digital can record at different rates and get the same results, then where are the cheaper interfaces failing in consistency?

The results of the tests were somewhat surprising for all parties involved. However, the consistency appears to follow each test location rather than any particular file or group of participants. So, if this is the case, does it imply that we should re-assess the Watkinson-Salter proposal from the Reproduced Sound 1999 conference? (See Section 1) Should we consider the resolution of the loudspeaker chains – all the way to the ears – as information channels? Also, if so, what part of the response could be responsible for the differences in resolution?

Furthermore, the audible detectability of the different sampling rates may also depend on the hearing systems of individual people, although not necessarily on the 'hearing acuity' in its conventionally-measured ways. The secondary question about the participants' *preferences* of the sounds surely raises this issue, as well as how much the concept of 'preference' gets interchanged with the concept of 'better quality'. In this respect, even though a difference may be heard by several participants, is there any justification in generalizing that the higher sampling rate is automatically 'better'?

In the Introduction of this report, it mentions that ever since the introduction of higher sampling-frequencies there has been a continuing debate as to whether there is really is any *improvement* of sound quality, or not. Perhaps some of the findings reported in this document can go some way to explaining why such a debate still continues

7 ACKNOWLEDGEMENTS

The research reported in this document took place during the undertaking of an MA at the University of Central Lancashire. The author would like to thank the following people for their help in facilitating and/or participating in the experiments: Ramon Galarza, Alessandro Cortini, Branko Neskov, Teresa Niza Braga, and Julius Newell in Lisbon, and Graham Lynch in the UK. Thanks also to Keith Holland and Philip Newell for consultation during the draft of the paper.

8 REFERENCES

Newell, J. *Sound System Analysis. The Flawed Science of Measuring the Unmeasurable*. Newellacousticengineering.com. 2021. [online] Available at: <http://www.newellacousticengineering.com/assets/sound-system-analysis.pdf> [Accessed 1 July 2021].

Newell, P. 2003. *Recording Studio Design*. 1st ed., Oxford, UK, Focal Press, pp.407,408

Newell, P., 2015. *Music Mastering - An Inside View*, Comora, Ukraine.

Newell, P. and Holland, K., 2007. *Loudspeakers*, 1st ed., Oxford, Focal Press.

Pras, A. and Guastavino, C., 2010. *Sampling Rate Discrimination: 44.1 kHz vs. 88.2 kHz*. [online] researchgate. Available at: <https://www.researchgate.net/publication/257068631_Sampling_Rate_Discrimination_441_kHz_vs_882_kHz> [Accessed 2 July 2021].

Shannon, C. and Weaver, W., 1999. *The Mathematical Theory of Communication*, Urbana: The University of Illinois Press

Watkinson, J., Salter, R., 1999, *Modelling and Measuring the Loudspeaker as an Information Channel*, Institute of Acoustics *Reproduced Sound* conference, Stratford upon Avon,

9 APPENDIX

9.1 Combined table including all results

Note 1: in the following table:

Green text indicates a confirmation that the listener has identified difference files

Blue (italics) indicates un unclear result.

Red (bold) text indicates no difference heard

Note 2:

All files included in tests are available to download with the following link:

https://drive.google.com/drive/folders/1OdyOlgFjesK6gZCpVk67FvTKiyuhbdYf?usp=sharing

Listener	Age	Profession	Water	Drums 1	Drums 2	Piano	Moog	TB303
A	45	Musician	YNYN	YNYN	YNYN	**NNNN**	**NNNN**	**NNNN**
E	35	Photographer	YNYN	YNYN	YNYN	**NNNN**	**NNNN**	**NNNN**
J	51	Audio expert	YNYN	YNYN	YNYN	**NNNN**	**NNNN**	**NNNN**
G	29	Mastering eng.	**NNNN**	**NNNN**	**NNNN**	**NNNN**	**NNNN**	**NNNN**
P	55	Lecturer	**NNNN**	**NNNN**	**NNNN**	YNYN	YNYN	**NNNN**
D	30	Sound Engineer	**NNNN**	**NNNN**	**NNNN**	YNYN	YNYN	**NNNN**
PD	40	musician	**NNNN**	YNYN	*NNYN*	YNYN	YNNN	**NNNN**
L	26	Musician	YNYN	YNYN	**NNNN**	YNYN	YNYN	**NNNN**
M	21	Photographer	YNYN	YNYN	**NNNN**	YNYN	YNYN	**NNNN**
JW	39	Deejay	YNYN	**NNNN**	**NNNN**	YNYN	*YNYY*	**NNNN**

9.2 Figures

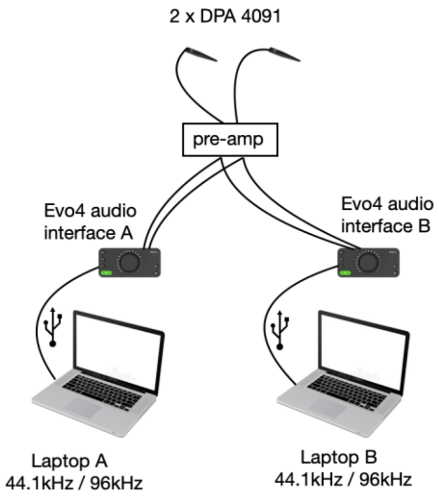


Figure 1. Diagram showing configuration of the recording system

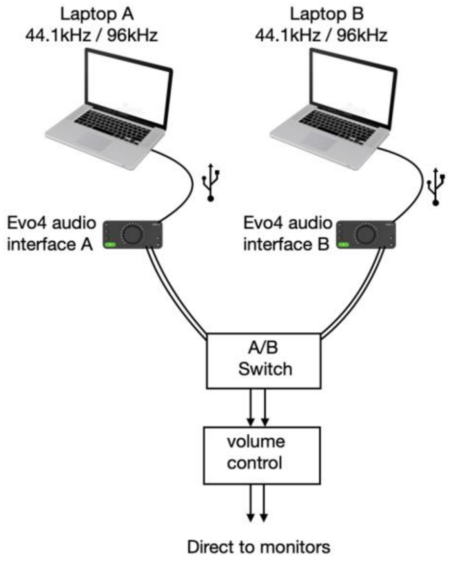


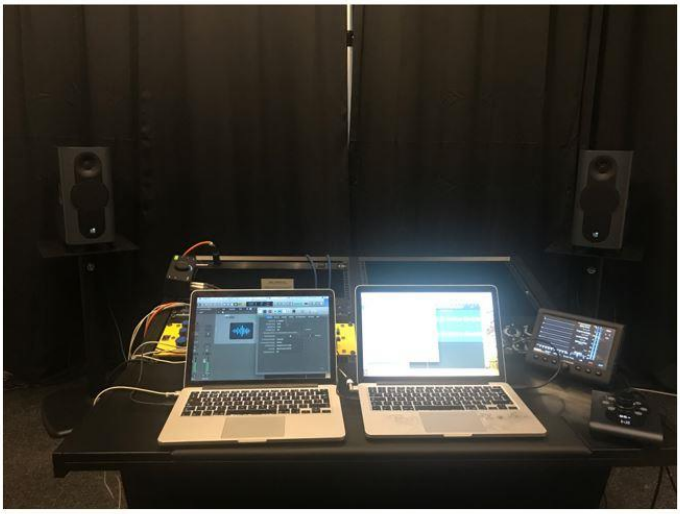
Figure 2 Diagram showing the configuration of the playback system

9.3 Photographs of the recording and playback systems

Recording the drums at Loudness Films

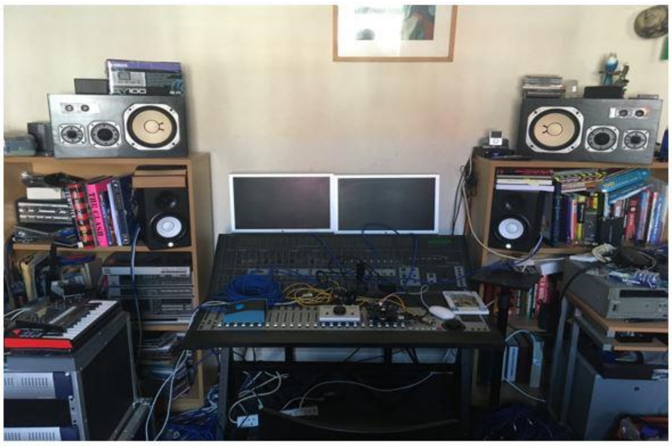
The recording set-up at Loudness Films

The set-up for Test 1



The set-up for Test 2

The set-up for Test 3



The set-up for Test 4