Forensic Resources Ltd

CBTC, Senghenydd Road
Cardiff CF24 4AY
Telephone: 02920 647 043
Facsimile: 02920 647 009
Email: info@forensicresources.co.uk
Website: www.forensicresources.co.uk

Audio Analysis

Amarjit Singh

Prepared by:

James Zjalic

CONTENTS

1. INTRODUCTION

1.01 Examiner

My name is James Zjalic. My specialist field is audio-visual forensics. I hold two degrees in this area, namely a Masters in Media Forensics and a 1st Class Bachelors in Audio Engineering. Experience includes peer-reviewed and published research on the subjects of audio authentication and enhancement, presentations at international conferences and multiple articles in magazines including Sound on Sound, Digital Forensics Magazine and Forensic Focus. I also held a position on an ongoing US Air Force / DARPA funded multimedia analysis research project (Codename: MEDIFOR) and have performed numerous media authentication examinations for organisations such as the United Nations. I attend and give presentations at national and international forensics conferences annually, and undertake CPD courses in a number of disciplines related to my field.

Full details of my qualifications entitling me to give expert opinion evidence are documented within the Appendix of this report.

1.02 Summary of the Case

The following is taken verbatim from the answers to an authentication questionnaire completed by the instructing party:

'The defence the perpetrator has claimed is that snippets of the recordings are placed together in a way which sounds incriminating but taken out of context.

The accusing victim maintains that the recordings have not been chopped/changed as per the defence claims and are genuine conversations.'

1.03 Information Provided Regarding Instruction

I was instructed on the 30th of January 2019 on behalf of Amarjit SINGH to perform audio authentication examinations on four (4) audio recordings.

1.04 Issues to be Addressed

1. Analysis of the provided exhibits to determine whether manipulation of any form has occurred.

1.05 Purpose of this Report

To provide an expert opinion in the field of audio forensics to assist the court in this case.

1.06 Technical Terms

All technical terms have been indicated in italicized typeface when first used. These terms are defined within a glossary located within the Appendix. A list of references which have been consulted in the creation of this report can also be found in the Appendix.

1.07 Exhibits

The following steps were taken on receiving the evidence.

1. Digital copies were created to ensure provenance is maintained;
2. The original files name, size, and HASH checksums were documented;
3. The digital copies name, size, and HASH checksums were documented;
4. Comparisons were then made between the documentation to verify that exact bit stream copies had been created.

Please refer to the Appendix 'List of Exhibits Examined' for full details on the documents and files on which the results of this examination rely.

1.08 Conflicts of Interest

There are no conflicts of interest of which I am aware in relation to any of the people or organisations involved in the case instructed.

I have no personal or business connection with any of the parties involved in this case. I am independent of the parties and intend to remain so throughout the course of this matter before the

Court. There are therefore no conflicts of interest, of which I am aware, in relation to any of the people or organisations involved in the case instructed.

1.09 Disclaimer

The results of this report only relate to the evidence audio analysed for this examination.
Any information supplied by the client and relied upon within this report can affect the validity of the results.

2. EXECUTIVE SUMMARY

Analysis resulted in four (4) conclusions, each related to a single exhibit:

1. It is my opinion, based on the analysis performed within this examination, that the recording entitled 'REC1.AAC' is *inconsistent* with an *original* recording. Further to that, it is *consistent* with a recording which has been *acoustically re-recorded* with *no identifiable areas of manipulation* within the recording.
2. It is my opinion, based on the analysis performed within this examination, that the recording entitled 'REC2.AAC' is *inconsistent* with an *original* recording. Further to that, it is *consistent* with a recording which has been *acoustically re-recorded* with *no identifiable areas of manipulation* within the recording.
3. It is my opinion, based on the analysis performed within this examination, that the recording entitled 'REC3.AAC' is *inconsistent* with an *original* recording. Further to that, it is *consistent* with a recording which has been *acoustically re-recorded* with *no identifiable areas of manipulation* within the recording.
4. It is my opinion, based on the analysis performed within this examination, that the recording entitled 'REC4AND5.M4A' is *inconsistent* with an *original* recording. Further to that, it is *consistent* with a recording which has been *acoustically re-recorded* with *no identifiable areas of manipulation* within the recording.

For further analysis, it is therefore of the utmost importance that the original versions of all recordings, before any form of re-recording has taken place, are provided.

3. FACTS, ASSUMPTIONS & LIMITATIONS

3.01 Facts

The following are facts relied upon for the analyses documented within this report:

1. According to the instructing party, the recordings are of mobile phone conversations;
2. According to the instructing party, *'REC1.aac'* and *'REC2.aac'* were originally captured using a Motorola Android application before being replayed through the speaker of the original device and captured by an iPhone device;
3. According to the instructing party, *'REC3.aac'* was captured through having a call conversation on loudspeaker whilst being captured by an iPhone device.

3.02 Assumptions

No assumptions are made within this report.

3.03 Limitations

The following are the limitations of the analysis documented within this report:

1. The range of analyses available are limited by the processing which the provided exhibits have already been accepted to have undergone;
2. Comparative analysis against reference recordings is not possible as:
    - The acoustic environment, positioning of devices and levels of playback are not known so cannot be reliably recreated.

4. ANALYSIS

4.01 Abbreviations

From this point forward:

- 'REC1.aac' will be referred to as Ex1;
- 'REC2.aac' will be referred to as Ex2;
- 'REC3.aac' will be referred to as Ex3;
- 'REC4AND5.m4a' will be referred to as Ex45.

4.02 Statement of Methodology

To perform an exhaustive assessment multiple analyses were performed following a framework taken from a research paper published by Dr. Catalin Grigoras, Daniel Rappaport and Jeff M. Smith entitled "Analytical Framework for Digital Audio Authentication".

4.03 Waveform Statistics

The specifications of the audio data are documented in Table 1. The data shows the recordings have not been normalised, a process in which recordings are normalised to an integer peak sample level such as -1dB. Normalisation can occur when recordings are processed, and generally wouldn't be expected from unprocessed recordings.

All recordings are captured at an amplitude beyond that of the capabilities of the recording device, causing clipping and a 0.00 dBFS peak sample rate. Recordings Ex1 – Ex3 show consistencies between Max RMS Level, Min RMS Level, Total RMS Level and DC Offset. This would be expected from recordings captured under the same conditions (for example the speaker to microphone geometry, room and type of recording).

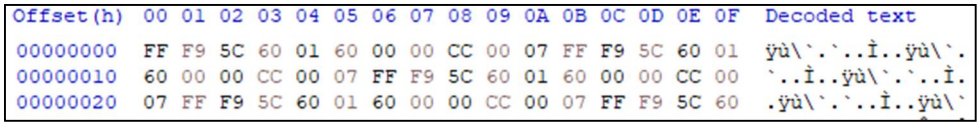
Table 1: Audio Data

Exhibit	Ex1	Ex2	Ex3	Ex45
Sample Peak Level	0.00 dB	0.00 dB	0.00 dB	0.00 dB
Max RMS Level	-3.40 dB	-4.27 dB	-4.77 dB	-4.44 dB
Min RMS Level	-81.59 dB	-81.91 dB	-81.76 dB	-65.90 dB
Total RMS Level	-14.19 dB	-16.15 dB	-16.70 dB	-21.87 dB
Possibly Clipped Samples	1961	337	25	43
DC Offset	+0.001%	0.000%	0.000%	-0.001%

4.04 Hex Offsets

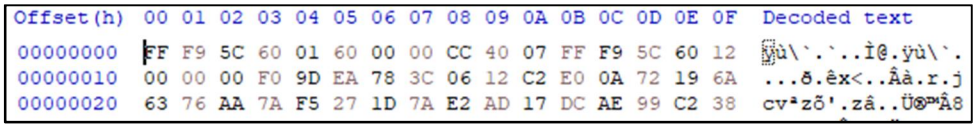
Hex data can reveal traces of editing software, which in this case, there are none. The structure of recordings Ex1 – Ex3 are consistent with one another, indicating coming from the same device.

The header and footer structure of exhibit Ex45 is consistent with that of an iPhone Voice Memo recording.



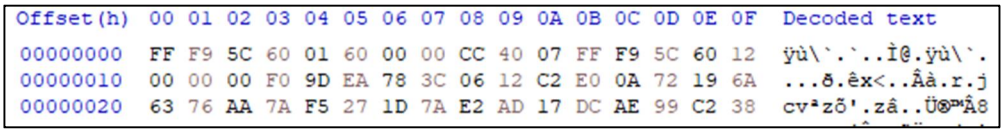
Offset(h)	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F	Decoded text
00000000	FF	F9	5C	60	01	60	00	00	CC	00	07	FF	F9	5C	60	01	ÿù\`.`..Ì..ÿù\`.
00000010	60	00	00	CC	00	07	FF	F9	5C	60	01	60	00	00	CC	00	`..Ì..ÿù\`.`..Ì.
00000020	07	FF	F9	5C	60	01	60	00	00	CC	00	07	FF	F9	5C	60	.ÿù\`.`..Ì..ÿù\`

Figure 1: Ex1 File header



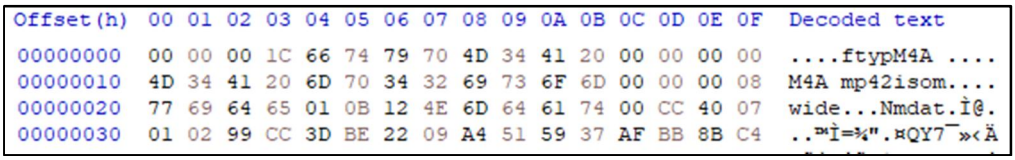
Offset(h)	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F	Decoded text
00000000	FF	F9	5C	60	01	60	00	00	CC	40	07	FF	F9	5C	60	12	ÿù\`.`..Ì@.ÿù\`.
00000010	00	00	00	F0	9D	EA	78	3C	06	12	C2	E0	0A	72	19	6A	...ð.êx<..Âà.r.j
00000020	63	76	AA	7A	F5	27	1D	7A	E2	AD	17	DC	AE	99	C2	38	cvªzõ'.zâ..Ü®™Â8

Figure 2: Ex2 File Header



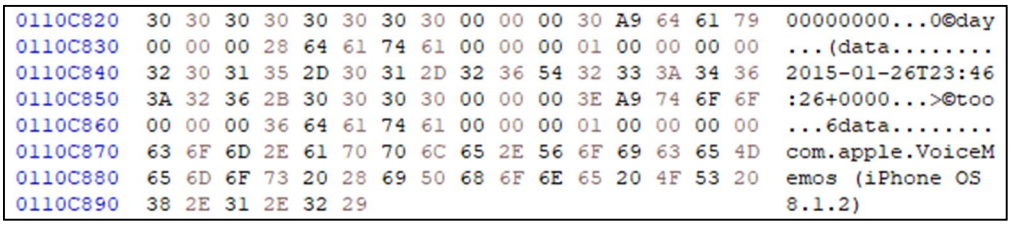
Offset(h)	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F	Decoded text
00000000	FF	F9	5C	60	01	60	00	00	CC	40	07	FF	F9	5C	60	12	ÿù\`.`..Ì@.ÿù\`.
00000010	00	00	00	F0	9D	EA	78	3C	06	12	C2	E0	0A	72	19	6A	...ð.êx<..Âà.r.j
00000020	63	76	AA	7A	F5	27	1D	7A	E2	AD	17	DC	AE	99	C2	38	cvªzõ'.zâ..Ü®™Â8

Figure 3: Ex3 File Header



Offset(h)	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F	Decoded text
00000000	00	00	00	1C	66	74	79	70	4D	34	41	20	00	00	00	00	....ftypM4A ....
00000010	4D	34	41	20	6D	70	34	32	69	73	6F	6D	00	00	00	08	M4A mp42isom....
00000020	77	69	64	65	01	0B	12	4E	6D	64	61	74	00	CC	40	07	wide...Nmdat.Ì@.
00000030	01	02	99	CC	3D	BE	22	09	A4	51	59	37	AF	BB	8B	C4	..™Ì=¾".¤QY7¯»‹Ä

Figure 4: Ex45 File Header



0110C820	30	30	30	30	30	30	30	30	00	00	00	30	A9	64	61	79	00000000...0©day
0110C830	00	00	00	28	64	61	74	61	00	00	00	01	00	00	00	00	...(data........
0110C840	32	30	31	35	2D	30	31	2D	32	36	54	32	33	3A	34	36	2015-01-26T23:46
0110C850	3A	32	36	2B	30	30	30	30	00	00	00	3E	A9	74	6F	6F	:26+0000...>©too
0110C860	00	00	00	36	64	61	74	61	00	00	00	01	00	00	00	00	...6data........
0110C870	63	6F	6D	2E	61	70	70	6C	65	2E	56	6F	69	63	65	4D	com.apple.VoiceM
0110C880	65	6D	6F	73	20	28	69	50	68	6F	6E	65	20	4F	53	20	emos (iPhone OS
0110C890	38	2E	31	2E	32	29											8.1.2)

Figure 5: Ex45 File Footer

4.05 Critical Listening

The following are findings from critical listening:

Ex1

1. Consistent broadband noise throughout;
2. There are no discontinuities between the flow of questions and answers;
3. No inconsistencies within the chronology of the conversation;
4. All audible artefacts are explainable as related to distortions;
5. No inconsistencies within the chronology of the conversation;
6. Reverberation profiles of speakers are consistent throughout;
7. Amplitude levels of speakers are consistent throughout;
8. Speaker to noise ratio consistent throughout;
9. All clicks and pops are explainable as they contain reverberation and/or speech across the section;
10. No breaks in recording consistent with an edit point;
11. Two capture channels: voice coming through phone consistent with quality of phone transmission capture, other is consistent with captured within the room;
12. Unnatural cut off to the recording, consistent with recording of a recording.

Ex2

1. Consistent broadband noise throughout;
2. There are no discontinuities between the flow of questions and answers;
3. No inconsistencies within the chronology of the conversation;
4. All audible artefacts are explainable as related to distortions;
5. Reverberation profiles of speakers are consistent throughout;
6. All audible artefacts are explainable as related to distortions;
7. Amplitude levels of speakers are consistent throughout;
8. Speaker to noise ratio consistent throughout;
9. All clicks and pops are explainable as they contain reverberation and/or speech across the section;
10. No breaks in recording consistent with an edit point;
11. Two capture channels: voice coming through phone consistent with quality of phone transmission capture, other is consistent with captured within the room;
12. Presence of third speaker provides reference of speaker within room;
13. Noise profile between 04:38 and 04:48 indicates change to speaker to microphone geometry.

Ex3

1. Consistent broadband noise throughout;
2. There are no discontinuities between the flow of questions and answers;
3. No inconsistencies within the chronology of the conversation;
4. All audible artefacts are explainable as related to distortions;
5. Reverberation profiles of speakers are consistent throughout;
6. Amplitude levels of speakers are consistent throughout;
7. Speaker to noise ratio consistent throughout;
8. All clicks and pops are explainable as they contain reverberation and/or speech across the section;
9. No breaks in recording consistent with an edit point;
10. Two capture channels: voice coming through phone consistent with quality of phone transmission capture, other is consistent with captured within the room;

Ex45

1. Consistent broadband noise throughout;
2. There are no discontinuities between the flow of questions and answers;
3. No inconsistencies within the chronology of the conversation;
4. All audible artefacts are explainable as related to distortions;
5. Reverberation profiles of speakers are consistent throughout;
6. Amplitude levels of speakers are consistent throughout;
7. Consistent speaker to noise ratio;
8. All clicks and pops are explainable as they contain reverberation and/or speech across the section;
9. No breaks in recording consistent with an edit point;
10. Two capture channels: voice coming through phone consistent with quality of phone transmission capture, other is consistent with captured within the room;
11. Second phone call within recording provides reference point as to the re-recording with has taken place.

4.06 Format and EXIF

Recordings Ex1 – Ex3 contained no EXIF metadata, so no analysis of such could be performed. This also indicates different post processing of these recordings compared to Ex45.

The specifications of Ex45 are consistent with a capture using the Voice Memos App within an iPhone running iOS 8.1.2. The recording bit-rate indicates no edits have been made within the Voice Memos application (edited recordings are transcoded to 256 kbps, the bit-rate in this case is 64.6 kbps, consistent with an original voice memo recording). The 1 second disparity between the create date and modify date is due to the time taken between the end of capture and the file being encoded in an M4A file container.

The length of the recording within the EXIF data is consistent with the actual length of the recording (00:36:06 and 36:05:957).

Capture date is shown as 2015:01:27 00:22:52, but it must be considered that any temporal data is taken from the internal clock of the device which is definable by the user within the settings. It must therefore be treated with caution. As the file was acoustically re-recorded, no temporal data in relation to the original can be determined.

Table 1: EXIF data

File Name	REC4AND5.m4a
File Size	17 MB
File Type	M4A
File Type Extension	m4a
MIME Type	audio/mp4
Major Brand	Apple iTunes AAC-LC (.M4A) Audio
Minor Version	0.0.0
Compatible Brands	M4A
Movie Data Size	17502790
Movie Data Offset	44
Movie Header Version	0
Create Date	2015:01:27 00:22:52
Modify Date	2015:01:27 00:22:53
Time Scale	44100
Duration	00:36:06
Preferred Rate	1
Preferred Volume	100.00%
Preview Time	0 s

Preview Duration	0 s
Poster Time	0 s
Selection Time	0 s
Selection Duration	0 s
Current Time	0 s
Next Track ID	2
Track Header Version	0
Track Create Date	2015:01:27 00:22:52
Track Modify Date	2015:01:27 00:22:53
Track ID	1
Track Duration	00:36:06
Track Layer	0
Track Volume	100.00%
Matrix Structure	1 0 0 0 1 0 0 0 1
Media Header Version	0
Media Create Date	2015:01:27 00:22:52
Media Modify Date	2015:01:27 00:22:53
Media Time Scale	44100
Media Duration	00:36:06
Media Language Code	und
Handler Description	Core Media Audio
Balance	0
Audio Format	mp4a
Audio Channels	2
Audio Bits Per Sample	16
Audio Sample Rate	44100
Handler Type	Metadata
iTunSMPB	0 840 0 5B17BC0 0 0 0 0 0 0 0 0 0
Content Create Date	2015:01:26 23:46:26+00:00
Encoder	com.apple.VoiceMemos (iPhone OS 8.1.2)
Avg Bitrate	64.6 kbps

4.07 Visual Analysis

Bandwidth Extensions

There are a number of bandwidth extensions within Ex1 – 3 [Fig. 6], which are above the low-pass filter used for AAC encoding. These are related to interpolation at a higher sample rate (in which an area of the signal is reconstructed based on the surrounding samples), and can be indicative of a cross fade applied at an edit point. In this case they are all consistent with clipping, in which the input signal is higher in amplitude than the capabilities of the recording device so interpolation is applied to these areas to reduce any artefacts which can be caused by such an occurrence.

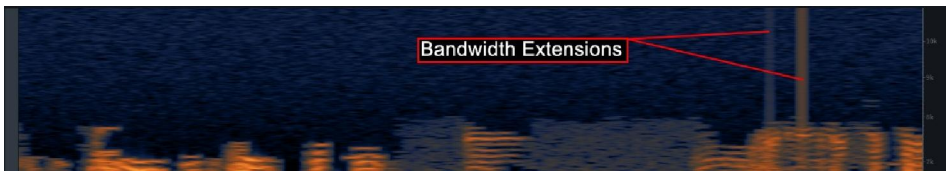


Figure 6: Bandwidth Extensions

Frequency Dropouts

The frequencies present within a recording at any one period of time are dependent on a myriad of factors. For recordings which are captured from speakers within rooms, the soundwaves interaction with the environment can cause gaps in the spectrum due to frequencies which have gone uncaptured. Movement of either the playback or capture device can cause the interactions of the sound with the room to change, resulting in a change in the areas at which frequencies are uncaptured. This change occurs in Ex2 between 04:38 and 04:48 and is identifiable both audibly through the change in broadband background noise and visually through the frequency spectrum [Fig. 7].

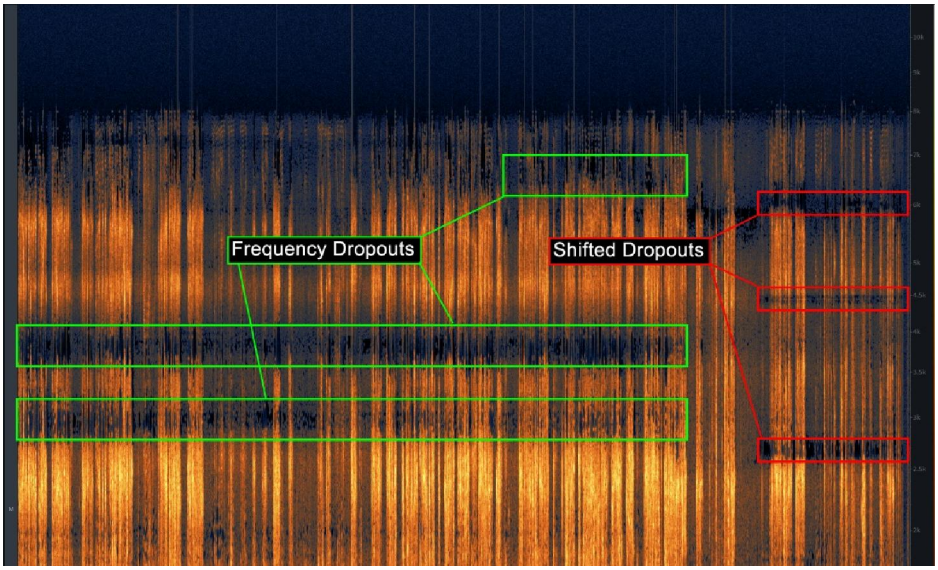


Figure 7: Frequency Dropouts

Speaker Frequencies

The bandwidths occupied by the speakers are consistent throughout the recordings.

Speaker at the other end of phone: 400Hz – 8000Hz;

Speaker at this end of phone: 200Hz – 8000Hz;

Speaker within room: 100Hz – 8000Hz.

As the bandwidth of telephone transmission only extends to 4000Hz, the extended range of the speaker at the other end of the phone is caused by the frequency response of the speaker playing the recording back and the interaction of such with the room.

4.08 Spectrum

Figure 8 shows the content of Ex1 - 3 in terms of frequencies present. For a recording captured at 22.05kHz, the highest frequency component would be expected to be approximately 11.025 kHz. This is consistent with the findings for these exhibits. AAC algorithms work by removing data which is perceived less by the human auditory system, in this case by applying a filter to remove all frequencies above 8kHz.This can be seen within Figure 4.

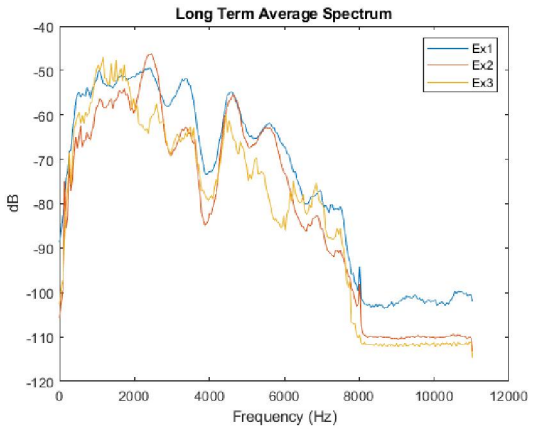


Figure 8: Ex1 - 3 LTAS

The Ex45 spectrum indicates capture at 44.1kHz with a perceptual encoding filter applied at approximately 15kHz [Fig. 9].

As neither recording is purported to have been captured directly, the acoustics of the room and frequency response of the speaker from which the recording was replayed will have an impact on the frequencies present.

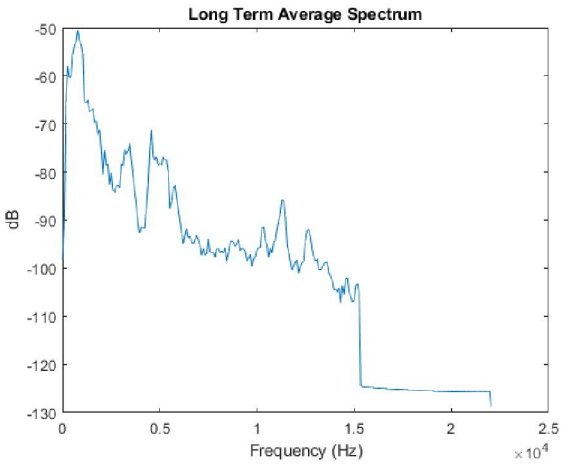


Figure 9: Ex45 LTAS

Any inconsistencies or non-uniform regions within the Long-Term Average Sorted Spectrum (LTASS) can be signs of manipulations. This is not the case for the exhibit recording.

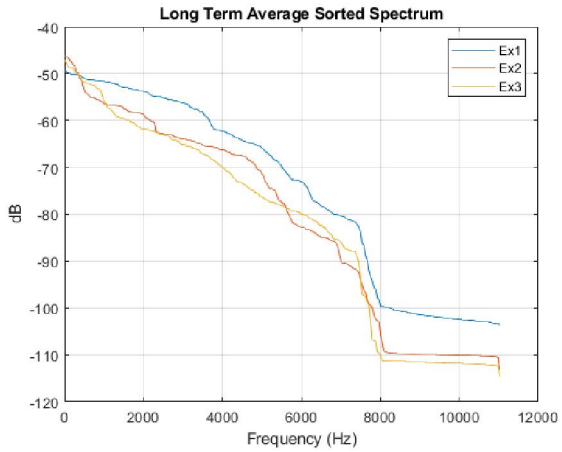


Figure 10: Ex1 - 3 LTASS

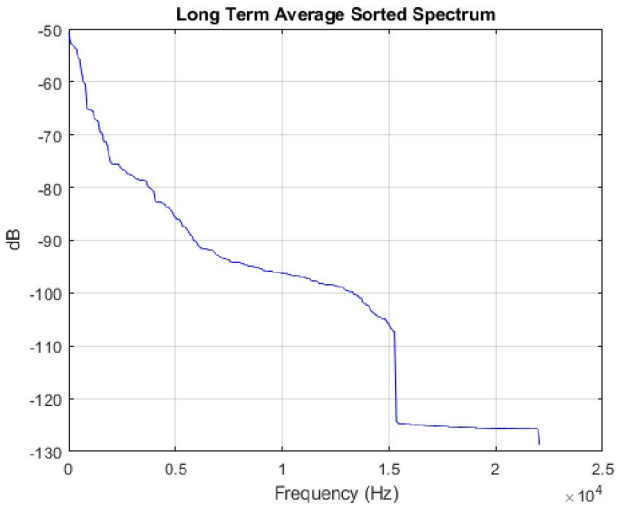


Figure 11: Ex45 LTASS

5. OPINION

5.01 Ex1

It is my opinion, based on the analysis performed within this examination, that the recording entitled 'REC1.AAC' is *inconsistent* with an *original* recording. Further to that, it is *consistent* with a recording which has been *acoustically re-recorded* with *no identifiable areas of manipulation* within the recording.

For further analysis, it is of the utmost importance that the original version of this recording, before any form of re-recording has taken place, is provided.

This is based on the following premises:

1. No visual spectrogram inconsistencies;
2. Speaker levels consistent throughout;
3. Reverberation profiles of speakers consistent throughout;
4. Speaker to Noise level consistent throughout;
5. Broadband noise consistent throughout;
6. LTAS consistent with sample-rate;
7. LTASS shows no inconsistencies;
8. File structure consistent with file extension;
9. No clicks or pops which would be consistent with a butt splice;
10. There are no discontinuities between the flow of questions and answers;
11. No inconsistencies within the chronology of the conversation;
12. All events are explainable;
13. Bandwidth extensions consistent with interpolation from clipped samples and not edit points.

5.02 Ex2

It is my opinion, based on the analysis performed within this examination, that the recording entitled 'REC2.AAC' is *inconsistent* with an *original* recording. Further to that, it is *consistent* with a recording which has been *acoustically re-recorded* with *no identifiable areas of manipulation* within the recording.

For further analysis, it is of the utmost importance that the original version of this recording, before any form of re-recording has taken place, is provided.

This is based on the following premises:

1. No visual spectrogram inconsistencies;
2. Speaker levels consistent throughout;
3. Reverberation profiles of speakers consistent throughout;
4. Speaker to Noise level consistent throughout;
5. Broadband noise consistent throughout;
6. LTAS consistent with sample-rate;
7. LTASS shows no inconsistencies;
8. File structure consistent with file extension;
9. No clicks or pops which would be consistent with a butt splice;
10. There are no discontinuities between the flow of questions and answers;
11. No inconsistencies within the chronology of the conversation;
12. All events are explainable;
13. Bandwidth extensions consistent with interpolation from clipped samples and not edit points;
14. Frequency dropout consistent with changes to speaker to microphone geometry.

5.03 Ex3

It is my opinion, based on the analysis performed within this examination, that the recording entitled 'REC3.AAC' is *inconsistent* with an *original* recording. Further to that, it is *consistent* with a recording which has been *acoustically re-recorded* with *no identifiable areas of manipulation* within the recording.

For further analysis, it is of the utmost importance that the original version of this recording, before any form of re-recording has taken place, is provided.

This is based on the following premises:

1. No visual spectrogram inconsistencies;
2. Speaker levels consistent throughout;
3. Reverberation profiles of speakers consistent throughout;
4. Speaker to Noise level consistent throughout;
5. Broadband noise consistent throughout;
6. LTAS consistent with sample-rate;
7. LTASS shows no inconsistencies;
8. File structure consistent with file extension;
9. No clicks or pops which would be consistent with a butt splice;
10. There are no discontinuities between the flow of questions and answers;
11. No inconsistencies within the chronology of the conversation;
12. All events are explainable;
13. Bandwidth extensions consistent with interpolation from clipped samples and not edit points.

5.04 Ex45

It is my opinion, based on the analysis performed within this examination, that the recording entitled 'REC4AND5.M4A' is *inconsistent* with an *original* recording. Further to that, it is *consistent* with a recording which has been *acoustically re-recorded* with *no identifiable areas of manipulation* within the recording.

.

For further analysis, it is of the utmost importance that the original version of this recording, before any form of re-recording has taken place, is provided.

This is based on the following premises:

1. No visual spectrogram inconsistencies;
2. Speaker levels consistent throughout;
3. Reverberation profiles of speakers consistent throughout;
4. Speaker to Noise level consistent throughout;
5. Broadband noise consistent throughout;
6. LTAS consistent with sample-rate;
7. LTASS shows no inconsistencies;
8. File structure consistent with file extension;
9. No clicks or pops which would be consistent with a butt splice;
10. There are no discontinuities between the flow of questions and answers;
11. No inconsistencies within the chronology of the conversation;
12. All events are explainable;
13. Bandwidth extensions consistent with interpolation from clipped samples and not edit points.

6. STATEMENTS AND DECLARATIONS

6.01 Statement of Compliance

I understand my duty as an expert witness is to the court. I have complied with that duty and will continue to comply with it. This report includes all matters relevant to the issues on which my expert evidence is given. I have given details in this report of any matters which might affect the validity of this report. I have addressed this report to the court. I further understand that my duty to the court overrides any obligation to the party from who I received instructions.

6.02 Declaration of Awareness

I confirm that I am aware of the requirements of Part 35 and Practice Direction 35, and the Guidance for the Instruction of Experts in Civil Claims 2014.

6.03 Statement of Truth

I confirm that I have made clear which facts and matters referred to in this report are within my own knowledge and which are not. Those that are within my own knowledge I confirm to be true. The opinions I have expressed represent my true and complete professional opinions on the matters to which they refer.

6.04 Statement of Conflicts

I confirm that I have no conflict of interest of any kind, other than any which I have already set out in this report. I do not consider that any interest which I have disclosed affects my suitability to give expert evidence on any issue on which I have given evidence and I will advise the party by whom I am instructed if, between the date of this report and the trail, there is any change in circumstances which affects this statement.

Signature

……

James Zjalic

15/02/2019

---------- End of Report ----------

APPENDIX 1 – GLOSSARY

Authentication Glossary

Audio Authentication: The process of determining whether an audio recording is an accurate representation of its origin.

Consecutive Zeros: Consecutive zero padding is performed at the beginning of compressed recordings such as the MP3 format. Different software apply a different number of zeros, and zeros are added each time a new save of the same recording is performed.

Format Analysis: Analysis of the format of the audio recording, including the file type, sample-rate and compression rate to determine if a recording of this type could be made with the system in question and/or to look for inconsistencies between recordings.

LTAS: Long Term Average Spectrum calculates the average frequencies across the audio recording, resulting in an overall representation of the signal. Similar recordings made on the same system are expected to have similar LTAS.

LTASS: Long Term Average Sorted Spectrum calculates the average frequencies across the audio recording, resulting in an overall representation of the signal before sorting them to provide an overall view of the frequency response of the recording.

Quantization Level Analysis: Analysis of the compression performed when a recording is created. Multiple compressed recordings (from multiple saves) will show different compression levels to original recordings.

Structure Analysis: Analysis of the recordings file structure, including the header and footer. Different software structure the data in different ways and any differences within the file structure can be indicative of being saved out of a different software.

Audio Quality Glossary

Bit Depth

The number of levels available to represent the volume of the samples captured, in a log 2 format. For example, there would be 65,536 increments in a 16-bit system ($2^{16} = 65,536$). Generally, the higher the bit depth, the higher the quality of audio.

Bit-rate

Bit rate is determined by the bit-depth, sampling frequency and number of channels. For example, a 16 bit, 44100Hz recording with two channels (CD quality) has 1411 kb/s, calculated by multiplying the specifications together.

Generally, a higher bit rate recording will result in higher quality audio.

Channels

Audio recordings are captured by microphones, of which there can be one or multiple. The most common scenarios are as follows:

- Monoaural: Captured by a single microphone, resulting in one channel;
- Double Mono: Captured by a single microphone but converted to two identical channels;
- Stereo: Captured by two microphones, resulting in two differing channels.

Clicks

Clicks and pops in audio are generally defined as undesired audible transients. They are perceived by the listener in many ways, but in the realm of digital audio they are often heard as tiny tick sounds and are caused by poorly concealed digital errors and timing problems within the ADC (Analogue to Digital Converter) of the recording device. These digital clicks and pops can not only be distracting but also cause masking of phenomes, which are the elements of speech which determine the meaning of sound. For example, the C or B in “Cat” or “Bat”.

Clipping

Clipping is caused when there are no more quantisation levels available for the system to store a higher amplitude signal, due to limitations in the systems dynamic range. For example, if the maximum quantization level is 65,536 in an unsigned 16-bit system, any signals that are above this upper limit will be represented as 65,536 as there are no more levels available. Common reasons for clipping are setting the record input level too high, an unexpected loud signal or using a bit depth that is too low to accommodate the type of recording.

Compression

Compression can be either lossless or lossy.

Lossless compression reduces the file size, but no data is lost in the process.

Lossy compression reduces the file size, but information is lost during the process and can never be recovered. This is achieved through the removal of information which has no dramatic effect on the listening experience, such as removing frequencies which the human auditory system is less sensitive to.

Distortion

Distortion can be caused by clipping, system noise or a complete or partial signal loss due to the electronic components of the capture device. For the former, it is not possible to recover any of the audio as it was not captured in the first place.

Hum

Hum is one of the more common noises present on a forensic recording and is caused by electrical sources. One of the most well-known causes is that of ENF (Electrical Network Frequency), which can be inadvertently captured on a recording if the power source is mains driven or in close proximity of a mains powered device, through electromagnetism. This phenomenon is caused by the alternating current (around 50Hz in the UK, 60 Hz in the US) and has harmonics, albeit, at a lower amplitude, that can exist all the way into the bandwidth occupied by speech. Other causes of hum are lights and powerlines.

Reverberation

Reverberation can be commonly thought as of echo. Every environment has a different reverberation profile, based on its size, the materials used to construct the walls and the number of objects within the room amongst many other factors. Reverb can cause the start of words to become unintelligible as there is a smudging of the sound as the reverb tail from the previous word masks the beginning of the current word.

Sampling Frequency

The sampling frequency of an audio recording determines the number of samples of the acoustic environment captured per second. For example, 10kHz would indicate 10,000 samples taken per second. Generally, the higher the sampling frequency, the higher the quality of audio.

APPENDIX 2 – TOOLS

All work took place on a forensic workstation running a Microsoft Windows 10 Operating System displayed on a single 43 inch, 4K Resolution monitor.

Critical Listening was performed using Beyerdynamic DT770 Pro Headphones and a Focusrite Saffire Pro 40 Interface.

Specific software and its application during the analysis featured in this report is as documented in the table below.

Software	Version	Task
Adobe Acrobat	Version 2018.11.2038	Report Output
Jacksum	Version 1.7.0	Calculation of Hash Sums
Microsoft Word	2016	Documentation
EXIF tool	10.51	Format Analysis
HXD	1.7.7.0	Header Analysis
Izotope RX Advanced	7.00	Critical Listening
Matlab	R2017A	Audio Analysis

APPENDIX 3 – EXHIBITS REVIEWED

The following is a list of exhibits reviewed during the work documented within this report to ensure a chain of custody is maintained and to aid in transparency in relation to the exhibits which may have influenced any conclusions reached.

Exhibit	Format	Delivery Method	Description	Date Provided
REC1.aac	AAC	Email Attachment	Exhibit Recording	17/01/2019
SHA256: 2e050d1172088a1615d3467d463f534387243879eb7549cdd02eea67b08d9f74				
REC2.aac	AAC	Email Attachment	Exhibit Recording	17/01/2019
SHA256: 6f17e2a81f8a33d27ecc53c3e162a0b4c84537f7b6670a15282ddcfdc7ac813b				
REC3.aac	AAC	Email Attachment	Exhibit Recording	17/01/2019
SHA256: a8724e05c5455ac8deb8f07ab3dac901d57a50c8ba14a95f79902eb4e6062157				
REC4AND5.m4a	AAC	Email Attachment	Exhibit Recording	17/01/2019
SHA256: 2e0d2b358b84fcd12ee8ac86aa50c1a8e01434f8dbeb56ebcbcc382eb8d459cb				

NOTE: No hash values were delivered with the exhibit, so cannot be verified.

APPENDIX 4 – BIBLIOGRAPHY

Guidance

SWGDE, “Digital and Multimedia Evidence Glossary Version 3.0.” 20-Jun-2016.

SWGDE, “Best Practices for Forensic Audio.” 08-Oct-2016.

Theory

C Grigoras and JM Smith, “Audio Enhancement and Authentication,” in *Encyclopaedia of Forensic Sciences*, Second Edition., Elsevier Ltd, 2013, pp. 315–326.

Bruce Koenig and Catalin Grigoras, “Digital Audio Authentication Workshop,” presented at the 2017 AES Audio Forensics Conference, Arlington, VA, Jun-2017.

Anthony T.S Ho and Shujun Li, *Handbook of Digital Forensics and Multimedia Data Devices*. UK: John Wiley & Sons, Ltd, 2015.

C. Grigoras, D. Rappaport, and J. M. Smith, “Analytical framework for digital audio authentication,” in *Audio Engineering Society Conference: 46th International Conference: Audio Forensics*, 2012.

ACPO, “Good Practice Guide for Digital Evidence.” 2012.

A. J. Cooper, “Detecting butt-spliced edits in forensic digital audio recordings,” in *Audio Engineering Society Conference: 39th International Conference: Audio Forensics: Practices and Challenges*, 2010.

B. E. Koenig, D. S. Lacey, C. Grigoras, S. G. Price, and J. M. Smith, “Evaluation of the average DC offset values for nine small digital audio recorders,” *J. Audio Eng. Soc.*, vol. 61, no. 6, pp. 439–448, 2013.

C. Grigoras, “Statistical tools for multimedia forensics,” in *Audio engineering society conference: 39th international conference: audio forensics: practices and challenges*, 2010.

APPENDIX 5 – EXAMINER EXPERIENCE AND QUALIFICATIONS

Professional Experience:

- Approximately three (3) years Media Forensics examination experience;
- Approximately two (2) years ongoing forensic imagery analysis for the US Pentagons DARPA agency;
- Ongoing Professional Field Advisor to Foclar, a forensic imagery software developer.

Certificates and Qualifications:

- Media Forensics MSc - National Centre for Media Forensics, Denver, CO, USA;
- Sound Engineering BSc (1st Class Hons) - Birmingham City University, Birmingham, UK.

Published, Presented and Participated Research:

- A Proposed Image Authentication Triage, Chartered Society of Forensic Science Autumn Conference, Northampton (Nov, 2018);
- European Network of Forensic Science Institutes (ENFSI) Facial Comparison Inter-Laboratory Proficiency Tests 2018 (June, 2018);
- A Proposed Audio Enhancement Framework for Forensic Purposes – Presentation, AAFS Annual Conference 2018 (Seattle, WA, USA);
- A Proposed Forensic Audio Enhancement Framework – MSc Thesis, 2017;
- A Low Cost, Cloud Based, Portable ENF System – Paper, AES Audio Forensics 2017 (Arlington, VA);
- Determining Dimension Specific Information for Monaural Sound Recordings – Poster, AES Audio Forensics 2017 (Arlington, VA);
- Medifor Project – DARPA (Department of Advanced Research Project Agency, USA).

Member of the Following Organisations:

- Audio Engineering Society;
- Forensic Image Analysis Division;
- The Chartered Society of Forensic Science;
- International Association of Identification.

Security Clearances

- UK Security Clearance;
- UK Non-Police Personnel Vetting Level 3.

Conferences and Continuing Education:

- Chartered Society of Forensic Sciences Autumn Conference (Northampton, November 2018);
- Bond Solon Criminal Law and Procedure (Manchester, September 2018);
- Bond Solon Cross Examination (Manchester, September 2018);
- Chartered Society of Forensic Sciences Standard Operating Procedure Workshop (July 2018);
- Bond Solon Witness Familiarisation (Manchester, June 2018);
- Bond Solon Excellence in Report Writing (Manchester, June 2018);
- Media Forensics Seminar for the Legal Profession (Presenter, Birmingham, March 2018);
- Forensic Image Processing Workshop (AAFS 2018, Seattle, Feb 2018);
- American Academy of Forensic Science Annual Conference (Seattle, Feb 2018);
- AES Audio Forensics Conference (Arlington, June 2017);
- American Academy of Forensic Science Annual Conference (New Orleans, Feb 2017);
- American Academy of Forensic Science Annual Conference (Las Vegas, Feb 2016);

Parliamentary Published Evidence:

- Admissibility of Forensic Image Comparison Evidence (House of Lords Select Committee, Oct 2018).

Published Articles:

- Attack of The Clones: A.I. in Media Forensics (Digital Forensics Magazine, Aug 2018);
- True Crimes: Why Forensic Audio Isn't Audio Engineering (Sound on Sound Magazine, June 2018);
- Charlatans in Digital Forensics (Forensic Focus, Jan 2018);
- The CSI Effect: Limitations Vs Expectations (Forensic Focus, Nov 2017);
- Iron Bars, Cement and Super Glue (Forensic Focus, Sept 2017);
- When No Conclusion is a Conclusion: Facial Comparison (Forensic Focus, Aug 2017);
- Audio Forensics in 2017 (Digital Forensics Magazine, Aug 2017);
- The Future of ENF Systems (Forensic Focus, Aug 2017).

APPENDIX 6 – QUALITY ASSURANCE

Scientific integrity is of the utmost importance to forensic examinations. In order for this to be maintained, it is our aim to minimize cognitive biases, errors during reports and the over-stating of any findings through a quality assurance program. Every report is reviewed to ensure that both the scientific principles and legal requirements are met. These include:

- Reproducibility. Is it possible that an equally qualified expert who reproduces the examination will obtain the same results?
- Repeatability. Is it possible for the expert to repeat the examination and obtain the same results?
- Has the ACE-V scientific order of operations been followed, namely Analysis, Comparison, Evaluation and Verification?
- Has the organisations Standard Operating Procedure been adhered to?
- Has the chain of custody of the evidence been maintained?
- Have the Criminal Rules of Procedure been met?

---------- End of Appendix ----------