THE JOURNAL OF DIGITAL FORENSICS, SECURITY AND LAW

Journal of Digital Forensics, Security and Law

Volume 10 | Number 3 | Article 2

2015

Identification and Exploitation of Inadvertent Spectral Artifacts in Digital Audio

N. C. Donnangelo
The MITRE Corporation

W. S. Kuklinski
The MITRE Corporation

R. Szabo
The MITRE Corporation

R. A. Coury
The MITRE Corporation

G. R. Hamshar
The MITRE Corporation

Recommended Citation

Donnangelo, N. C.; Kuklinski, W. S.; Szabo, R.; Coury, R. A.; and Hamshar, G. R. (2015) "Identification and Exploitation of Inadvertent Spectral Artifacts in Digital Audio," *Journal of Digital Forensics, Security and Law*: Vol. 10 : No. 3 , Article 2.
DOI: https://doi.org/10.15394/jdfsl.2015.1204
Available at: https://commons.erau.edu/jdfsl/vol10/iss3/2

IDENTIFICATION AND EXPLOITATION OF INADVERTENT SPECTRAL ARTIFACTS IN DIGITAL AUDIO

N. C. Donnangelo , W. S. Kuklinski, R. Szabo, R. A. Coury and G. R. Hamshar
The MITRE Corporation
Bedford, MA 01730-1420

ABSTRACT

We show that modulation products from local oscillators in a variety of commercial camcorders are coupled into the recorded audio track, creating narrow band time invariant spectral features. These spectral features, left largely intact by transcoding, compression and other forms of audiovisual post processing, can encode characteristics of specific camcorders used to capture the audio files, including the make and model. Using data sets both downloaded from YouTube and collected under controlled laboratory conditions we demonstrate an average probability of detection (P_d) approaching 0.95 for identification of a specific camcorder in a population of thousands of similar recordings, with a probability of false alarm (P_{fa}) of about 0.11. We also demonstrate an average P_d of about 0.93 for correct association of make and model of camcorder based on comparison of audio spectral features extracted from random YouTube downloads compared to a reference library of spectral features captured from known makes and models of camcorders, with a P_{fa} of 0.06. The method described can be used independently or synergistically with image plane-based techniques such as those based upon Photo Response Non-Uniformity.

Keywords: camcorder identification, audio forensics, digital forensics, photo response non-uniformity

1. INTRODUCTION

Previous research efforts have developed numerous approaches for forensic camera identification, including exploiting nonhomogeneous dark currents in charge coupled devices detected by integrating multiple images [1], and similar approaches for digital images [2]. Photo Response Non-Uniformity (PRNU) has also been applied to digital video [3]. Techniques that exploit digital post processing, such as MPEG compression, have been reported for tamper and forgery detection [4]. While PRNU is robust and useful, exploitation of the image plane is susceptible to the deleterious effects of compression, transcoding, and other forms of post processing, such as cropping or rotating, that lower PRNU estimation accuracy and remove content details that carry PRNU noise [5–9]. Some tools available as free downloads are specifically designed to alter the PRNU patterns without having a negative impact on the underlying image quality. An independent (i.e., non-image plane) measure of camcorder identification would supplement or confirm classification by PRNU-based camcorder identification techniques.

In this paper we introduce a novel technique to identify camcorder make and model by matching audio spectral artifacts produced by the operation of camcorder internal electronic circuits that are inadvertently coupled into the camcorder audio track as a device-specific signature. We examine the physical mechanism that gives rise to these spectral artifacts in the audio and assess the feasibility of using these features to perform forensic identification of the make and model of camcorder used to record the subject audio file or to link a particular camcorder with a specific audio file.

2. BACKGROUND

Over the last decade the proliferation of low-cost, high-resolution audio-video capture devices, combined with the ease of digital dissemination over the internet, has revolutionized the way we document life and communicate with others. Few have recognized the magnitude of this trend: estimates indicate that 72 hours of video are uploaded to YouTube every minute, and in 2011 YouTube had more than 1 trillion views, about 140 views for every person on Earth [10]. Unfortunately criminals employ the same media to promote nefarious and deeply harmful activities such as child exploitation.

Dozens of different camcorders are commercially available worldwide, with popular models including Sony, Canon, JVC, Panasonic, Samsung, Aiptek, and Vivitar. These cameras range in price from tens to tens of thousands of dollars. However, all modern camcorders incorporate digital logic that implements numerous signal-processing steps. For example, the low-cost, consumer-grade Aiptek A-HD 720p captures HD (1280 x 720 – 16:9 aspect ratio) video clips at 30 frames per second with H.264 video compression.

Figure 1 contains images of the Aiptek A-HD 720p circuitry. There are three visible local oscillators which operate at 32.768 kHz, 14.3 MHz, and 27.0 MHz, respectively.

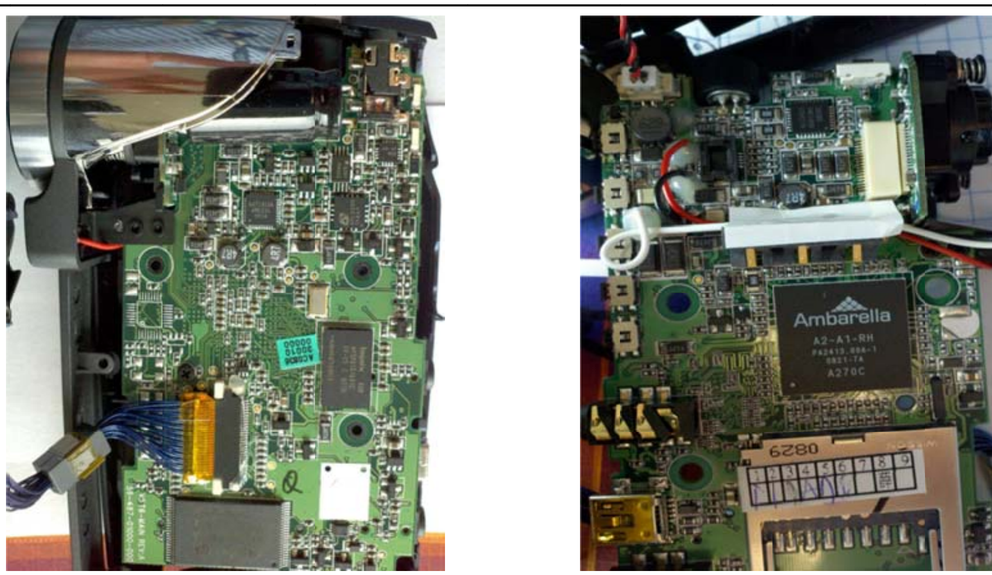

Figure 1. Circuit Boards from the Aiptek A-HD 720p, front (left), back (right), Showing Local Oscillators

These local oscillators are down converted into a plurality of frequencies that control all the functional aspects of the camcorder. For example, the 32.728 kHz clock provides a reference that likely serves as a real-time clock in standby mode to run the logic that initializes the camcorder electronics when a user switches from "standby" to "on." The 14.3 MHz oscillator is probably used for communication between the Ambarella video processing chip, used in this Aiptek device, and flash memory [11, 12]. The 27 MHz local oscillator probably controls numerous and diverse video timing functions in the camcorder, from luminance sampling (13.5 MHz) to PAL sub-carrier frequency (6.25 Hz) [13].

These oscillators and associated circuitry are highly integrated packages. The drive toward miniaturization and cost containment often conflicts with electronic design practices, including the reduced use of shielding to diminish electronic cross talk. The absence or reduced level of inter-substrate shielding can give rise to weak but measurable spectral artifacts as modulation products of various clocks are aliased onto recorded audio. It is important to note that these "self-noise" spectral artifacts do not degrade the performance of any camcorder in a commercially relevant way, because in any normal circumstances the electronically coupled audio spectral features fall well below the audible noise level or lie outside the audible frequency range. However, long integration makes it possible to extract and exploit these temporally invariant spectral features. Because these features are related to both the physical layout of the circuit board and characteristics of component choices, we conjecture the observed spectral artifacts are essentially watermarks characteristic of a given make and model of audio-visual recording device. Given these artifacts are internally generated and time invariant they are essentially independent of the actual recorded audio signal and the specific microphone used on a given camcorder.

To evaluate this hypothesis, we captured audio clips from a variety of prosumer- and consumer-grade camcorders, stripped the audio from the video files using AVSVideo Converter 6, and saved the associated audio in .wav file format [14]. We subjected the resulting audio files to spectral analysis using a short-time Fourier transform, generally integrated for 10 to 30 seconds, with a 0.1 second overlap. The resulting spectrograms clearly show narrowband time-invariant artifacts at tens to hundreds of discrete frequencies (see Figure 2 as an exemplar). We then applied a peak-picking algorithm to the spectrogram and frequency bands and declared detections in 20 percent or more of the spectrograms as persistent bands.

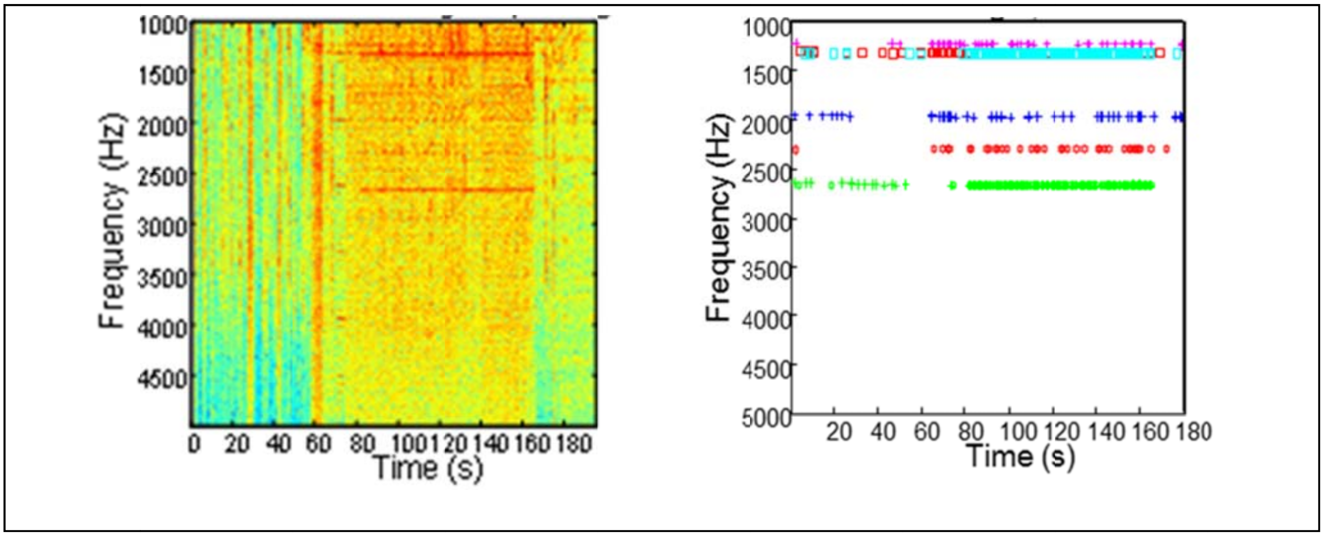

Figure 2. Lofargram of Audio Spectra Captured Using an Aiptek HD Showing Time-invariant Spectral Features between 1000 Hz and 5000 Hz (left) and Associated Persistent Bands (right)

We evaluated a total of 11 different camcorder models from six (6) manufacturers, including multiple serial numbers of the same make and model, as enumerated in Table 1. In all cases we found spectral artifacts in the captured audio. Occasionally, we also observed Electric Frequency Noise (ENF) at 50 or 60 Hz caused by coupling with the power supply [15]. We filtered out these bands and their harmonics and excluded them from the present analysis as nuisance bands. Observed narrow band spectral features remained consistent within a given make and model but showed marked differences among different makes and models.

Table 1
Camcorders Evaluated for Self-Excitation Audio Artifacts

Make	Model	Manufacturer serial #	Label #	Tests -isolation	Tests-Sound
Aiptek	HSHD-v2t6	bru80006331	7	11	0
Aiptek	HSHD-v2t6	bru80002943	8	11	0
Aiptek	HD1PRO-z5x2	bod70009675	9	11	0
Canon	Vixia HFR300	522464108369	10	10	0
Canon	Vixia HFR300	522444111360	11	10	0
Canon	Vixia HFR300	522474110541	12	10	0
Samsung	hmx-q20	A23ecn0c3000t1	13	10	0
Samsung	hmx-q20	a23ecn0c30006B	14	10	0
Samsung	hmx-q20	a23ecn0c30002w	15	10	0
Panasonic	HCV500	e2tw00232	16	0	0
Panasonic	HCV500	b2hg00718	17	0	0
JVC	GYHM100U	155v6179	18	21	9
Panasonic	HMC40P	K9HK00076	19	18	8
Panasonic	HMC40P	GOHK00238	20	20	7
Sony	SR15E	1520221	21	13	9
Sony	SR15E	1521505	22	12	9
Sony	FX7	10005146	23	14	4
Sony	HVRHD1000	242902	24	11	0
Sony	HC62	130052	26	2	4

Having demonstrated the existence of self-excitation spectral features in all captured digital audio segments, we undertook a more systematic survey of the different camcorders to (1) minimize any environmental influences in obtaining a catalogue of self-excitation noise artifacts for different makes and models of equipment, (2) determine the stability and reproducibility of those features over time, and (3) attempt classification based only on recovered self-excitation features.

The spectral artifacts we observe are generally narrow band and temporally invariant at frequencies ranging from hertz to several kilohertz. While very few pure tones are found in environmental background to minimize any chance for environmental influence we constructed an acoustic isolation chamber to host controlled measurements. The chamber, shown in Figure 3, consists of a one-meter cubic sheet steel enclosure with welded seams completely lined on the inside with melamine foam composite sound-absorbing material. The front lid, also lined with sound-absorbing foam, is firmly sealed by four latches [16]. In operation, the enclosure is grounded through a lug and wire to the laboratory ground bus bar. Further electromagnetic isolation is achieved by turning off all power to the lights, HVAC, and outlets in the RF screen room where the tests took place. The enclosure was not lined with Mu Metal and was not intended to completely eliminate low-frequency mains coupling. Multiple tests over a span of several weeks were required to build a database of 204 isolation chamber files.

Figure 3. Acoustic Isolation Test Chamber

2.1. PHYSICAL MECHANISM

An initial assessment revealed that all isolation chamber data files contained time-invariant spectral artifacts. We then undertook a more detailed assessment of the underlying physical mechanism.

The low cost of the consumer-grade Aiptek 720 HD made this camcorder attractive for physical assessment. We removed the housing and mounted the exposed printed circuit boards on a bench, as shown in Figure 4. Using a Detectus EMC scanner, we applied a Lagrange LB 3 probe to the Aiptek 720 HD camcorder circuit board at fixed positions relative to the board definition file [17]. Signals detected by the probe were amplified using 20 dB wideband (10 kHz–2.5 GHz) Sonoma 330 instrument amplifiers and recorded on a Gage 12400 USB memory device [18]. As this test did not require precise temporal synchronization, we manually initiated *record* on the camcorder at the same approximate time as the start of digitizer capture. In other words, the camcorder was recording normally during the probe placement and data capture. The probe capture rate was set at 100 MHz for 0.1 seconds. Figure 5 shows a representative spectrum of the recorded probe signal.

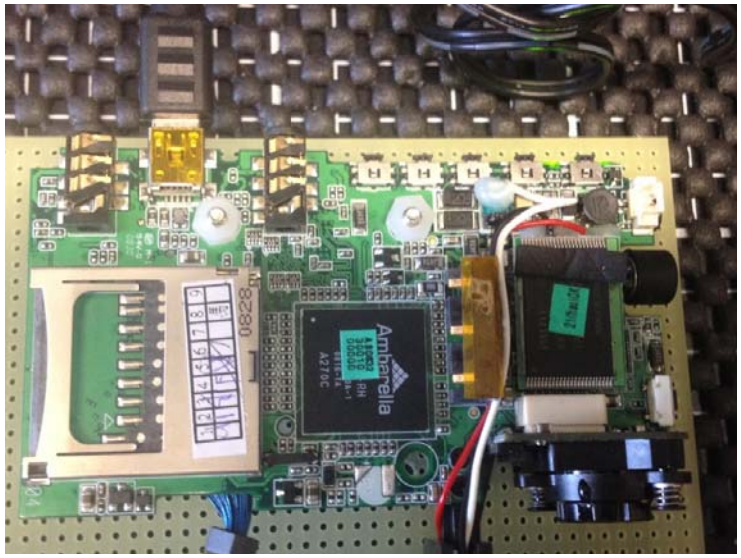

Figure 4. Disassembled Aiptek 720HD Camcorder on the Detectus EMC Scanner Test Bench[1].

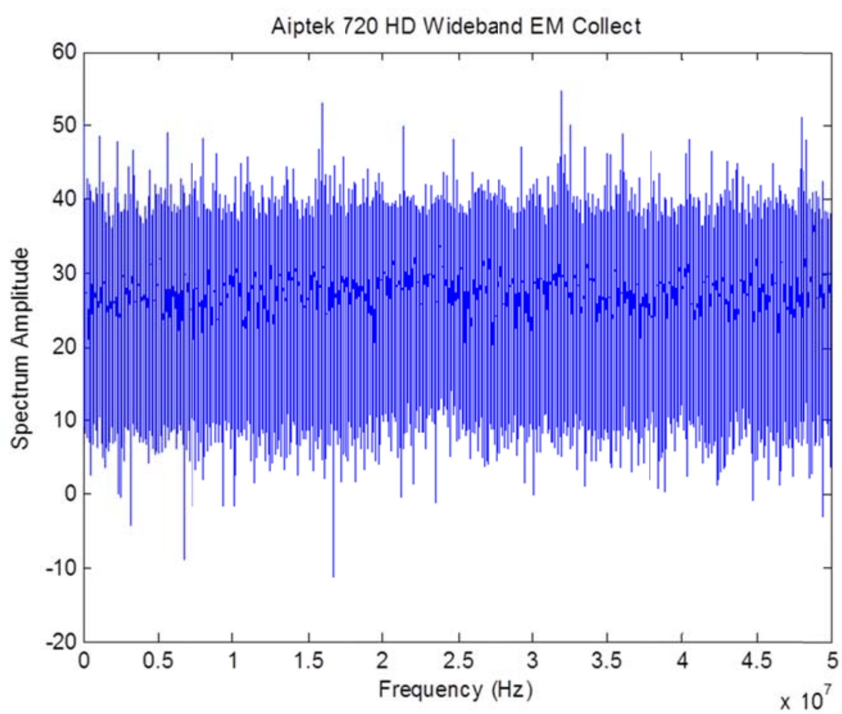

Figure 5. Spectrum of Lagrange Probe Data Collected at a Single Point on Aiptek 720 HD During *record* Status

[1] The camera lens is in the lower right corner of the image; the microphone is on the lower right side corner. The bundle of blue cables at the lower left of the image (below the SD card slot) goes to the display, which is not shown in this image.

We performed peak detection on the probe data using a sliding window noise estimator and a 10 dB detection threshold and then calculated the corresponding aliased values expected for a 48 kHz data rate. If the signals coupled to the audio recording components, they were necessarily aliased due to the 48 kHz recording frequency of the camera. We then compared the measured persistent bands in the audio spectrogram to the computed aliased probe bands. Spectral bands that matched within 5 Hz (corresponding to the frequency resolution of the electromagnetic spectrum) were counted as matches.

Figure 6 is a histogram showing Aiptek HD720 persistent bands extracted from the aliased signals detected by the probe and the bands that had matches in those signals. Of the persistent bands found in the Aiptek 720 HD audio track, 33 out of 62 were correlated with an aliased band found in the probe measurements of coupled oscillator signals. Not surprisingly, we did not find a match for all bands. Filters and amplifiers may result in changes in relative signal amplitudes between the EM and audio so that a signal detected in the EM spectra may not be observed in the audio track recording.

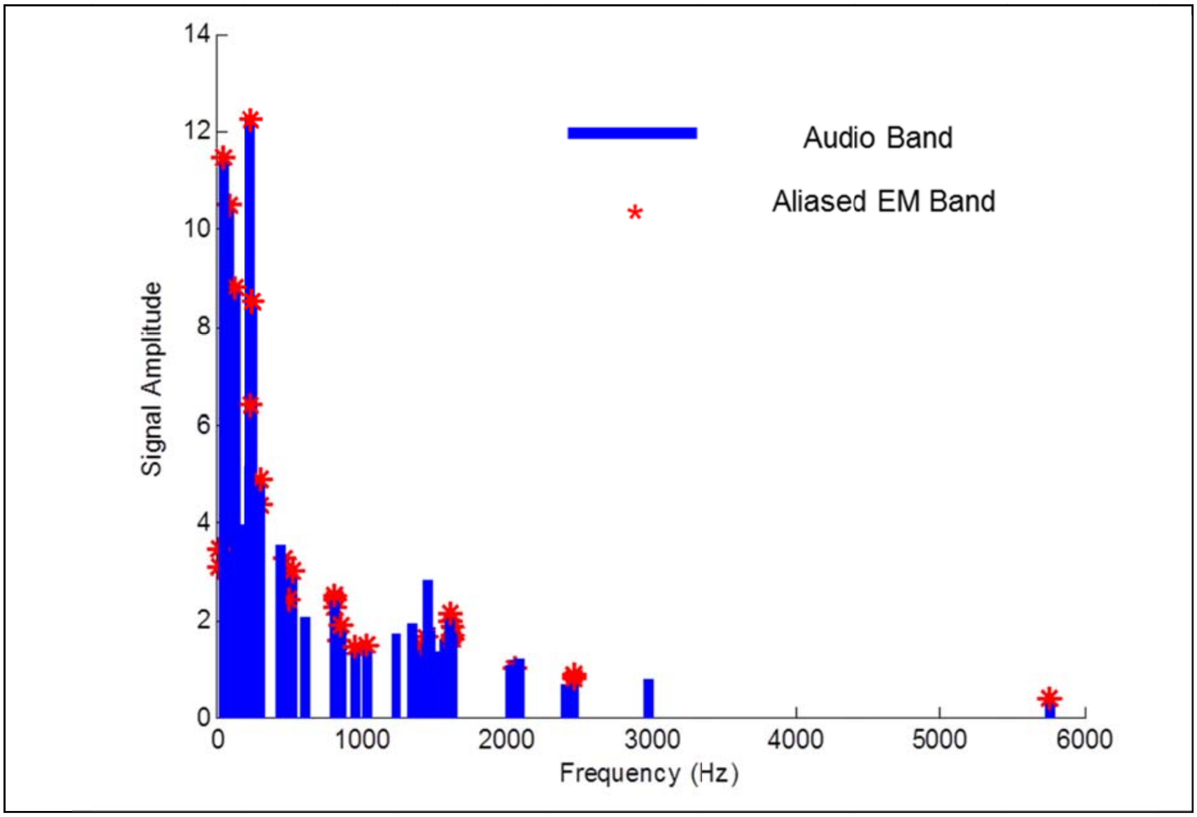

Figure 6. Aiptek 720HD Persistent Audio Bands with Matching EM Signals Extracted from Probe Measurements

While the results presented here are specific to the Aiptek 720 HD, we believe the conclusion is generalizable to any other camcorder or digital recording device. The self-excitation persistent bands are a function the particular mix of oscillators and resonators in a given camcorder. Each oscillator operates at a particular frequency within a given tolerance. Non-linear coupling effects and aliasing tend to accentuate the differences. We assess that the probability of the identical oscillators in any two different cameras having the same exact nominal frequency is very low, and the probability of a match among multiple oscillating circuits in a camera is even lower. Changing a single component, such as a processor in a particular make or model device, would likely result in frequency or amplitude shifts in coupled spectral features. Indeed, we observed subtle changes in multiple camcorders of the same make and model.

3. IMPACT OF TRANSCODING

Having determined that all of the commercial camcorders tested exhibit narrow band spectral artifacts in captured audio files, and having quantified the underlying physical mechanism that links components to the observed features, we next explored the robustness of the observable features by applying transcoding and audio post processing to captured files. Only if the self-excitation features were robust to transcoding and audio post processing did it make sense to consider the information embedded in the artifacts and the feasibility of forensic classification.

In this portion of the study we examined two cameras: the Sony SR15E and the Sony FX7. The SR15E produced clips in the MPEG-2 format, while the FX7 camera produced clips in the WMV format. We performed the bulk of the analysis on the clips recorded by the SR15E, although some tests were conducted on clips recorded by the FX7 as a consistency check.

The first examination considered a recording, hereafter denoted T67, produced by the Sony SR15E in MPEG-2 format and just over 19 minutes in length. Approximately 18 minutes of T67 were recorded while the camera was in the isolation chamber; we denoted this portion of the clip as T67Q. The period while the camera recorded outside the chamber included some human vocalization as well as some limited ambient noise; this portion of the recording is designated T67V. The Sony SR15E showed predominant, temporally invariant spectral features at a variety of frequencies, including: 159 Hz, 490.5 Hz, and 770 Hz.

We evaluated the effect of two different audio-visual post processing software packages: NV11 and CPD10.[2] We loaded the original MPEG-2 file into these packages and exported it into other formats using two different options: (1) no compression and (2) a pre-defined setting labeled 'MS Video 1.' No processing other than transcoding was performed on the file. We calculated spectra over the duration of the entire T67 recording, in the various formats, from both inside and outside the isolation chamber. For simplicity and proof of concept, this analysis focused on the region of the audio below 1000 Hz.

Figure 7 shows three overlaid curves for T67Q, illustrating (1) the audio spectrum of the original MPEG-2 file (shown in blue), (2) the spectrum of a compressed AVI file (shown in green), and (3) the spectrum for an uncompressed AVI file (shown in red). In all cases the frequency response has been averaged over the duration of the T67Q data set. It is

[2] The Nero Video 11 (NV11) editing software is fully featured and allows clips to be exported in Audio Video Interleave (AVI), Windows Media Video (WMV), and Flash Video (FLV) formats, in addition to some others that were not examined. The software allowed the impact of bass and treble boost/reduction, reverb, noise reduction, as well as some pre-defined audio equalization profiles on self-excitation audio artifacts to be assessed. Cyberlink PowerDirector 10 (CPD10) produced clips exported as Digital Video (DV) AVI, WMV, and MPEG-2 formats, in addition to some others that were not examined. Unlike NV11, CPD10 did not have the option to apply audio effects options to the clips.

clear from the figure that transcoding did not introduce or distort observed SR15 self-excitation features.

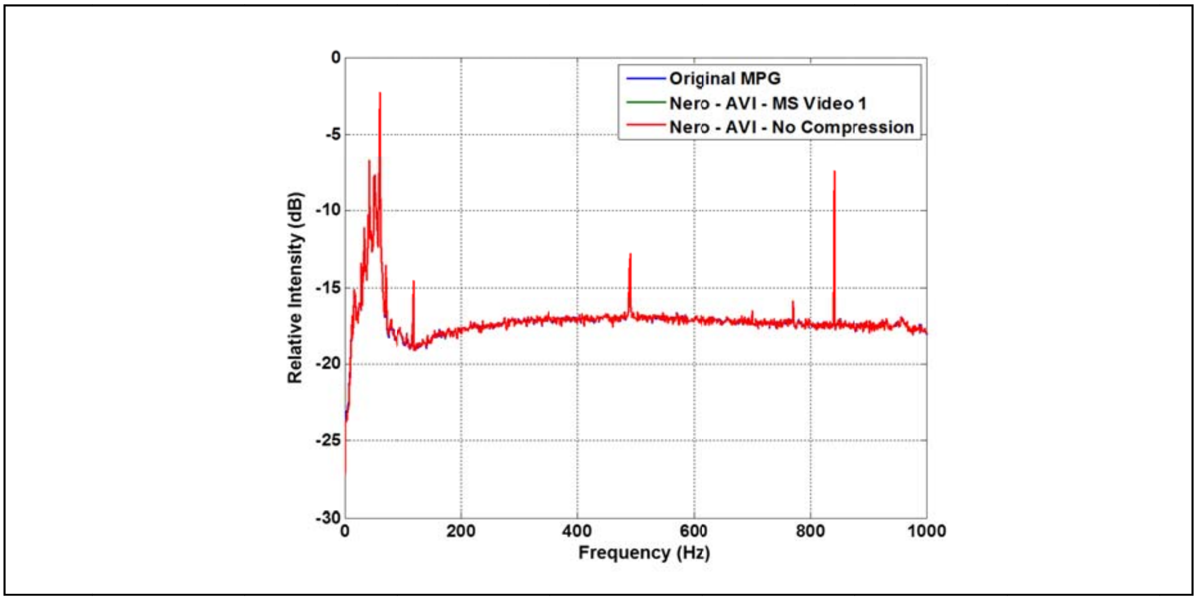

Figure 7. Low-frequency Portion of the Audio Spectrum for the Portion of Clip T67 Recorded by a Sony SR15E, Transcoded and Compressed

We followed a similar process by ingesting T67 into Nero Video 11 and exporting the recording in WMV format, using a predefined set of parameters to export the highest-quality WMV file available. The clip was then truncated at the 4:48 mark and again exported into WMV format. Figure 8 depicts the spectra for the three files, with the blue curve showing the spectrum of the original MPEG-2 file and the green and red curves indicating the spectra of the WMV files. As before, no significant difference appears between the observed SR15 audio spectra in either the location or strength of the spectral lines.

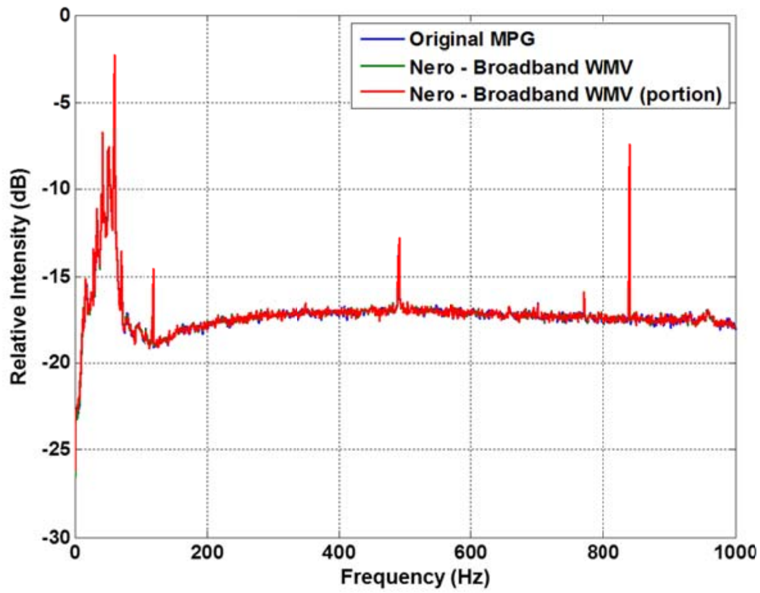

Figure 8. Low-frequency Portion of the Audio Spectrum for Clip T67Q, Transcoded in WMV Format

We adjusted and repeated this process by ingesting the T67 files into NV11 and exporting the recording in FLV format, using both the default settings as well as a pre-defined set of parameters labeled 'Maximum Volume.' Figure 9 shows the results of this process. In this figure, the blue curve again represents the spectrum of the MPEG-2 file for T67Q, while the green and red curves represent the spectra of the FLV files. The figure shows differences between the spectra of the original SR15 audio and transcoded files. Most noticeable is the roll-off of frequencies above 200 Hz. Still, the spectral features associated with self-excited noise in the SR15 in the original file remain visible in the transcoded file. As an example, consider the narrow band self-excitation feature near 700 Hz. The original MPEG-2 file contains a small peak approximately 1 dB above the neighboring frequencies. In the FLV files, however, the relative strength of this line increased by approximately 1.5 dB. Similarly, the FLV files exhibit a low-intensity line near the 900 Hz mark that was not noticeable in the original MPEG-2 recording. Since the red and green curves are nearly coincident, the specific parameters used in the FLV encoding do not affect the modification of the self-excitation spectrum during the transcoding process.

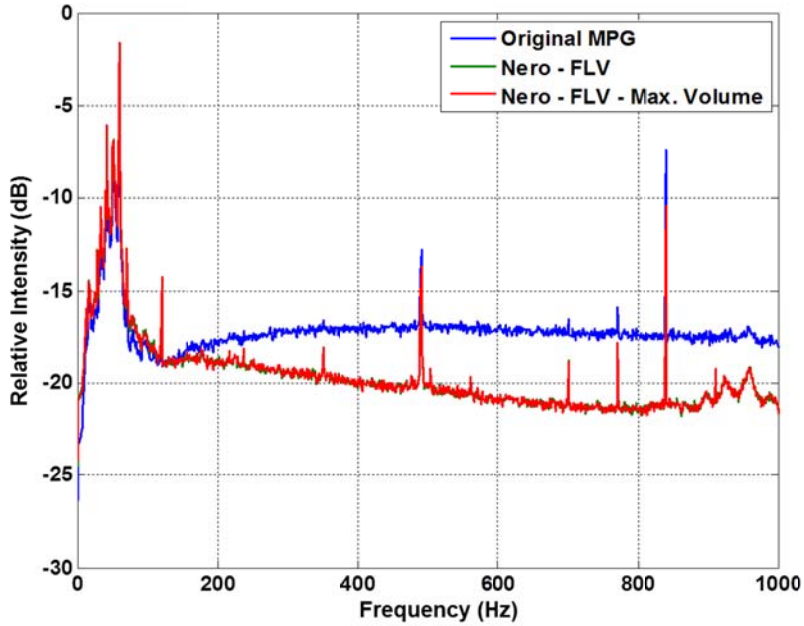

Figure 9. Low-frequency Portion of the Audio Spectrum for Clip T67Q, Exported in FLV Format

We repeated his process using the CPD10 software. In this case the MPEG-2 files were loaded into CPD10 and exported into WMV, DV AVI, and MPEG-2 formats. The results of this process are shown in Figure 10 and Figure 11. Figure 10 depicts the CPD10 conversion to WMV format, including the application of Dolby 5.1 processing to the soundtrack. Dolby 5.1 processing did not introduce any modification in the location or strength of the spectral features of interest. Similarly, the CPD10 transcoding to DV AVI and MPEG-2 (Figure 11) did not significantly affect the T67Q spectrum.

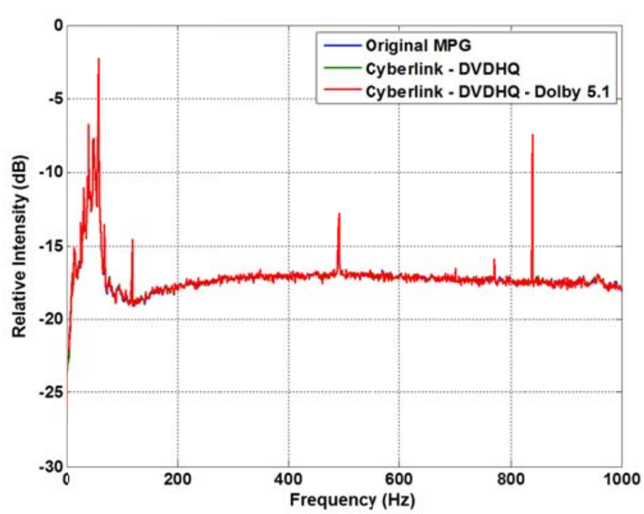

Figure 10. Low-frequency Audio Features Associated with SR15E; CPD10 Conversion to WMV Format

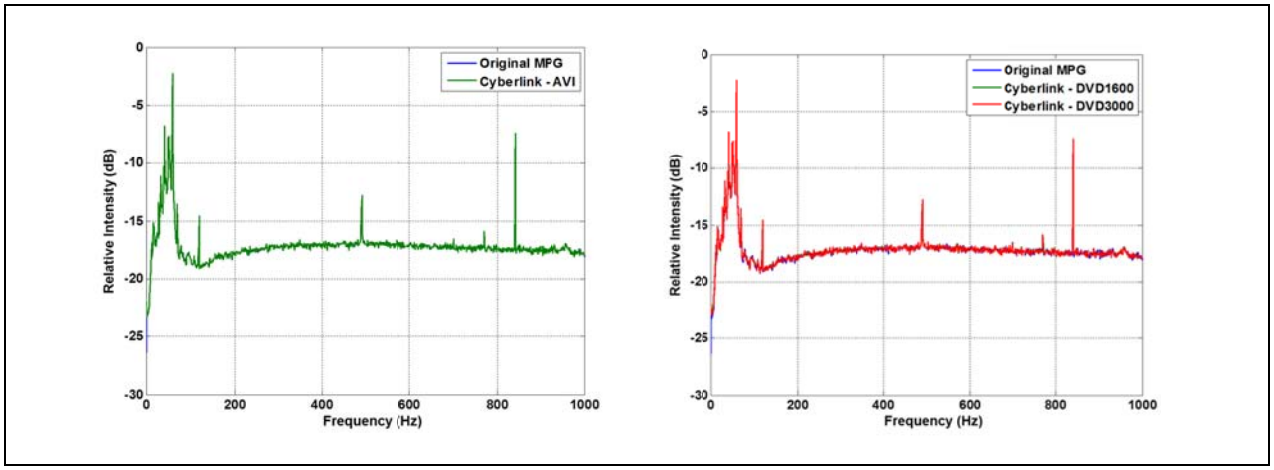

Figure 11. Low-frequency Audio Features Associated with SR15E CPD10 Transcoding to DV AVI (left) and MPEG-2 right)

We replicated this analysis with audio spectra extracted from a Sony FX7, which produces output files in WMV format. The clip designated T51 was just over 15 minutes in length, of which approximately 14 minutes were recorded while the camera was in the isolation chamber. The isolation chamber portion of the FX7 recording is denoted as T51Q. The period while the camera was recording outside of the chamber included some human vocalization as well as some limited ambient noise; this portion of the recording is designated T51V. The original T51 WMV file was loaded into NV11 and exported into WMV files using several different pre-defined quality settings. As can be seen in Figure 12, the transcoding process produced virtually no observable change in the characteristic self-excitation spectra for the FX7.

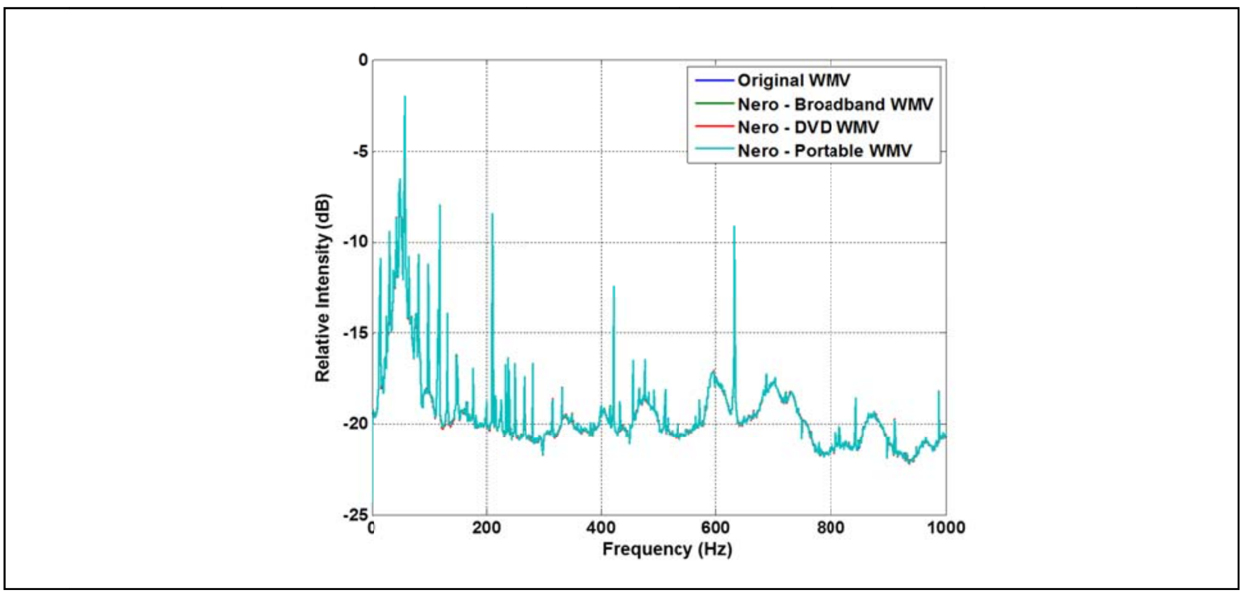

Figure 12. Low-frequency Portion of the Audio Spectrum for the Portion of Clip T51 Recorded by a Sony FX7 in an Isolation Chamber as Transcoded by CD11 and Nero

4. EFFECT OF AUDIO POST PROCESSING

Some video editing software, such as NV11, enables investigators to apply various effects to the audio track of a recording. The NV11 software used in this analysis had several pre-defined filters that could be applied to the audio tracks. We exercised several of these filters to examine their impact upon the narrow band spectral features.

The T67 recording (Sony SR15E) was ingested into NV11, the noise reduction filter was applied using both the (pre-defined) minimum and maximum settings, and the result was exported into WMV format. The blue curve in Figure 13 illustrates the spectrum of the original MPEG-2 file, while the green and red curves depict the spectra of the noise-reduced files, for minimum and maximum reduction, respectively. The figure shows that the minimum noise reduction filter had no significant impact on the characteristic of the audio spectrum (and in fact completely overlays the spectrum of the unmodified file), while the maximum noise reduction filter lowered the overall amplitude in frequencies above approximately 150 Hz by 2–3 dB. While some noticeable changes in the audio spectrum are associated with the maximum noise reduction case, for example in the range of 800–1000 Hz, the location of the original SR15 self-excitation spectral lines is preserved, although the strength of those lines above 150 Hz diminishes somewhat.

We also modified the audio track of the T67 recording by applying a 'large room reverb' effect using NV11. Figure 14 shows the impact of this processing.

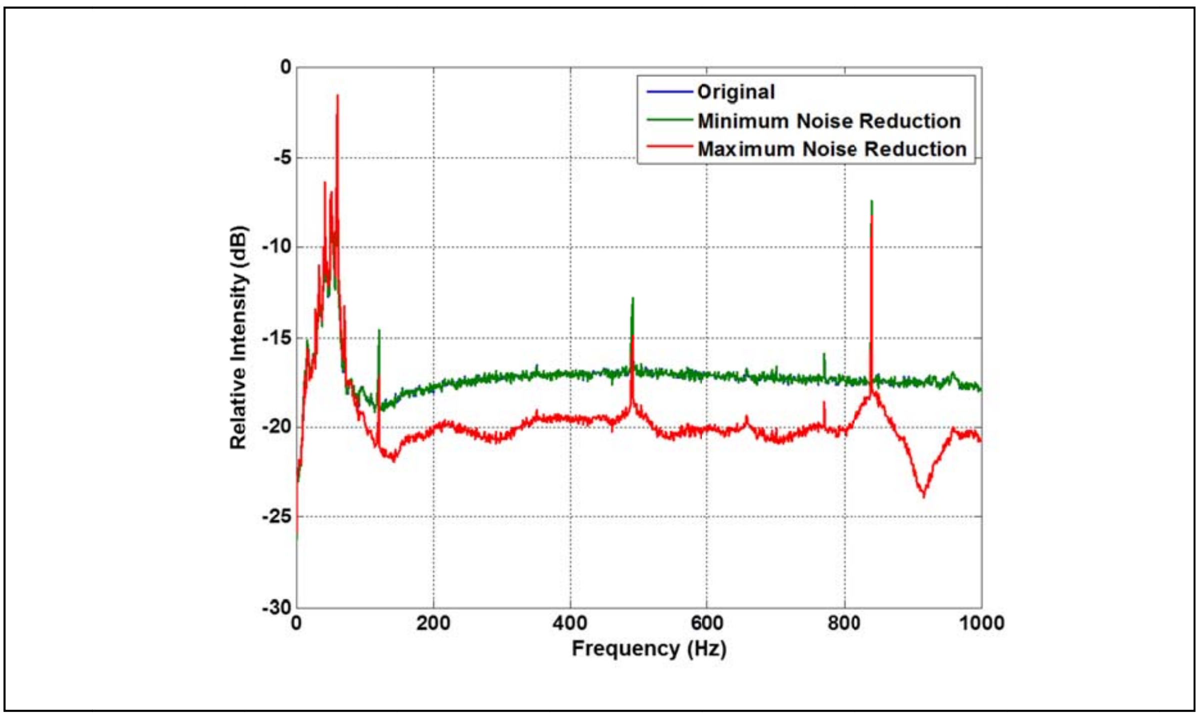

Figure 13. Low-frequency Portion of the Audio Spectrum for Clip T67Q with Different Levels of Noise Reduction

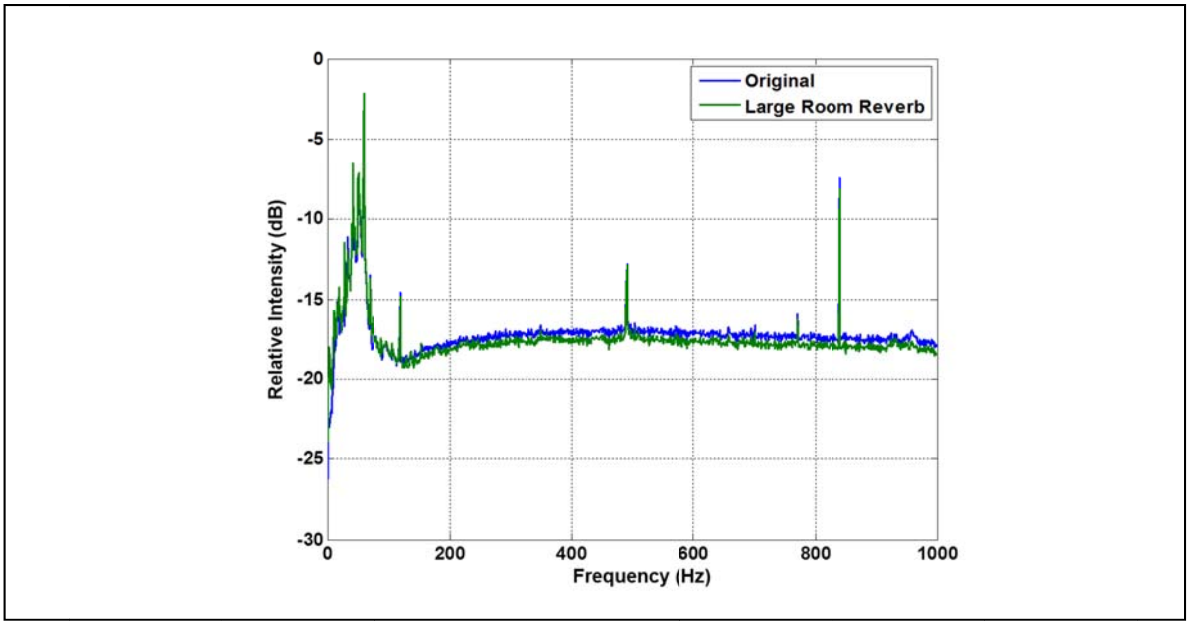

Figure 14. Low-frequency Portion of the Audio Spectrum for Clip T67Q with 'Large Room Reverb' Effect

Figure 15 shows the effect of transcoding over the range from 0 Hz to 10 kHz. While the addition of the Dolby 5.1 processing to the transcoding process caused an upswing in the frequency response above 7 kHz, transcoding over the full band left the SR15E spectral features largely unaltered in both strength and frequency.

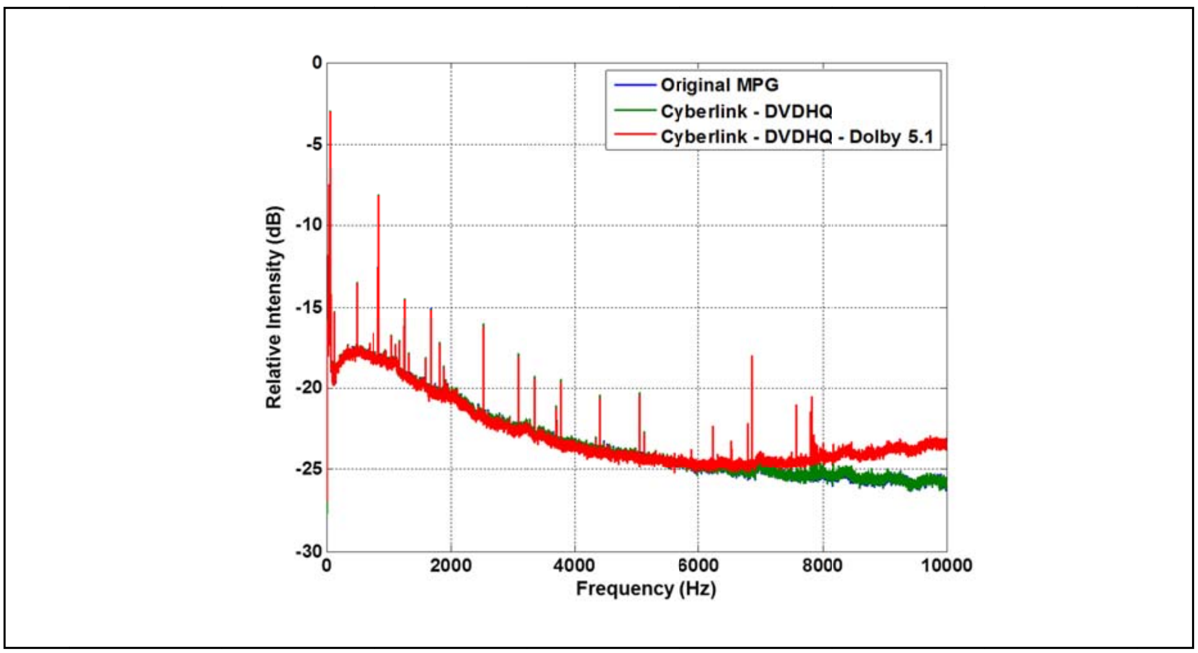

Figure 15. Partial Audio Spectrum for Sony SR15E; Male Speaking in Normal Voice and Low-level Ambient Noise

5. CLASSIFICATION: MAKE AND MODEL

The study showed that background noise, transcoding, compression, and audio effects have minimal impact on the self-excitation features of the Sony FX7 and SR15. We believe this conclusion is also extensible to other camcorders.

We next considered the feasibility of feature classification for forensic device identification based on narrow band spectral artifacts recovered from audio. The data set for classification experiments consisted of the isolation chamber measurements discussed above, as well as videos downloaded from YouTube. This analysis used only downloads from camcorders in our collection with labeled make and model. All of the clips downloaded from YouTube had the camera make/model information embedded in the title (e.g., “Sony NexVg Panning Test”). The YouTube videos were downloaded in .mp4 format [19]. Table 2 lists the makes and models represented in our data set.[3]

We used FFMPEG to strip audio from the downloaded YouTube clips and save them as .wav files. The first step in the classification process consisted of feature extraction. To support this, we generated spectrograms from each of the audio tracks using 5-second, non-overlapping sliding windows of audio data. This yielded a frequency resolution of 0.2 Hz.[4] We did not attempt to remove speech or background noise that might interfere with spectral artifacts.

We used a local thresholding technique to detect persistent bands. Each spectrogram consisted of N_f cells, where N_f is the number of frequency bins resulting from the Fourier transform of the audio data. Consider frequency bin q. We computed the average of the amplitude of bins q–15 through q–6 and q+6 through q+15, thus averaging 10 bins on each side of q and ignoring the 10 cells directly adjacent to the bin of interest (5 on each side). If the test bin amplitude exceeded the average of the 20 windowing bins by at least 6 dB, we declared a detection. Figure 16 shows a spectrogram alongside the results of detection band processing.

The green pixels in the right-hand figure represent frequency bins in which detections were declared. Note that the detected signals may have been environmentally generated or come from the camera hardware itself. However, most environmentally generated signals evince temporal variations, whereas the self-excitation features are temporally invariant. We then identified persistent peaks by comparing consecutive spectrograms.

Some of the recordings contained audio impulses that uniformly populated a wide range of frequencies for a short period of time. During these broad spectrum impulsive sounds, the bands of interest to us were sometimes hard to detect, but they reappeared at the end of the audio impulse. We also used Nero software to verify that we could detect the same bands when music was added to the audio tracks using Nero software.

[3] Most of the data had audio tracks sampled at 48 kHz. A few data sets were sampled at 41kHz.

[4] For this analysis we made no attempt to remove speech or background noise that may interfere with some spectral artifacts.

Table 2

Data Sets Used in Make/Model Classification Analysis

Make	Model	Number of Cameras in House	Number of Cameras from YouTube Collects
Sony	NexVg20	0	10
Sony	HVRHD1000	1	10
JVC	GYHM100U	1	10
Sony	FX7	1	14
Panasonic	HMC40P	2	10
Panasonic	HCV700	0	10
Sony	hdrcx190	0	10
Canon	Vixia HF R300	0	10
JVC	HM440	0	10
Aiptek	ahd 720p	2	10
Samsung	hmx q20	0	10

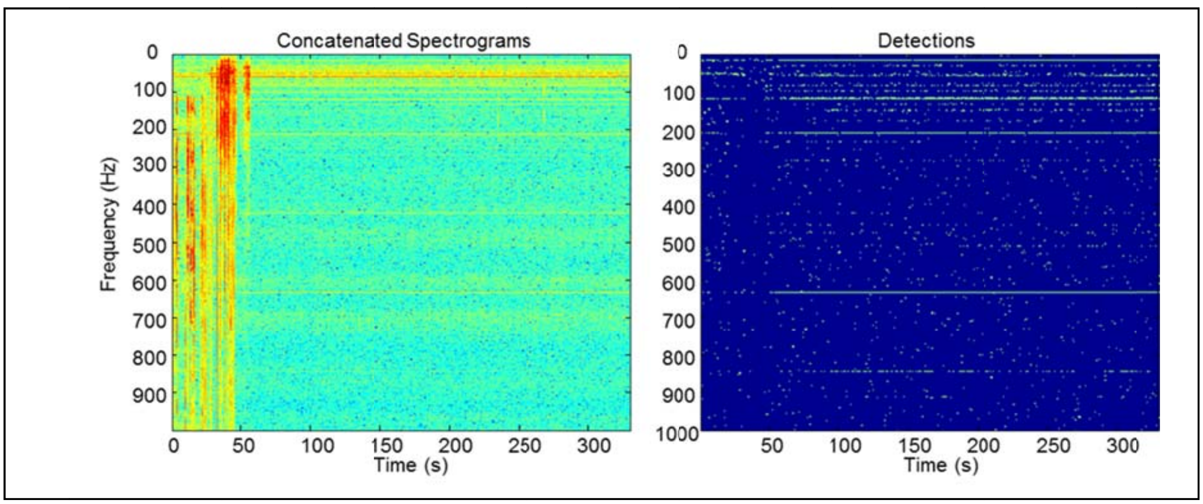

Figure 16. Spectrogram and Resulting Detections

We attempted to exploit three features of the devices: (1) frequencies of the persistent bands, (2) fundamental frequencies of bands with many harmonics present, and (3) frequencies of bands in which persistent streaks were not detected. For a given make/model, we considered only features that appeared in at least 50 percent of the recordings by devices of that make/model.

Although it consisted of several hundred recordings, our sample set was still limited in a statistical sense. In general, the data requirement for building a useful classifier is a linear function of the number of problem-specific parameters used to describe the classifier. In the case of a simple Bayesian classifier, the number of training samples needed to produce a 10 percent uncertainty in classifier performance, when the class separation is σ, is $10*N_H*N_F$, where N_H is the number of hypotheses and N_F is the number of features. This linear requirement produced an incentive to minimize the number of hypotheses and features used in a classifier. Creating a single classifier with hypotheses corresponding to every available camcorder in the world would have been impractical. The classification problem considered therefore had only two hypotheses: the make/model is of interest and the make/model is not of interest. We grouped all camcorders other than the make/model of interest into the latter category. To test for classification utility, we used a three-layer perceptron with a hidden layer sigmoid function and evaluated all combinations of candidate features five at a time. The algorithm selected the combination that yielded the best separation between the make/model under consideration and all other makes/models.

The three-layer perceptron used the features that produced the best performance during the feature selection process described above. We obtained statistical performance estimates by using a leave one/some out approach. We trained the classifier on a cross section of prosumer- and consumer-grade camcorders and tallied correct and incorrect decisions. We repeated this process 30 times per make/model. Table 3 shows the resulting performance statistics. The average over all makes and models was 96 percent correct identification with 6 percent false alarms.

Table 3
Statistical Make/Mode Classification Performance Results

Make/Model	P_d	STD $_{Pd}$	P_{fa}	STD $_{Pfa}$
Panasonic HCV700	0.99	0.06	0.03	0.06
Sony FX7	0.99	0.02	0.03	0.04
Sony NexVg20	0.98	0.13	0.06	0.09
Samsung hmx q20	0.98	0.15	0.04	0.20
JVC HM440	0.96	0.07	0.02	0.04
Aiptek A-HD	0.95	0.09	0.00	0.05
JVC GYHM100	0.92	0.14	0.13	0.14
Sony HVRHD1000	0.92	0.13	0.14	0.18
Panasonic HMC40P	0.88	0.12	0.13	0.19
Canon Vixia HF R300	0.85	0.12	0.13	0.15
Sony HDRCX190	0.80	0.11	0.00	0.04

6. CLASSIFICATION: SAME CAMCORDER

The second classification problem resembled the make/model identification problem except that we now assessed the ability to classify a specific device based on audio artifacts from camcorders of the same make and model. We conjectured that allowable manufacturing tolerances in oscillators and changes of secondary components in the camcorder bill of material over time could alter the mutual coupling and cross talk by shifting some frequency lines while attenuating or amplifying others.

We augmented the data set described earlier with additional recordings known or believed to come from the same exact device. We collected some of the same-device recordings from in-house devices and harvested others from YouTube. To find same-camera recordings, we looked for multiple YouTube recordings from the same uploader where the recordings were identified as having come from the same device (e.g., “Here’s another test recording from my new Sony FX7”). Table 4 lists the same-camera recordings used for this part of the analysis.

Table 4
Data Sets Used in Same Camcorder Classification Analysis

Make	Model	Number of In-house Recordings Used	Number of YouTube Recordings Used
Canon	Vixia_HF_R300_12	3	0
Canon	Vixia_HF_R300_12	3	0
Canon	Vixia_HF_R300_12	3	0
Canon	Vixia_HF_R300_12	0	2
Samsung	Hmx-q20	5	0
Samsung	Hmx-q20	5	0
Samsung	Hmx-q20	5	0
Sony	nexVg20	0	2
Panasonic	HMC40P	2	0
Panasonic	HMC40P	2	0
JVC	HM440	0	3
Sony	Hdrcx190	0	2
Panasonic	HCV700	0	3
JVC	GYHM100	4	0
Sony	FX7	4	0
Aiptek	A-HD	4	0
Aiptek	A-HD	3	0

Feature extraction, selection, and classification began with the identification of persistent bands, as discussed previously. In this case, however, the primary feature we used was a correlation measure to determine whether these two recordings came from the same camera. For each possible pair of recordings, we created a vector of binary values V_i, of length equal to the number of bands in the spectrogram. We assigned an entry in V_i the value 1 if the corresponding band was occupied and a value of 0 if it was unoccupied. We calculated three quantities:

$S_{i,i} = \sum V_i \bullet V_i$, $S_{i,j} = \sum V_i \bullet V_j$, and $S_{j,j} = \sum V_j \bullet V_j$.

$S_{i,j}$ and $\min(S_{i,i}, S_{j,j})$ were used as features. The analysis used a three-layer perceptron as described previously, where the two hypotheses were “same camera” and “different camera.” We generated statistics using 100 iterations of the leave one/some out strategy. Table 5 summarizes the performance results, showing only the features that performed better than the strategy used in the make/model classification.

Table 5
Statistical Same Camcorder Classification Performance Results

P_d	STD_{Pd}	P_{fa}	STD_{Pfa}
0.93	0.06	0.11	0.08

7. LIMITATIONS OF THE ANALYSIS AND DIRECTION OF FUTURE WORK

This paper reports on the existence of modulation products electromagnetically coupled and aliased onto the digital audio track of a variety of consumer and prosumer grade camcorder. The electromagnetically coupled modulation products are manifested as multiple (order tens to hundreds) narrow band, time invariant spectral tones. These audio tones are of no practical consequence to the normal operation of a camcorder, but with long integration they can be detected and extracted. We show that the spectral location of these tones is entirely defined by the local oscillators and impedance and circuit board geometry of the capture instrument and as such encodes identifying information about the device used to capture the audio file.

While we have observed these spectral artifacts in every digital audio recording device we have analyzed, the population of audio capture devices we consider in this paper is small and much more work needs to be done to assess if electromagnetically coupled audio spectral features are universally found in all digital audio recording apparatus.

Our initial analysis suggest the spectral artifacts encode make, model and manufacturing batch information, however this analysis is limited by small sample size and much more work needs to be done to assess the uniqueness of these features and their info-theoretic content.

While we propose a methodology to exploit the spectral audio artifacts for camcorder identification, statistically designed experiments have yet to be performed to assess the performance of our technique across an operationally relevant range of conditions, including environmental noise. We conjecture that aliased audio artifacts could be co-processed with video data to enhance overall classification but we have yet to propose a specific algorithm or methodology for implementing this conjecture, and have not derived any quantitative estimate of the info-theoretic content of the audio features vis features derived from video data such as PRNU.

Elements of the processing chain we describe, such as the peak picking algorithm, could themselves introduce artifacts or attenuate legitimate features and should be more carefully assessed. Since the spectral features we report are generally both highly dimensioned and temporally invariant we did not explore the effect of background noise or microphone characteristics on classification accuracy. We anticipate a more systematic follow on analysis to include characterization of the effect of background noise and microphone transfer function on feature extraction and classification.

A preliminary analysis demonstrated that many temporally invariant features can be attributed to coupling of modulation products created by the aliasing of multiple oscillators in the highly integrated electronic package. We anticipate future electromagnetic modeling of the modulation products and coupling pathways will inform more optimal processing method. This more complete understanding could also inform better processing approaches to detect, extract and exploit the audio spectral features and potentially uncover other classes of features.

8. CONCLUSIONS

While audio cross talk is a well-known phenomenon, the literature has not previously reported the presence of self-excitation noise features in camcorder audio. We observed a plurality of these narrow band spectral artifacts in a sample of 11 different camcorders of differing make/model/year of manufacture and serial number.

These spectral features, left largely intact by transcoding, compression, and other forms of audiovisual post processing, can form highly dimensioned feature vectors that encode characteristics of the specific camcorder used in the audio capture, as well as make- and model-specific characteristics of the device. In a small sample, we demonstrated an average P_d approaching 0.95 for identification of a specific camcorder in a population of similar recordings with a P_{fa} of about 0.11. We also demonstrated an average P_d of about 0.93 for correct association of make and model of camcorder based on comparison of audio spectral features extracted from a small sample of random YouTube downloads compared to a reference library of spectral features captured for known makes and models of camcorders, with an associated P_{fa} of about 0.06. Audio spectral features could be processed independently or synergistically with PRNU-based approaches to enhance camcorder identification.

REFERENCES

[1] Kuroki, K., Kurosawa, K., & Saitoh, N. (January 2002). An Approach to Individual Video Camera Identification. *Journal of Forensic Sciences*, *47*(1).

[2] Lukáš, J., Fridrich, J., & Goljan, M. (June 2006). Digital Camera Identification from Sensor Pattern Noise. *IEEE Transactions on Information Forensics and Security*, *1*(2), 205–214.

[3] Chen, M., Lukáš, J., Fridrich, J., & Goljan, M. (2007). Source Digital Camcorder Identification Using Sensor Photo-Response NonUniformity. *Proc. of SPIE Electronic Imaging, Photonics West*

[4] Wand, W., & Farid, H. (2009). Exposing Digital Forgeries in Video by Detecting Double Quantization. MM&Sec'09 *Proceedings of the 11th ACM workshop on Multimedia and Security*, 39–48.

[5] Rosenfeld, K., Sencar, T., & Memon, N. A Study of the Robustness of PRNU-based Camera Identification," Retrieved on November 1, 2014, from http://isis.poly.edu/~forensics/PAPERS/17.pdf

[6] McCloskey, S. (2008).Confidence Weighting for Sensor Fingerprinting. Honeywell Labs. Retrieved on November 1, 2014, from http://www.cim.mcgill.ca/~scott/papers/WVU_2008.pdf

[7] Sencar, H. T., & Memnon, N. (2008).Overview of State-of-the-Art in Digital Image Forensics. Statistical Science and Interdisciplinary Research. World Scientific Press.

[8] Swaminathan, A., Wu, M., & Liu, K. J. Ray (May 2006). Non-Intrusive Forensic Analysis of Visual Sensors Using Output Images. *Proc. of IEEE Conference on Acoustic, Speech and Signal Processing (ICASSP)*, *5*, 401–404.

[9] Retrieved on [date] from http://www.winsite.com/Multimedia/Image-Viewers/PRNU-Decompare

[10] YouTube, Statistics, Retrieved on November 1, 2014, from http://www.youtube.com/yt/press/statistics.html

[11] dDForum, homepage, Retrieved on November 1, 2014, from http://www.3dphoto.net/forum/index.php?topic=1034.0

[12] Ambarella™, Technology – The Ambarella Difference, Retrieved on November 1, 2014, from http://www.ambarella.com/technology/anatomy-of-quality.html

[13] dDForum, homepage, Retrieved on November 1, 2014, from http://www.3dphoto.net/forum/index.php?topic=1034.0

[14] AVS Video Converter 9.0, homepage, Retrieved on November 1, 2014, from http://www.avs4you.com/AVS-Video-Converter.aspx?type=GoogleAdWordsSearch&gclid=CPrH7JPPurwCFWYaOgodX3oA0A

[15] Grigoras, C. (2005). Digital Audio Recording Analysis – The Electric Network Frequency Criterion. *International Journal of Speech Language and the Law, 12*(1), 63–76.

[16] Acoustical Surfaces, Inc., homepage, Retrieved on November 1, 2014, from http://www.acousticalsurfaces.com/foam_stop/mel_comp.htm

[17] Detectus AB, Products, EMC Scanner, Retrieved on November 1, 2014, from http://detectus.se/products_emc.html

[18] Sonoma-Instrument.com, Specifications, 330 Broadband Scanner, Retrieved on November 1, 2014, from http://www.sonoma-instrument.com/pdf/330ds.pdf

[19] Keepvid: Download Online, Retrieved on November 1, 2014, from www.keepvid.com