Audio Engineering Society

Convention Paper 9897

Presented at the 143rd Convention
2017 October 18–21, New York, NY, USA

Challenges of audio forensic evaluation from personal recording devices

Robert C. Maher

Electrical & Computer Engineering, Montana State University, Bozeman, MT USA 59717-3780

Correspondence should be addressed to R.C. Maher (rob.maher@montana.edu)

ABSTRACT

Typical law enforcement audio forensic investigations involve audio evidence recorded under less-than-ideal circumstances by mobile phones, surveillance systems, and personal audio recorders. Moreover, the audio information is often transmitted and stored using a data compression algorithm such as a speech coder (e.g., VSELP) or a wideband audio coder (e.g., MP3). There are few systematic studies of the signal behavior of these systems for forensically-relevant audio material, and this may discourage a forensic examiner from using such acoustic evidence to draw reliable conclusions. This paper includes simulation and evaluation of personal audio recording systems in the context of audio forensics for gunshot sounds. The results indicate areas of strength and weakness in the forensic realm.

1 Introduction

Audio forensic evidence is of increasing importance in law enforcement investigations because of the growing use of personal audio/video recorders carried by officers on duty. These recording systems capture speech, background environmental sounds, and in some cases, gunshots and other firearm sounds.

While our recent work involved collecting gunshot recordings with specialized recording equipment and procedures [1, 2, 3, 4], typical law enforcement investigations involve gunshot sounds recorded under less-than-ideal circumstances by mobile phones and personal audio recorders. To our knowledge, there are few systematic studies of the response of these systems to intense acoustical sounds, and this lack of knowledge seriously limits the ability of an audio forensic examiner to use such acoustic evidence to draw reliable conclusions about gunshot evidence even when a database of recorded gunshot sounds is available. Common personal audio recorders are designed to capture human speech conversations, and the forensic impact of speech-optimized processing and lossy perceptual encoding algorithms on unanticipated sounds such as gunshots is not well understood [5].

The work for this paper focuses on the challenge of gunshot sounds, as this type of audio forensic evidence is perhaps the most challenging for personal recording devices [6]. Specifically, this paper reports upon two simulation strategies. For the first strategy, we start with a wideband recording of a gunshot sound obtained using the quasi-anechoic recording technique at a 500 kHz sampling rate [3]. This is considered the reference recording. Next, the signal is filtered with a response approximating the microphone and audio recording subsystem of a mobile phone device in speech telephony mode, and the resulting signal is compared to the wideband original. For the second strategy, we use a similar approach, but also apply MP3 perceptual audio coding to the speech bandwidth version of the gunshot sound. We then compare the reconstructed MP3 signal to the original, and also to recordings made with a personal memo recorder [5].

2 Simulation 1: effect of recording system bandwidth

Personal recording devices such as mobile phones can operate in two or more different microphone audio modes. One mode typically intended for telephone speech conversations has a bandpass characteristic, rolling off the low frequencies below 200 Hz and the high frequency range above about 4 kHz. A common audio equivalent sampling rate would be approximately 10 kHz. Figure 1 is a sketch of the approximate audio recorder bandwidth in this mode.

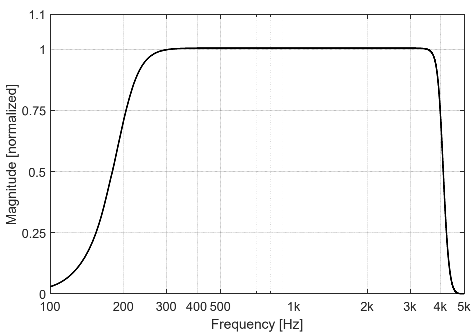

Figure 1. Example phone speech response.

Figure 2 shows a recorded gunshot sound from a pistol (Colt M1911, 0.45 caliber). The recording microphone was located 3 meters from the firearm and off to the side (82° off-axis.) The recording used a 500 kHz sampling rate and the microphone system bandwidth was approximately 6.5 Hz to 70 kHz [3].

The wideband recording of Figure 2 shows an extremely brief and abrupt onset of the muzzle blast sound. The acoustic peak pressure momentarily reaches 3.5 kPa (165 dB peak SPL re 20 μPa), rising in approximately 6 microseconds.

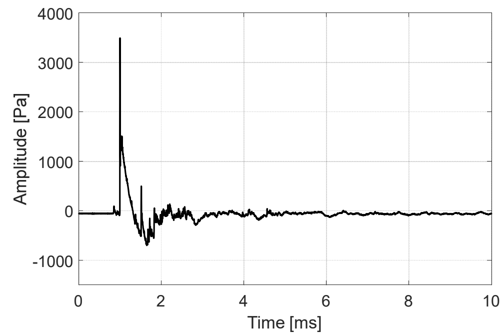

Figure 2. Gunshot sound, Colt M1911, 0.45 caliber handgun (82° azimuth, 3 meters, 500 kHz sampling rate).

It is possible to simulate at least some of the characteristics of a gunshot recording from a mobile phone set in the telephone speech mode by filtering and downsampling this high sampling rate recording. We expect that the abrupt rise time and other high frequency details of the signal will be lost, and Figure 3 shows the resulting signal.

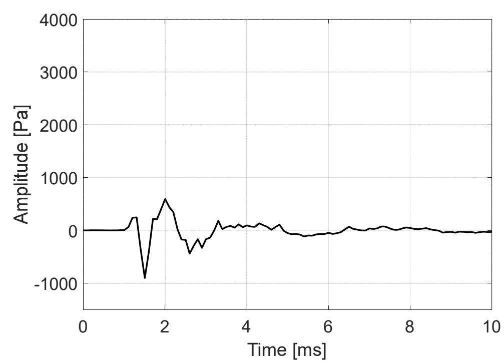

Figure 3. Gunshot sound of Fig. 2, processed with filter response of Fig. 1, and downsampled to 10 kHz sampling rate.

As expected, the simulated speech-mode recording demonstrates the reduced signal resolution and detail of a speech bandwidth representation. It is important to note that the waveshape of the speech-bandlimited recording in Fig. 3 is substantially different from the wideband reference recording.

A recording of a different handgun (Glock 19, 9mm caliber) is shown in Figure 4.

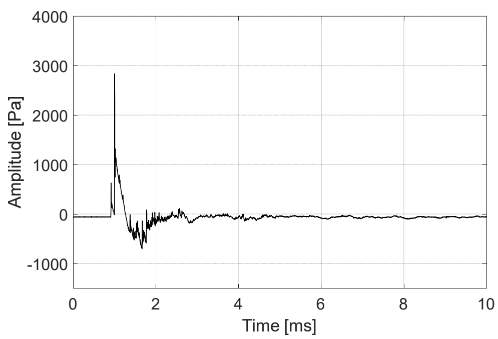

Figure 4. Gunshot sound, Glock 19, 9 mm caliber handgun (82° azimuth, 3 meters, 500 kHz sampling rate).

There are subtle but distinctive differences between the Colt handgun (Figure 2) and the Glock firearm (Figure 4). Filtering and downsampling the reference recording as before, the resulting waveform is shown in Figure 5.

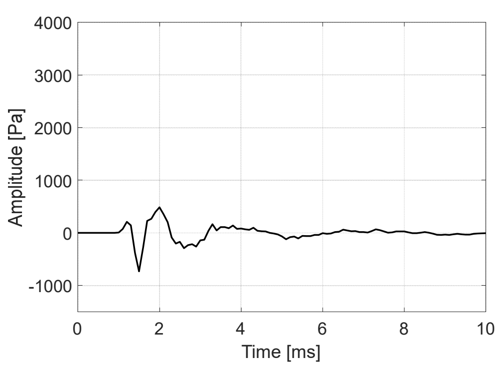

Figure 5. Gunshot sound of Fig. 4, processed with filter response of Fig. 1, and downsampled to 10 kHz sampling rate.

Comparing Fig. 5 and Fig. 3, the waveshape differences are difficult to discern. This leads to the assessment that it is unlikely for an audio forensic examiner to distinguish between these two handguns based on the speech-band audio representation, even when the initial waveform is recorded from the same position and without the usual acoustic reflections, reverberation, and background noise that is generally present in a forensic recording.

Finally, a recording of a rifle (AR15) is shown in Figure 6. Note the greater vertical scale in this figure to accommodate the louder report of this rifle compared to the handguns shown previously.

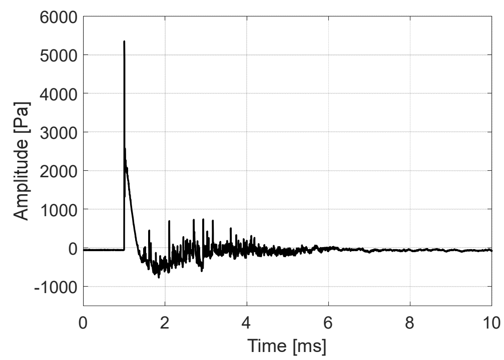

Figure 6. Gunshot sound, AR15 rifle (82° azimuth, 3 meters, 500 kHz sampling rate).

The AR15 rifle signal represented in the mobile phone speech bandwidth is shown in Figure 7.

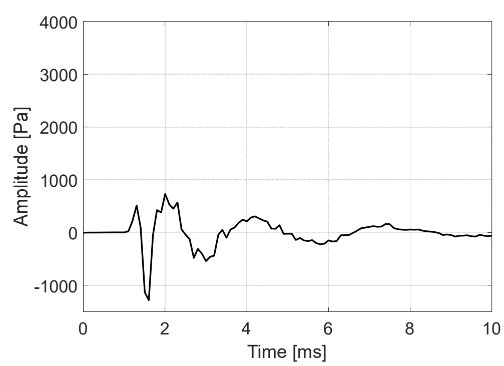

Figure 7. Gunshot sound of Fig. 6, processed with filter response of Fig. 1, and downsampled to 10 kHz sampling rate.

Now comparing Fig. 7, Fig. 5, and Fig. 3, the primary observation is that the signals appear to be impulse responses of the audio speech filter rather than being highly dependent upon the details of the gunshots themselves. An impulse sent through the filter of Fig. 1 results in the response in Figure 8.

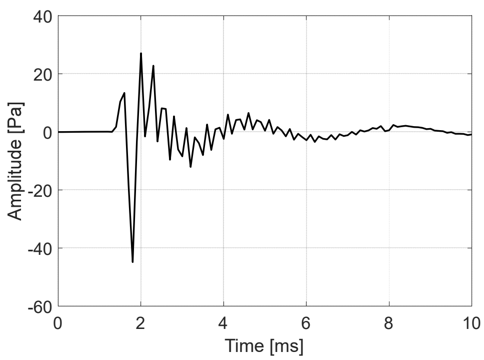

Figure 8. Impulse response of speech filter (Fig. 1), downsampled to 10 kHz sampling rate.

At the low sampling rate, the gunshot details are essentially impulses. In other words, the speech bandwidth recording of the various gunshots may tell more about the frequency response of the recorder than about the details of the gunshots.

3 Simulation 2: perceptual coding

In order to understand more about the characteristics of gunshot audio recordings likely to be encountered in forensic investigations, we apply a perceptual audio coder MP3 (MPEG-1, Layer 3) [7] to the AR15 rifle example presented in the previous section. The result of applying the perceptual coder to the waveform of Fig. 6 is shown in Figure 9. Note the expanded time scale in Fig. 9 to show the pre-echo effect of the lossy coder.

Comparing Fig. 9 and Fig. 7, the MP3 coder has retained much of the waveform detail, but the pre-echo due to the MP3 block length is evident. Audio forensic questions often involve timing and relative amplitude considerations, so care must be taken when examining evidence that has been filtered and perceptually compressed.

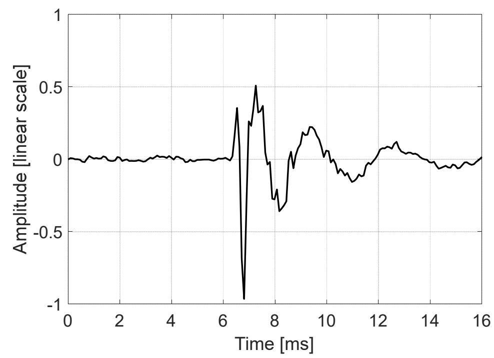

Figure 9. Gunshot sound of Fig. 6, processed with filter response of Fig. 1, downsampled to 11.025 kHz sampling rate, MP3 encoded, and reconstructed. Expanded time scale.

4 Conclusions

As has been noted in previous publications [4, 5], audio forensic examiners must be reminded to exercise care when drawing conclusions from audio evidence obtained from common portable personal recording devices such as mobile phones.

The advice is to use caution when attempting time waveform-based processing, such as correlation analysis, with speech devices due to the inherent bandwidth limitations, perceptual coding, and the resulting change in waveform details and overall waveshapes. Similarly, analyzing the relative timing of sounds, echoes, and other audio events must take into account the effects of the recording device, recorded bandwidth, and lossy data compression.

The examples presented here are wideband gunshot recordings processed to simulate the effects of the limited bandwidth and processing of a mobile speech device. In these examples, it is tacitly assumed that the device was capable of accommodating the intense sound of a firearm's muzzle blast without clipping the microphone or the device's electronics.

In many actual forensic cases when the recording device is in proximity to the firearm, the gunshot sounds may be of such intensity that the recording is overloaded for many milliseconds, revealing little detail about the firearm or any concurrent sounds.

References

[1] R.C. Maher and T.K. Routh, "Gunshot acoustics: pistol vs. revolver," *Proc. 2017 Audio Engineering Society International Conference on Audio Forensics*, Arlington, VA (2017).

[2] R.C. Maher and T.K. Routh, "Wideband audio recordings of gunshots: waveforms and repeatability," Preprint 9634, *Proc. 141st Audio Engineering Society Convention*, Los Angeles, CA (2016).

[3] T.K. Routh and R.C. Maher, "Recording anechoic gunshot waveforms of several firearms at 500 kilohertz sampling rate," *Proc. Mtgs. Acoust.* 26 (2016); http://dx.doi.org/10.1121/2.0000262 .

[4] R.C. Maher and T.K. Routh, "Advancing forensic analysis of gunshot acoustics," Preprint 9471, *Proc. 139th Audio Engineering Society Convention*, New York, NY (2015).

[5] R.C. Maher and S.R. Shaw, "Gunshot recordings from digital voice recorders," *Proc. Audio Engineering Society 54th Conference, Audio Forensics—Techniques, Technologies, and Practice*, London, UK (2014).

[6] S.D. Beck, H. Nakasone and K.W. Marr, "Variations in recorded acoustic gunshot waveforms generated by small firearms," J. Acoust. Soc. Am., vol. 129, no. 4, pp. 1748-1759 (2011).

[7] ISO/IEC 11172-3: *Information technology -- Coding of moving pictures and associated audio for digital storage media at up to about 1,5 Mbit/s -- Part 3: Audio*, International Organization for Standardization (1993).