Audio Engineering Society

Convention Express
Paper 265

Presented at the AES 157th Convention
2024 October 8-10, New York, NY, USA

Interpreting user-generated audio from war zones

Robert C. Maher

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

Correspondence should be addressed to rmaher@montana.edu

ABSTRACT

Increasingly, civilian inhabitants and combatants in conflict areas use their mobile phones to record video and audio of armed attacks. These user-generated recordings (UGRs) often provide the only source of immediate information about armed conflicts because access by professional journalists is highly restricted. Audio forensic analysis of these UGRs can help document the circumstances and aftermath of war zone incidents, but consumer off-the-shelf recording devices are not designed for the circumstances and sound levels of warfare, nor do the battlefield circumstances provide clear, noise-free audio. Moreover, as with any user-generated material that generally does not have a documented chain-of-custody, there are forensic concerns about authenticity, misinformation, and propaganda that must be considered. In this paper we present several case studies of UGRs from armed conflict areas and describe several methods to assess the quality and integrity of the recorded audio. We also include several recommendations for amateurs who make UGRs so that the recorded material is more easily authenticated and corroborated.

1 Introduction

The emergence of professional journalists serving as war correspondents dates to the mid-19th century. The profession grew in prominence and importance as photography, the telegraph, and later motion picture film and audio recording equipment became available. By World War II, most military leaders promulgated policies about the presence of authorized independent journalists embedded with military units [1].

The views of international political leaders about the role of professional journalists in combat zones is changing rapidly [2]. This change in attitude, coupled with the ubiquity of handheld recording devices, civilian satellite communications networks, and rapid dissemination via social media platforms, has created a situation in which reports coming from war zones are increasingly generated by participants and amateur bystanders rather than third-party professionals.

This user-generated content, no longer curated in the traditional sense, presents many opportunities and challenges. The presence of so many recording devices makes it likely that many more incidents will be captured by digital still photographs and moving pictures with sound, and this can only increase the amount of information available from areas of conflict. However, digital photos and audiovisual recordings released from war zones frequently come from anonymous and unverified sources, and this may allow skilled purveyors of forged content to release fraudulent material for the purposes of spreading propaganda or sowing conspiracy theories.

This paper is organized as follows. We describe four example cases in which professional journalists attempting to report on a combat incident had received unverified user-generated recordings for

which they wanted audio forensic examination. For each example, we consider several types of journalist questions that can be addressed through audio forensics. We conclude the paper with comments and recommendations regarding audio forensic interpretation of user-generated material from war zones.

2 Case Example 1: Killing of Journalist Abelardo Liz in Colombia (13 August 2020)

In August 2020, an incident took place between land rights protesters and government soldiers in rural Colombia. A journalist, Abelardo Liz, was among several individuals shot and killed during the confrontation. Following the incident, statements by the army and statements by civilian witnesses differed in many aspects of the conflict, such as who was shooting and from where. The Netherlands-based investigative journalism group Bellingcat, and the Colombian media outlet Cerosetenta, began an inquiry into the circumstances of the shootings [3].

While no *official* videos are known to exist, the reporters were able to obtain at least 16 user-generated videos from mobile phone cameras that were filming at the time of the fatal gunfire. The cameras were spread out over a distance of a few hundred meters from where Abelardo Liz was shot. Thus, an important step in the assessment of the user-generated material is to compare the sound events present in the various videos to identify the degree to which the audible material is consistent.

The available user-generated videos have no inherent synchronization, and in most cases, the hand-held cameras were shifting around because the person filming was turning, bending, running, or otherwise in motion (see Figure 1).

Figure 1: Still photo from video prior to the fatal shooting, showing armed soldiers at the scene. [3]

Among the types of information to be studied are timelines, identification of likely sound sources, and geometrical orientation observed from the recorded soundscape. For example, the recordings from the Abelardo Liz incident included the sound of numerous gunshots, and the key investigative questions dealt with when and from where those shots occurred. In this case, many of the audible gunshot sounds included both a ballistic *shock wave* signature (caused by the passage of a supersonic bullet), and the *muzzle blast* sound from the firearm, creating a so-called *crack-pop* sequence (see Figure 2). These observations can be used to estimate the distance between the firearm and the camera making the recording [4][5][6].

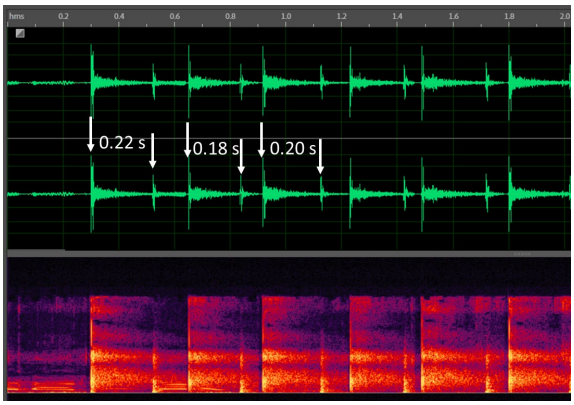

Figure 2: Example of several gunshot sounds from one of the Abelardo Liz incident recordings, showing a ballistic shock wave *crack* sound followed ~200ms later by the arrival of the muzzle blast *pop* sound.

Sound recorded by common consumer mobile phones can be difficult to interpret for a number of reasons. First, the audio portion of a cell phone movie generally comes from the phone's built-in microphone system, which may have one, two, or more microphones on the exposed outer surface of the phone. These microphones could be affected by the position of the user's hands holding the phone, the vertical or horizontal orientation of the phone, whether or not the user has the phone in a protective case, and also any electronic aspects of the audio system, such as automatic gain control, noise canceling features, and perceptual audio coding.

While many mobile phones produce a stereo (2-channel) audio recording, it is not clear whether the stereo field follows a conventional left-right orientation that might allow inter-channel comparison, especially as the orientation of the phone

changes from vertical (portrait) to horizontal (landscape). Because of these uncertainties, the general recommendation is not to attribute much veracity to the stereo details.

In the case of the Abelardo Liz shooting, the journalists were able to use a combination of the user-generated audio, video, and witness accounts to develop an understanding of the incident [3].

3 Case Example 2: Moscow explosion (9 August 2023)

An explosion in Sergiev Posad, a suburb located 50 km northeast of Moscow, on August 9, 2023, resulted in at least one death and more than 60 injuries. The explosion at the Zagorsk Optical-Mechanical Plant, reportedly an optical electronics factory, raised questions about whether the incident was an accident, sabotage, or perhaps a drone attack connected to the ongoing war with Ukraine [7].

After the incident, footage appeared on social media outlets showing the aftermath of the explosion, as well as at least three segments of what appeared to be closed-circuit television (CCTV) surveillance video, with audio, from cameras at a residential apartment complex approximately 0.5 km from the factory.

An example CCTV recording posted by users of social media had its camera pointing in the direction of the optical plant. A few frames of the video are shown in Figure 3. The onset of the blast flash is visible in the video, followed approximately 1.5 seconds later by the abrupt arrival of the blast's shock wave, which jarred loose window screens and other debris. A column of smoke is then visible rising from the explosion site.

Analyzing the audio that accompanied the video, there are several audible background sounds that precede the arrival of the extremely loud explosive blast sound (see Figure 4). These sounds include footsteps of an individual walking past the camera, a tonal hum sound that appears to be due to ventilation equipment at the apartment complex, and then a set of low-frequency rumbling sounds that precede the arrival of the main explosion approximately 1.5 seconds after the first indication of a flash in the video. The time lag between the observed flash and the arrival of the explosion is consistent with the shock wave traveling at the speed of sound over the distance of approximately 500 meters between the factory and the camera.

The journalists investigating the incident raised questions about the presence of low-frequency rumble sounds prior to the arrival of the main shock wave. The audio forensic analysis cannot provide a definitive answer, but one possibility is that the speed of sound in the soil is substantially faster than the speed of sound in air, and so the ground surface vibration of a large explosion could conceivably out-pace the arrival of the corresponding sound through the air. Another possibility is that the main visible flash of the explosion was preceded by initial smaller explosions that were not visible in the camera view.

Figure 3: Frames from CCTV recording of the 9 Aug 2023 explosion in Sergiev Posad, Russia.

Some of the investigative journalists questioned whether the tonal sound attributed to ventilation equipment near to the camera might actually have been due to a drone aircraft overhead. An audible aircraft sound would support the theory that the

explosion was triggered as a hostile action related to the Russia-Ukraine hostilities. Although the audio from this single video could be considered ambiguous in this regard, the fact that the other concurrent videos did not capture the same tonal sound, and the fact that no Doppler or amplitude modulation is present that might indicate a moving or maneuvering aircraft, reduces the likelihood that the sound was attributable to a drone.

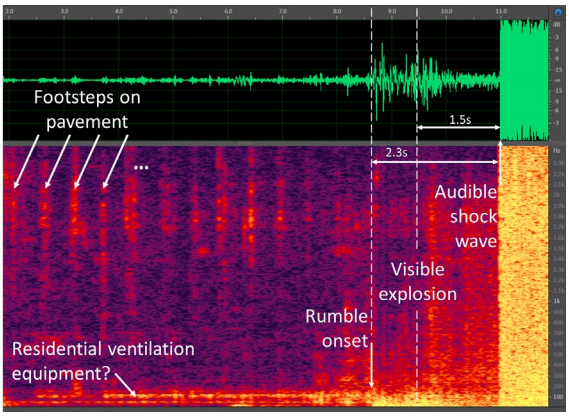

Figure 4: CCTV recording waveform and spectrogram for 9 seconds prior to the audible explosion.

4 Case Example 3: South Lebanon investigation of fatal explosion (13 October 2023)

Journalists covering strikes and counterstrikes by Hezbollah and Israeli Defense Forces (IDF) near Alma al-Shaab in southern Lebanon encountered shell fire on 13 October 2023. At least six journalists were injured, and Lebanese video journalist Issam Abdallah, working for Reuters, was killed by what has been reported to be a tank-fired shell. After the first shell killed the journalist, a second shell impacted the same area (see Figure 5).

The journalists from several different agencies were reportedly located out in the open on a ridge line well away from any ongoing hostilities, wearing garments identifying them as journalists, and not in the presence of any active hostilities.

Figure 5: Still frame from video following the two shell explosions near journalists in Lebanon on 13 Oct 2023.

Following the tragic incident, Human Rights Watch, among other organizations, collected at least five audiovisual recordings from the scene, and sought to analyze the audio characteristics of the explosions [8]. The audio forensic information included an observation of approximately 37 seconds between the first impact and the second impact. The available user-generated recordings also included the sound (and images) of a helicopter and possibly drone aircraft circling the area in the hours before the strikes. The motion of the aircraft with respect to the camera produced Doppler frequency shifts observable as the aircraft circled in front of the camera (see Figure 6). The recordings of the explosions also contained apparent muzzle blast sounds following the two explosive impacts. Assuming a tank-fired supersonic shell, the timing between the explosion and the arrival of the muzzle blast sound gave an estimate of 1.5 – 1.8 km as the distance from which the shell was fired.

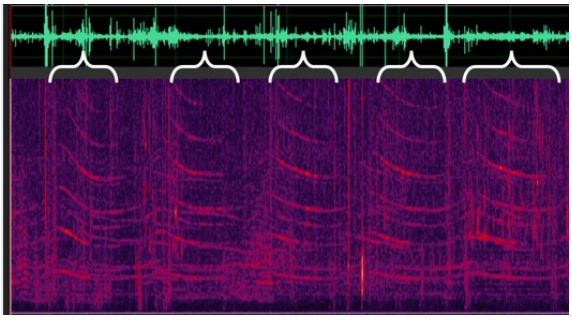

Figure 6: Spectrogram of approximately 8 minutes of audio, showing the cyclical Doppler frequency shift as an aircraft slowly circles in the area in front of the camera.

5 Case Example 4: Al-Ahli Hospital explosion in Gaza (17 October 2023)

On the evening of 17 October 2023, an explosion occurred in the courtyard of the Al-Ahli Hospital in Gaza City. The hospital, managed by the Episcopal Church in Jerusalem, was in the midst of hostilities associated with the Israel-Hamas war that started with a surprise attack by Hamas-led fighters on 7 October 2023. The opposing forces cast blame upon each other following the Al-Ahli Hospital explosion, and the high level of distrust made it impossible to have an authoritative determination. Nonetheless, several user-generated recordings, including a location not far from the explosion (see Figure 7), captured sounds of the scene before and after the explosion [9].

Figure 7: Still frame from user-generated handheld cell phone video of Al-Ahli Arab Hospital explosion on 17 Oct 2023.

Among the forensic questions about the explosion was the direction of approach of the explosive device. One partisan assertion was that the explosion was caused by an Israeli air-dropped bomb, others claimed it was an Israeli artillery shell, a Hamas improvised explosive device detonated at the scene, or an errant Hamas rocket that malfunctioned and fell to earth at the hospital. A variety of videos from many locations around the scene added to the mystery [10].

Audio forensic analysis of the user-generated handheld cell phone video identified a "whoosh" sound preceding the explosion, with an increasing and then decreasing audible pitch (see Figure 8).

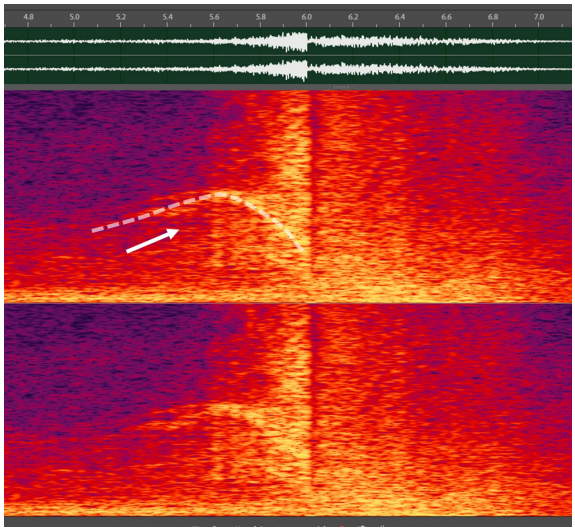

Figure 8: Spectrogram of sound 1.4s before to 1.4s after the al-Ahli explosion, showing the rising and falling frequency pattern just prior to the blast.

One theory that might explain a Doppler-related frequency increase and then decrease would be a sound source initially accelerating toward the microphone (frequency increasing) and then passing by the microphone (frequency decreasing). For example, the sound from an object falling and accelerating under the force of gravity would show an increasing frequency observed on the ground. If the object was falling some distance from the microphone, the Doppler shift would decrease as the trajectory neared the ground due to the decreasing radial velocity, as indicated in Figure 9.

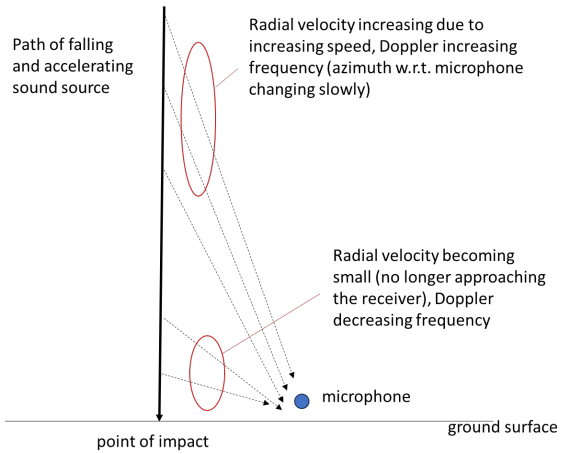

Figure 9: Sketch of expected Doppler profile of a falling and accelerating sound source.

A related audio forensic question would be whether the observed rising and falling frequency preceding the explosion would be consistent with an object approaching the microphone and impacting the ground, or consistent with an object passing over the microphone before impact (see Figure 10). The radial velocity change with respect to the microphone appears to favor the approaching trajectory compared to the overpassing trajectory.

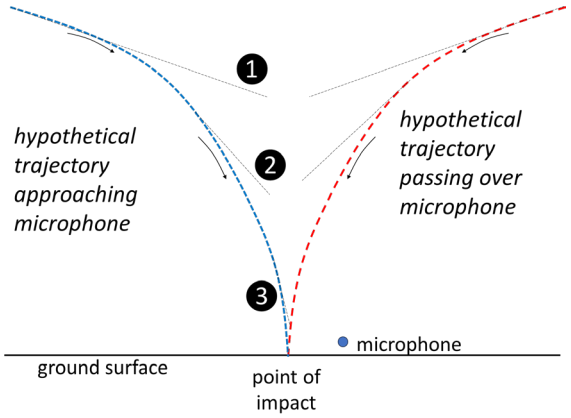

Figure 10: Hypothetical profile of descending sound source approaching vs. passing over the microphone.

The investigation of the entire incident involves many sources of information, such as examination of the physical evidence at the scene, video surveillance over the region, radar tracking information, and witness reports. The audio forensic examination from the user-generated recordings cannot answer every possible question, but the interpretation can help support or refute the various theories about the circumstances.

6 Conclusion

Each of the examples of user-generated material described in this paper, and many other examples from recent conflicts around the world, demonstrate the application of audio forensic interpretation in journalists' reports and formal incident investigations. The audio forensic viewpoint cannot answer every question, but certain details of timing, sound identification, and geometric orientation can be discerned from these recordings. It is increasingly common to have several concurrent user recordings from different locations, and the spatial diversity provides additional options for assessing direction and movement.

Although not from a war zone, the recent attempted assassination of presidential candidate Donald Trump on 13 July 2024 at a campaign rally in Butler, Pennsylvania, was documented by no fewer that 20 user-generated recordings made by spectators in the audience and in areas surrounding the rifle-wielding perpetrator. Most of the recordings were localizable by observing the mobile phone's video material with respect to known landmarks at the scene, and this provided useful comparisons of the gunshot sounds at multiple locations [11].

The examples described in this paper did not involve any assertions of forgery or misrepresentation of the recordings. Our current era of emerging generative artificial intelligence systems for audio and video will present an authenticity challenge for future use of user-generated recordings. A skilled adversary may soon be able to render fake or modified recordings that are indistinguishable from real evidence, and other means of determining authenticity will become essential.

7 Acknowledgements

The author gratefully acknowledges reporters from Bellingcat, CNN, Human Rights Watch, New York Times, and Washington Post, who provided the example material described in this paper.

References

[1] R. Morris, Jr., "The Pen & the Sword: A Brief History of War Correspondents," *Military Heritage*, vol. 8, no. 7 (2007).

[2] A. Balguygallois, "Protection of Journalists and Media Professionals in Time of Armed Conflict," (English translation), International Committee of the Red Cross, URL: https://casebook.icrc.org/case-study/protection-journalists, accessed 26 Aug 2024 (2004).

[3] "The Sound of Bullets: The Killing of Colombian Journalist Abelardo Liz," Bellingcat Investigations, URL: https://www.bellingcat.com/news/2023/12/11/the-sound-of-bullets-the-killing-of-colombian-journalist-abelardo-liz/, accessed 29 Aug 2024 (2023).

[4] D.R. Begault, S.D. Beck, and R.C. Maher, "Overview of forensic gunshot analysis

techniques," elib 20475, Proc. 2019 Audio Engineering Society International Conference on Audio Forensics, Porto, Portugal (2019).

[5] R.C. Maher, "Interpretation of audio forensic information from the shooting of journalist Shireen Abu Akleh," Express Paper 22, Proc. 153rd Audio Engineering Society Convention, New York, NY (2022).

[6] R.C. Maher, "Examining tell-tale sounds in forensic gunshot recordings," elib 22634, AES 8th International Conference on Audio Forensics, Denver, CO (2024).

[7] "One dead, eight missing in blast at factory northeast of Moscow," Reuters, URL: https://www.reuters.com/world/europe/sixteen-injured-blast-factory-northeast-moscow-tass-2023-08-09/, accessed 29 Aug 2024 (2024).

[8] "Israel: Strikes on Journalists in Lebanon Apparently Deliberate," Human Rights Watch, URL: https://www.hrw.org/news/2023/12/07/israel-strikes-journalists-lebanon-apparently-deliberate , accessed 31 Aug 2024 (2023).

[9] "Gaza: Findings on October 17 al-Ahli Hospital Explosion: Evidence Points to Misfired Rocket but Full Investigation Needed," Human Rights Watch, URL: https://www.hrw.org/news/2023/11/26/gaza-findings-october-17-al-ahli-hospital-explosion, accessed 4 Sep 2024 (2023).

[10] "A barrage and a midair explosion: What visual evidence shows about the Gaza hospital blast," Washington Post, URL: https://www.washingtonpost.com/investigations/2023/10/26/gaza-hospital-blast-evidence-israel-hamas/, accessed 4 Sep 2024 (2023).

[11] "Speculation swirls about what hit Trump. An analysis suggests it was a bullet," New York Times, URL: https://www.nytimes.com/2024/07/26/us/politics/trump-shooter-bullet-trajectory-ear.html?unlocked_article_code=1._U0.xNxX.CWDTVlrG2bXu&smid=url-share , accessed 3 Sept 2024 (2024).