Audio Engineering Society

Conference Paper

Presented at the International Conference on
Audio Forensics
2019 June 18–20, Porto, Portugal

This paper was peer-reviewed as a complete manuscript for presentation at this conference. This paper is available in the AES E-Library (http://www.aes.org/e-lib) all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Who Fired When: Associating Multiple Audio Events from Uncalibrated Receivers

Steven D. Beck[1]

[1] *Beck Audio Forensics, Austin, TX 78731, USA*

Correspondence should be addressed to stevendbeck@alumni.rice.edu

ABSTRACT

Audio recordings from crime scenes often involve multiple gunshots shots fired from multiple firearms. To further complicate crime scene audio analysis, there may be multiple recordings from spatially distributed recorders that are not time synchronized and have uncertain locations. In this paper we describe the general Time Difference of Arrival (TDOA) problem with multiple sources and multiple receivers. From these equations, we show that for each source location, the time differences between any two impulsive events are identical for all non-synchronized (uncalibrated) receivers, assuming all locations remain stationary. Assuming the receivers are not collinear, the time differences between events from different sources are not identical at different receivers, and can be used to identify the presence of multiple sources. This timing analysis method is shown to be applicable in some cases where the recordings are noisy, distorted, and have interference.

1 Introduction

Some of the most significant problems encountered when analyzing crime scene audio recordings are identifying each event, determining the time or order of emission, and relating those events among multiple recordings. For the case of a single audio recording, a standard method of identifying transient events like gunshots is by cross correlating the events with each other, or with known examples. The problem with correlation is that there is no single threshold or set of thresholds indicating a match, no match, or uncertain match condition. For gunshots in particular, there are many variations that can affect the correlation between events, even if they are generated from the same source [1-3]. The recording under test may have noise, interference, or distortions that can significantly reduce the correlation values [4]. For example, in high quality recordings, a correlation value of 0.9 or higher may indicate a match. On the other hand, the House Committee investigating the assassination of President Kennedy states that in comparing two gunshot recordings, a sample correlation coefficient greater than 0.6 may indicate significant waveform similarity [5]. Correlation may not provide reliable similarity measurements for field recordings obtained under uncontrolled conditions.

When multiple sound sources are present, those sources most likely have different physical locations. We would like to separate sounds based on their location as well as their temporal and spectral characteristics. If those sounds are recorded on multiple recorders or recording channels, where the receiving sensors are physically separated, their locations known, and they are time synchronized, then time delay of arrival (TDOA) techniques are available for separating sounds based on source

location. For instance, if we have an array of receivers with short separation distances (<0.5 wavelength separation), then TDOA techniques can be used to calculate the sound source angle of arrival. Likewise if the sounds are transient in nature and the time synchronized receivers are widely separated and with known location, then TDOA and multilateration techniques can be used to calculate the location of the sources [6,7].

Time synchronized array processing and recording systems are becoming more common. The military has developed gunshot detection and localization systems that fit on vehicles and some small enough to be worn by solders. There are also commercial systems with widely spaced receivers that use TDOA multilateration to locate transient gunshot-like sounds, like the ShotSpotter system [8]. Recordings from these systems are generally not available without a request through legal channels.

The most common recordings from crime scenes are single-channel recorders like cell phones or police recorders. Most police cars are equipped with dash cams that include an audio recorder. Many police are now required to wear body cams that include a single audio recording channel. If multiple recordings of a crime scene were made, the problem is that the recorder location is usually unknown or uncertain, the recordings are not time-synchronized, and the recordings are often poor quality (heavily compressed or clipped if gunshots are involved). While having these crime scene recordings is beneficial, extracting additional information like source locations can be almost impossible. As will be shown, the order of the recorded events may also change from recorder to recorder depending on the source-receiver geometry corresponding to differences in propagation path length.

This focus of this paper is about associating impulse gunshot sounds, referred to here as events, recorded on non-synchronized receivers. These events can also be other sounds with an identifiable start or peak time. Section 2 is a review of the general distance and TDOA equations for multiple sources and multiple non-synchronized receivers. Source localization under these conditions is a non-convex problem with many solutions [9, 10]. In Section 3, we derive a technique that can uniquely associate events from the same location using multiple non-collinear recorders, and discriminate them from events generated from different source locations. Section 4 shows the application of this technique to an example problem. Finally, Section 5 shows the results of applying this technique to a real example of three independent recordings of the same gunshots in the presence of interference and distortion.

2 Review of Generalized Multi-Source Multi-Receiver TDOA

For source localization, the problem is to determine the coordinates of a source emitter using the time difference of arrival (TDOA) measurements from a distributed array of receivers. Following [9,10], let the i^{th} receiver r_i be at location (x_i, y_i, z_i) and the m^{th} source s_m be at location (x_m, y_m, z_m), and assume that all receiver and source locations are stationary. The squared distance between the i^{th} receiver and the m^{th} source is then:

$$R_{im}^2 = \|r_i - s_m\|^2 = (x_i - x_m)^2 + \left(y_i - y_m\right)^2 + (z_i - z_m)^2 \quad (1)$$

If we let η_m be the time offset (for synchronization) from the m^{th} source, let δ_i be the time offset for the i^{th} receiver, and let v be the speed of sound, then the time of arrival (TOA) at the receiver is:

$$t_{im} = \frac{R_{im}}{v} + \eta_m + \delta_i \quad (2)$$

The time difference of arrival (TDOA) between the i^{th} and j^{th} receiver for the m^{th} sensor event is:

$$\tau_{jmim} = t_{jm} - t_{im} = \frac{(R_{jm} - R_{im})}{v} + \delta_j - \delta_i \quad (3)$$

The TDOA between the m^{th} and n^{th} source arriving at the i^{th} receiver is:

$$\tau_{inim} = t_{in} - t_{im} = \frac{(R_{in} - R_{im})}{v} + \eta_n - \eta_m \quad (4)$$

Setting up the coordinate system, the receivers r_1 and r_2 can be set to the following locations:

$$r_1 = [0, 0, 0],\ r_2 = [x_{r2}, 0, 0] \quad (5)$$

Since all receivers and sources are non-synchronous, we will assume time $\delta_1 = 0$ for the first event

arriving at the first receiver. Working with (4) and setting $\eta_{nm} = \eta_n - \eta_m$:

$$0 = \tau_{inim} + \frac{(R_{im}-R_{in})}{v} - \eta_{nm} \quad (6)$$

Assuming two sources (m=1, n=2), then the equations for receivers i=1 and j=2 are:

$$0 = \tau_{1211} + \frac{(R_{11}-R_{12})}{v} - \eta_{21} \quad (7)$$

$$0 = \tau_{2221} + \frac{(R_{21}-R_{22})}{v} - \eta_{21} \quad (8)$$

Subtracting (7) from (8) eliminates the unknown source times yielding:

$$0 = (\tau_{2221} - \tau_{1211}) + \frac{(R_{21}-R_{22}-R_{11}+R_{12})}{v} \quad (9)$$

The terms τ_{2221} and τ_{1211} are measured from the data and v is assumed to be known. If the receiver locations are known, the source locations can be simultaneously estimated via the maximum likelihood method:

$$\hat{S} = \underset{S}{\arg min}\left((\tau_{2221} - \tau_{1211}) + \frac{(R_{21}-R_{22}-R_{11}+R_{12})}{v}\right)^2 \quad (10)$$

Equation (10) has many variables and nonlinear terms. Minimization is non-convex and sensitive to initial parameter selection - resulting in a large number of local solutions.

3 Source Association Among Unsynchronized Receivers

The analysis above indicates that if the sources and the receivers all have unsynchronized times, even if the receiver locations are known, accurately estimating the source locations is almost impossible without further information or constraints. In many forensic situations, the receivers have unknown or uncertain locations as well as being unsynchronized. It turns out that even under these conditions, multiple events from a single location can be associated, and can be distinguished from events emanating from different locations, assuming the receivers are not collinear or not mirror images about a collinear array.

Expanding on (2), if we let μ_{mk} be the time offset to the k^{th} event from the m^{th} source, and μ_{ml} be the time offset to the l^{th} event from the same source, then the times of arrival at the i^{th} receiver are:

$$t_{imk} = \frac{R_{im}}{v} + \delta_i + \eta_m + \mu_{mk} \quad (11)$$

$$t_{iml} = \frac{R_{im}}{v} + \delta_i + \eta_m + \mu_{ml} \quad (12)$$

The time difference of arrival between the k^{th} and the l^{th} events from the m^{th} source at the i^{th} receiver with receive offset time δ_i is then:

$$\tau_{imk,iml} = t_{iml} - t_{imk} = \mu_{ml} - \mu_{mk} \quad (13)$$

The time difference of arrival between the k^{th} and the l^{th} events from the m^{th} source at the j^{th} receiver with receive offset time δ_j is:

$$\tau_{jmk,jml} = t_{jml} - t_{jmk} = \mu_{ml} - \mu_{mk} \quad (14)$$

Equations (13) and (14) show that the time difference between multiple events from the same source is the same, even though the receivers are unsynchronized (different receive offset times) with unknown locations and the source location is unknown. This is graphically illustrated in Figure 1.

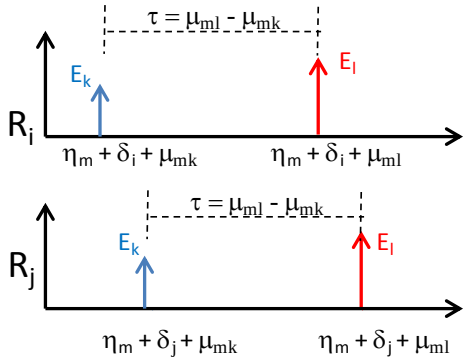


Figure 1. The TDOAs τ between the k^{th} and l^{th} events from the m^{th} source is the same at both the i^{th} and j^{th} receivers.

If the l^{th} event is generated by the n^{th} source (at a different location), then equation (12) becomes:

$$t_{inl} = \frac{R_{in}}{v} + \delta_i + \eta_n + \mu_{nl} \quad (15)$$

The TDOAs between the k^{th} and the l^{th} events from the two sources at the i^{th} and the j^{th} receivers are:

$$\tau_{imk,inl} = t_{inl} - t_{imk} = \frac{R_{in}}{v} + \eta_n + \mu_{nl} - \frac{R_{im}}{v} - \eta_m - \mu_{mk} \quad (16)$$

$$\tau_{jmk,jnl} = t_{jnl} - t_{jmk} = \frac{R_{jn}}{v} + \eta_n + \mu_{nl} - \frac{R_{jm}}{v} - \eta_m - \mu_{mk} \quad (17)$$

Equations (16) and (17) are the same except for the times $(R_{in}–R_{im})/v$ and $(R_{jn}-R_{jm})/v$. This result means that the TDOAs between events generated by different sources will be different at different receivers.

In summary, equations (13) and (14) show that multiple events from one source location at arriving at multiple unsynchronized receivers can be associated by their similar TDOAs. Equations (16) and (17) show that under these conditions, events emanating from sources at different locations can be distinguished.

4 An Example with Two Sources and Three Unsynchronized Receivers

The following example illustrates the technique of associating multiple source events from multiple unsynchronized receivers. To set the problem up, assume there are two sources that generate acoustic events at the following times as shown in Figure 2:

Source #	1	2	1	1	2
Event Times (s)	1	2	3	4	10

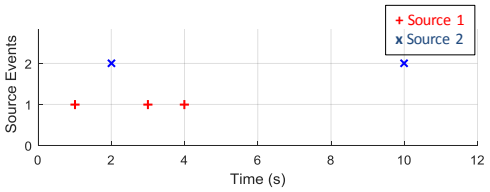


Figure 2. Example of multiple ground truth event times generated from multiple sources.

The relative locations for the two sources and the three receivers are shown in Figure 3.

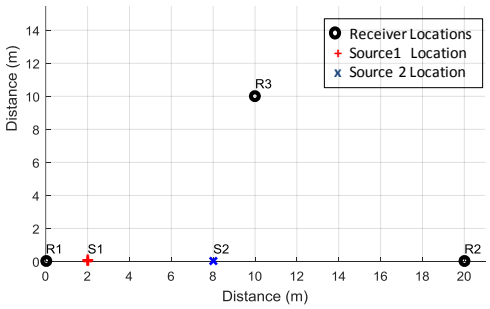


Figure 3. Example ground truth source (S) and receiver (R) locations on a 2-D grid.

For this example, assume an event wave traveling at a speed of v=1 (a normalized velocity). To generate this example, the ground truth reception times for the three receivers are given in Table 1 and shown in Figure 4. Events generated by source 1 are shown by a red "+" and those by source 2 are a blue "x".

R1	R2	R3
3.0 (S1)	14.0 (S2)	12.2 (S2)
5.0 (S1)	19.0 (S1)	13.8 (S1)
6.0 (S1)	21.0 (S1)	15.8 (S1)
10.0 (S2)	22.0 (S1)	16.8 (S1)
18.0 (S2)	22.0 (S2)	20.2 (S2)

Table 1. Ground truth event received times listed for each receiver.

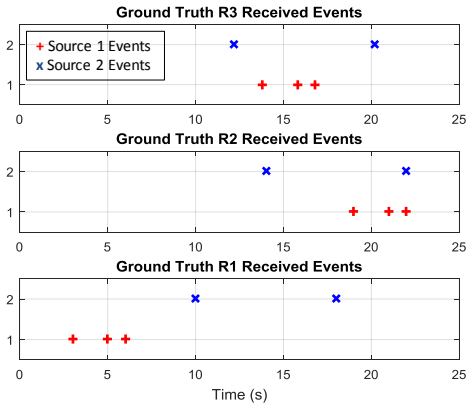


Figure 4. Ground truth (synchronized) receiver event times for each of the three receivers

Figures 2-3 illustrate how the example problem was set up and how the events were generated. We are

interested in the data from unsynchronized receivers, where the starting times of the first event is independent among all receivers. Subtracting the first event time in the file gives the received event times (*what the analyst starts with*) for each unsynchronized receiver, as shown in Figure 5.

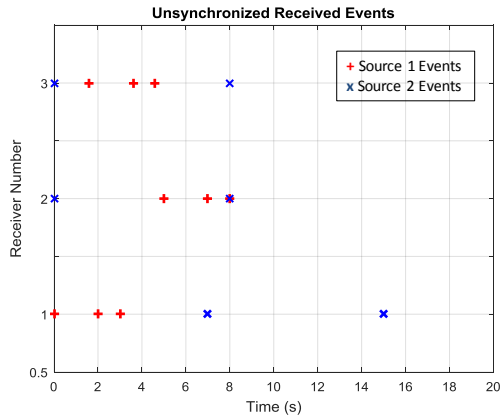


Figure 5. Unsynchronized event times at each of the three receivers.

If the distributed receivers are unsynchronized, then the analyst is presented with the following:

- The start of the first event at each receiver is assumed to be time 0.
- The order and timing of the sequence of events may not be the same for each receiver
- The first event arriving at each receiver may not be from the same source

In order to demonstrate association of events from the same source, we can slide the individual receiver plots and show that the TDOA's per source are the same. In order to align the events from each source, the time series from receiver R1 was correlated with the time series from both receivers R2 and R3. When R1 was correlated with R2, the time offset of the highest coefficient indicated R2 was delayed 5.0 s relative to R1 (aligning source 1 events), and the time offset of the second highest coefficient indicated that R1 was delayed 7.0 s relative to R2 (aligning source 2 events). When R1 was correlated with R3, the time offset of the highest coefficient indicated R3 was delayed 1.6 s relative to R1 (aligning source 1 events), and the time offset of the second highest coefficient indicated that R1 was delayed 7.0 s relative to R2 (aligning source 2 events).

Figure 6 shows the alignment of the events generated by source #1. The vertical dashed lines indicate that the time differences are the same for all three receivers.

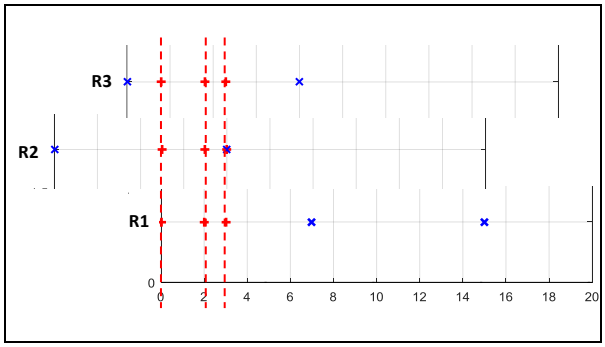


Figure 6. TDOA alignment and association of source 1 events.

Figure 7 shows the alignment of the events generated by source #2. Again, the vertical dashed lines indicate that the time differences are the same for all three receivers.

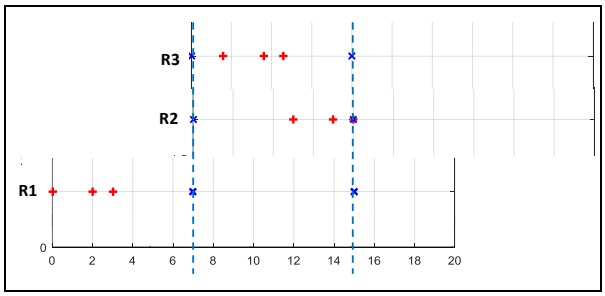


Figure 7. TDOA alignment and association of source 2 events.

Notice that the fourth event on receiver 2 aligns with the events from both source #1 and source #2. This and the fact that there are five detected events at receivers 1 and 3, but only four at receiver 2 indicate that there are overlapping events on receiver 2.

5 A Real Example with Noisy and Distorted Signals

On March 30, 1981, U.S. President Reagan left a speaking engagement at the Washington Hilton Hotel. As President Reagan approached his car, John Hinkley Jr. produced a Rohm .22 caliber 6-shot revolver and fired multiple times in the direction of the President [11]. The three major news services CBS, ABC, and NBC were on the scene filming the President and recorded the gunshots on audio. As part of a pilot series in 2013, a major cable network company questioned if all of the recorded gunshots were from one source location, or if there was a possibility of multiple shooters. The three audio sets were recorded in a crowd of people yelling at the President, and all recordings had wind and crowd noise. John Hinkley Jr. was standing very close to the recorders, and all gunshot events were highly distorted and did not look like the classic muzzle blast waveform. The source and receiver locations were unknown and the recorders were unsynchronized.

After analyzing the audio recordings, eight gunshot-like events were found in the CBS file, eight events were found in the ABC file, and four events were found in the NBC file (the sound faded after the fourth event). The audio data time series plots with labels for each detected event are shown in Figure 8.

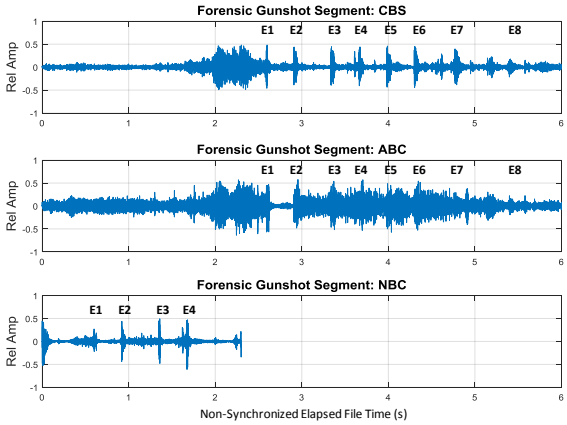


Figure 8. CBS, ABC, and NBC audio data with labeled events

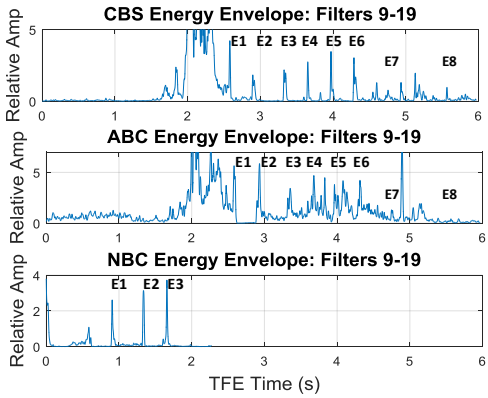


Figure 9. Energy envelopes calculated from the 200-2000 Hz bandpass region.

These times series data are noisy and the events are distorted, which makes identifying the events and start times difficult. Event start times were measured from the energy envelope, shown in Figure 9. The energy envelope was computed by accumulating the energy in the 200 Hz – 2000 Hz bands from a third octave spectrogram.

The time difference between successive detected events is shown in Figure 10. The time differences for each successive pair are almost identical (less than a millisecond difference likely due to small movement and measurement error) for all three audio files, indicating that they all originated from the same location. Additional analysis indicated that the first six events are firearm muzzle blasts and the last two events are revolver hammer clicks.

Event Time Differences	E2-E1	E3-E2	E4-E3	E5-E4	E6-E5	E7-E6	E8-E7
CBS Time Diff	0.3182	0.4284	0.3269	0.3181	0.3179	0.4590	0.6377
ABC Time Diff	0.3180	0.4290	0.3270	0.3185	0.3170	0.4600	0.6365
NBC Time Diff	0.3176	0.4285	0.3267				

Figure 10. Successive events time differences for three non-synchronized audio recordings.

Correlation analysis of the signal envelopes from two receivers can be used to find the time delays between each pair of event sequences. A correlation peak will occur when the event sequences line up. The CBS and ABC envelope signals between 2.5 to

3 seconds were correlated using the FFT-based sample correlation technique described in [4]. The resulting correlogram, shown in the top plot in Figure 11, shows a peak correlation value of 0.56 with the ABC events lagging the CBS events by 6 ms. Correlations between the CBS and NBC envelope, shown in the bottom plot of Figure 11, show a peak correlation value of 0.54 with the CBS events lagging the NBC events by 2.0s.

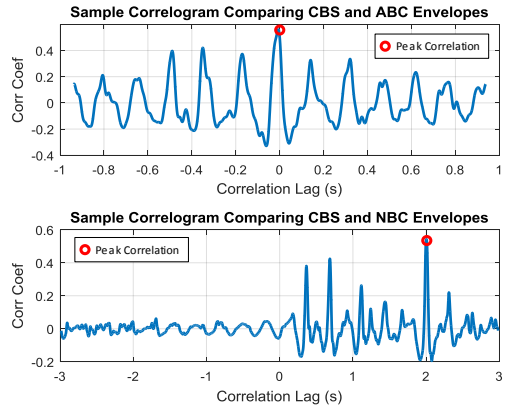


Figure 11. Sample correlations comparing receiver envelope event sequences for maximum alignment.

Visual inspection of the aligned signals, as shown in Figure 12, is advised to avoid ambiguity from multiple peaks and aid in understanding the events, particularly when they are noisy and/or distorted.

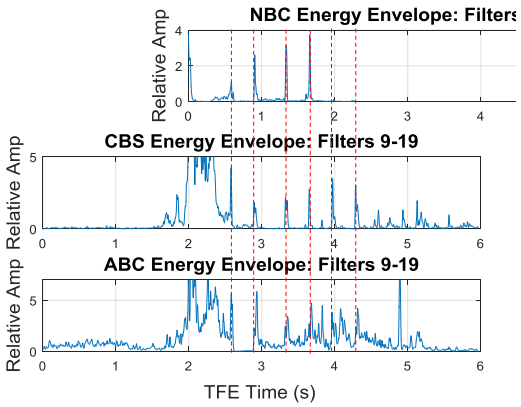


Figure 12. The near perfect alignment of the CBS, ABC, and NBC gunshot sequences indicate they were all fired from one source location.

6 Discussion and Conclusions

In many audio forensic cases, especially involving gunshots, there are multiple sound sources and multiple independent and uncalibrated (non-synchronized) receivers. The locations of the sources and receivers are usually unknown or at least uncertain. The recorded signals are often corrupted with noise, interference, and distortion, making association with correlation unreliable. Without additional constraints, the source locations cannot be reliably calculated using TDOA methods. However, if the sources emit multiple events while all locations remain stationary and the receivers are not collinear, the TDOA analysis described here can be applied to align and associate the events from each source, while distinguishing the events from other sources.

Acknowledgements

I would like to acknowledge producers Nihar Patel and Tod Mersirow for providing me with the ABC, CBS, and NBC data and encouraging my work on this project. I would also like to acknowledge the helpful comments provided by Prof. Rob Maher and the AES reviewers.

References

[1] B. Koenig, S. Hoffman, H. Nakasone, and S. Beck, "Signal Convolution of Recorded Free-Field Gunshot Sounds" *J. Audio Eng. Soc*. Vol. 46, No. 7/8, (1998).

[2] S. Beck, H. Nakasone, K. Marr, "Variations in Recorded Acoustic Gunshot Waveforms Generated by Small Firearms," *J. Acous. Soc. Am*. 129 (4), pp. 1748-1759 (2011).

[3] R.C. Maher, "Acoustical characterization of gunshots," Proc. IEEE SAFE 2007: Workshop on Signal Processing Applications for Public Security and Forensics, Washington, DC, pp. 109-113, (April, 2007).

[4] S.D. Beck, "A Short-Time Cross Correlation with Application to Forensic Gunshot Analysis", *Proceedings of the AES*

International Conference on Audio Forensics, (2019) – submitted.

[5] *Report of the Select Committee on Assassinations of the U.S. House of Representatives, Washington, DC: United States Government Printing Office (1979*).

[6] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*, Springer-Verlag, Berlin, (2008).

[7] R.C. Maher and E.R. Hoerr, "Audio Forensic Gunshot Analysis and Multilateration," Preprint 10100, Proc. 145th Audio Engineering Society Convention, New York, NY, (October, 2018)

[8] J. R. Aguilar, "Gunshot detection systems in civilian law enforcement," *Journal of the Audio Engineering Society*, vol. 63, no. 4, pp. 280–291,(2015).

[9] A. Plinge, F. Jacob, R. Haeb-Umbach, and G. Fink, "Acoustic Microphone Geometry Calibration", *IEEE Signal Processing Magazine*, Vol 33, No 4, (July 2016).

[10] L. Wang, T-K. Hon, J. Reiss, and A. Cavallaro, "Self-Localization of Ad-Hoc Arrays Using Time Difference of Arrivals*", IEEE Transactions On Signal Processing*, Vol. 64, No. 4, (Feb. 2016).

[11] D.Q.Wilber, *Rawhide Down, the Near Assassination of Ronald Reagan*, Henry Holt, NY (2011).