**ShotSpotter**

ShotSpotter is one of the most widely used gunshot detection systems. It is designed to detect, pinpoint, and report gunfire incidents in real time.

**1. Acoustic Sensor Array**

* **Microphones**: Arrays of microphones mounted on rooftops, utility poles, or other elevated locations.
* **Omnidirectional Sensing**
* **Sensor Density**: The sensors are spaced between 0.5 to 2 kilometres apart. A minimum of three sensors are required to triangulate the location of a gunshot.

**2. Signal Processing and Detection Algorithms**

* **Acoustic Signature Recognition**: Gunshots produce a specific acoustic signature that differs from other sounds. ShotSpotter uses proprietary algorithms to analyze the sound wave's characteristics, such as the shockwave and muzzle blast.
* **Multi-Sensor Detection**
* The system continuously records ambient sound but only activates detailed analysis when a potential gunshot is detected. This helps reduce false alarms and limits the amount of data processed.

**3. Location and Triangulation**

* **Triangulation**: By using the data from at least three sensors, the system can calculate the gunshot's location with high accuracy. Advanced algorithms take into account variables like the speed of sound and environmental conditions to refine the calculation.
* **GPS and Clocks**: Each sensor is equipped with a GPS receiver and a highly accurate clock, which ensures that the time-of-arrival data is synchronized across sensors.

**4. Data Transmission and Cloud Processing**

* **Edge Computing**: Some level of initial processing is done at the sensor level to filter out irrelevant sounds and reduce the burden on the central system. Only relevant sound events are sent to the cloud for further analysis.
* **Cloud Processing**: Once the data is transmitted to ShotSpotter's cloud-based servers, advanced algorithms continue processing the data, applying filtering techniques to ensure the detected sound is indeed gunfire.
* **Artificial Intelligence (AI)**: AI-powered algorithms are used to improve detection accuracy, minimizing false positives by distinguishing gunfire from other loud sounds.

**5. Human Verification**

**6. Real-Time Alerts and Integration**

**7. Gunfire Location**

**Audio Waveform Visualization**: The system creates an audio waveform of the gunshot event, which helps in analyzing the sound characteristics. This data can be stored and used in investigations and legal proceedings.

* **Geolocation Accuracy**: By leveraging data from multiple sensors, ShotSpotter can determine the gunshot’s location with accuracy typically within 25 meters.

**8. Environmental Considerations**

* **Weather Adjustments**
* **Urban Obstacles**: Buildings and other structures can reflect or block sound waves. The system's algorithms are designed to account for these obstacles when determining the location of the gunfire.

**9. Integration with Law Enforcement Systems**

* **Dispatch Systems**:
* **Evidence Collection**

**10. Cloud-Based Data Management**

* Historical Data for Crime Analysis

**Algorithms Used in Gunshot Detection:**

**Acoustic Signature Recognition**

* **Waveform Pattern Recognition**: Gunshots produce specific acoustic signatures characterized by a rapid rise in sound intensity, followed by a gradual decay. This "impulse" is distinct from most other noises. Algorithms are trained on the sound patterns of various types of gunfire to recognize their signature.
  + **Muzzle Blast**: The muzzle blast is the sound produced by expanding gases as a bullet leaves the firearm. It has a specific frequency and amplitude pattern.
  + **Shockwave**: Some high-velocity bullets create a sonic boom (shockwave) as they travel faster than the speed of sound. This produces a sharp acoustic signal that can help identify a gunshot.

**Signal Processing (Fourier Transform)**

* **Fast Fourier Transform (FFT)**: This algorithm converts the sound wave from the time domain to the frequency domain. This allows the system to analyze the frequency spectrum of the sound and separate gunshot-like sounds from other noise.
  + **Frequency Filtering**: The system applies filters to focus on the frequency range typically associated with gunshots. Gunshots usually fall within specific high-frequency ranges that differentiate them from other urban noises.
  + **Windowing**: By applying windowing functions (such as Hamming or Hanning windows), the system isolates the gunshot impulse from surrounding noise and analyzes only the segment of the audio signal that likely contains the gunshot.

**Cross-Correlation for Sound Source Matching**

Cross-correlation also helps identify whether the same sound event is being detected at different locations, validating that the event is indeed a gunshot.

**Triangulation for Gunshot Localization**

**Time-of-Arrival (TOA) Measurement**

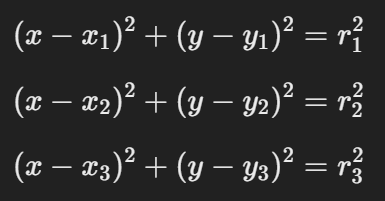
The system measures the time difference between when each sensor detects the gunshot. This time difference is then used to calculate the relative distances from the source to each sensor.

**Triangulation**

* **Distance Calculation**: Once the TOA is measured for each sensor, the system calculates the distance between the gunshot and each sensor using the formula:

D = V . T

* **Circle Intersection Method**:
  + **Three-Sensor System**: In a 2D plane, the three sensors' calculated distances form three circles. The intersection of these three circles gives the precise location of the gunfire.
  + **Hyperbolic Positioning (Multilateration)**: Hyperbolic lines are plotted for each sensor pair, and the intersection of these hyperbolas gives the source location.



Where:

* (x1​,y1​), (x2,y2),and (x3​,y3​) are the coordinates of the sensors,
* r1​, r2​, and r3​ are the calculated distances from each sensor to the gunshot.

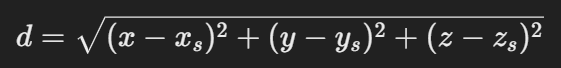
**Z-axis Calculation:**

The fundamental principle behind calculating the Z-axis position (elevation) is the **Time Difference of Arrival (TDOA)**, which is also used for X and Y coordinates. The process follows these steps:

* To accurately calculate the Z-axis, at least four sensors are needed. These sensors need to be positioned at varying heights or in locations with different Z-axis elevations to detect vertical differences.

The system uses these TDOA measurements to generate **hyperbolas** in 3D, instead of simple circles in 2D. The intersection of these hyperbolas across multiple sensors gives the 3D coordinates (X, Y, Z) of the gunshot source.

**Sound Travel Time**: The sound's travel time to each sensor is affected by its 3D distance from the gunshot source. This distance can be broken down into:



The system uses the geometric relationships between sensors to calculate the Z-axis. For example, a gunshot occurring at ground level will produce shorter time delays between sensors placed on the same plane, while a gunshot fired from a higher elevation will introduce larger time differences between ground-level sensors and elevated sensors**.**

**Random Forests**: A decision-tree-based ensemble algorithm. Each tree in the forest is trained on a subset of features (frequency spectrum, time duration, rise/fall pattern) and votes on whether a sound is a gunshot. The final decision is made by aggregating the votes across all trees.

**Support Vector Machines (SVM)**: An SVM classifier finds the optimal hyperplane that separates gunshot sounds from non-gunshot sounds. Features like the sound's frequency range, time-domain characteristics, and amplitude rise time are used as input data points.

**Artificial Neural Networks (ANN)**: ANN models, especially Convolutional Neural Networks (CNNs), are often used for audio classification tasks. They can automatically extract relevant features from raw sound waveforms, such as the temporal and frequency-based characteristics of a gunshot. CNNs can be trained on spectrograms (graphical representations of the sound signal's frequency spectrum over time) to identify patterns unique to gunfire.

**Recurrent Neural Networks (RNNs)**: Given that sound data has a time-sequential nature, RNNs, or their variant **Long Short-Term Memory (LSTM)** networks, are useful for detecting patterns over time. They can identify temporal dependencies in the sound data to help differentiate gunshots from other impulsive noises. These networks analyze the sequence of sound frames, allowing the system to make decisions based on how the sound evolves.

**Feature Extraction**: AI models typically use features such as:

* **Frequency and Amplitude Peaks**
* **Duration**: Small Duration

**Spectrograms**

A spectrogram is a visual representation of the spectrum of frequencies in a sound signal as it varies with time. It's essentially a heat map that shows how the energy of a sound signal is distributed across various frequencies over a given period.

The key components of a spectrogram include:

* Time (X-axis): The horizontal axis represents time, showing how the sound signal evolves.
* Frequency (Y-axis): The vertical axis represents frequency, with lower frequencies at the bottom and higher frequencies at the top.
* Amplitude/Intensity (Color or Brightness): The color or brightness at any point in the spectrogram represents the intensity or amplitude of the sound at a particular frequency and time. Brighter colors (or more intense colors) usually represent higher energy, while darker colors indicate lower energy.

To create a spectrogram, we use a process called the **Short-Time Fourier Transform (STFT)**. Here's how it works:

1. **Windowing the Signal**: The sound signal (a time-domain waveform) is divided into short segments or windows. This is done because the Fourier transform works best for analyzing stationary signals, and by windowing the sound, we can treat short segments as approximately stationary.
2. **Applying Fourier Transform**: The **Fourier Transform** is applied to each windowed segment of the signal, converting it from the time domain to the frequency domain. This process gives us the frequency components present in each window.
3. **Combining the Windows**: The result is a sequence of frequency spectra for each time window, which are then combined to form the spectrogram. The amplitude (or power) of each frequency component at each time step is plotted as color or brightness on the spectrogram.

This process generates a visual map of how the frequency content of the signal evolves over time, which is the key advantage of spectrograms over simple time-domain waveforms.

AI models, particularly **deep learning models**, take spectrograms as input because spectrograms provide a highly informative view of the sound data. By treating the spectrogram like an image, AI models can extract useful patterns and features for classification tasks.

* **CNNs (Convolutional Neural Networks)**: CNNs are widely used for analyzing spectrograms because they are highly effective in identifying spatial patterns in images. In the case of a spectrogram, these "spatial" patterns correspond to time-frequency patterns of sound. For example:
  + **Vertical patterns** might correspond to frequency bands excited by certain sound events (like gunshots).
  + **Horizontal patterns** could represent sustained frequencies or harmonics over time.

CNNs use convolutional layers to detect local patterns in the spectrogram (like edges or texture in an image). These patterns help identify unique features of gunshots, such as sharp, high-frequency bursts that are characteristic of gunfire.

**Spectrogram Types Used in AI**

There are a few types of spectrograms that may be used, depending on the task:

**1. Mel Spectrogram**

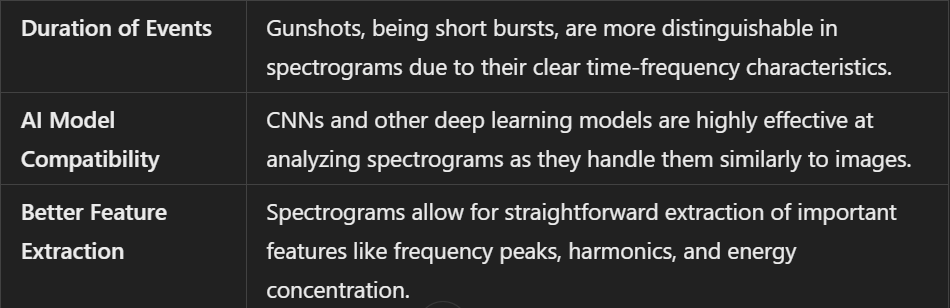
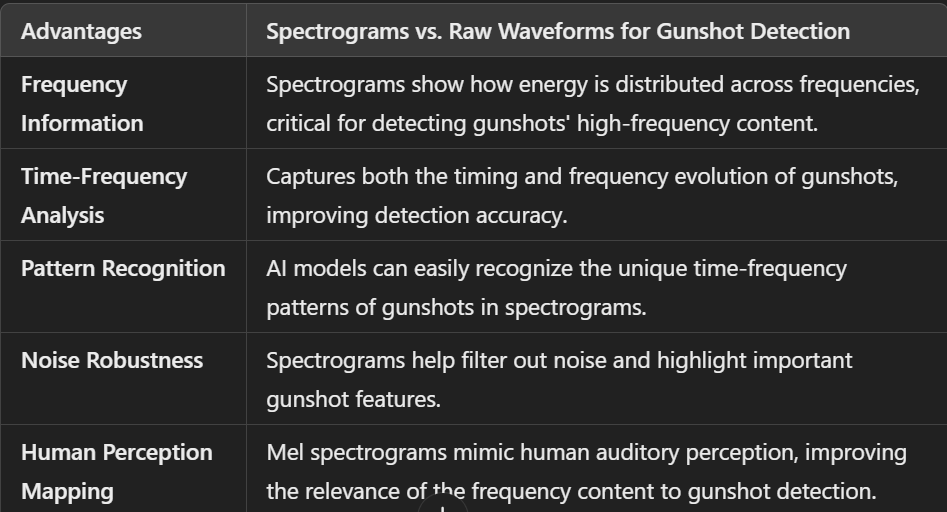
* The **Mel scale** is a perceptual scale that mimics how humans perceive sound, focusing on lower frequencies. A **Mel spectrogram** is commonly used in audio processing because it emphasizes frequencies that are more relevant to human hearing.
* Mel spectrograms are particularly useful in tasks like gunshot detection, where the goal is to distinguish sounds based on human-audible characteristics.

**2. Log-Frequency Spectrogram**

* A **logarithmic frequency scale** is sometimes used instead of a linear one. This scale is useful because it gives more resolution to lower frequencies, which are typically more important in audio classification tasks.

**3. Power Spectrogram**

* A **power spectrogram** shows the squared magnitude of the signal at each frequency, emphasizing the intensity of different frequency components over time. This can help in identifying louder, more energetic sounds like gunshots.



<https://www.mdpi.com/2079-9292/11/23/3859>

<https://digital.auraria.edu/files/pdf?fileid=e562b890-41f9-4170-a13b-27b07d1f1626>

<https://scholarworks.indianapolis.iu.edu/server/api/core/bitstreams/52ce1fe4-9fea-47eb-84a9-4631b0beade9/content>

* Include 3rd axis(z-axis)
* Spectrogram analysis
* Train models based on spectrograms(CNN’s)