White Paper

Intelligent Audio Analysis Technology

Classification, Direction Detection White Paper & Installation Guide

February 2026

Contents

1. Introduction

This white paper introduces Hanwha Vision's AI-based audio analysis technology, which integrates sound classification, direction detection, and aggression detection into a unified multimodal security solution. These AI-driven technologies work in synergy to maximize situational awareness and set a new standard for next-generation security systems. The solution is designed for system integrators, IT administrators, security operators, and professionals seeking to go beyond traditional video surveillance to enable proactive sound-based threat detection.

This white paper provides an in-depth look into Hanwha Vision's deep learning–based **sound classification and direction detection technologies**, offering optimized application guidelines tailored to diverse environments.

By evolving from simple detection to full situational awareness, these capabilities serve as a preventive tool that enhances early threat recognition and supports crime deterrence.

2. Background

In the face of unseen threats, sound remains one of the most powerful yet often overlooked tools in surveillance. While traditional video security systems have focused on visually capturing what is happening, today's security environments have evolved to the stage of identifying what type of sound events occur and where they originate. As the boundaries of public safety and asset protection continue to expand, audio analysis technology has emerged as more than a supplemental tool—it now holds significant potential for crime prevention and rapid incident response.

With the growing demand for multimodal security, AI-powered audio analytics is becoming an essential element of proactive threat detection.

As AI technology advances, the security industry has moved beyond merely recording and listening to audio. Cutting-edge systems can now detect and analyze abnormal sounds such as screams or breaking glass in real time, effectively addressing the blind spots of video-only monitoring. This capability proves especially valuable in environments where visual surveillance is restricted or limited, such as out-of-hours buildings, restrooms, and medical facilities.

This autonomous technology is designed to proactively detect events that may not be easily perceived through visual information alone, using sound-based analysis to improve situational awareness. The detected audio events can be integrated with other devices or monitoring systems, allowing operators to recognize and respond to situations more quickly and effectively.

As concerns over privacy continue to grow, reluctance toward video exposure is increasing. In this context, detection technologies that leverage various sensors, particularly AI audio analysis minimizing personally identifiable elements, are gaining attention as a privacy-conscious alternative for intelligent security systems.

3. AI-Based Audio Analysis Technology

3.1. Sound Classification

Hanwha Vision's Sound Classification technology is built on a core deep learning model: the **Convolutional Neural Network (CNN)**. The process begins by converting audio signals, an abstract form of data, into **spectrograms**[1], a visual representation optimized for CNN processing. Through this approach, the system **identifies and classifies various sound events with a high degree of confidence**, including screams, breaking glass, car horns, tire skidding, and **sounds of aggression**.

A spectrogram serves as an "acoustic fingerprint" that distinctly reveals the unique frequency patterns of each sound type. The CNN excels at automatically learning and recognizing subtle acoustic patterns and features within these spectrogram images—details often indistinguishable to the human ear.

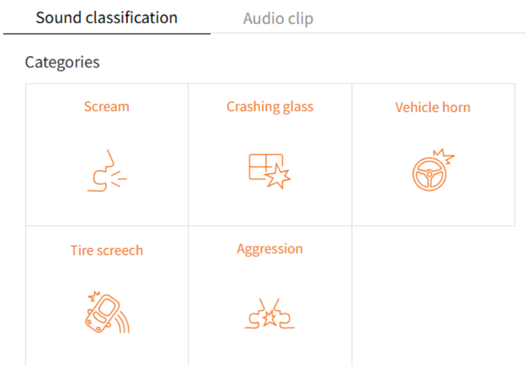
Among these capabilities, the aggression sound detection technology analyzes audio data in real time to detect vocal expressions that indicate aggressive emotion, such as shouting in anger or raising one's voice to attract attention. Unlike conventional methods that rely solely on measuring sound volume (dB), this technology uses a deep learning model to analyze nonverbal acoustic patterns in the frequency domain.

By evaluating features such as energy fluctuation, frequency distribution, and spectral density—regardless of speech content—it determines whether a sound should be classified as aggression. This enables effective detection of aggressive vocal patterns even in everyday environments, supporting proactive response to unexpected or potentially hazardous situations.

Once sound detection and classification are completed, the system automatically performs a data extraction process across the full audio stream.

Since the audio sampling and preprocessing are already accomplished, the separated sound segments are generated as audio clip files, accompanied by metadata for easy download and utilization.

This sound classification technology is available on products that support the Hanwha Vision's WiseSound App.



[1] A spectrogram is a visual representation of the frequencies of a signal as they change over time.

3.2. Sound Direction Detection

Hanwha Vision's Sound Direction Detection technology supports a rapid response by identifying and notifying users of the direction of a specified audio event. The technology determines this direction by measuring the **Time Difference of Arrival (TDoA)** of the sound signal as it reaches multiple, physically separated microphones.

The TDoA algorithm works by analyzing the phase difference in the time it takes for a sound to reach each microphone, thereby estimating the actual distance to the source. This information is then used to calculate the angle of the sound source. As illustrated in Figure 1, a multi-microphone system with microphones (MIC1,MIC2,MIC3,MIC4) arranged in a circle can determine the distance differences (d1,d2,d3,d4) between the sound source and each microphone. Calculating the time difference of arrival based on these distance differences is the core of the TDoA algorithm.

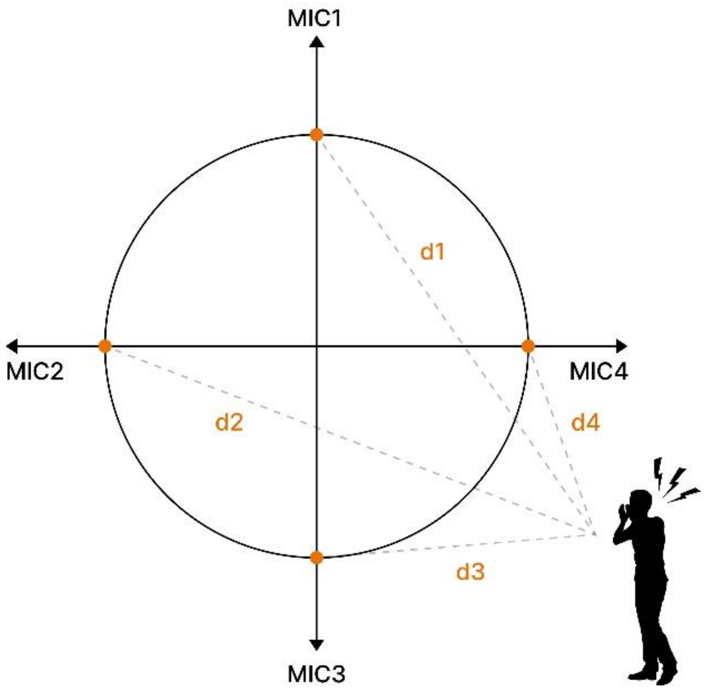


Figure 1: Sound Direction Detection Using Multiple Microphones

Figure 2 visually demonstrates the time difference (τij) in the arrival of a sound signal at two microphones (brown and blue waveforms). By precisely measuring these arrival time differences, the system can accurately triangulate the direction of the sound source.

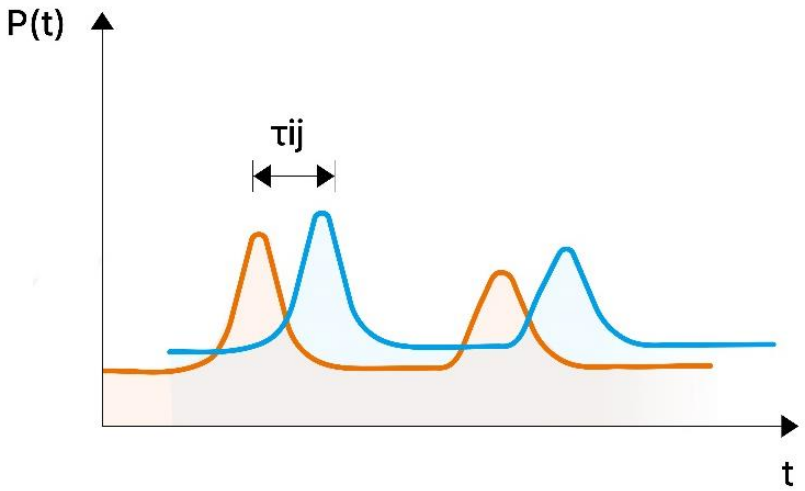


Figure 2: The Concept of Time Difference Measurement in the TDoA Algorithm

The sound direction detection process is broken down into four main steps:

1. **Signal Collection**: Simultaneously collect sound signals via multiple microphones.
2. **Signal Processing**: Analyze the collected signals using a specialized algorithm.
3. **Direction Estimation**: Estimate the sound's direction based on the processed signal.
4. **Result Output**: Display the final detected direction as a bearing angle.

This technology is available on Hanwha Vision products that support multiple microphones, such as Audio Beacon (SPS-A100M) and certain Wisenet 9 SoC-equipped cameras.

Figure 3: Microphone Locations on Audio Beacon SPS-A100M (Left) and Wisenet 9 Cameras

4. Installation and Environment: A Guide to Optimal Performance

The effectiveness of Hanwha Vision's AI Audio Solution is closely tied to its installation environment. By actively considering the following points, you can maximize the system's potential and ensure stable performance.

4.1. Selecting the Optimal Installation Location

For reliable Sound Classification and Direction Detection performance, the following conditions are recommended:

- **Sound Classification**: The system operates most reliably when the distance between the product and the sound source is at least 2m. This distance is based on the height of a sound source. If the distance is too close (within 2m), even a seemingly low-volume sound like a clap can become excessively loud, leading to false positives. **Ceiling installation in an indoor setting is an ideal method** for sound classification as it minimizes acoustic reflections and allows for uniform sound detection across a wide area.

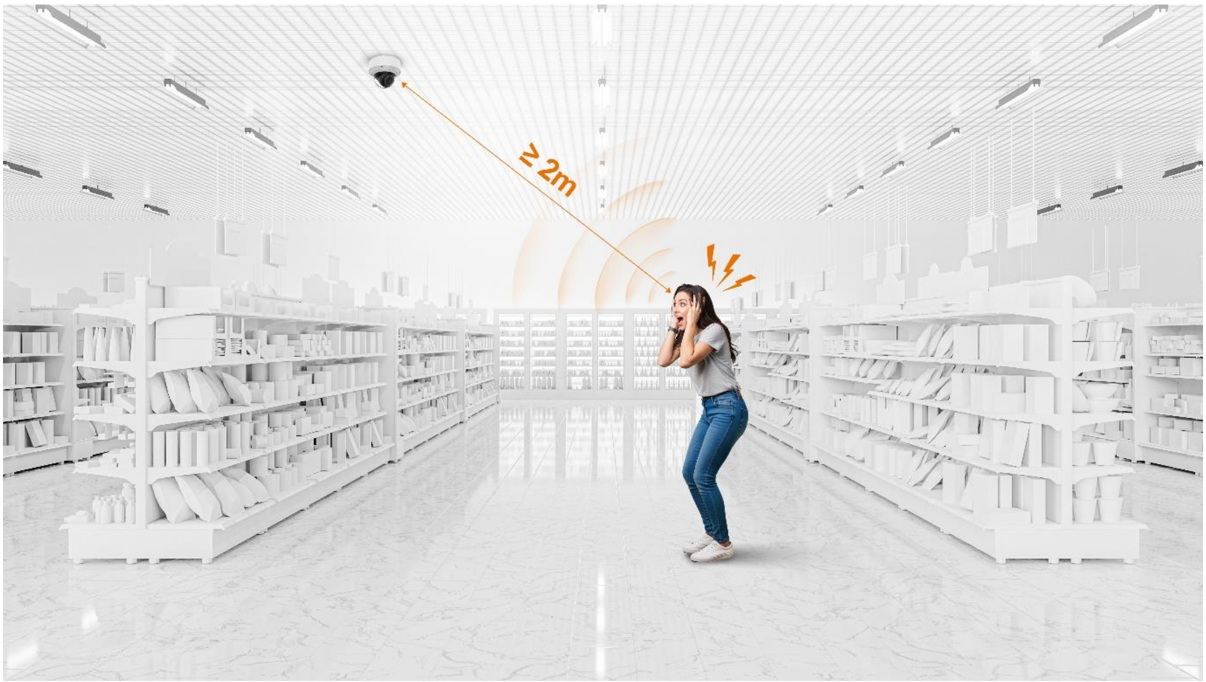


Figure 4: Effective Installation Space for Sound Classification

- **Ensuring a Clear Sound Path:** Physical obstacles like walls, glass, or thick curtains between the sound source and the product can weaken or distort the signal. To achieve maximum performance, ensure a clear, direct path for the sound.

- **Sound Direction Detection:** For accurate direction detection, a minimum space of at least 6.0m wide by 6.0m long is recommended. This minimizes the effects of sound reflections and reverberations and ensures sufficient space for signal analysis between multiple microphones.

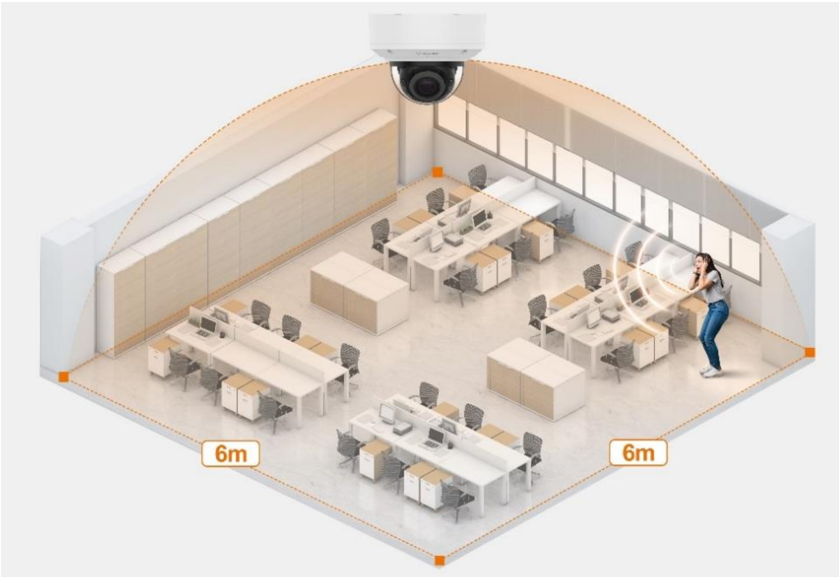


Figure 5: Ideal Installation Space for Accurate Sound Direction Detection

- **Maintaining Proper Distance and Incident Angle:** The distance and angle between the event sound source and the product are critical for detection accuracy. If the incident angle of the event sound is too large (exceeding 20◦) or the distance is too short, the detection accuracy may decrease. The table below provides recommended minimum distances based on the product's installation height.

Product Installation Height	Minimum Direction Detection Distance
2.3m	≥ 2.2m
2.5m	≥ 2.7m
2.7m	≥ 3.3m
2.9m	≥ 3.8m
3.1m	≥ 4.4m
3.3m	≥ 4.9m
3.5m	≥ 5.5m
3.8m	≥ 6.3m
4m	≥ 6.9m
5m	≥ 9.6m

Table 1: Minimum Sound Direction Detection Distance based on Product Installation Height

4.2. Environmental Analysis for Effective Sound Detection and Classification

For accurate sound detection and classification, consider the following acoustic conditions and surrounding environmental factors.

Sound Type	dB Threshold	Predicted Distance
Screaming, Tire Skidding, Aggression	>65dB	2m ~ 16m
Glass Breaking	>60dB	2m ~ 7m
Car Horns	>60dB	2m ~ 20m

Table 2: Minimum Acoustic Conditions for Sound Types

For example, a screaming sound can be accurately classified and directionally detected when its volume is above 65dB. The event sound's volume must also be significantly louder than the surrounding background noise (recommended: at least 15dB louder). For accurate measurement and classification, the background noise should ideally not exceed 60dB, which ensures a clear distinction between the event and ambient noise.

Since ambient noise can affect performance, it's good practice to analyze the following in advance:

- **Outdoor Environments**: Be aware of natural noises (wind, rain, thunder) and artificial sounds (traffic, impacts, car jerks). In unpredictable environments, a thorough analysis can help you select the optimal installation location.
- **Indoor Environments**: Sound reflections and reverberations can be significant depending on the materials (walls, ceilings, floors) and room size. Sounds that are similar to a target event, such as a balloon popping or a heavy box being dropped, can create reverberation that leads to false alarms. Installation should account for the acoustic properties of the indoor space.

4.3. Configuring Sound Classification dB Thresholds

To optimize the Sound Classification function, you can configure the dB threshold to suit your specific environment.

- In a noisy environment, set the threshold higher to reduce false alarms.
- In a quiet environment where events are subtle, set the threshold lower to avoid missing important alerts.
- After checking the average background noise dB, it is recommended to set a threshold at least 15dB higher than that average.

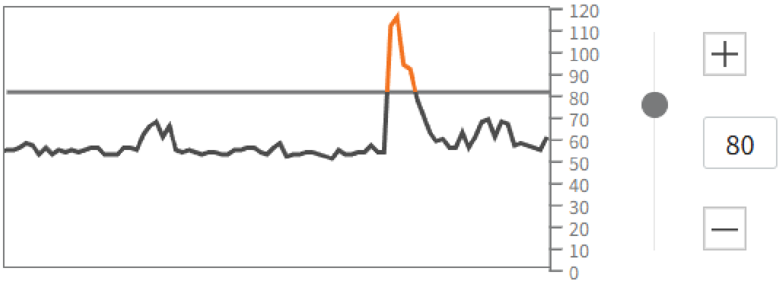


Figure 6: Example of Sound Classification dB Threshold Settings

As shown in Figure 6, the dB threshold can be adjusted intuitively using a slider or a number input field, directly impacting the real-time detection sensitivity. The graph visually represents the change in sound dB over time (black line) and the configured threshold (gray line), making it easy to see when a sound event (orange peak) exceeds the threshold.

4.4. Sound Direction Calibration and System Configuration

Hanwha Vision products provide events as audio clips, which include both the sound classification and direction detection results.

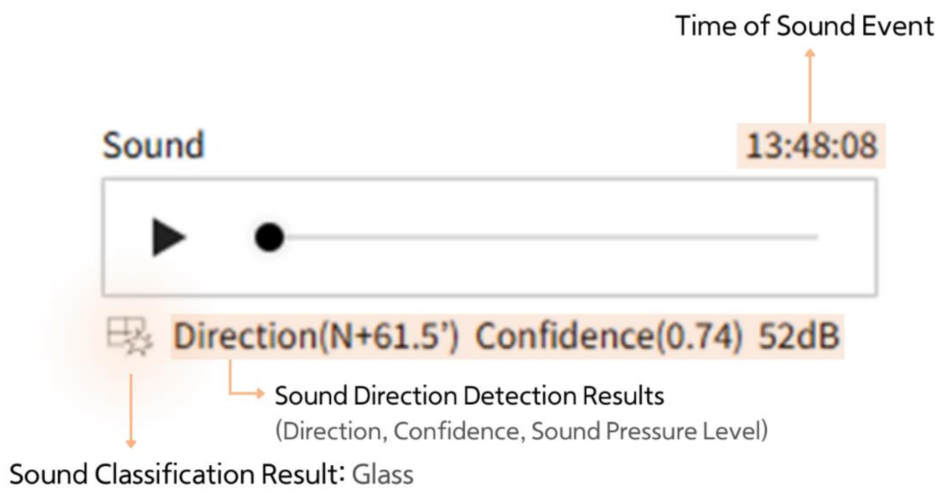


Figure 7: Example of Sound Direction Detection Results

As shown in Figure 7, the sound classification result is displayed with an intuitive icon at the bottom, along with the sound direction detection result. 'Direction (N+301.8◦)' means the sound source is located 301.8◦ clockwise from North (N).

The accompanying 'Confidence (0.74)' value indicates a 74% confidence level. This, along with the sound pressure level (52dB), helps users accurately assess the situation and respond quickly.

The system's sound direction information may deviate from true North over time or due to installation. Since accurate direction information is essential, it's important to calibrate the North reference point as needed. This can be done using one of three methods:

1. Install the product to face true North as a compass indicates.
2. In the product menu, navigate to [System] > [Product Info] > [Mounting Mode] and directly enter the angle measured clockwise from compass North to the camera's reference point.
3. Use the compass feature included in the Wisenet Installation tool for a more convenient and accurate initial setup.

4.5. Tips for Complex Acoustic Environments

- **Complex Acoustic Environments**: In an environment with multiple simultaneous sounds, the AI model may classify them as a single sound or misclassify them. This is a natural phenomenon; a comprehensive analysis of the information provided by the system will help ensure accurate situational awareness.
- **Environmental Analysis for Accurate Alarms**: The sound classification model may generate alarms for sounds that are similar to event sounds but are not in the classification categories—such as the friction of metal objects, animal calls, musical instruments, or other sudden, powerful noises. Understanding this characteristic of the model allows you to anticipate and prepare for alarms from these exceptional sounds, effectively reducing unnecessary confusion.

5. Conclusion

Hanwha Vision's AI Audio Solution goes beyond visual detection to establish a comprehensive early warning system through intelligent sound analysis—capturing potential threats through autonomous monitoring that conventional systems might miss. By integrating sound classification and direction detection, the solution delivers seamless 360° security coverage and continues to evolve to provide even more advanced capabilities.

This white paper provides practical guidelines to help users fully leverage the solution's potential and design a customized security system optimized for their specific environments. From installation to performance optimization, the document aims to minimize trial and error and offer clear guidance for achieving the best results from the very first deployment.

Hanwha Vision will continue to advance its AI audio analysis technology to meet the evolving demands of the market—empowering users to maintain a safer, more reliable, and more efficient security environment even in unpredictable situations.

Hanwha Vision Co.,Ltd
13488 Hanwha Vision R&D Center,
6 Pangyo-ro 319-gil, Bundang-gu, Seongnam-si, Gyeonggi-do
TEL 070.7247.8771-8 **FAX** 031.8018.3715
www.HanwhaVision.com