

White Paper

AI Sound Classification & Sound Direction Detection

September 2025

1. Introduction

2. AI-Based Audio Analysis Technology

- 2.1. Sound Classification
- 2.2. Sound Direction Detection

3. Installation and Environment: A Guide to Optimal Performance

- 3.1. Selecting the Optimal Installation Location
- 3.2. Environmental Analysis for Effective Sound Detection and Classification
- 3.3. Configuring Sound Classification dB Thresholds
- 3.4. Sound Direction Calibration and System Configuration
- 3.5. Tips for Complex Acoustic Environments

4. Conclusion



1. Introduction

Sound is often an overlooked yet powerful surveillance tool amid invisible threats. While conventional video surveillance systems have focused on visually capturing what is happening, today's security environment has evolved to recognize not only the types of sound events but also their exact sources. As the boundaries of public safety and asset protection expand, audio analytics technology holds the potential to contribute beyond simple assistance to crime prevention and rapid incident response.

In this context, Hanwha Vision's deep learning-based **Sound Classification** technology provides intelligent functions that accurately recognize specific audio events—such as pre-trained screams and glass breaking—triggering immediate alerts. Furthermore, **Sound Direction Detection** technology identifies the location of the audio source, delivering decisive information on not only 'what the sound is' but also 'where the sound originated.' These two technologies work synergistically to maximize integrated situational awareness capabilities, setting a new benchmark for next-generation security systems.

This white paper delves into these technologies, providing practical guidance for optimal implementation and use in diverse environments.



2. AI-Based Audio Analysis Technology

2.1. Sound Classification

Hanwha Vision's Sound Classification technology is built on a core deep learning model: the **Convolutional Neural Network (CNN)**. This technology begins by transforming abstract sound information into a visual form known as a **spectrogram**¹.

A spectrogram acts as an acoustic "fingerprint," clearly displaying the unique patterns of a specific sound. The CNN excels at automatically learning and recognizing the subtle acoustic features and patterns within these spectrogram images that are often difficult for the human ear to distinguish. This process enables the accurate identification and classification of a wide range of sound events, including screaming, glass breaking, car horns, and tire skids.

Once a sound has been detected and classified, the system automatically extracts data from the audio stream. Since the audio data is already pre-processed and sampled, the classified sound is then generated as an audio clip file, complete with metadata for easy download and review.

This technology is available on select Hanwha Vision products.

2.2. 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.

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

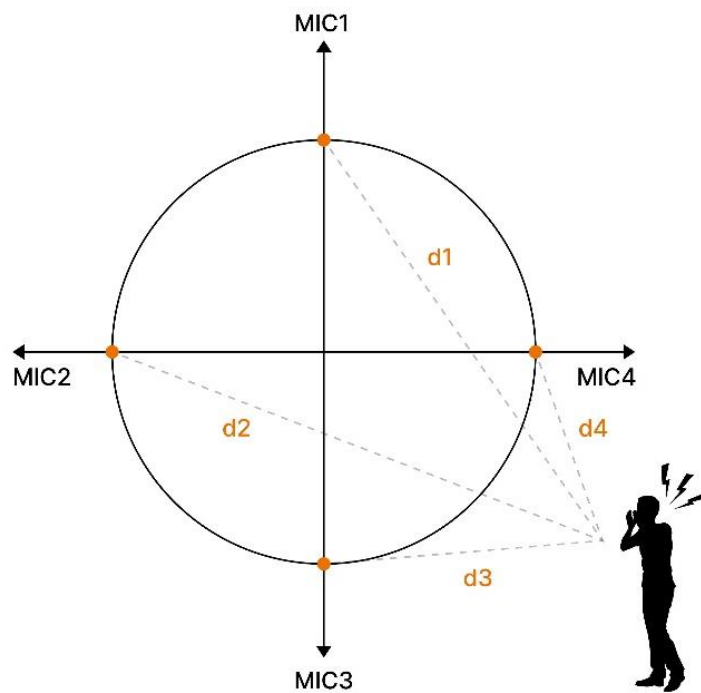


Figure 1: Sound Direction Detection Using Multiple Microphones

Figure 2 visually demonstrates the time difference (τ_{ij}) 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.

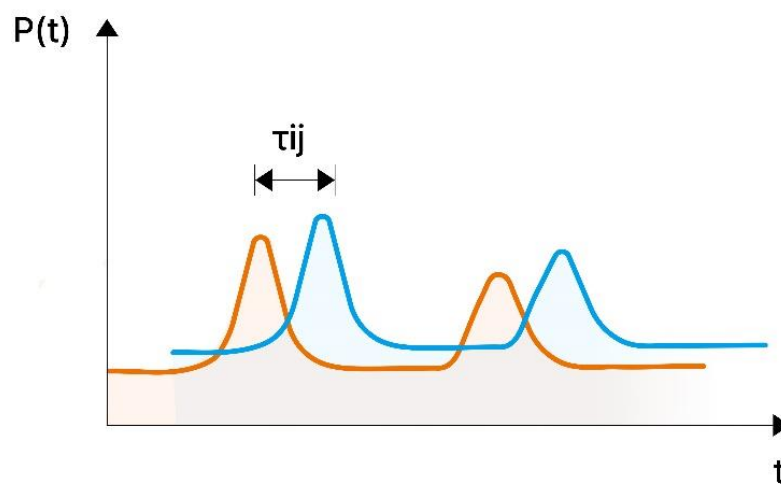


Figure 2: The Concept of Time Difference Measurement in the TDoA Algorithm

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

3. 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.

3.1. 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.

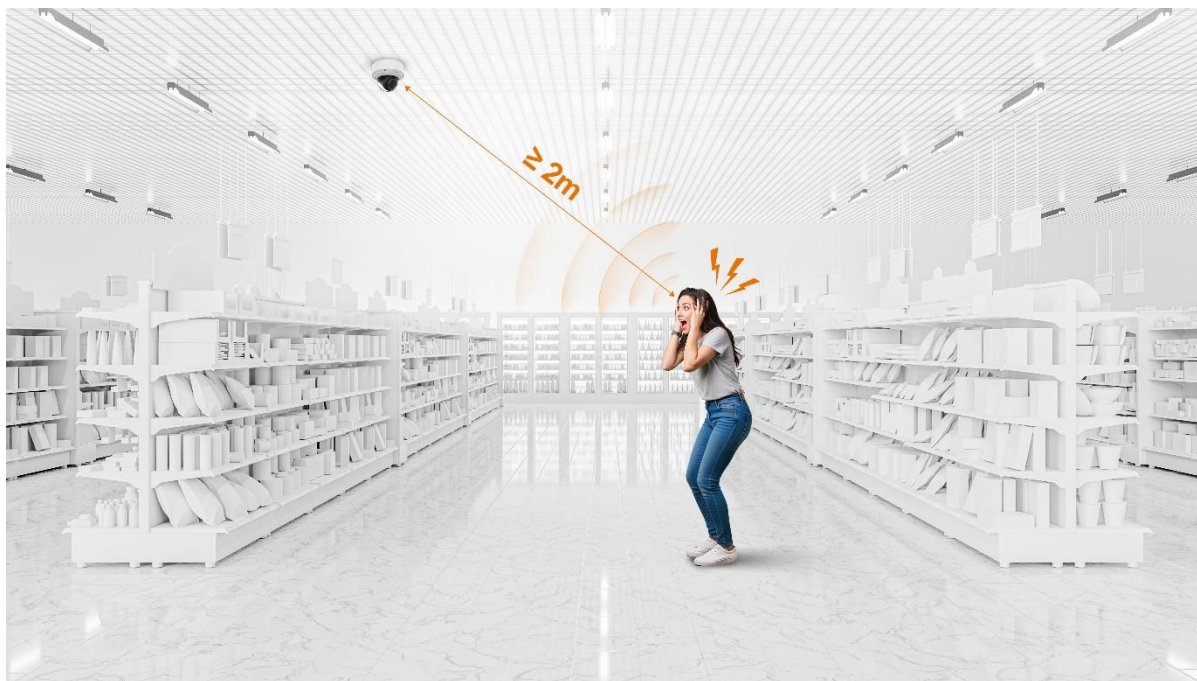


Figure 4: Effective Installation Space for Sound Classification

- **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

- **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.

3.2. 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	>70dB	2m~20m
Glass Breaking, Car Horns, Tire Skidding	>80dB	2m~16m

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 70dB. The event sound's volume must also be significantly louder than the surrounding background noise (recommended: at least 30dB 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.

3.3. 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 55dB higher than that average.

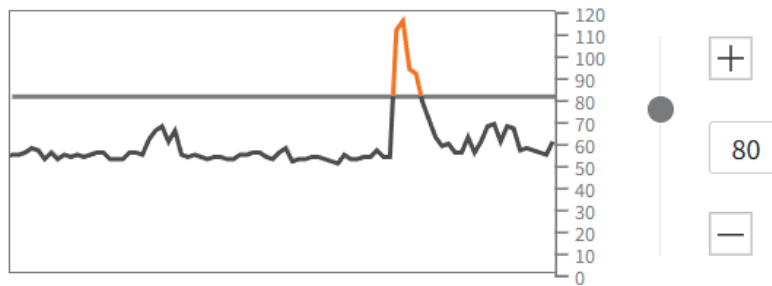


Figure 6: Example of Sound Classification dB Threshold Settings

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.

3.4. Sound Direction Calibration and System Configuration

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

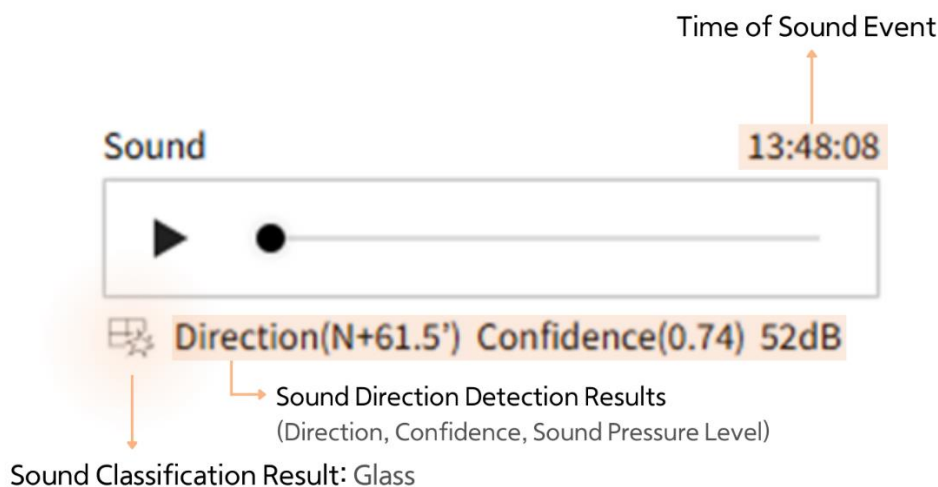



Figure 7: Example of Sound Direction Detection Results

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.

3.5. 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.



4. Conclusion

By moving beyond the limitations of visual observation, Hanwha Vision's AI Audio Solution creates a truly comprehensive early-warning system that intelligently analyzes sound.

This white paper serves as a practical guide, empowering you to implement and optimize the technology for your specific environment—from initial installation to fine-tuning for peak performance.

As security challenges evolve, Hanwha Vision remains committed to advancing its audio analysis capabilities, ensuring a more stable, efficient, and proactive security experience in any situation.

Hanwha Vision

13488 Hanwha Vision R&D Center,
6 Pangyo-ro 319-gil, Bundang-gu, Seongnam-si, Gyeonggi-do, Korea
www.HanwhaVision.com

Copyright © 2025 Hanwha Vision. All rights reserved.

