

White Paper

# AI による音源識別・方向検知技術

2025 年 9 月

## 1 序論

## 2 AI ベースのオーディオ分析技術

### 2.1. 音源分類（Sound Classification）技術

### 2.2. 音源方向検知（Sound Direction Detection）技術

## 3 最適な性能のための設置および環境ガイド

### 3.1. 最適な設置位置の選定

### 3.2. 効果的な音源検知および分類のための環境分析

### 3.3. 音源分類における dB 閾値の設定

### 3.4. 音源方向の補正およびシステム設定

### 3.5. 複雑な音響環境および特殊音源の活用ヒント

## 4 結論



## 1. 序論

目に見えない脅威の中で、「音」はしばしば見過ごされがちな、強力な監視手段の一つです。従来の映像セキュリティシステムが「何が起きているか」を視覚的に捉えることに重点を置いていたのに対し、今日のセキュリティ環境は「どんな音が、どこで発生したのか」までを認識する段階へと進化しています。公共の安全や資産保護の領域が拡大するにつれ、オーディオ分析技術は単なる補助的手段にとどまらず、犯罪の予防や迅速な状況対応を支援する重要な要素としての可能性を備えるようになりました。

Hanwha Vision の音源分類技術は、悲鳴やガラス破損音などの特定の音を事前学習によって正確に認識し、それに応じたイベントを即座に発生させるインテリジェントな機能を提供します。

さらに、音源方向検知技術は、発生した音の正確な位置を特定することで、「どんな音か」だけでなく、「どこでその音が発生したのか」という決定的な情報をユーザーに提供します。

これら二つの技術は相互に連携し、統合的な状況認識能力を最大化します。

そして次世代セキュリティシステムにおける新たな基準を提示します。

本ホワイトペーパーでは、Hanwha Vision のディープラーニングベース音源分類技術および音源方向検知技術を詳しく解説し、ユーザー環境に最適化された活用方法を紹介します。

## 2. AIベースのオーディオ分析技術

### 2.1. 音源分類（Sound Classification）技術

Hanwha Vision の音源分類技術は、ディープラーニングに基づく CNN（畳み込みニューラルネットワーク）を中心にしています。この技術は、音という抽象的な情報を、CNN が処理しやすい視覚的な形式であるスペクトログラムに変換する過程から始まります。

スペクトログラムはオーディオ波形の“指紋”のように、特定の音の種類が持つ固有のパターンを明確に示します。

CNNはこのスペクトログラム画像から、人間の耳では識別が難しい微細な音響的特徴やパターンを自動で学習・認識することに優れた性能を発揮し、悲鳴、ガラス破損音、車のクラクション、タイヤのスリップ音などの音イベントを正確に識別・分類します。

音源の検知・分類が完了すると、音源ストリーム全体からデータ抽出の工程が自動で行われます。既にオーディオデータの前処理によるサンプリングが完了しているため、分類結果とともに分離された音源はオーディオクリップ形式のファイルとして生成され、メタデータとともに提供されるので、簡単にダウンロードして利用できます。

この音源分類技術は、Hanwha Vision の一部製品に搭載されています。

### 2.2. 音源方向検知（Sound Direction Detection）技術

Hanwha Vision の音源方向検出技術は、ユーザーが指定したイベント音源が認識された際に、その音源の方向を検出してユーザーに状況を通知し、迅速な対応を支援する。この技術は、物理的に分離された複数のマイクそれぞれに到達する音源信号の時間差（TDoA：Time Difference of Arrival）を測定することで方向を推定する。

TDoA アルゴリズムは、音源発生後に各マイクに到達する時間の位相差を解析し、実際に音がマイクに到達する距離を推定する。そしてこれにより、音源発生位置の角度を計算する。図 1 に示すように、複数マイクシステムでは、複数のマイク（MIC1、MIC2、MIC3、MIC4）を円形に配置し、音の発生地点と各マイク間の距離差（ $d_1$ 、 $d_2$ 、 $d_3$ 、 $d_4$ ）を把握する。この距離差に基づいて音が各マイクに到達する時間差を計算することが、TDoA アルゴリズムの核心である。

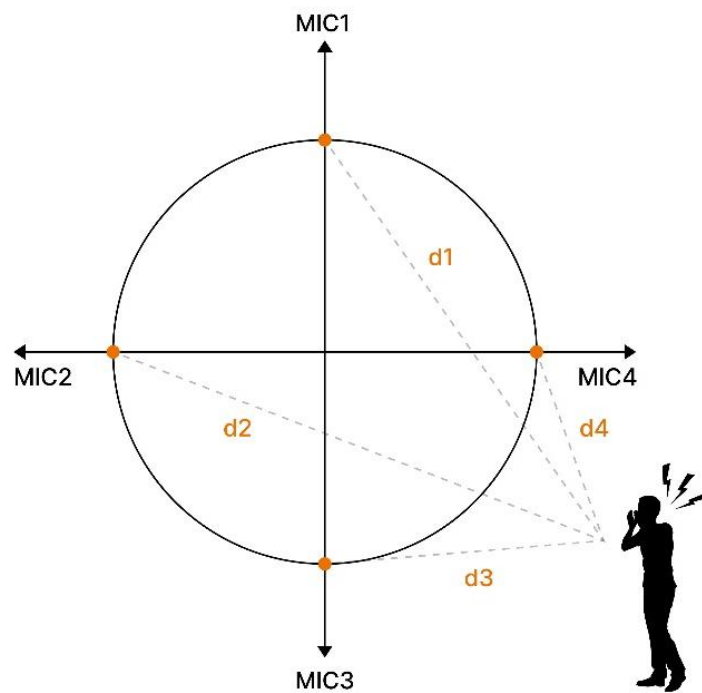


図 1：複数マイクを活用した音源方向検知

図 2 は、TDoA アルゴリズムにおいて、2 つのマイク（茶色と青の波形）に音信号が到達する時間差（ $\tau_{ij}$ ）を視覚的に示している。このように音波形の到達時間差を精密に測定することで、システムは音源発生地点の正確な方向を逆算することができる。

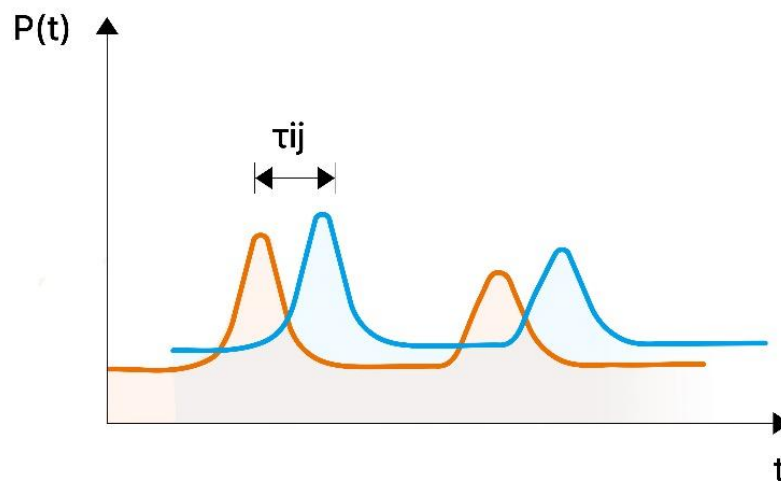


図 2 TDoA アルゴリズムにおける時間差測定概念

音源方向検出のプロセスは大きく 4 つの段階に分けられる。

1. **信号収集段階**：複数のマイクを通じて音源信号を同時に収集する。
2. **信号処理段階**：収集した音源信号を信号処理アルゴリズムを用いて解析する。
3. **方向推定段階**：処理された信号を基に音の発生方向を推定する。
4. **結果出力段階**：最終的に検出された方向を方位角として表示・出力する。

この音源方向検出技術は、複数マイクに対応したオーディオビーコン（AudioBeacon）や、一部の Wisenet 9 SoC を搭載したカメラなど、Hanwha Vision の特定製品で提供されている。



図 3 : オーディオビーコン SPS-A100M（左）と Wisenet 9 搭載カメラのマイク位置

### 3. 最適な性能のための設置および環境ガイド

Hanwha Vision の AI オーディオソリューションの性能は、設置環境の特性と密接に関連しています。

次の事項を積極的に考慮することで、システムの潜在能力を最大限に活用し、安定した性能を確保することができます。

#### 3.1. 最適な設置位置の選定

安定した音源分類および方向検知性能を保証するための最適な設置条件は次のとおりです。

- **音源分類**：製品と音の発生地点との距離が少なくとも 2m 以上ある場合に、最も安定して動作します。

この距離は音源（例：人の身長の高さ）を基準としています。

もし距離が 2m 以内と近すぎる場合、拍手のような音が過度に大きく認識され、誤検知が発生する可能性があります。

室内の天井設置は音の反響を最小化し、広い空間の音を均一に感知できるため、音源分類機能に最も適した設置方法です。

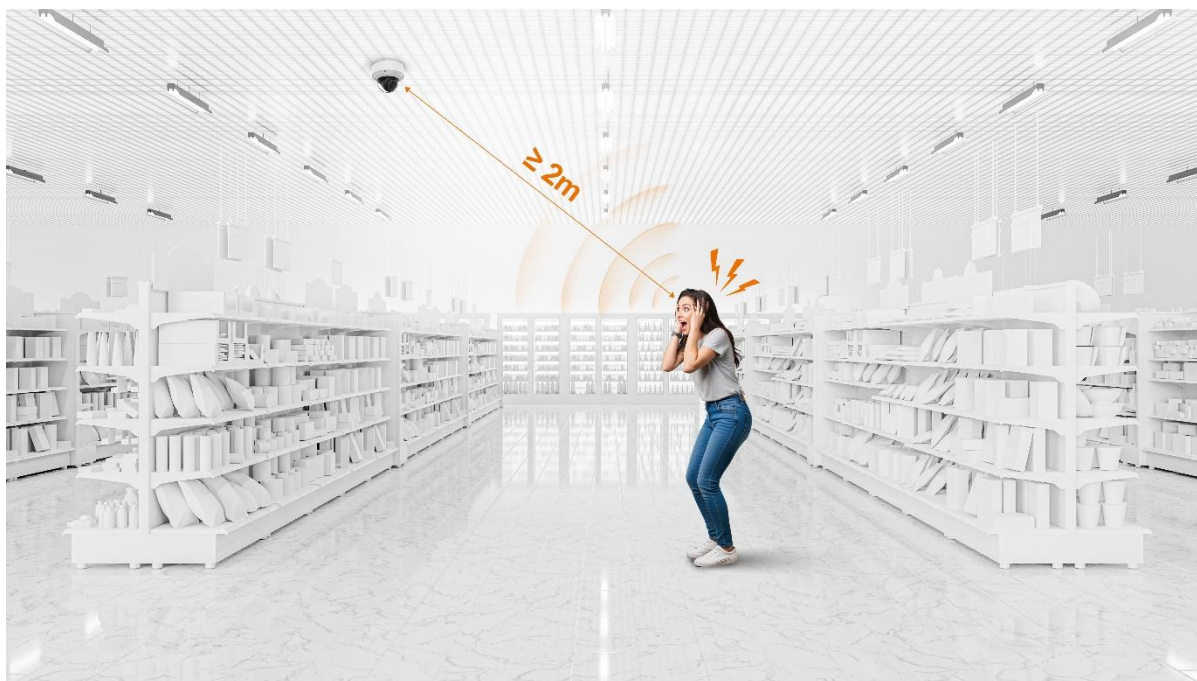


図 4：効果的な音源分類のための設置空間

- **音源方向検知**：正確な方向検知性能を得るためには、設置場所の最小空間サイズが幅 6.0m 以上 × 奥行き 6.0m 以上であることを推奨します。

これは、音の反響や残響の影響を最小限に抑え、複数マイク間での信号解析に必要な十分な空間を確保するため



図 5：正確な音源方向検出のための設置空間

**距離および入射角の遵守：** イベント音源の発生地点と製品との距離および角度は、正確な検知性能に大きく影響する重要な要素です。

イベント音源の入射角が20度以上と大きすぎたり、製品との距離が近すぎる場合、方向検知の精度が低下する可能性があります。

以下の表は、製品の設置高さに応じた推奨最小距離を示しています。この距離を遵守することで、より信頼性の高い方向検知が可能となります。

| 製品設置高さ | 方向検知可能な最小距離 |
|--------|-------------|
| 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      |



表 1：製品設置高さに応じた音源方向検知の最小距離

- **直接的な音源経路の確保**：音源の発生地点と製品の間に、壁、ガラス、厚手のカーテンなどの物理的障害物が存在する場合、音源信号が減衰したり歪む可能性があります。システムの最大性能を引き出すためには、直接的な音源経路を確保することが重要です。

### 3.2. 効果的な音源検知および分類のための環境分析

正確な音源検知および分類のためには、次の音圧条件および周囲環境を考慮する必要があります。

| 音源の種類                   | 限界 dB | 限界予測距離 |
|-------------------------|-------|--------|
| 悲鳴                      | >70dB | 2m~20m |
| ガラス割れ、車のクラクション、タイヤのスリップ | >80dB | 2m~16m |

表 2：音源の種類ごとの最小音圧条件

例えば、悲鳴は 70dB 以上であれば、正確な分類および方向検知が可能です。また、イベント音源の dB レベルは周囲の背景雑音より一定以上大きいことが必要です（推奨：30dB 以上）。

さらに、正確な測定および分類を行うためには、周囲の背景雑音が 60dB を超えない場所に設置することが重要です。これは、イベント音源と雑音を明確に区別するためです。

周囲雑音は設置環境によって性能に影響を与える可能性があるため、以下の点を事前に分析するといでしょう。

- **屋外環境**：風、雨、雷などの自然音や、交通、物体の衝撃、車の反動などの人工音が存在する場合がある。予測不可能な雑音が多い環境では、追加の環境分析を通じて最適な設置位置を選定することが重要である。
- **室内環境**：周囲の素材（壁、天井、床）や部屋の大きさによって、音の反響や残響が大きく発生することがあります。  
風船の破裂音や重い箱を落とす音など、特定の音源と類似した特徴を持つ音は、残響によってエネルギーが蓄積され、誤検知を引き起こす可能性があるため、室内空間の音響特性を考慮した設置が必要です。

### 3.3. 音源分類におけるdB閾値の設定

音源分類機能の性能を最適化するため、ユーザーの環境に応じて dB しきい値（Threshold）を設定することができる。

- 騒音が大きい環境では、しきい値をデフォルトより高く設定することで、不必要なアラーム（誤分類）を減らすことができます。
- イベント音源が小さい環境では、しきい値をデフォルトより低く設定することで、重要なアラームの見逃し（未分類）を減らすことができます
- 設置環境の背景雑音 dB を確認した上で、平均値より最低でも 55dB 以上のしきい値を設定することを推奨します。

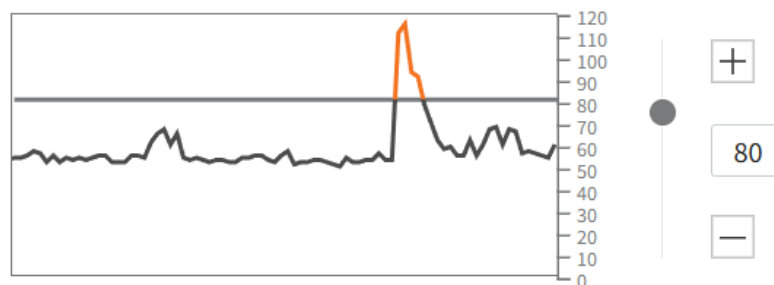


図 6：音源分類 dB しきい値設定の例

図 6 のように、ユーザー設定では dB しきい値をスライダーや数値入力フィールドで直感的に調整することができ、この設定はリアルタイムで音源検知の感度に直接影響します。

グラフは、時間経過に伴う音源の dB 変化（黒線）と設定されたしきい値（Threshold、灰色線）を視覚的に示しており、音源レベルがしきい値を超えたとき（オレンジ色のピーク）にイベントが発生することを、容易に理解できるようになっています。

### 3.4. 音源方向の補正およびシステム設定

Hanwha Vision の製品は、イベントをオーディオクリップとして提供し、そこには音源分類および方向検知の結果も併せて表示されます。

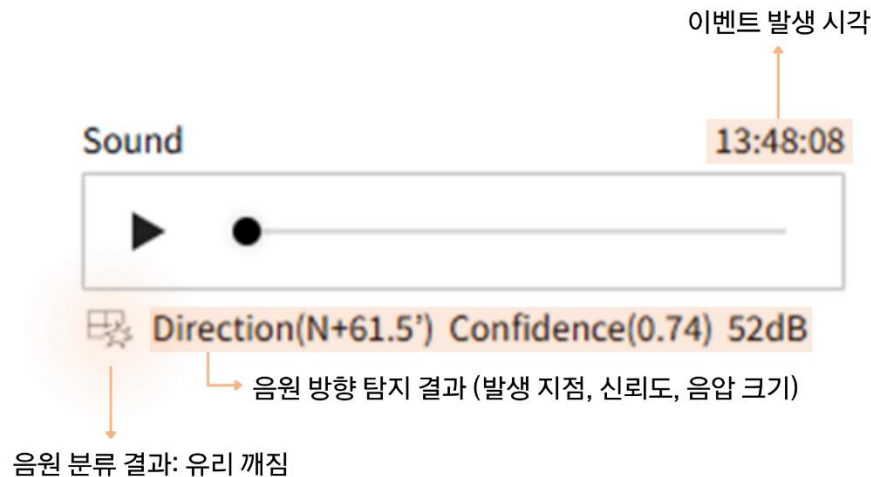


図 7：音源方向検知結果の例

図 7 のように、下部には音源分類結果を直感的に示すアイコンとともに、音源方向検知結果が表示されます。

ここで「Direction(N+301.8°)」は、音源発生地点の方向が北（N）から 301.8 度の位置にあることを意味します。

併せて提供される「Confidence(0.74)」は、検知結果の信頼度が 74%であることを示し、音圧レベル（52dB）とともに、ユーザーが状況を正確に把握し、迅速に対応できるよう支援します。

システムが提供する音源方向情報は、初期設置環境や時間経過により、実際の北方向と誤差が生じる可能性があります。

正確な方向情報はシステムの活用度を高めるために不可欠であるため、必要に応じて基準となる N 値（北）を補正することが重要です。そのために、次の 3 つの方法を利用できます。

1. 製品をコンパスの N 極が指す真北方向に設置する。
2. 製品内の [System] > [製品情報] > [マウンティングモード] メニューで、「北と機器基準点の間の角度」フィールドに、コンパスの N 極を基準としてカメラ基準点までの時計回りの角度を直接入力する。
3. Wisenet Installation ツールに含まれるコンパス機能を活用する。この機能により、より便利かつ正確に初期方向の設定を行うことができる。

### 3.5. 複雑な音響環境および特殊音源の活用ヒント

- **複合音響環境での活用：**2 種類以上の音源が同時に発生する環境では、AI モデルが 1 つの音源として分類したり誤分類する場合があります。



これは自然な現象であるため、システムが提供する情報を総合的に判断することで、正確な状況把握に役立ちます。

- **正確なアラームのための環境分析：**音源分類モデルは、金属がぶつかる摩擦音、動物の鳴き声、楽器の音などイベント音源と類似した音や、分類カテゴリに含まれない突発的かつ強い音に対してアラームを発生させる場合があります。

このモデルの特性を理解することで、例外的な音でも発生し得るアラームを事前に予測・対策し、不必要な混乱を効果的に減らすことができます。

## 結論

Hanwha Vision の AI オーディオソリューションは、単に「見る」ことを超えて、高度な音声分析を通じて、従来のシステムでは検知できなかった脅威まで捕捉し、隙間のない早期検知体制を構築します。

本ホワイトペーパーは、ソリューションの潜在能力を最大限に活用できる実用的な指針を提供し、ユーザーがそれぞれの環境に最適化されたカスタマイズ可能なセキュリティシステムを構築できるよう支援します。システムの設置から性能最適化に至るまで、すべての段階でユーザーが不必要な試行錯誤を減らし、即座に最良の結果を得られるよう、明確なガイドラインを提示しています。

Hanwha Vision は今後も、市場の要求に応え、ユーザーが予測不可能な状況下でもより安定的かつ効率的なセキュリティ環境を体験できるよう、AI オーディオ分析技術の発展に努めてまいります。

Hanwha Vision  
13488 Hanwha Vision R&D Center,  
6 Pangyo-ro 319-gil, Bundang-gu, Seongnam-si, Gyeonggi-do, Korea  
[www.HanwhaVision.com](http://www.HanwhaVision.com)

Copyright © 2025 Hanwha Vision. All rights reserved.

