

백서

# 보안 및 안전을 위한 오디오 분석 애플리케이션

2월 2024

## 요약

보안 및 안전을 위한 오디오 분석 애플리케이션은 사운드 패턴을 감지하고 라이브 오디오에서 예기치 않은 사운드를 강조할 수 있습니다. 예를 들어, 분석 애플리케이션은 폭언 상황을 식별하여 상황의 확대와 폭행을 방지하고, 유리 파손을 감지하여 침입을 방지하며, 곤란한 상황에 놓인 사람에 대한 조기 경고를 제공할 수 있습니다.

비디오 및 오디오(카메라와 마이크)와 같은 다양한 유형의 센서를 사용하면 감지 결과에 대한 신뢰도가 높아지고 실행 가능한 인사이트를 더 많이 얻을 수 있습니다.

AXIS Audio Analytics는 Axis 장치 소프트웨어에 통합된 에지 기반 분석 애플리케이션을 위한 소프트웨어 플랫폼입니다. 이러한 분석 애플리케이션은 원본 오디오 스트림을 저장하지 않고 사운드를 캡처하고 감지합니다. 이는 개인정보를 보호하는 방법이며, 분석 애플리케이션이 에지 기반이고 오디오 메타데이터를 제공하기 때문에 작동합니다.

# 목차

<b>1</b>	<b>서론</b>	4
<b>2</b>	<b>기술</b>	4
2.1	사운드 이벤트 감지	4
2.2	오디오 캡처 및 처리	4
2.3	에지 기반 또는 서버 기반 분석	5
2.4	메타데이터	6
<b>3</b>	<b>AXIS Audio Analytics</b>	7
3.1	개인정보 보호	7
3.2	사용 사례	8
<b>4</b>	<b>법적 제한</b>	12
<b>5</b>	<b>면책 고지</b>	12

# 1 서론

보안 및 안전을 위한 오디오 분석 애플리케이션은 사운드 패턴을 감지하고 라이브 오디오에서 예기치 않은 사운드를 강조할 수 있습니다. 예를 들어, 분석 애플리케이션은 공격적 행위와 관련된 사운드를 식별하거나 유리 파손을 감지하거나 곤란한 상황에 놓인 사람에 대한 조기 경고를 제공할 수 있습니다.

영상 감시와 결합된 오디오 분석 애플리케이션은 운영자에게 진행 중인 잠재적 사고를 경고하고 관련 카메라 화면으로 안내할 수 있습니다. 이를 통해 조기 발견, 신속한 개입, 그리고 많은 경우 추가 확대를 방지할 수 있습니다.

이 백서에서는 보안과 안전을 위해 오디오 분석 애플리케이션을 어떻게 활용할 수 있는지 설명합니다. 카메라나 마이크에서 직접 AI 기반 사운드 분류와 같은 실시간 에지 분석에 중점을 두고 오디오를 캡처하고 처리하는 기술에 대해 설명합니다. 추가적으로, 에지에서 오디오 분석 애플리케이션을 통해 오디오 메타데이터를 사용하여 개인정보 보호를 유지하기 위한 여러 가지 옵션을 어떻게 가능하게 하는지 보여줍니다.

이 백서는 법률 자문을 제공하지 않습니다. 감시 시스템을 설치하기 전에, 해당 지역과 사용 사례에 적용되는 법률 및 규정을 조사해야 합니다. 시스템이 해당 지역의 법률, 규정 및 권장 사항을 준수하는지 확인하는 것은 시스템 소유자의 책임입니다.

## 2 기술

### 2.1 사운드 이벤트 감지

사운드 이벤트는 비명 소리나 유리 깨지는 소리와 같이 일반적으로 사람이 고유한 개념으로 식별하는 오디오 세그먼트를 말합니다. 이러한 종류의 개념적 사운드는 비디오 분석에서 객체 클래스를 감지하고 라벨을 지정하는 것과 유사한 방식으로 감지하고 라벨을 지정할 수 있습니다.

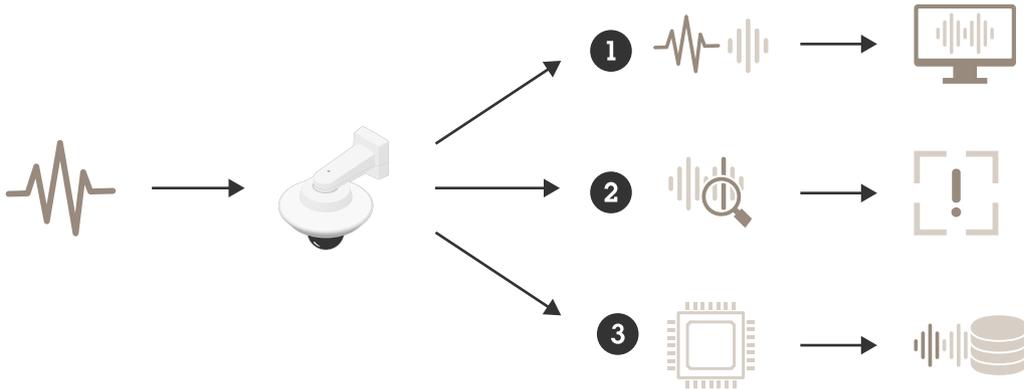
사운드 패턴을 인식하도록 훈련된 분석 애플리케이션은 일반적으로 시간 경과에 따라 데시벨 레벨부터 다양한 주파수의 에너지에 이르기까지 다양한 특성의 조합을 청취합니다. 특정 사운드 패턴이 감지되면, 시스템은 시각적 경고를 통해 또는 알람을 트리거하여 직원에게 자동 알림을 보낼 수 있습니다.

AI 기반 알고리즘을 사용하면, 대량의 데이터로 알고리즘을 학습시킬 수 있습니다. 예를 들어 알고리즘은 수천 개의 사운드로 학습한 후 사람의 비명을 신뢰성 있게 감지할 수 있습니다.

### 2.2 오디오 캡처 및 처리

오디오 분석은 캡처된 오디오 데이터를 사용하고 관련 사운드 특성을 분석하여 비 오디오 출력을 생성합니다. 오디오 캡처는 기본적으로 오디오를 디지털화하여 소프트웨어에서 사용할 수 있도록 하는 것을 의미합니다. 이것은 마이크를 사용하여 공기 중의 사운드 진동을 수집하고, 이러한 아날로그 신호를 디지털 신호로 변환하여 처리 장치에 전달하는 방식으로 이루어집니다. 캡처한 오디오는 플래시 메모리나 하드 드라이브와 같은 영구 매체에 저장되지 않으면 녹음되지 않습니다. Axis 카메라에서는 오디오 스트리밍 및 녹음이 기본적으로 꺼져 있습니다.

초기 오디오 캡처 후, 캡처된 정보는 다음 처리 단계를 위해 준비됩니다. 이와 병행하여 또는 별도로 다른 준비를 할 수 있습니다.



### 오디오 준비

#### 1 변환

#### 2 실시간 에지 분석

3 스트리밍 또는 저장을 위한 처리 및 인코딩 — Axis 카메라를 사용하는 경우, 오디오 스트리밍을 직접 켜지 않으면 오디오가 스트리밍되거나 저장되지 않습니다.

- **변환.** 사운드가 추상화 되어 사운드 스펙트럼을 보여주는 그래프와 같은 시각적 정보로 변환됩니다. 이 과정은 되돌릴 수 없습니다. 스펙트럼 그래프에서 원래의 사운드를 가져올 수 없습니다.

- **실시간 에지 분석.**

사운드가 에지에서 처리되는 경우 **사운드 분류기**를 사용할 수 있습니다. 이를 통해 사운드의 특성을 설명하는 메타데이터를 얻을 수 있습니다. 원래의 사운드를 메타데이터에서 다시 만들 수 없습니다.

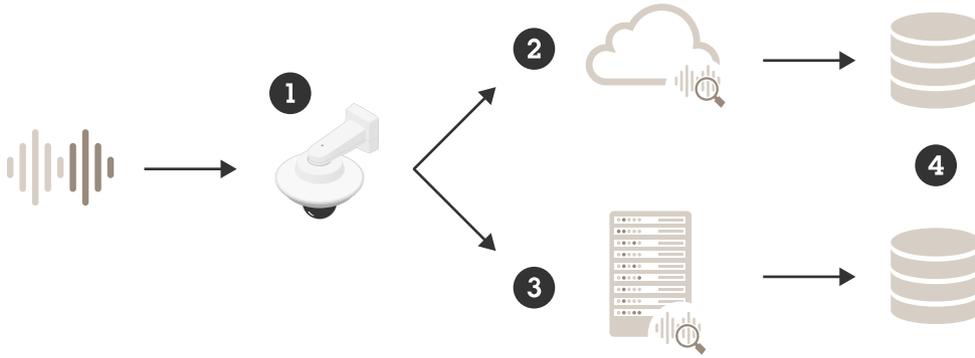
**사운드 감지기**를 사용하여 패턴, 레벨 또는 주파수를 인식하고 상태 정보를 제공할 수 있습니다. 다시 말하지만 원래의 사운드는 복원할 수 없습니다.

- **처리 및 인코딩.** 원래의 오디오가 사용되는 경우(변환 또는 분석되지 않음), 의도된 사용 사례에 맞게 오디오 데이터를 준비하기 위해 일반적으로 일부 처리 및 인코딩이 수행됩니다. 이러한 사용 사례에는 에지에 오디오 데이터 저장하기, 추가 처리(서버 또는 클라우드에서)를 위해 외부 클라이언트로 스트리밍하기 또는 외부 저장이 포함될 수 있습니다. Axis 카메라를 사용하려면 먼저 개인정보 보호를 위해 기본적으로 꺼져 있는 오디오 스트리밍을 직접 켜야 합니다(오디오 개인정보 제어).

## 2.3 에지 기반 또는 서버 기반 분석

시스템에서 분석 엔진의 위치는 여러 가지 이유로 중요합니다. 특히 개인정보 보호 문제를 관리하고 개인정보에 관한 규정을 준수하려면, 소프트웨어 알고리즘이 오디오 데이터를 분석하는 위치가 중요합니다. 오디오 데이터를 네트워크를 통해 전송할 수 없는 상황이 있기 때문에, 캡처된(저장되지 않

은) 오디오 데이터를 로컬에서 분석할 수 있어야 합니다. 에지에서 실행할 수 없는 컴퓨팅 집약적인 알고리즘이 필요한 경우, 디지털 오디오 데이터를 클라우드 또는 서버로 전송해야 할 수 있습니다.



분석을 실행할 수 있는 위치

1 에지

2 클라우드

3 서버

4 저장소 — 오디오 스트리밍이 활성화된 경우에만 원본 오디오를 저장할 수 있습니다.

- **에지 분석.** 에지에서 분석이 실행되면, 오디오 스트림을 카메라에서 내보낼 필요가 없습니다. 수행된 분석의 출력, 즉 메타데이터 또는 트리거만 카메라에서 내보냅니다. AXIS Audio Analytics 내의 분석은 에지 기반입니다.
- **서버 분석.** 분석이 서버에서 실행될 때는, 카메라에서 오디오 데이터를 전송해야 합니다. 카메라 내에서 사전 처리된 경우, 추상화된 메타데이터 또는 비개인화된 메타데이터로 제한될 수 있습니다. 서버는 일반적으로 폐쇄된 시스템의 일부이므로(시스템 소유자가 제어), 전송된 오디오의 개인정보 보호 문제를 관리할 수 있습니다. 그럼에도 불구하고 해당 규칙 및 규정을 준수해야 합니다.
- **클라우드 분석.** 클라우드 맥락에서 오디오 데이터는 서버로 전송될 수도 있습니다. 서버 분석의 경우와 마찬가지로 오디오 정보를 메타데이터로 사전 처리할 수 있습니다. 클라우드 사용은 종종 분산되므로, 개인정보 보호 문제를 해결하고 규정을 준수하도록 하는 것이 훨씬 더 중요합니다.

## 2.4 메타데이터

오디오 분석 애플리케이션은 오디오 레벨 데이터의 지속적인 메타데이터 스트림을 생성합니다. 또한 분석 애플리케이션은 분류 기능이 감지한 이벤트 및 적응형 오디오 감지를 통해 감지된 이벤트를 기반으로 메타데이터를 생성합니다. 에지에서 실행되는 분석 애플리케이션은 카메라 내의 오디오 정보를 분석합니다. 분석 애플리케이션은 어디에서도 실제 오디오 스트림을 전송할 필요가 없으며, 장면에서 일어나는 상황에 대한 인사이트를 제공하는 메타데이터만 전송할 수 있습니다.

메타데이터 스트림을 사용하면 오디오 이벤트와 비디오 이벤트를 정렬하여 함께 볼 수 있는 대시보드 등에 표시되는 오디오 엔벨로프를 시각적으로 검사할 수 있습니다. 이렇게 하면 메타데이터를 통해 특정 이벤트와 예기치 않은 사운드를 효율적이고 빠르게 검색할 수 있습니다. 이를 통해 조사관은 방대한 양의 영상을 검색하는 데 드는 시간을 절약할 수 있습니다.

## 3 AXIS Audio Analytics

AXIS Audio Analytics는 새롭고 업그레이드된 분석 애플리케이션 및 기능을 지속적으로 소개하는 소프트웨어 플랫폼입니다. 이 분석 애플리케이션은 장치 소프트웨어 AXIS OS에 통합되어 있으며, 카메라 및 기타 Axis 장치에 무료로 탑재되어 있습니다.

플랫폼 내의 분석 애플리케이션은 에지 기반이며, 카메라에서 알고리즘이 직접 실행됩니다. 이는 최적의 확장성, 낮은 데이터 트래픽, 개인정보 보호를 제공합니다. 분석 애플리케이션의 출력(메타데이터 또는 트리거)만 저장됩니다. 카메라에서 오디오가 녹음되거나 스트리밍되지 않으며, 메타데이터에서 원본 사운드를 재생성할 수 없습니다.

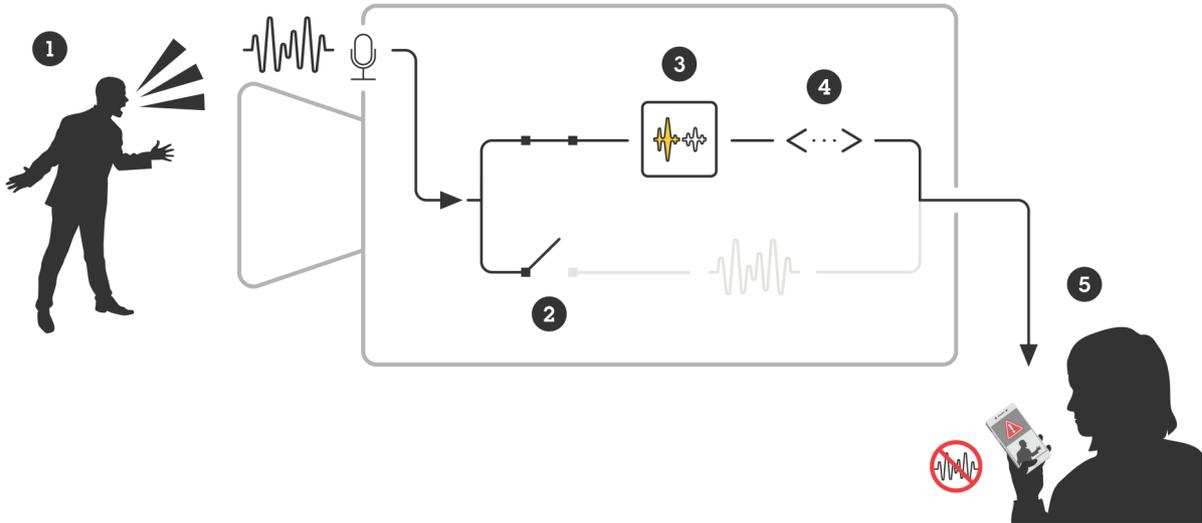
- **오디오 분류(비명 및 고함 소리)** 비명 및 고함 소리 등의 특정 사운드를 감지하고 분석하는 AI 기반 사운드 분류 기능입니다. 유리 깨지는 소리 등 더 많은 오디오 분류가 향후 추가될 예정입니다. 분류 분석의 결과는 사운드의 특성을 설명하는 메타데이터입니다.
- **적응형 오디오 감지** 오디오 레벨에 갑작스러운 변화가 있을 때 이벤트를 생성하는 사운드 감지 기능입니다. 모든 종류의 사운드 피크를 감지하며, 소음 레벨이 달라져도 주변 소음에 적응할 수 있다는 장점이 있습니다.

### 3.1 개인정보 보호

일반적으로 오디오 분석 애플리케이션은 입력되는 오디오를 녹음하거나 카메라에서 전송하지 않습니다. 추가 조사를 위한 대시보드나 운영자에게 경고하기 위한 영상 관리 소프트웨어와 같이 수신 시스템에서 특정 이벤트, 패턴 또는 사운드 레벨을 검색할 수 있도록 사운드를 처리할 뿐입니다. 오디오 데이터를 재구성할 수 없으며 사적인 대화를 녹음할 수 없습니다. 이는 이러한 분석 애플리케이션이 에지 기반이며 오디오 메타데이터를 제공하기 때문입니다.

AXIS Audio Analytics는 오디오를 녹음하거나 스트리밍하지 않고 메타데이터만 전송합니다. 개인정보 보호를 위해, Axis 장치에서는 모든 오디오 스트리밍도 기본적으로 꺼져 있으므로(오디오 개인정보 제어), 오디오를 스트리밍하거나 녹음하거나 재생성할 수 없습니다. 필요한 경우 오디오 스트리밍을 켤 수 있지만, 오디오 스트리밍이 꺼져 있어도 관련 사항이 발생하면 알림을 받게 됩니다.

아래 그림은 사운드를 포착하고 메타데이터를 사용하여 경보를 생성할 때, AXIS Audio Analytics가 오디오 개인정보 보호 제어와 함께 어떻게 작동하는지 간략하게 보여줍니다.



비명 소리(1)부터 경고음(5)에 이르기까지, Axis 카메라의 오디오 개인정보 보호 제어 및 AXIS Audio Analytics.

- 1 마이크가 사운드를 포착합니다.
- 2 기본적으로 오디오 스트리밍은 꺼져 있습니다.
- 3 AXIS Audio Analytics의 오디오 분류 기능은 입력되는 오디오에서 비명 소리나 고함을 감지합니다.
- 4 오디오 분류 기능이 이벤트 알림을 포함한 메타데이터를 생성합니다.
- 5 이해 관계자는 이벤트 알림 및 메타데이터를 기반으로 알림을 받습니다. 비디오 스트림을 확인하여 어떤 종류의 사운드였는지 확인할 수 있습니다. 오디오 스트림은 제공되지 않습니다.

## 3.2 사용 사례

AI 기반 분석 애플리케이션은 관련 없는 소음을 걸러낼 가능성이 높지만, 배경 소음이 많은 경우 허위 경보를 제공할 수 있습니다. 유리창에 부딪치는 빗소리, 천둥, 사이렌, 음악, 사람들이 시골벽적으로 떠드는 장면 등이 허위 경보를 유발할 수 있습니다. 따라서 일반적인 사용 사례에는 은행이나 안내 창구와 같은 조용한 공간뿐만 아니라 매장, 식당, 계단 또는 사무실과 같이 업무 시간 이후의 다양한 유형의 실내 공간도 포함됩니다.

### 3.2.1 경보를 동반한 사운드 감지

은행이나 안내 창구에서는 사운드 분류 분석 애플리케이션을 통해 해당 구역을 모니터링하고 공격적 행위나 유리 파손과 관련된 사운드를 감지할 수 있습니다. 사운드가 감지되면, 시스템의 이벤트

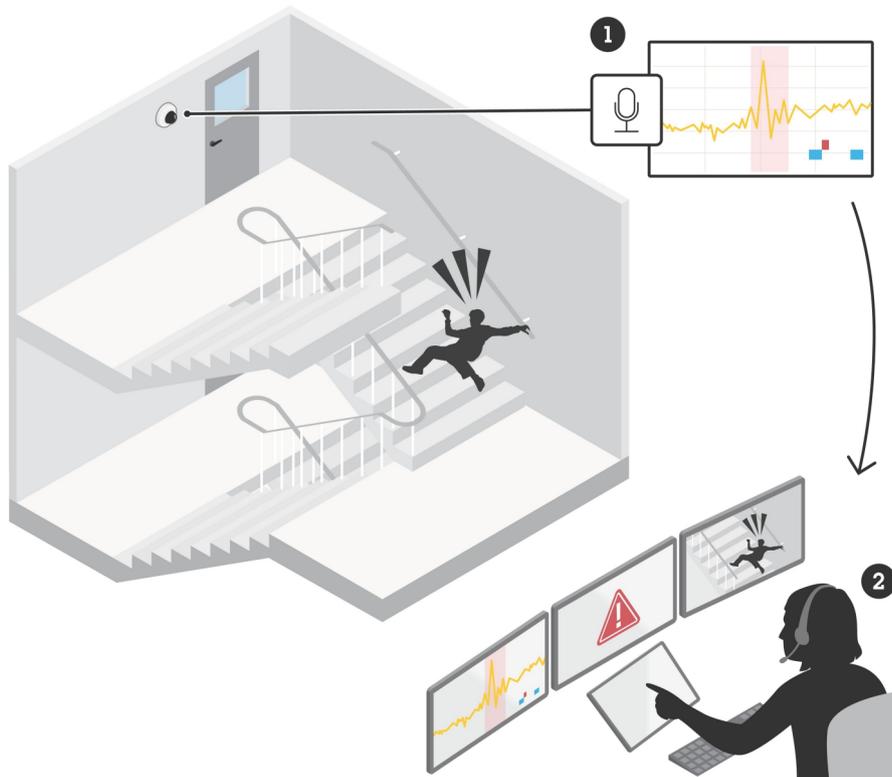
데이터와 오디오 메타데이터가 시각적 경고를 통해 또는 알람을 트리거하여 통해 직원에게 자동 알림을 보냅니다. 이는 빠른 대응과 개입을 가능하게 하는 조기 경고를 제공합니다.



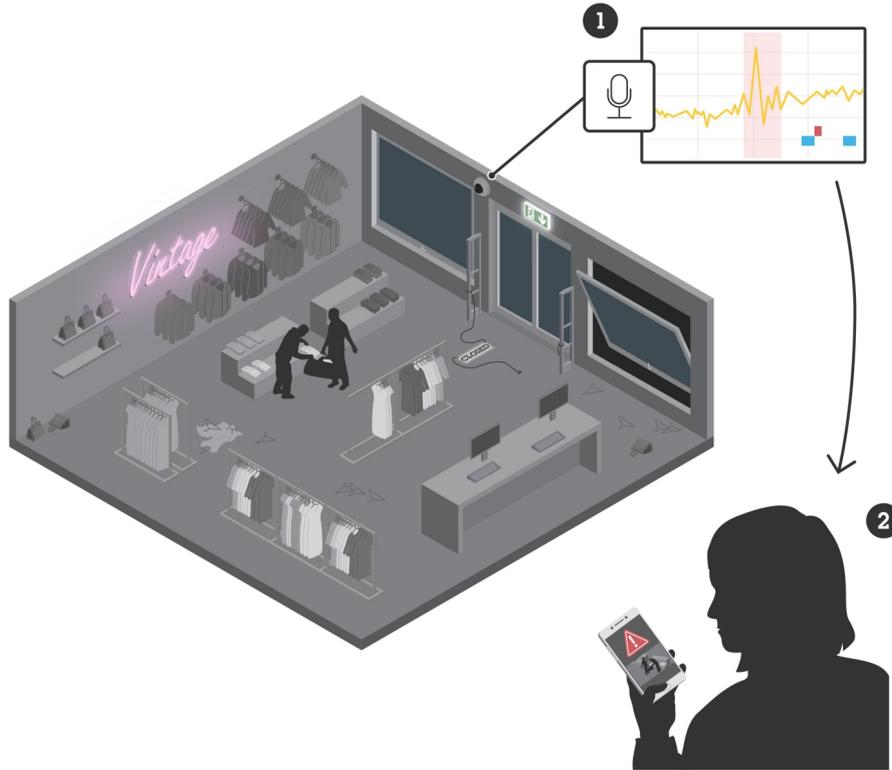
- 1 사운드 분류 분석 애플리케이션이 탑재된 카메라가 안내 창구에서 비명을 지르거나 고향을 지르는 것을 감지합니다.
- 2 운영자는 경보를 수신한 후, 추가 조치를 취하기 전에 비디오 스트림을 확인하여 어떤 종류의 사운드였는지 확인할 수 있습니다.

적응형 오디오 감지 분석 애플리케이션을 사용하여 업무 시간 외에 예기치 않게 발생하는 사운드를 감지할 수 있습니다. 이 분석 애플리케이션은 주변 사운드를 분석하여 음성, 유리창이 깨지는

소리 또는 기타 갑작스럽고 짧은 소음을 감지하면 응답합니다. 이벤트가 감지되면, 그에 따라 분석 애플리케이션이 메타데이터를 전달하여 운영자에게 알립니다.



- 1 사운드 분류 분석 애플리케이션이 탑재된 카메라가 업무 시간 이후 예상치 못한 사운드를 감지합니다.
- 2 운영자는 경보를 수신한 후, 추가 조치를 취하기 전에 비디오 스트림을 확인하여 어떤 종류의 사운드였는지 확인할 수 있습니다.



- 1 사운드 분류 분석 애플리케이션이 탑재된 카메라가 영업 시간 이후 매장에서 발생하는 사운드를 감지합니다.
- 2 매장 업주는 경보를 수신한 후, 추가 조치를 취하기 전에 비디오 스트림을 확인하여 어떤 종류의 사운드였는지 확인할 수 있습니다.

### 3.2.2 센서를 결합하여 감시 시스템을 최대한 활용

감시 시스템은 종종 여러 유형의 센서를 통합합니다. 카메라의 이미지 센서는 당연히 장면의 시각적 측면을 기록하는 센서입니다. 레이더 기술 또는 적외선 방사 방출을 기반으로 하는 모션 디텍터와 같은 비 시각적 센서도 일반적으로 사용됩니다. 비시각적 센서는 다른 유형의 정보 입력을 추가하여 카메라 시스템을 보완합니다.

또한 감시 설비에 오디오 센서(마이크)를 사용함으로써 가능한 모든 사용 사례의 대부분이 강화됩니다. 오디오 기능이 없는 시스템에 오디오 기능과 오디오 분석 애플리케이션을 추가하면 멀티 센서 상호작용이 가능해집니다. 비디오 분석 애플리케이션을 사용하는 경우, 오디오 분석 애플리케이션을 추가하면 감지 신뢰도를 높일 수 있습니다. 예를 들어 저조도 조건으로 인해 비디오 분석에 문제가 있는 경우 특히 그렇습니다.

예를 들어 영상 관리 소프트웨어에서 비디오 분석 애플리케이션과 오디오 분석 애플리케이션이 모두 반응한 경우에만 조치를 트리거하도록 시스템을 설정할 수 있습니다. 예를 들어 오디오 분석

애플리케이션은 비명을 감지하고 비디오 분석 애플리케이션은 카메라 화각에 있는 사람을 감지합니다. 일부 환경에서는 이 조합이 적절한 수준의 보안을 제공합니다.

### 3.2.3 대시보드에 입력 제공

오디오 메타데이터는 메타데이터를 수집하여 시각적으로 표시하는 VSaaS(서비스로서 영상 감시), IoT(사물 인터넷) 또는 비즈니스 인텔리전스 플랫폼과 같은 분석 대시보드에 입력될 수 있습니다. 이러한 분석 대시보드는 실시간 및 과거 추이를 분석하여 즉각적인 오버뷰와 실행 가능한 인사이트를 생성합니다. 고객 흐름 또는 고객 경험에 기반한 통계 분석을 통해 데이터 기반 의사 결정을 내리고 운영을 개선할 수 있습니다.

대시보드를 사용하면, 실제 오디오를 듣거나 원음을 검색할 필요 없이 결과를 확인할 수 있습니다. 대신, 이벤트 계수 등에서 실행 가능한 인사이트를 얻을 수 있으며, 데이터 개인정보 보호에 대한 우려도 없습니다. 오디오가 녹음되었는지 또는 그저 캡처되었는지에 따라 법적 제한이 다를 수 있다는 점에 유의하십시오.

## 4 법적 제한

많은 사람들이 영상 감시에서 마이크를 사용하는 것에 대해 우려하고 있습니다. 이러한 우려는 일반적으로 영상 자료와 함께 평범한 대화를 녹음하는 것과 관련이 있습니다. 하지만 오디오 분석 애플리케이션을 사용하면 일반적으로 오디오를 녹음하거나 스트리밍하지 않습니다. 감시를 규제하는 법률은 지역 및 국가마다 다르므로, 감시 시스템에서 오디오를 사용하기 전에 허용되는 사항을 확인하십시오.

오디오 캡처 및 오디오 녹음은 국가 입법 또는 다양한 유형의 현지 규칙 및 규정에 따라 여러 가지 이유로 금지되거나 특별한 고려가 필요할 수 있습니다. 특정 지역 또는 환경에서 오디오 캡처를 허용할 수도 있지만, 여전히 오디오 녹음을 금지할 수 있습니다. 기업이 자체 구내에서 오디오 감시를 금지할 수도 있습니다.

## 5 면책 고지

이 문서와 그 내용은 Axis가 무상으로 제공하는 것이며, 문서에 대한 모든 권리 또는 그와 관련된 모든 지적 재산권(상표, 상표명, 로고 타입 및 유사한 마크를 포함하되 이에 국한되지 않음)은 법률로 보호되고 문서에 대한 모든 권리, 제목 및/또는 이해관계 또는 이와 관련된 모든 지적 재산권은 Axis Communications AB에 귀속됩니다.

이 문서는 어떤 종류의 보증도 없이 정보 목적으로만 "있는 그대로" 제공됩니다. 이 문서에서 제공하는 정보는 법률 자문을 구성하지 않으며 법률 자문을 의도하지도 않습니다. 이 문서는 Axis Communications AB 및/또는 그 계열사의 법적 의무를 생성하기 위한 것이 아니며 그렇게 해서도 안 됩니다. Axis 제품과 관련된 Axis Communications AB 및/또는 그 계열사의 의무는 Axis와 해당 제품을 Axis에서 직접 구매한 법인 간의 약관을 전적으로 따릅니다.

확실성을 기하기 위해, 이 문서의 사용, 결과 및 이행에 대한 모든 위험은 문서 사용자가 부담하고 AXIS는 법률이 허용하는 최대 범위 내에서 법적, 명시적 또는 묵시적 보증을 포함한 모든 보증에 대해 면책되고 그러한 보증을 배제합니다. 이러한 보증은 상품성, 특정 목적에 대한 적합성, 소유

권 및 비침해 및 제품 책임에 대한 묵시적 보증 또는 이 문서와 관련한 제안, 규격 또는 샘플과 관련하여 발생하는 보증을 포함하되 이에 국한되지 않습니다.

# Axis Communications 정보

Axis는 보안 및 새로운 비즈니스 성과를 개선하기 위한 솔루션을 창조하여 더 스마트하고 안전한 세상을 가능하게 합니다. 네트워크 기술 회사이자 업계 리더인 Axis는 비디오 감시, 접근 제어, 인터콤, 오디오 시스템 솔루션을 제공합니다. 이러한 솔루션은 지능형 분석 애플리케이션으로 향상되고, 고품질 교육의 지원을 받습니다.

Axis에서는 50개 이상의 나라에 약 4,000명의 전담 직원이 있으며 전 세계 기술 및 시스템 통합 파트너와 협력하여 고객 솔루션을 제공합니다. Axis는 1984년에 설립되었으며 본사는 스웨덴 룬드에 있습니다