

백서

"오디오 입력"의 잠재력

장면 인식 및 증거를 위한 사운드 캡처 및 처리

6월 2021

목차

1	요약	3
2	서론	4
	2.1 녹음하지 않고 캡처	4
3	장애물 살펴보기	4
	3.1 법률과 규정에서는 무엇이라고 말할까요?	5
	3.2 가능성을 조사하십시오.	5
4	설치 문제	6
5	오디오 준비	7
6	분석 토폴로지	7
7	사용 사례 및 예	8
	7.1 오디오 분석을 사용하여 사고를 감지하십시오.	9
	7.2 영상에서 사운드의 시각화	10
	7.3 청취 및 상호작용	11
	7.4 녹음 및 저장	12
	7.5 감시 시스템을 최대한 활용하십시오.	12
8	모니터링 및 감지	12
	8.1 사운드 특성	12
	8.2 신호 처리	13
	8.3 인간의 청력	14
9	면책 고지	15
	부록 1 오디오 품질 용어	16

1 요약

비디오 카메라에 통합되어 바로 사용 가능하거나 외부 마이크에서 제공하는 오디오 캡처 기능은 다양한 중요한 사용 사례를 가능하게 합니다. "오디오 입력"을 책임감 있고 전문적으로 사용하면 보안 시스템에 중요한 가치와 이점이 추가될 수 있습니다. 예를 들어, 포렌식 조사에서 누락된 증거를 제공하거나 보안 요원 또는 병원 직원의 즉각적인 주의가 필요한 이벤트를 실시간으로 감지할 수 있습니다. 오디오 감시가 진행되고 있다는 사실만으로도 억제 효과를 발휘해 범죄를 예방할 수 있습니다.

오디오 캡처(종종 즉각적인 분석 작업과 결합됨)를 독립형 기술로 배치하여 범죄 예방, 보호 및 포렌식에서 여러 사용 사례를 가능하게 할 수 있습니다.

그러나 오디오 캡처는 영상과 결합될 경우 기존 감시 사용 사례의 대부분을 강화할 수 있는 잠재력도 갖고 있습니다. 예를 들어, 보안 운영자는 비디오 스트림이 오디오 스트림으로 보완되는 경우 장면 이벤트에 대한 훨씬 더 나은 개요를 얻을 수 있습니다.

시각적 감지를 기반으로 하는 자동 이벤트 감지 및 알람 실행을 위해 여러 유형의 비디오 분석을 사용할 수 있는 것과 같이, 오디오 분석은 오디오 스트림을 모니터링하고 주목할 만한 이벤트가 있을 때 반응할 수 있습니다.

오디오 분석 소프트웨어는 마이크가 사람들이 외치는 소리, 유리 깨지는 소리 또는 총소리와 관련된 사운드를 캡처할 때 자동 알람 및 기타 작업을 트리거하도록 설정할 수 있습니다. 이는 빠른 대응과 개입을 가능하게 하는 조기 경고를 제공합니다.

또한 오디오 분석은 예상치 못한 소리가 왼쪽에서 나는지 또는 오른쪽에서 나는지 감지하고 PTZ 카메라를 소리가 나는 곳으로 자동으로 향하도록 할 수도 있습니다. 병원 또는 요양 시설에서는 오디오 분석을 사용하여 환자가 고통스러운 상태임을 암시하는 높은 사운드 레벨을 감지하고 간호사에게 자동 알림을 보낼 수 있습니다. 이 사용 사례는 여러 위치의 사운드를 동시에 더 쉽게 모니터링할 수 있는 사운드 시각화 분석의 이점을 누릴 수 있습니다.

사운드 캡처와 녹음에는 차이가 있습니다. 많은 유형의 사용에서는 목표를 달성하기 위해 오디오를 녹음할 필요가 없으며, 이는 개인정보 보호 문제를 관리하고 개인 정보에 관한 규정을 준수하는 데 도움이 될 수 있습니다. 일반적으로 오디오 분석 애플리케이션은 사운드를 지속적으로 녹음하지 않습니다. 일반적으로 입력 오디오를 처리하여 특정 패턴, 레벨 또는 주파수를 검색합니다. 분석이 에지에서(카메라에서) 실행될 때 디지털 오디오 데이터를 카메라에서 내보낼 필요가 없습니다. 수행된 분석의 결과, 즉 메타데이터 또는 트리거만 내보냅니다.

Axis는 법률 자문을 제공하지 않습니다. 감시를 규제하는 법률은 지역, 주 및 국가마다 다르며, 모든 감시가 규정을 준수하여 수행되는지 확인할 책임은 제품 사용자(일반적으로 최종 고객)에게 있습니다. 영상 감시의 경우와 마찬가지로 오디오 감시 시스템을 설치하기 전에 오디오 감시 시스템 설치의 법적 측면에 대한 조사와 이해가 선행되어야 합니다.

법적 조건을 충족하기 위해 필요한 조치를 취한 후에는, 장비 배치 및 구성과 관련하여 설치를 신중하게 고려해야 합니다. 이것은 약간의 계획이 필요할 수 있지만 일반적으로 어렵지 않으며 간단한 조치는 오디오 사용성을 달성하는 데 큰 도움이 됩니다.

2 서론

오디오 정보는 범죄 예방, 보호 또는 포렌식 사용을 위한 중요한 자산이 될 수 있습니다. 캡처된 오디오는 활동, 행동 또는 이벤트를 감지하기 위한 매우 효율적인 오디오 모니터링을 가능하게 하는 분석 소프트웨어에 의해 실시간으로 처리될 수도 있습니다.

이 백서에서는 일반적인 사용 사례의 예를 통해 보안에서 오디오가 갖는 잠재력을 설명합니다. 다양한 유형의 오디오 분석이 오디오 분석의 작동 방식에 대한 간략한 개요와 함께 제공됩니다.

이 백서에서는 법적 조언을 제공하지 않지만 시스템 설정에 도움이 될 수 있는 다양한 기술 솔루션을 제공합니다. 오디오 분석을 구현하는 방법에 따라 지역 법률 및 권장 사항을 살펴보고 필요한 경우 이 강력한 도구를 사용할 수 있습니다.

이 백서의 범위는 오디오, 즉 오디오 *입력의 캡처 및 가능한 녹음*으로 제한됩니다. 보안 솔루션에서 오디오의 또 다른 일반적인 용도는 일반적으로 침입자 또는 절도범을 막기 위해 음성 메시지 또는 알람을 재생하기 위한 오디오 *브로드캐스트*, 즉 오디오 *출력*과 관련되어 있습니다. 보안을 위한 오디오 브로드캐스트에 대한 자세한 내용은 www.axis.com/products/audio에서 확인할 수 있습니다.

2.1 녹음하지 않고 캡처

오디오를 녹음하지 않고도 캡처하고 사용할 수 있습니다. 오디오 캡처는 기본적으로 오디오를 디지털화하여 소프트웨어에서 사용할 수 있도록 하는 것을 의미합니다. 이것은 마이크를 사용하여 공기 중의 소리 진동을 기록하고, 이러한 아날로그 신호를 디지털 신호로 변환(A/D 변환 장비를 사용)하여 처리 장치에 전달하는 방식으로 이루어집니다.

캡처한 오디오는 플래시 메모리나 하드 드라이브와 같은 영구 매체에 저장되지 않으면 녹음되지 않습니다. 인간 운영자가 캡처한 오디오를 실시간으로 듣고 있는 경우와 같은 일부 사용 사례에서는 녹음이 필요하지 않을 수 있습니다. 어떤 상황에서는 오디오를 녹음하지 않는 특정한 이유가 있습니다. 오디오가 녹음되었는지 또는 그저 캡처되었는지에 따라 법적 제한이 다를 수 있습니다.

일반적으로 오디오 분석 애플리케이션은 사운드를 지속적으로 녹음하지 않습니다. 제대로 작동하기 위해 일시적으로 사운드를 버퍼링합니다. 많은 시스템이 감지 직전과 직후에 버퍼링된 내용을 기록하도록 설정되어 보안 요원이 감지를 확인하고 가능한 경우 포렌식 증거를 위해 사운드를 보존하도록 할 수 있습니다.

3 장애물 살펴보기

많은 사람들이 영상 감시 상황에서 마이크를 사용하는 것에 대해 우려하고 있습니다. 이러한 우려는 일반적으로 영상 자료와 함께 평범한 대화를 녹음하는 것과 관련이 있습니다.

단순한 오디오 녹음보다 "오디오 입력"에 더 많은 가능성이 있음을 이해한다면 이 초기 장애물을 넘어설 수 있습니다. 사운드 정보를 녹음할 필요가 없는 많은 사용 사례가 있습니다.

감시를 규제하는 법률은 지역 및 국가마다 다르므로, 감시 시스템에 오디오를 추가하기 전에 허용되는 사항을 확인하십시오.

3.1 법률과 규정에서는 무엇이라고 말할까요?

영상 감시의 경우와 마찬가지로 오디오 감시 시스템을 설치하기 전에 오디오 감시 시스템 설치의 법적 측면에 대한 조사와 이해가 선행되어야 합니다. 해당되는 경우 적절한 신청 서류를 제출하고 허가를 받아야 합니다. 필요한 경우 표지판이나 공고문을 사용해야 합니다.

오디오 사용 및/또는 녹음은 국가 입법 또는 다양한 유형의 현지 규칙 및 규정에 따라 여러 가지 이유로 금지되거나 특별한 고려가 필요할 수 있습니다. 한 지역 또는 환경에서 오디오 캡처를 허용할 수 있지만 여전히 오디오 녹음을 금지할 수 있습니다. 기업이 자체 구내에서 오디오 감시 사용을 금지할 수도 있습니다.

3.1.1 미국의 예

미국의 법률과 규정은 주마다 다릅니다.

일부 주에서는 오디오 녹음을 위한 한쪽 당사자의 동의가 필요합니다. 즉, 대화의 한쪽 당사자만 동의하면 감시가 합법적으로 이루어질 수 있습니다.

다른 주에서는 모든 당사자(또는 양쪽 당사자)의 동의가 필요합니다. 즉, 모든 당사자가 녹음에 동의해야만 녹음을 수행할 수 있습니다. 모든 당사자의 동의에 대한 예외는 개인의 사생활 보호를 기대할 수 없는 공공 장소에서 적용될 수 있습니다.

법적 평가는 오디오를 녹음하지 않는 오디오 분석 애플리케이션을 사용할 때 일부 지역에서 다른 결과로 이어질 수도 있습니다. 따라서 해당 주에서 적용되는 법률과 규정을 조사해야 합니다.

3.1.2 유럽의 예

오디오 감시는 유럽 국가의 국내 법률에 따라 규제됩니다. 따라서 해당 국가에서 적용되는 법률과 규정을 조사해야 합니다.

오디오 녹음에는 GDPR(General Data Protection Regulation)이 적용되는 개인 정보가 포함될 수 있습니다. GDPR은 반드시 오디오 녹음을 금지하지는 않지만 오디오 캡처 또는 녹음에는 특별한 고려 사항이 필요합니다. 기존 영상 감시에 오디오를 추가할 때 GDPR에 따른 개인 정보 처리를 위한 법적 근거가 여전히 적용되는지 고려해야 합니다.

3.2 가능성을 조사하십시오.

오디오가 감시에서 절대 허용되지 않는다는 일반적인 오해가 있습니다. 이러한 오해는 너무나 널리 퍼져 있어서 많은 경우에 오디오로 감시 시스템을 강화할 가능성이 전혀 고려되지 않습니다.

그러나, 예를 들어 사람들에게 알릴 경우, 사용자가 동의하는 경우 등에 여러 유형의 감시 시스템 설치가 허용될 수 있습니다. 해당 지역과 해당 사용 사례에 적용되는 법률과 규정을 조사해야 합니다. *녹음 및 저장 사용 사례가 보안 시스템에서 허용되지 않더라도 청취 및 상호작용, 청취 및 목적, 오디오 분석을 사용한 사고 감지*와 같은 많은 사용 사례를 개인정보 보호 권리를 침해하지 않도록 조정할 수 있습니다.

4 설치 문제

장면에서 마이크의 위치는 잠재적인 응용 분야를 정의합니다. 오디오 장비를 설치하기 전에 배치 및 구성을 신중하게 고려해야 합니다. 이것은 약간의 계획이 필요할 수 있지만 일반적으로 어렵지 않으며 간단한 조치는 오디오 사용성을 높이는 데 큰 도움이 됩니다.

적절한 마이크 배치를 고려하고 음향이 좋은 지점을 선택하면 감시 목표를 달성할 확률이 높아집니다. 물론 마이크는 관련 사운드를 쉽게 캡처할 수 있도록 배치해야 합니다. 일반적으로는, 카메라와 함께 룸 중앙에 배치되거나 특정 관심 이벤트가 발생할 수 있는 위치에 가까운 곳에 배치되는 경우가 많습니다. 마이크를 환기 장치나 기계와 같은 소음 발생원 가까이에 배치해서는 안 됩니다. 이는 약하거나 멀리서 들리는 소리를 가릴 수 있습니다.



일반적 마이크 배치

- 1 관심을 끄는 행동이 발생할 수 있는 장소
- 2 카메라
- 3 룸 중앙

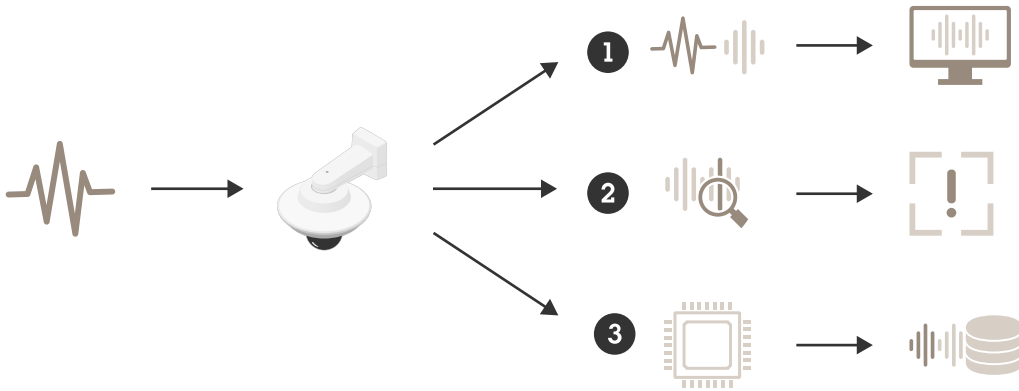
벽이나 천장/바닥의 흡음 특성과 매우 긴 복도와 같은 입체적 복잡성과 같은 음향 환경은 특정 위치의 음장에 심각한 영향을 미칠 수 있는 다양한 잔향과 울림을 생성합니다. 예를 들어, 음성은 교회나 타일로 완전히 마감된 욕실과 비교하여 심하게 감쇄된 장소(예: 방음 처리된 회의실)에서 매우 다르게 들립니다. 음향 조건이 까다로운 상황에서는 마이크 배치가 중요해질 수 있습니다.

장비의 설치 및 구성(예: 오디오 게인 설정)은 물론 오디오 장비와 감시 시스템의 통합이 모두 중요합니다. 시스템 설치업체 및 통합업체가 특정 상황에 대한 권장 사항을 제공할 수 있습니다.

오디오 분석의 경우, 특정 권장 사항이 적용되는 경우가 있으며 이는 일반 오디오 녹음을 위한 권장 사항과 다를 수 있습니다. 항상 사용자 문서를 검토하여 해당 전제 조건을 파악하십시오.

5 오디오 준비

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



1 변환

2 실시간 에지 분석

3 처리 및 인코딩

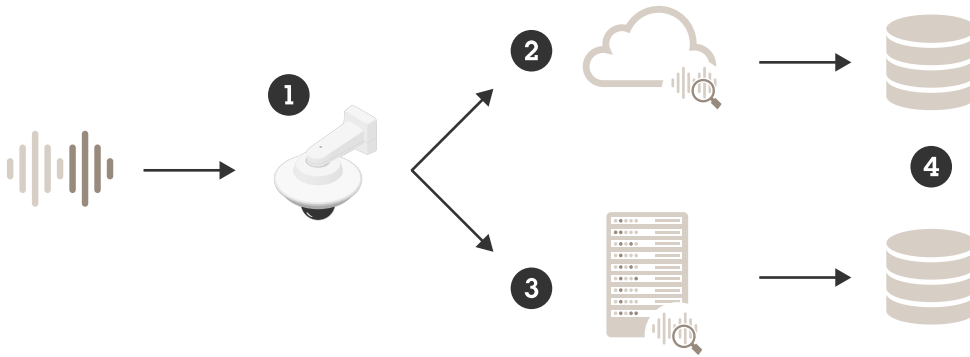
- **변환.** 사운드가 추상화 되어 사운드 스펙트럼을 보여주는 그래프와 같은 시각적 정보로 변환됩니다. 이 과정은 되돌릴 수 없습니다. 스펙트럼 그래프에서 원래의 사운드를 가져올 수 없습니다.
- **실시간 에지 분석.** 사운드가 에지에서 처리되는 경우 **사운드 분류기**를 사용할 수 있습니다. 이를 통해 사운드의 특성을 설명하는 메타데이터를 얻을 수 있습니다. 원래의 사운드를 메타데이터에서 다시 만들 수 없습니다. **사운드 감지기**를 사용하여 패턴, 레벨 또는 주파수를 인식하고 상태 정보를 제공할 수 있습니다. 다시 말하지만 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.

6 분석 토폴로지

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

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.

- **에지 분석.** 에지에서 분석이 실행되면 디지털 오디오 데이터를 카메라에서 내보낼 필요가 없습니다. 녹음하지 않고 오디오를 캡처하는 경우, 수행된 분석의 결과, 즉 메타데이터 또는 트리거만 전송됩니다.
- **서버 분석.** 서버에서 분석이 실행될 때는 디지털 오디오를 카메라에서 내보내야 합니다. 카메라(에지)에서 사전 처리된 경우 이 데이터는 추상화거나 비개인화된 메타데이터가 될 수 있습니다. 서버는 일반적으로 폐쇄형 시스템(시스템 소유자가 제어하는)의 일부이므로 전송된 오디오의 개인정보 보호 문제를 관리 할 수 있습니다. 그럼에도 불구하고 해당 규칙 및 규정을 준수해야 합니다.
- **클라우드 분석.** 디지털 오디오는 클라우드 맥락에서 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.

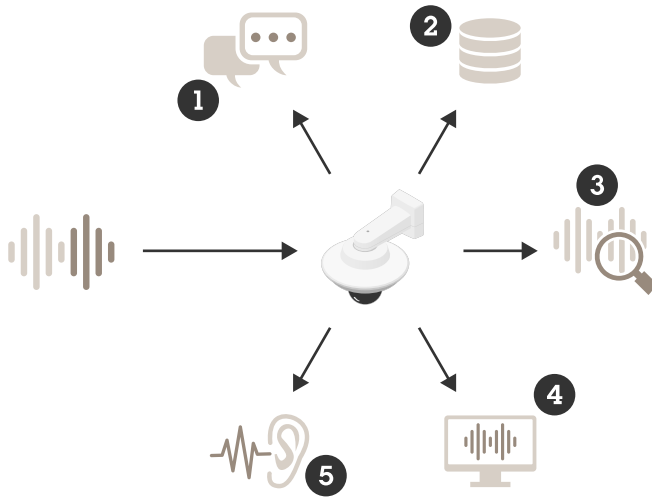


- 1 에지
- 2 클라우드
- 3 서버
- 4 저장

7 사용 사례 및 예

오디오 기능은 종종 통합되어 비디오 카메라에 바로 사용할 수 있습니다. 책임 있고 전문적인 "오디오 입력" 사용이 중요한 가치와 몇 가지 잠재적 이점을 제공할 수 있는 다양한 사용 사례가 있습니다. 예를 들어, 포렌식 조사에서 누락된 증거를 제시하거나 보안 요원 또는 병원 직원의 즉각적

인 주의가 필요한 이벤트를 실시간으로 감지하는 데 사용할 수 있습니다. 오디오 감시가 진행되고 있다는 사실도 억제 효과를 발휘해 범죄를 예방할 수 있습니다.



오디오 캡처의 일반적인 목적:

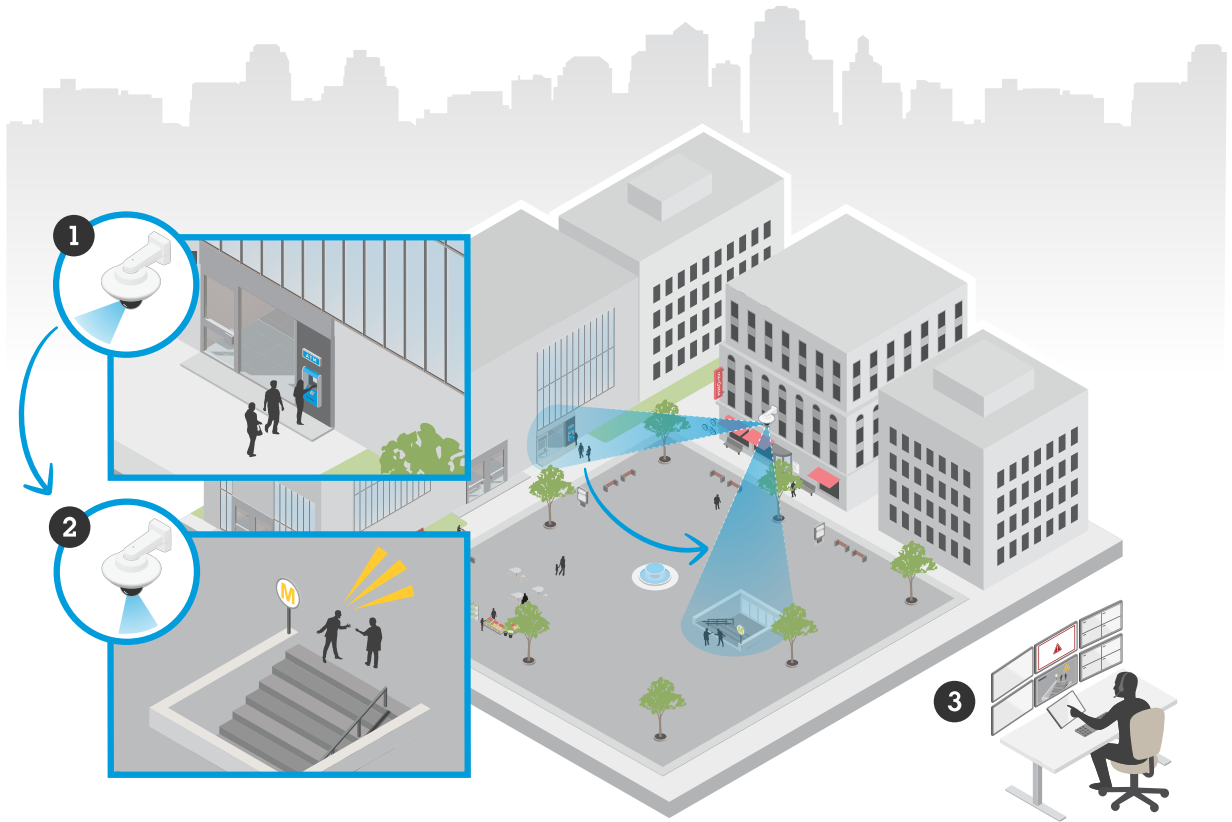
- 1 통신
- 2 녹음
- 3 분석
- 4 시각화
- 5 청취

7.1 오디오 분석을 사용하여 사고를 감지하십시오.

오디오 분석 애플리케이션은 특정 정보를 찾고 추출하기 위해 캡처된 오디오를 처리하는 소프트웨어 프로그램입니다. 오디오 분석 애플리케이션은 총성, 유리 파손 또는 공격과 같은 이벤트를 감지하는 데 사용됩니다. 예를 들어, 오디오 분석 애플리케이션은 입력 오디오를 처리하여 "창이 깨졌습니까?"라는 질문에 '예' 또는 '아니요'라는 대답을 제공할 수 있습니다. 감지 시, 시스템은 일반적으로 시각적 경보를 통해 또는 알람을 트리거하여 직원에게 자동 알림을 보냅니다. 이는 빠른 대응과 개입을 가능하게 하는 조기 경고를 제공합니다.

7.1.1 카메라 방향 변경

오디오 분석의 또 다른 예는 PTZ 카메라 방향 변경 애플리케이션입니다. 이것은 오디오 소스를 감지하고 카메라를 오디오 소스를 향하도록 방향을 자동으로 변경하여 오디오 및 비디오 기능을 결합합니다.



- 1 PTZ 카메라가 ATM 기기를 모니터링하고 있습니다.
- 2 카메라 마이크가 시끄럽고 갑작스러운 소음을 포착하면 카메라가 즉시 사고 지점으로 방향을 변경합니다.
- 3 운영자는 알람을 수신하고 사고를 확인합니다.

7.2 영상에서 사운드의 시각화

영상에서 캡처된 사운드를 시각화하여 모니터에 사운드 스펙트럼 다이어그램으로 표시할 수 있습니다. 설정된 임계값을 초과하면 다이어그램이 알람을 표시하기 시작합니다.

이러한 사운드 시각화는 병원의 여러 병실과 같이 한 번에 여러 소스의 사운드를 모니터링해야 하는 상황에서 유용할 수 있습니다. 많은 오디오 소스(많은 병실에서 들리는 소리)를 동시에 듣는 것은 너무 어

렵지만, 간호사실의 모니터에서 많은 시각적 정보를 보는 것이 훨씬 더 쉬울 것입니다. 병실의 영상 피드를 사용할 수 있는 경우, 시각적 정보를 비디오 이미지에 오버레이로 추가할 수 있습니다.



병원의 영상 피드에 오버레이로 추가된 사운드 시각화.

7.3 청취 및 상호작용

아마도 가장 기본적이고 직관적인 사용 사례는 장면 인식을 높이기 위해 운영자와 직접 상호작용하는 오디오 감시일 것입니다. 이에 대한 예는 일반적으로 의심스러운 대화를 감지하는 것입니다. 이러한 경우 보안 요원을 보내 추가로 조사할 수 있습니다. 또는 병원이나 요양 시설에서 환자가 곤란을 겪고 있는지 듣고 간호사를 부르는 것입니다. 또한 '이상한' 소리가 왼쪽에서 나는지 또는 오른쪽에서 나는지 감지하고 PTZ 카메라가 소리가 나는 곳을 향하도록 하는 것일 수도 있습니다.

이러한 사용 사례에는 관제실에서 또는 모바일 장치의 보안 애플리케이션을 통해 오디오 환경에 액세스할 수 있는 한 명 이상의 운영자가 포함됩니다. 사람의 귀는 소리를 캡처하고 뇌는 장면이나 상황과 관련된 것을 추출합니다. 영상 감시와 함께 사용하는 경우, 오디오는 의사 결정을 위한 또 다른 차원의 정보를 추가합니다. 경우에 따라, 오디오 소스가 카메라 화각 밖에 있거나 조명 조건이 까다로운 경우, 실제로 오디오가 유일한 정보가 됩니다.

7.3.1 청취 및 목적

오디오 감시는 목격관(들은) 이벤트에 기반한 직접적인 증언 목적으로도 사용할 수 있습니다. 이 사용 사례는 목적이 의사 결정이 아니기 때문에 **청취 및 상호작용** 사용과 다르지만, 사용 사례가 종종 공존합니다. 예를 들어, 유죄를 입증하는 발언으로 격화되는 언쟁을 듣게 되면, 운영자는 경비원 뿐만 아니라 나중에 들은 것에 대한 증인도 보낼 수 있습니다.

7.4 녹음 및 저장

적절한 경우 오디오 데이터를 캡처하고 녹음하는 사용 사례는 추가 증거를 제공할 수 있습니다. 이것은 유죄를 입증하는 발언이나 총기 발사와 관련이 있을 수 있습니다. 녹음된 오디오는 누가 무슨 말을 했는지, 몇 발의 총성이 들렸는지 또는 이와 유사한 포렌식 이벤트에 대한 증거를 제공할 수 있습니다.

오디아가 포렌식 맥락에서 녹음되는 경우, 원본 데이터를 보존하고 처리(다른 맥락에서는 필요하거나 유익할 수 있음)를 피하기 위해 주의를 기울여야 합니다. 포렌식 녹음 기록의 경우 모든 유형의 처리가 증거 변조로 간주될 수 있습니다. 음성 향상 알고리즘을 사용하여 녹음된 음성의 가청도를 높일 수 있습니다. 이를 통해 포렌식 가치를 향상시킬 수 있습니다. 그러나 이러한 알고리즘은 나중에 녹음된 자료의 사본에 적용되어야 합니다. 녹음 기록을 되도록 처리되지 않은 상태로 유지하면, 나중에 자료를 여러 가지 방법으로 사용할 수 있습니다.

7.5 감시 시스템을 최대한 활용하십시오.

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. 때로는 영상 감시가 적절하지 않아 비 시각적 센서가 독립형 장치로 사용됩니다. 그러나 대부분의 경우 비 시각적 센서는 다른 유형의 정보를 추가하여 카메라 설치를 완료하는 데 사용됩니다.

또한 감시 설비에 오디오 센서(마이크)를 사용함으로써 가능한 모든 사용 사례의 대부분이 강화됩니다. 비 오디오 시스템에 오디오 기능을 추가하면 분석 또는 운영자 상호 작용을 통해 다중 센서 상호 작용이 가능합니다.

청취 및 상호 작용 사용 사례는 운영자가 오디오 스트림을 수신할 때도 장면 이벤트에 대한 훨씬 더 나은 개요를 얻는 간단한 예입니다. 사람을 바라보는 것만으로는 공격적인 행동을 감지하기 어려울 수 있지만, 사람의 목소리를 들을 수도 있다면 감지가 훨씬 더 쉽습니다.

또 다른 일반적인 예는 비디오 모션 디텍션과 같은 영상 분석을 사용하는 것입니다. 예를 들어 분석 애플리케이션이 저조도 조건으로 인해 적절한 분석 결과를 제공하지 못할 경우, 오디오 분석이 사용될 경우 감지 신뢰도가 높아질 수 있습니다.

8 모니터링 및 감지

오디오에는 모니터링 및 오디오 분석에 모두 사용할 수 있는 여러 종류의 정보가 포함되어 있습니다. 다양한 유형의 처리 및 특성 분류는 더 쉽게 사용하고 주변 시스템과 더 쉽게 상호작용 하도록 이 정보를 추출하고 정제하는 데 도움이 됩니다.

8.1 사운드 특성

소리의 세기 및 음조와 같은 특성은 감시 맥락에서 중요한 정보를 구성할 수 있습니다. 소리가 얼마나 들렸는지, 움직이는지, 혹은 가까이서 들렸는지 멀리서 들렸는지 여부는 우리가 들은 소리에 대한 결론을 도출할 때 퍼즐에 추가되는 모든 조각들의 예입니다. 오디오 모니터링 및 감지를 위한 하드웨어와 소프트웨어는 데시벨 레벨에서 시간 경과에 따른 다양한 주파수의 에너지까지의 복잡한 특성 조합을 "청취"하면서 동일한 유형의 정보로 작동하도록 설계되었습니다.

- **공간적 정보.** 이것은 위치, 방향 및 거리와 같은 개념을 포함하여 우리 주변의 물리적 세계와 관련이 있습니다. 공간적 정보를 사용하여 오디오 캡처에 초점을 맞추거나 오디오 캡처를 여러 방향으로 확대하여 더 나은 녹음을 할 수 있습니다. 또한 분석에서 소리가 어느 방향에서 오는지 또는 소리 발생원이 얼마나 멀리 떨어져 있는지 확인하는 데 사용할 수 있습니다.
- **시간적 정보.** 시간적(시간) 정보는 동적 의미(시간에 따른 변화)와 절대적 의미(무엇이 발생했습니까?) 모두에서 중요하고, 영상과 같은 다른 센서에서 제공하는 정보와 관련하여 종종 볼 수 있습니다. 시간적 정보는 어떤 일이 언제 얼마나 오래 일어났는지 알기 위한 행동 분석에서 중요한 역할을 합니다.
- **스펙트럼 정보.** 이것은 소리의 높은 음조 또는 더 복잡한 소리의 음조 조합과 같은 주파수와 관련이 있습니다. 오디오 감시에 사용되는 마이크는 평탄한 주파수 응답을 갖도록 설계되었습니다. 즉, 가청 범위(20 Hz ~ 20 kHz) 내의 모든 주파수를 캡처하려고 합니다. 이것은 인간의 청각 시스템이 작동하는 방식과 다릅니다. 사람은 사람의 말에서 일반적으로 발생하는 주파수를 다른 주파수보다 더 쉽게 감지할 수 있기 때문입니다.
- **진폭 정보.** 이것은 소리의 세기 또는 크기에 관한 것입니다. 진폭 정보는 스펙트럼 정보를 보완할 수 있고, 들어오는 오디오가 구성되는 방식에 대한 이미지를 그리는 데 함께 사용될 수 있습니다.

8.2 신호 처리

오디오 감시 내에서 신호 처리는 일반적으로 전송, 저장 효율 또는 주관적 품질을 개선하는 것에 관한 것이거나, 관심 있는 구성 요소를 강조 또는 감지하기 위한 것입니다. 이는 다양한 방식으로 오디오를 수정하거나 분석하는 소프트웨어 알고리즘을 통해 수행됩니다.

8.2.1 신호 변경

알고리즘은 일반적으로 다음과 같은 특정 목적을 위해 신호를 변경하는 데 사용할 수 있습니다.

- 신호 개선, 예를 들어 자동 게인 제어를 통해 가청도를 높임.
- 예를 들어, 이퀄라이저로 상대 주파수 내용을 변경하여 신호를 변경.
- 특정 주파수 또는 진폭을 제거하여 신호를 제한. 이는 압축을 통해 데이터 양을 감소시키거나 음성 스크램블링을 통해 개인 정보를 보호하는 것에 관한 것일 수 있습니다.

8.2.2 신호 분석

오디오 분석은 캡처된(그러나 일반적으로 녹음되지 않은) 오디오 데이터를 사용하고 관련 사운드 특성을 분석하여 비 오디오 결과를 생성합니다. 오디오 분석 애플리케이션은 기본적으로 오디오 데이터를 다른 형식의 더 실행 가능한 자산으로 변환합니다. 예를 들어 공격, 총성, 유리 파손 또는 차량 알람 등을 감지하기 위해 특별히 개발된 분석 애플리케이션이 있습니다.

머신 러닝 알고리즘을 사용하는 경우, 분석 애플리케이션은 명시적으로 프로그래밍되지 않아도 예측을 수행하는 방법을 배우기 위해 대량의 데이터에서 학습할 수 있습니다. 오디오 맥락에서의 한 가지 예는 수천 개의 사운드로 훈련된 후 도어가 닫히는 소리를 신뢰성 있게 감지할 수 있는 알고리즘일 수 있습니다.

8.3 인간의 청력

인간의 귀는 오디오를 감지하고 분석하는 데 사용할 수 있는 최고의 도구 중 하나입니다. 매우 시끄러운 환경에서 인간의 귀와 뇌는 여전히 대부분의 알고리즘이 감지 및 해석하지 못할 수 있는 음성을 감지하고 해석할 수 있습니다.

인간은 귀를 사용하여 소리가 나는 위치와 오디오 소스의 이동 여부와 같은 공간적 정보를 장면에서 추출할 수 있습니다. 인간은 귀가 두 개이기 때문에 소리가 왼쪽에서 나는지 오른쪽에서 나는지 또는 양쪽 사이에서 나는지 알 수 있습니다. 또한 인간의 귀와 뇌는 소리가 위에서 나는지 아래에서 나는지, 앞에서 나는지 또는 뒤에서 나는지 알 수 있도록 되어 있습니다. 뇌의 여러 "필터 단계"는 귀 사이의 시간적 차이를 처리하여 마이크로초 단위의 작은 편차를 즉시 감지하여 인간이 특정 유형의 이벤트를 인식하도록 합니다. 인간은 특히 인간의 목소리뿐만 아니라 과거의 위험과 관련된 소리와 관련된 오디오 신호 분석을 할 수 있는 잘 발전된 능력을 가지고 있습니다.

적절한 상황(예: 좋은 음질, 입체 음향, 지연이 너무 많지 않음)에서 인간 운영자는 강력한 "분석 도구"가 되어 감지 하드웨어 또는 소프트웨어를 보완할 수 있습니다. 운영자는 마이크가 두 개만 있는 오디오 감시 제품을 사용하여 소리가 나는 위치와 소리의 이동과 같은 공간적 정보를 장면에서 도출할 수 있습니다.

9 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.

의심을 피하기 위해서 이 문서의 사용, 결과 및 이행과 관련한 모든 위험은 문서의 사용자가 부담하고, Axis는 법적으로 허용되는 최대 범위 내에서 상품성, 특정 목적에 대한 적합성, 소유권 및 비침해 및 제조물 책임에 대한 일체의 묵시적 보증을 포함하되 이에 국한되지 않는 모든 보증 또는 이 문서와 관련된 일체의 제안, 사양 또는 샘플과 관련하여 발생하는 모든 보증으로부터 면제되고 배제됩니다.

부록 1 오디오 품질 용어

디지털 오디오:

디지털 오디오는 디지털 형식으로 녹음된 아날로그 오디오(종종 마이크로 캡처된 음향 신호)를 나타냅니다. 디지털 오디오에서 오디오 신호의 음파는 일반적으로 연속적인 숫자 샘플 시퀀스로 인코딩됩니다. 정확도는 인코더가 기록하는 유효 자릿수에 따라 다릅니다. 예를 들어, CD 오디오에서 샘플은 각각 16비트 샘플 심도로 초당 44,100번 수집됩니다.

노이즈:

노이즈는 음량 범위의 조용한 쪽을 정의하거나 제한하는 원치 않는(때로는 피할 수 없는) 소리입니다. 노이즈는 녹음된 소스(예: 실내 팬)에서 마이크(예: 자체 소음, 진동, 바람), 케이블 연결(예: 간섭, 혼선) 및 캡처 장치(예: 자체 노이즈, 디지털 샘플링 노이즈)에 이르는 오디오 체인의 모든 부분에서 생성됩니다. 이러한 것들이 모두 결합되어 일반적으로 노이즈 플로어라고 하는 것을 생성합니다.

노이즈는 일반적으로 정의된 레벨(때로는 시스템이 처리할 수 있는 가장 큰 소리)에서 노이즈 플로어까지의 전체 범위인 SNR(신호 대 잡음비)에 의해 정의됩니다.

영상에 해당하는 것은 무작위적(일반적으로) 정적 픽셀 패턴인 "스노우"처럼 보이는 영상 노이즈입니다. 이것은 어두운 이미지에서 볼 수 있는 것을 제한합니다(무음 신호의 경우 듣는 것을 제한하는 것처럼).

왜곡:

신호의 모든 원치 않는 변경은 원래의 "참값"에서 차감되고 이를 왜곡이라고 합니다(위에서 설명한 것처럼 노이즈는 일반적으로 왜곡 사양에서 제외됩니다). 왜곡은 주관적 품질을 떨어뜨리고(일반적으로 "좋은" 소리로 들리는 왜곡이 있음) 객관적인 정보 콘텐츠를 모호하게 하여 특히 콘텐츠 분석 시 신호를 더 듣기 어렵게 만들고 분석 기능을 저하시킵니다.

THD(Total Harmonic Distortion) 및 IMD(Inter Modulation Distortion)는 일반적으로 왜곡을 정량화하는 데 사용되는 두 가지 속성입니다.

왜곡은 색수차, 비네팅, 블러 등과 같은 아티팩트로서 영상과 관련이 있고, 이미지를 "나쁘게" 보이게 하고 볼 수 있는 세부 정보의 양을 제한합니다.

샘플 레이트 및 주파수 응답:

디지털 시스템에서 오디오는 초당 정해진 횟수만큼 샘플링됩니다. 이것이 샘플 레이트입니다(일반적으로 초당 8000 ~ 48,000 회 또는 Hz). 신호 이론(특히 Nyquist Shannon 샘플링 정리)에서는 사운드를 적절하게 캡처하려면 샘플 레이트가 아날로그 신호에서 원하는 또는 필요한 가장 높은 주파수의 최소 두 배가 되어야 한다고 말합니다.

정상적인 사람의 귀는 연령 및 기타 요인에 따라 20 Hz에서 약 15~20 kHz의 주파수를 듣습니다. 대략적으로 말하면 수백 Hz 이상의 저주파 범위는 종종 특정 사운드의 기초(음성의 기초와 같은)를 정의하는 반면, 수천 Hz 이상의 고주파 범위에는 더 많은 '세부 정보'가 포함됩니다.

오디오의 주파수 범위는 영상의 해상도 및 프레임 레이트와 관련이 있습니다. 낮게 설정할수록 세부 정보가 줄어듭니다.

비트 심도:

오디오가 샘플링될 때마다 아날로그 값이 캡처되어 디지털 신호로 변환됩니다. 디지털 도메인에는 무한대가 없으므로 세부 정보의 양은 정의된 비트 심도로 제한됩니다. 모든 비트는 정의된 진폭 범위(예: 선택한 전압 또는 음압 레벨)와 결합되어 이 범위의 일부를 생성하는 2의 인수(예: 0 또는 1, 낮음 또는 높음)를 나타냅니다. 2비트는 4개의 소수를 생성하고 3비트는 8개의 소수를 생성하는 식입니다. 간단히 말해서, 3개의 비트로 샘플링된 1 볼트 신호는 1/8 볼트 스텝으로 분할 및 표시됩니다.

충분한 오디오 품질을 얻으려면 적어도 사람의 귀에는 16 비트면 충분합니다(65,536 스텝을 나타내는). 이것이 CD 오디오가 사용하는 것입니다. 분석 또는 더 까다로운 사용의 경우 24 비트가 더 적절합니다.

비트 심도는 영상의 대비, 각 픽셀이 재현할 수 있는 휘도 또는 색차 범위와 상호 관련되어 있습니다.

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