

WHITEPAPER

Sicherheit durch Audio-Analysefunktionen

Februar 2024

Zusammenfassung

Audio-Analysefunktionen für Sicherheitsanwendungen können Klangmuster erkennen und unerwartete Geräusche in Live-Audio markieren. Die Analyse kann beispielsweise verbale Aggressionen erkennen, um Eskalationen und Übergriffe zu verhindern, Glasbruch erkennen, um Einbrüche zu verhindern, oder Frühwarnungen bei Personen in Not ausgeben.

Die Verwendung unterschiedlicher Sensortypen, wie z. B. Video- und Audiosensoren (Kamera und Mikrofon), erhöht die Zuverlässigkeit der Erfassung und liefert aussagekräftigere Erkenntnisse.

AXIS Audio Analytics ist eine Software-Plattform für in die Axis Gerätesoftware integrierte Edge-basierte Analysefunktionen. Diese Analysefunktionen erfassen und erkennen Geräusche, ohne den ursprünglichen Audiostream zu speichern. Das ist eine Möglichkeit, die Privatsphäre zu schützen. Das funktioniert, weil die Analysefunktionen Edge-basiert sind und lediglich Audio-Metadaten liefern.

Inhalt

1	Einführung	4
2	Technologie	4
	2.1 Erfassung von Geräuscheignissen	4
	2.2 Audio-Erfassung und -Bearbeitung	4
	2.3 Edge-oder Server-basierte Analysefunktionen	5
	2.4 Metadaten	6
3	AXIS Audio Analytics	7
	3.1 Datenschutz	7
	3.2 Anwendungsbeispiele	8
4	Gesetzliche Beschränkungen	12
5	Haftungsausschluss	12

1 Einführung

Audio-Analysefunktionen für Sicherheitsanwendungen können Klangmuster erkennen und unerwartete Geräusche in Live-Audio markieren. Die Analysefunktion kann zum Beispiel Geräusche erkennen, die auf Aggression hindeuten, Glasbruch erkennen oder frühzeitig warnen, wenn Menschen in Bedrängnis geraten.

Audioanalysefunktionen in Verbindung mit Videosicherheit können Sicherheitsmitarbeiter auf möglicherweise entstehende Vorfälle hinweisen und sie zu den relevanten Kameraansichten leiten. Dies kann eine frühe Erkennung und schnelles Eingreifen ermöglichen und in vielen Fällen eine weitere Eskalation verhindern.

Dieses Whitepaper zeigt, wie Audio-Analysefunktionen in Sicherheitsanwendungen eingesetzt werden können. Wir beschreiben die Technologie zur Erfassung und Verarbeitung von Audio, mit Schwerpunkt auf Echtzeit-Edge-Analysefunktionen wie KI-basierte Klangklassifizierung direkt in der Kamera bzw. im Mikrofon. Wir zeigen auch, wie die Audioanalyse direkt im Endgerät verschiedene Optionen zur Wahrung der Privatsphäre durch die Verwendung von Audio-Metadaten ermöglicht.

Dieses Whitepaper stellt keine Rechtsberatung dar. Vor der Installation eines Sicherheitssystems müssen Sie selbst prüfen, welche Gesetze und Vorschriften in Ihrer Region und für Ihren Anwendungszweck gelten. Der Besitzer des Systems muss sicherstellen, dass es den geltenden Gesetzen, Vorschriften und Empfehlungen entspricht.

2 Technologie

2.1 Erfassung von Geräuschereignissen

Ein Geräuschereignis ist ein Audiosegment, das ein Mensch als eigenständiges Konzept identifizieren würde, wie etwa Schreien oder zersplitterndes Glas. Videoanalysefunktionen können diese Art von spezifischen Schallmustern auf ähnliche Weise erkennen wie Objektklassen.

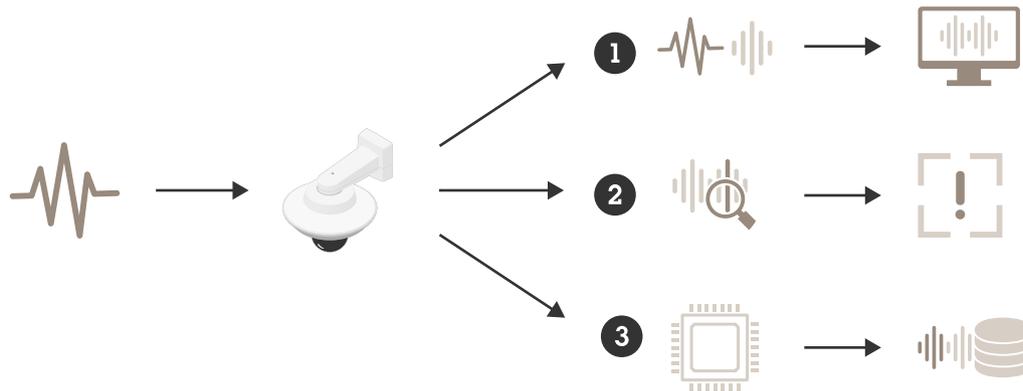
Analysefunktionen, die auf die Erkennung von Schallmustern trainiert sind, suchen meist nach einer Kombination aus Merkmalen, wie der Lautstärke in Dezibel kombiniert mit der Energie unterschiedlicher Frequenzen im Laufe der Zeit. Wird ein spezifisches Schallmuster erkannt, kann das System eine automatische Benachrichtigung in Form einer optischen Warnung an die Mitarbeiter senden oder einen Alarm auslösen.

KI-basierte Algorithmen können an einer großen Menge von Daten trainiert werden. So kann ein Algorithmus beispielsweise menschliche Schreie zuverlässig erkennen, nachdem er mit Tausenden solcher Geräusche trainiert wurde.

2.2 Audio-Erfassung und -Bearbeitung

Bei der Audioanalyse werden erfasste Audiodaten verwendet und die relevanten Klangeigenschaften analysiert, um eine nicht-akustische Ausgabe zu erzeugen. Bei der Erfassung von Audiodaten geht es im Prinzip darum, diese zu digitalisieren und für eine Nutzung durch Software bereitzustellen. Dazu werden die Schallschwingungen in der Luft von einem Mikrofon aufgefangen. Diese analogen Signale werden in digitale Signale umgewandelt und an eine Verarbeitungseinheit weitergeleitet. Wenn der aufgenommene Ton nicht auf einem dauerhaften Medium wie einem Flash-Speicher oder einer Festplatte gespeichert wird, wird er nicht aufgezeichnet. Audiostreaming und Aufzeichnung sind in Axis Kameras standardmäßig deaktiviert.

Audiodaten werden zunächst erfasst und anschließend auf die folgenden Verarbeitungsstufen vorbereitet. Die unterschiedlichen Vorbereitungsschritte können gebündelt oder einzeln erfolgen.



Audiodaten vorbereiten

- 1 *Umwandlung*
- 2 *Edge-Analyse in Echtzeit*
- 3 *Verarbeitung und Codierung für Streaming oder Speicherung – bei einer Axis Kamera wird Audio weder gestreamt noch gespeichert, wenn Audiostreaming nicht explizit aktiviert wird.*

- **Umwandlung.** Die akustische Information wird abstrahiert und z. B. als spektrales Klangdiagramm visualisiert. Dieser Schritt lässt sich nicht rückgängig machen: Sie können den Originalton aus einem spektralen Klangdiagramm nicht wiederherstellen.

- **Edge-Analyse in Echtzeit.**

Bei einer Analyse der Audiodaten „on the edge“ kann eine **Geräuschklassifizierung** zum Einsatz kommen. Das Ergebnis sind Metadaten, die Informationen zu bestimmten Klangeigenschaften liefern. Auch aus diesen Metadaten lässt sich der Originalton nicht wiederherstellen.

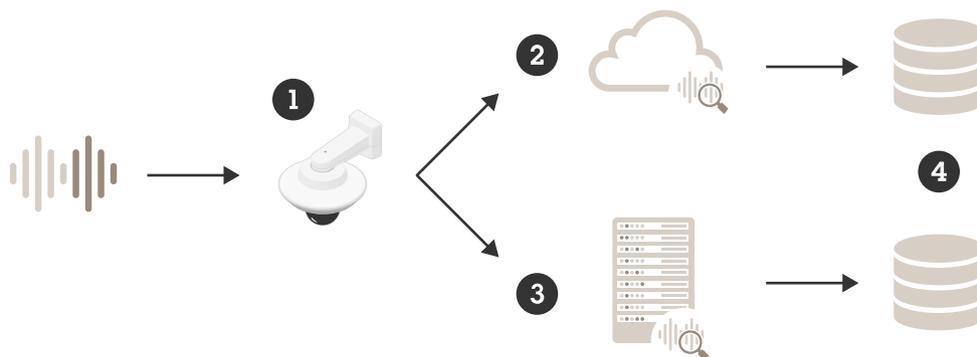
Mithilfe von **Geräuscherfassung** lassen sich Muster, Lautstärken oder Frequenzen erkennen und Informationen zum Status ermitteln. Auch hier gilt: Der Originalton lässt sich nicht wiederherstellen.

- **Verarbeitung und Kodierung.** Wenn die Audiodaten im Original (also weder umgewandelt noch analysiert) verwendet werden, finden bestimmte Verarbeitungs- und Kodierschritte statt, um die Daten für das jeweilige Schutzziel nutzbar zu machen. In solchen Fällen können Audiodaten „on the edge“ gespeichert, zur weiteren Verarbeitung (auf einem Server oder in der Cloud) an externe Kunden gestreamt oder extern gespeichert werden. Bei einer Axis Kamera muss Audiostreaming aktiv eingeschaltet werden. Standardmäßig ist es aus Datenschutzgründen (Audio-Datenschutz) ausgeschaltet.

2.3 Edge-oder Server-basierte Analysefunktionen

Es ist aus mehreren Gründen wichtig, wo im System sich das Analysemodul befindet. Insbesondere in Bezug auf Datenschutzbedenken und die Einhaltung des Datenschutzrechts spielt es eine große Rolle, wo der Softwarealgorithmus die Audiodaten analysiert. Manchmal können Audiodaten nicht über das Netzwerk versendet werden, und die Analyse der erfassten (aber nicht gespeicherten) Daten muss direkt vor Ort

möglich sein. Bei sehr rechenintensiven Algorithmen, die nicht „on the edge“ erfolgen können, müssen digitale Audiodaten vielleicht an die Cloud oder einen Server gesendet werden.



Mögliche Standorte für die Durchführung der Analysen

- 1 Edge
- 2 Cloud
- 3 Server
- 4 Speicherung – nur wenn Audio-Streaming aktiviert ist, kann der Originalton gespeichert werden.

- **Edge-Analyse.** Wenn die Analyse direkt im Gerät erfolgt, braucht kein Audiostream die Kamera zu verlassen. Nur das Ergebnis der durchgeführten Analyse, also Metadaten oder Trigger, wird übermittelt. Die Analysefunktionen in AXIS Audio Analytics sind Edge-basiert.
- **Server-Analyse.** Bei der Ausführung von Analysefunktionen in einem Server müssen Audiodaten von der Kamera übertragen werden. Bei einer Vorverarbeitung in der Kamera können diese auf abstrahierte oder anonymisierte Metadaten beschränkt werden. Ein Server ist normalerweise Teil eines geschlossenen Systems (ein Systembesitzer hat die Kontrolle), so dass die Geheimhaltung der übertragenen Audiodaten gewährleistet werden kann. Dennoch ist unbedingt sicherzustellen, dass die geltenden Gesetze und Vorschriften eingehalten werden.
- **Cloud-Analyse.** Audiodaten können auch an einen Cloud-Server gesendet werden. Wie bei der Server-Analyse lassen sich die Audiodaten auch hier vorab in Metadaten umwandeln. Da Cloud-Lösungen häufig dezentral organisiert sind, ist in diesem Fall umso mehr darauf zu achten, dass alle geltenden Datenschutzvorschriften eingehalten werden.

2.4 Metadaten

Die Audioanalyse erzeugt einen konstanten Metadatenstrom von Audiopegeldaten. Die Funktionen erzeugen außerdem Metadaten auf der Grundlage von Ereignissen, die von Analysefunktionen und adaptiver Audioerfassung erkannt werden. Edge-Analysefunktionen analysieren die Audiodaten direkt in der Kamera. Sie brauchen den eigentlichen Audiostream nicht zu übertragen – sie können nur die Metadaten senden, die Informationen zum Geschehen in der Szene liefern.

Der Metadatenstrom ermöglicht eine visuelle Inspektion der Audio-Hüllkurve, die beispielsweise in einem Dashboard dargestellt wird, in dem Audio- und Video-Ereignisse abgeglichen und gemeinsam betrachtet werden können. So ermöglichen Metadaten eine effiziente und schnelle Suche nach bestimmten Ereignissen und unerwarteten Geräuschen. Das kann den Ermittlungsbeamten viele Stunden Zeit beim Durchforsten umfangreicher Videodaten ersparen.

3 AXIS Audio Analytics

AXIS Audio Analytics ist eine Software-Plattform, die in Zukunft um weitere und aktualisierte Analysefunktionen und Merkmale erweitert wird. Die Analysefunktionen sind in die Gerätesoftware AXIS OS integriert und werden kostenlos mit Axis Kameras und anderen Geräten mitgeliefert.

Die Analysefunktionen und die Algorithmen laufen direkt in der Kamera. Das sorgt für optimale Skalierbarkeit, minimalen Datenverkehr und zuverlässigen Schutz der Privatsphäre. Nur das Ergebnis der Analyse (Metadaten oder Trigger) wird gespeichert. Es wird kein Audio von der Kamera aufgezeichnet oder gestreamt, und die ursprünglichen Geräusche können aus den Metadaten nicht wiederhergestellt werden.

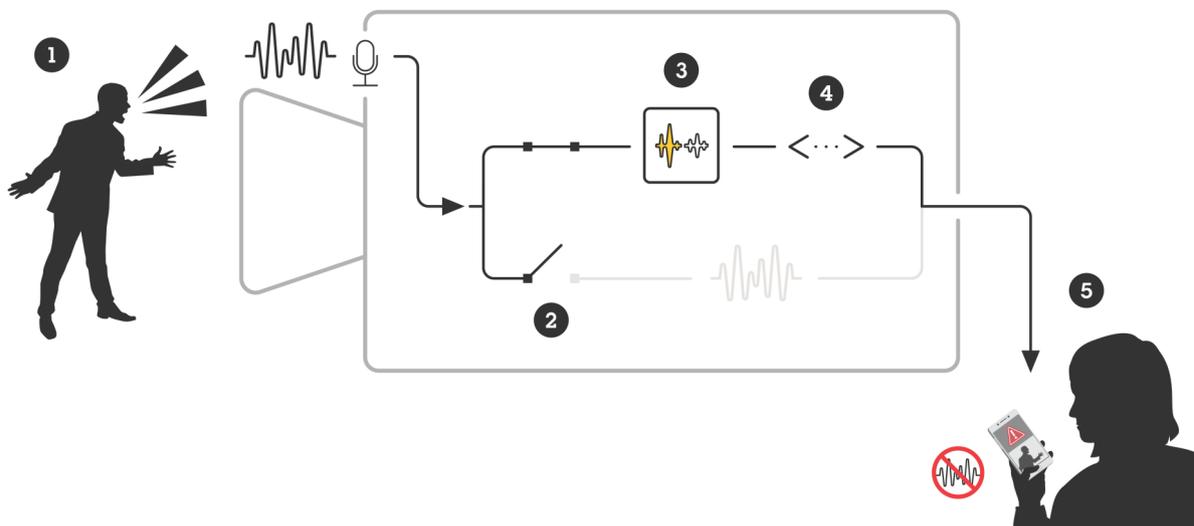
- **Audioklassifizierung (Rufe und Schreie)** Dies ist eine KI-basierte Klangklassifizierung, die bestimmte Geräusche wie Schreie erkennt und analysiert. In Zukunft werden weitere Audioklassifizierungen wie splitterndes Glas hinzukommen. Klassifizierungs-Analysefunktionen erzeugen Metadaten, die die Merkmale des Geräuschs beschreiben.
- **Adaptive Audioerkennung.** Dies ist ein Klangdetektor, der bei einer plötzlichen Veränderung der Lautstärke ein Ereignis erzeugt. Er erkennt Geräuschspitzen jeder Art und bietet den Vorteil, dass er sich auch bei veränderlicher Lautstärke an den Umgebungsgeräuschpegel anpasst.

3.1 Datenschutz

Bei der Audioanalyse wird der eingehende Ton im Allgemeinen nicht aufgezeichnet und auch nicht von der Kamera übertragen. Sie verarbeiten lediglich Geräusche, um in einem Empfängersystem wie etwa einem Dashboard nach bestimmten Ereignissen, Mustern oder Schallpegeln suchen zu können, um diese weiter zu untersuchen, oder einer Video Management Software, die Sicherheitsmitarbeiter alarmiert. Weder können die Audiodaten wiederhergestellt noch vertrauliche Gespräche aufgezeichnet werden. Dies liegt daran, dass diese Analysefunktionen Edge-basiert sind und Audio-Metadaten liefern.

AXIS Audio Analytics zeichnet keine Audiodaten auf oder überträgt sie. Es überträgt ausschließlich Metadaten. Aus Gründen des Datenschutzes ist außerdem jegliches Audiostreaming in Axis Geräten standardmäßig ausgeschaltet (Audio-Datenschutz). Das Audio wird also weder gestreamt noch aufgezeichnet und kann nicht wiederhergestellt werden. Audiostreaming kann bei Bedarf eingeschaltet werden, aber auch bei ausgeschaltetem Audio erfolgt eine Benachrichtigung, sobald etwas Relevantes geschieht.

Die Abbildung zeigt eine Übersicht der Funktionsweise von AXIS Audio Analytics in Verbindung mit Audio-Datenschutz beim Erfassen von Geräuschen sowie die Verwendung von Metadaten für die Erzeugung einer Warnung.



Audio-Datenschutz und AXIS Audio Analytics in Axis Kameras, von einem Schrei (1) bis zu einer Warnung (5).

- 1 Das Mikrofon erfasst Geräusche.
- 2 Audio-Streaming ist standardmäßig ausgeschaltet.
- 3 Die Audioklassifizierung von AXIS Audio Analytics erkennt Rufe oder Schreie im eingehenden Audiomaterial.
- 4 Die Audioanalyse erzeugt Metadaten einschließlich einer Ereignisbenachrichtigung.
- 5 Abhängig von der Ereignisbenachrichtigung und den Metadaten werden die zuständigen Personen in Kenntnis gesetzt. Diese können die Vorgänge im Videostream überprüfen. Es steht kein Audio-Stream zur Verfügung.

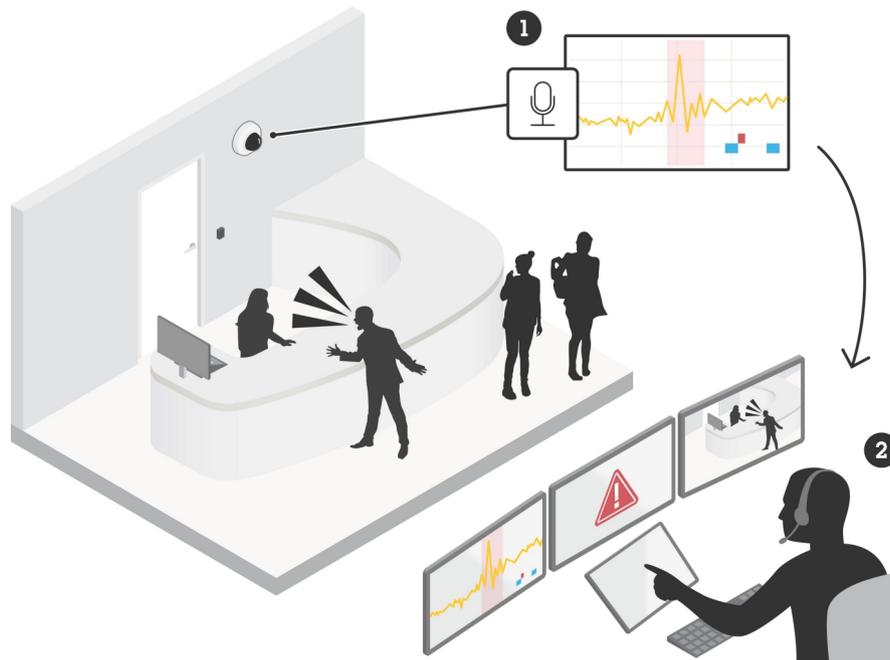
3.2 Anwendungsbeispiele

KI-basierte Analysefunktionen haben ein großes Potenzial, irrelevante Geräusche herauszufiltern, können aber bei vielen Hintergrundgeräuschen Fehlalarme auslösen. Regen, der gegen Fensterscheiben prallt, Donner, Sirenen, Musik oder belebte Szenen mit sich unterhaltenden Menschen können Fehlalarme auslösen. Typische Anwendungsbeispiele sind also beispielsweise ruhige Lokalitäten wie Banken und Empfangstresen, aber auch viele andere Arten von Innenräumen außerhalb der Geschäftszeiten, wie Geschäfte, Restaurants, Treppenhäuser oder Büros.

3.2.1 Geräuscherfassung mit Warnungen

Analysefunktionen zur Geräuschklassifizierung können den Bereich von Bankschaltern oder Empfangstresen überwachen und Geräusche erkennen, die mit Aggression oder zersplitterndem Glas in Verbindung stehen. Werden diese erkannt, alarmieren die Ereignis- und Audio-Metadaten des Systems automatisch Mitarbeiter

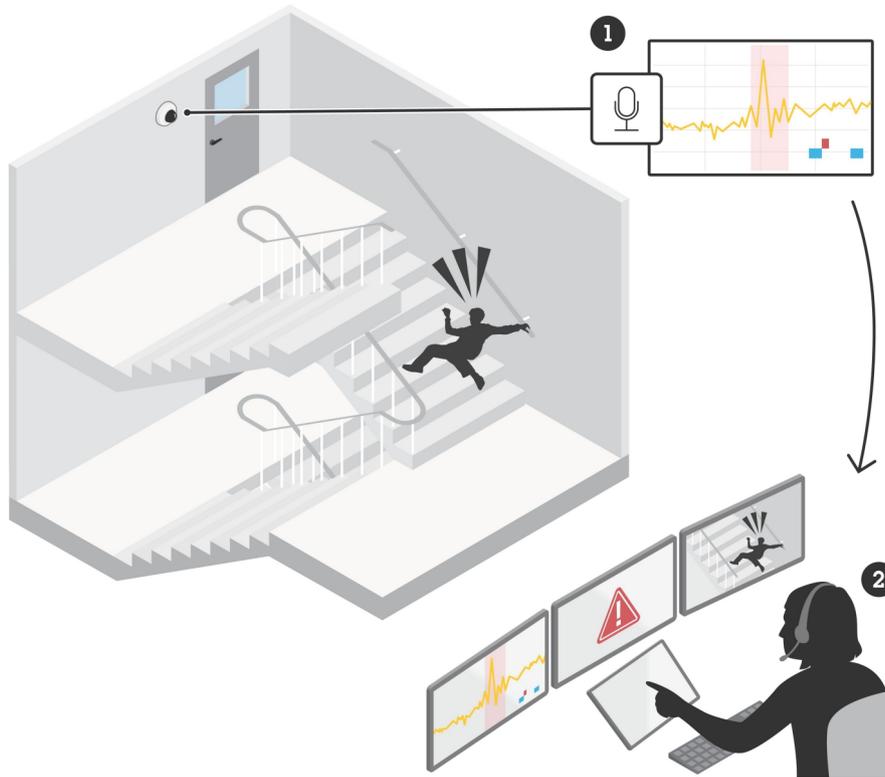
über eine optische Warnung oder lösen einen Alarm aus. Dies ermöglicht frühe Warnungen und damit schnelles Reagieren und Eingreifen.



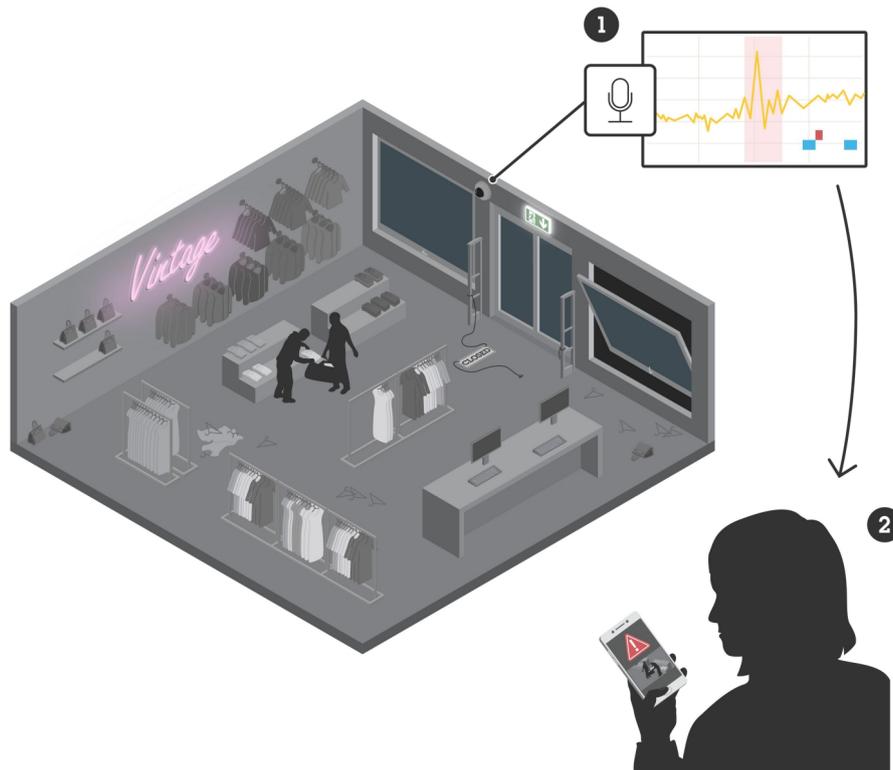
- 1 Eine Kamera mit Analysefunktionen zur Geräuschklassifizierung erkennt Beschimpfungen und Schreie am Empfangstresen.
- 2 Ein Sicherheitsmitarbeiter erhält eine Warnung und kann den Videostream überprüfen, bevor er weitere Maßnahmen einleitet.

Mithilfe adaptiver Audioerkennungsanalytik können unerwartete Geräusche außerhalb der Geschäftszeiten erfasst werden. Die Analysefunktionen analysieren die Umgebungsgeräusche und reagieren, sobald sie Stimmen, brechendes Fensterglas oder andere plötzliche, kurzzeitige Geräusche erkennen. Sobald

ein Ereignis erkannt wird, leiten die Analysefunktionen die Metadaten weiter und informieren die Sicherheitsmitarbeiter.



- 1 Eine Kamera mit Analysefunktionen zur Klangklassifizierung erkennt nach Büroschluss unerwartete Geräusche.
- 2 Ein Sicherheitsmitarbeiter erhält eine Warnung und kann den Videostream überprüfen, bevor er weitere Maßnahmen einleitet.



- 1 Eine Kamera mit Analysefunktionen zur Klangklassifizierung erkennt nach Geschäftsschluss Geräusche in einem Ladengeschäft.
- 2 Der Ladenbesitzer erhält eine Warnung und kann den Videostream überprüfen, bevor er weitere Maßnahmen einleitet.

3.2.2 Kombinieren von Sensoren für noch bessere Nutzung Ihres Sicherheitssystems

Überwachungssysteme enthalten häufig verschiedene Arten von Sensoren. Einer ist natürlich der Bildsensor der Kamera, der den visuellen Charakter einer Szene erfasst. Auch nicht-visuelle Sensoren kommen häufig zum Einsatz, z. B. in Bewegungsmeldern, die Radarwellen oder Infrarotstrahlung nutzen. Nicht-visuelle Sensoren ergänzen die Kamerainstallation, indem sie zusätzlich dazu weitere Arten von Informationen liefern.

Die allermeisten Nutzungsszenarien profitieren davon, wenn eine Überwachungslösung um Geräuschsensoren (Mikrofone) erweitert wird. Eine Erweiterung eines Systems ohne Audio um Audiofunktionen und Audioanalyse ermöglicht eine Interaktion zwischen unterschiedlichen Sensoren. Zusätzliche Audioanalysefunktionen können bei der Videoanalyse die Erfassungsgenauigkeit verbessern. Das ist besonders dann der Fall, wenn die Videoanalyse zum Beispiel durch schlechte Lichtverhältnisse erschwert wird.

Das System kann beispielsweise in Video Management Software eingerichtet werden, so dass es nur dann Aktionen auslöst, wenn beide Analysefunktionen (für Video und Audio) reagieren. Die Audioanalyse

erkennt zum Beispiel einen Schrei und die Videoanalyse eine Person im Sichtfeld der Kamera. In manchen Umgebungen liefert diese Kombination genau das richtige Maß an Sicherheit.

3.2.3 Lieferung von Input für Dashboards

Die Audio-Metadaten können in Analyse-Dashboards wie VSaaS (Video Surveillance as a Service), IoT (Internet der Dinge) oder Business-Intelligence-Plattformen eingegeben werden, die die Metadaten sammeln und visuell darstellen. Sie analysieren Echtzeit- und historische Trends und erzeugen daraus sofort eine gute Übersicht und hilfreiche Erkenntnisse. Statistische Analysen auf der Grundlage von Kundenbewegungen oder Kundenerfahrungsdaten ermöglichen datengestützte Entscheidungen zur Verbesserung der Betriebsabläufe.

Dashboards liefern Ergebnisse, ohne dass man sich das eigentliche Audiomaterial anhören muss. Das Original-Audiomaterial kann dabei nicht abgerufen werden. Stattdessen lassen sich zum Beispiel durch Zählen von Ereignissen praktisch nutzbare Erkenntnisse gewinnen, ohne dass es zu datenschutzbezogenen Bedenken kommen könnte. Man muss jedoch beachten, dass unterschiedliche rechtliche Regelungen gelten können, je nachdem, ob Audiodaten nur erfasst oder auch aufgezeichnet werden.

4 Gesetzliche Beschränkungen

Viele Menschen haben Bedenken hinsichtlich der Verwendung von Mikrofonen bei der Videoüberwachung. Die entsprechenden Bedenken beziehen sich häufig darauf, dass neben dem Videomaterial auch gesprochener Klartext aufgezeichnet wird. Doch bei der Audioanalyse wird Audio in der Regel weder aufgezeichnet noch gestreamt. Je nach Land und Region unterliegen Überwachungslösungen unterschiedlichen Rechtsvorschriften – prüfen Sie also, was erlaubt ist, bevor Sie Audiofunktionen in Ihrem Sicherheitssystem nutzen.

Nationales Recht oder lokal geltende Regelungen und Vorschriften können die Erfassung und Aufzeichnung von Audiodaten aus unterschiedlichen Gründen verbieten oder an bestimmte Auflagen knüpfen. Auch könnte in einer Region oder Umgebung die Erfassung von Audiodaten erlaubt, ihre Aufzeichnung aber verboten sein. Weiterhin können Unternehmen Audioüberwachung auf ihrem Gelände untersagen.

5 Haftungsausschluss

Dieses Dokument wird mit freundlicher Genehmigung von Axis zur Verfügung gestellt, und alle Rechte am geistigen Eigentum oder sonstigen Rechte an diesem (wie z. B. Marken, Handelsnamen, Logos und andere darin verwendete Zeichen) sind gesetzlich geschützt und liegen bei Axis Communications AB.

Beachten Sie bitte, dass dieses Dokument im vorliegenden Zustand, nur zu Informationszwecken und ohne jegliche Gewährleistung zur Verfügung gestellt wird. Die in diesem Dokument bereitgestellten Informationen stellen keine Rechtsberatung dar und sind nicht als solche konzipiert. Dieses Dokument bewirkt keine rechtliche Verpflichtung für Axis Communications AB und/oder seine Partner und ist auch nicht dafür vorgesehen. Verpflichtungen von Axis Communications AB und/oder seiner Partner in Bezug auf Produkte von Axis unterliegen ausschließlich den Bedingungen der Vereinbarung zwischen Axis und der Einheit, die solche Produkte direkt von Axis erworben haben.

ZUR KLARSTELLUNG: DAS GESAMTE RISIKO IM HINBLICK AUF DIE NUTZUNG, ERGEBNISSE UND UMSETZUNG DIESES DOKUMENTS ÜBERNIMMT DER NUTZER. AXIS SCHLIESST IM GESETZLICH MAXIMAL ERLAUBTEN RAHMEN SÄMTLICHE GEWÄHRLEISTUNGEN AUS, OB GESETZLICH, AUSDRÜCKLICH ODER STILLSCHWEIGEND, UNTER ANDEREM SÄMTLICHE STILLSCHWEIGENDE GEWÄHRLEISTUNGEN DER VERMARKTBARKEIT, EIGNUNG FÜR EINEN BESTIMMTEN ZWECK, RECHTSANSPRUCH UND

NICHTVERLETZUNG SOWIE PRODUKTHAFTUNG, ODER GEWÄHRLEISTUNGEN, DIE SICH IN BEZUG AUF DIESES DOKUMENT AUS EINEM VORSCHLAG, EINER SPEZIFIKATION ODER EINEM MUSTER ERGEBEN.

Über Axis Communications

Axis ermöglicht eine intelligente und sichere Welt durch Lösungen zur Verbesserung der Sicherheit und Geschäftsperformance. Als Unternehmen für Netzwerktechnologie und Branchenführer bietet Axis Lösungen in den Bereichen Videosicherheit, Zutrittskontrolle sowie Intercoms und Audiosysteme. Sie werden verstärkt durch intelligente Analyseanwendungen und unterstützt durch gute Schulungen.

Axis beschäftigt rund 4.000 engagierte Mitarbeiter in über 50 Ländern und arbeitet weltweit mit Technologie- und Systemintegrationspartnern zusammen, um den Kunden Lösungen anbieten zu können. Axis wurde 1984 gegründet und der Hauptsitz befindet sich in Lund, Schweden