

# Das Potenzial der Audioerfassung

Audioerfassung- und -verarbeitung für Überwachung  
und Nachweis

Juni 2021

# Inhalt

1	Zusammenfassung	3
2	Einführung	4
	2.1 Erfassen ohne Aufzeichnung	4
3	Hindernissen begegnen	4
	3.1 Welche Gesetze und Vorschriften gelten?	5
	3.2 Möglichkeiten prüfen	5
4	Die Installation	6
5	Audiodaten vorbereiten	7
6	Topologie der Analysefunktionen	7
7	Anwendungsfälle und Beispiele	8
	7.1 Ereignisse mittels Audioanalyse erkennen	9
	7.2 Klang in Videos visualisieren	10
	7.3 Hören und handeln	11
	7.4 Aufzeichnen und speichern	12
	7.5 Mehr aus Ihrem Überwachungssystem machen	12
8	Überwachung und Erkennung	12
	8.1 Klangeigenschaften	13
	8.2 Signalverarbeitung	13
	8.3 Menschliches Hören	14
9	Haftungsausschluss	15
Anhang 1	Begriffe zur Audioqualität	16

# 1 Zusammenfassung

Ob als integrierte Funktion einer Videokamera oder über ein externes Mikrofon: Die Erfassung von Audiodaten erweitert eine Überwachungslösung um wichtige Zusatzfunktionen und damit Einsatzmöglichkeiten. Ein verantwortungsvoller, professioneller Einsatz von „Audio in“ kann den Wert und die Vorteile einer Sicherheitsinstallation entscheidend erhöhen. Häufig liefert sie z. B. das fehlende Teil in einem Ermittlungsverfahren, oder sie ermöglicht die Echtzeiterkennung kritischer Situationen, die das sofortige Eingreifen von Sicherheitskräften oder medizinischem Personal erfordern. Außerdem kann bereits das Wissen darum, dass eine Audioaufzeichnung stattfindet, eine abschreckende Wirkung haben.

Schon als autonomes System kann die Audioerfassung (häufig in Verbindung mit einer unmittelbar anschließenden Analysefunktion) mehrere Schutzziele bei der Vorbeugung, Verhinderung und Ermittlung von Straftaten erreichen.

In Verbindung mit einem Videosystem kann sie zudem die allermeisten vorhandenen Überwachungssysteme noch leistungsstärker machen. So kann das Sicherheitspersonal Situationen sehr viel besser einschätzen, wenn der von ihnen beobachtete Videostream zusätzlich mit einem Audiostream versehen ist.

So wie die *Videoanalyse* verschiedene Möglichkeiten zur automatischen Erkennung und Meldung von Vorfällen mithilfe visueller Verfahren bietet, überwacht die *Audioanalyse* die akustischen Informationen und kann bei außergewöhnlichen Vorfällen entsprechende Reaktionen auslösen.

Software zur Audioanalyse lässt sich so konfigurieren, dass ein automatischer Alarm oder eine andere Aktion folgt, sobald ein Mikrofon Schreien, Glasbruch oder Schüsse erkennt. Dies ermöglicht frühe Warnungen und damit schnelles Reagieren und Eingreifen.

Weiterhin lässt sich mithilfe von Audiodaten erkennen, ob ein unerwartetes Geräusch von links oder rechts kommt, und eine PTZ-Kamera automatisch in Richtung der Geräuschquelle ausrichten. In einem Krankenhaus oder Pflegeheim kann die Audioanalyse ungewöhnlich hohe Lautstärken erkennen, die häufig auf eine Notsituation hindeuten, und automatisch das Pflegepersonal benachrichtigen. Solche Anwendungen können auch von Analysefunktionen zur Klangvisualisierung profitieren, die die akustische Überwachung mehrerer Orte vereinfachen.

Bei der Erfassung und Aufzeichnung akustischer Informationen handelt es sich um zwei grundsätzlich unterschiedliche Vorgänge. In vielen Szenarien ist gar keine Aufzeichnung von Audiodaten erforderlich, um das jeweilige Schutzziel zu erreichen. Dies kann Bedenken in Sachen Datenschutz ausräumen und die Einhaltung der Vorschriften zum Umgang mit personenbezogenen Daten erleichtern. In aller Regel zeichnen Systeme zur Audioanalyse akustische Daten nicht permanent auf. Normalerweise durchsuchen sie die erfassten Audiodaten lediglich auf bestimmte Muster, Lautstärken oder Frequenzen. Wenn Analysefunktionen „on the edge“ (z. B. in der Kamera) integriert sind, brauchen die digitalisierten Audiodaten die Kamera gar nicht zu verlassen, sondern nur die Ergebnisse der durchgeführten Analyse (also die Metadaten oder Auslöser).

Axis spricht keine juristisch relevanten Empfehlungen aus. Je nach Region, Bundesstaat oder Land unterliegen Überwachungssysteme unterschiedlichen Gesetzen und Vorschriften. Grundsätzlich sind die Benutzer entsprechender Produkte (in der Regel also die Endkunden) dafür verantwortlich, dass jegliche Überwachung rechtskonform erfolgt. Ob Video- oder Audioüberwachung: In beiden Fällen ist im Vorfeld genau zu prüfen, welche rechtlichen Aspekte für eine solche Installation relevant sind.

Nachdem die Einhaltung des geltenden Rechts sichergestellt ist, sollten Anordnung und Konfiguration der Geräte sorgfältig überlegt werden. Dies kann einige Planung erfordern, stellt aber eine einfache Maßnahme dar, um die Nutzbarkeit der erfassten Audiodaten zu erhöhen.

## 2 Einführung

Akustische Informationen können bei der Vorbeugung, Verhinderung und Ahndung von Straftaten eine wesentliche Rolle spielen. Mithilfe von Analysesoftware lassen sich erfasste Audiodaten in Echtzeit verarbeiten, was eine höchst effiziente Erkennung von Aktivitäten, Verhalten oder Vorfällen ermöglicht.

Dieses White Paper zeigt anhand von typischen Nutzungsbeispielen, welches Potenzial Audiolösungen für Sicherheitsanwendungen haben. Es beschreibt unterschiedliche Arten der Audioanalyse und ihre Funktionsweise.

Das vorliegende White Paper enthält keine juristischen Handlungsempfehlungen, sondern stellt verschiedene technische Lösungen vor, deren Kenntnis für die Planung einer Installation hilfreich ist. Je nachdem, in welcher Form Sie die Audioanalyse nutzen möchten, sind bezüglich der vor Ort geltenden Gesetze und Empfehlungen bestimmte Aspekte zu beachten, um ihr enormes Potenzial ausschöpfen zu können.

In diesem White Paper wird nur die *Erfassung und ggf. Aufzeichnung* von Audiodaten behandelt, also die *Audio-Eingabe*. Daneben beinhalten viele Sicherheitslösungen die *Wiedergabe* von Audiodaten, also eine *Audio-Ausgabe*. Diese erfolgt meist in Form von Sprachdurchsagen oder Alarmen, die ein unberechtigtes Betreten oder Ladendiebstähle verhindern sollen. Weitere Informationen zur Wiedergabe von Audiodaten in Sicherheitslösungen finden Sie unter [www.axis.com/de-de/products/audio](http://www.axis.com/de-de/products/audio).

### 2.1 Erfassen ohne Aufzeichnung

Um Audioinhalte zu erfassen und zu verwenden, müssen diese nicht aufgezeichnet werden. 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 mit einem Mikrofon registriert, diese analogen Signale in digitale Signale umgewandelt (mithilfe eines A/D-Wandlers) 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. Nicht immer ist eine Aufzeichnung erforderlich. Dies gilt z. B. dann, wenn eine Person die erfassten Audioinhalte in Echtzeit abhört. Manchmal sprechen auch ganz spezifische Gründe *gegen* eine Aufzeichnung von Audiodaten. Je nachdem, ob Audiodaten nur erfasst oder auch aufgezeichnet werden, können unterschiedliche rechtliche Regelungen gelten.

In aller Regel zeichnen Systeme zur Audioanalyse akustische Daten nicht permanent auf. Damit sie richtig funktionieren, puffern sie den Ton vorübergehend. Viele Systeme lassen sich so einrichten, dass nur die Daten aufgezeichnet werden, die sich kurz vor und kurz nach einem Vorfall im Pufferspeicher befanden. So können Sicherheitskräfte den Vorfall überprüfen und die Daten bei Bedarf als Beweismittel sicherstellen.

## 3 Hindernissen begegnen

Viele Menschen bringen dem Einsatz von Mikrofonen in Verbindung mit einer Videoüberwachung eine gewisse Skepsis entgegen. Die entsprechenden Bedenken beziehen sich häufig darauf, dass neben dem Videomaterial auch gesprochener Klartext aufgezeichnet wird.

Um diesem Hindernis zu begegnen, sollte uns klar sein, dass das Feld der „Audioeingabe“ weit mehr als nur die Aufzeichnung von Audiodaten umfasst. In vielen Anwendungsfällen besteht gar kein Bedarf, akustische Informationen aufzuzeichnen.

Je nach Land und Region unterliegen Überwachungslösungen unterschiedlichen Rechtsvorschriften – prüfen Sie also vor der Erweiterung eines vorhandenen Systems um Audiofunktionen, was erlaubt ist.

### 3.1 Welche Gesetze und Vorschriften gelten?

Ob Video- oder Audioüberwachung: In beiden Fällen ist im Vorfeld genau zu prüfen, welche rechtlichen Aspekte für eine solche Installation relevant sind. Gegebenenfalls sind die vorgeschriebenen Anträge einzureichen und entsprechende Genehmigungen einzuholen. Wo dies vorgeschrieben ist, sind Schilder oder Hinweise zu verwenden.

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

#### 3.1.1 Beispiele aus den USA

In den US-amerikanischen Bundesstaaten gelten unterschiedliche Rechtsvorschriften.

In manchen Bundesstaaten reicht für Audioaufzeichnungen ein einseitiges Einverständnis aus. Die Überwachung ist also rechtmäßig, wenn nur eine am Dialog beteiligte Partei ihr Einverständnis erklärt.

In anderen Bundesstaaten müssen alle (bzw. beide) Parteien ihr Einverständnis mit der Aufzeichnung erklären, bevor diese stattfinden darf. Ausnahmen von einem solchen allseitigen Einverständnis können für öffentliche Plätze gelten, an denen Personen keine Privatsphäre voraussetzen können.

Auch kann die Bewertung der Rechtslage zum Einsatz von Analysefunktionen, bei denen keine Aufzeichnung von Audiodaten erfolgt, in manchen Regionen zu einem anderen Ergebnis führen. Es ist daher zu prüfen, welche Gesetze und Vorschriften im jeweiligen Bundesstaat gelten.

#### 3.1.2 Beispiele aus Europa

In Europa unterliegt die Audioüberwachung dem im jeweiligen Staat geltenden Recht. Es ist daher zu prüfen, welche Gesetze und Vorschriften im jeweiligen Land gelten.

Audioaufzeichnungen können personenbezogene Daten enthalten, die der DSGVO (Datenschutz-Grundverordnung) unterliegen. Die DSGVO verbietet Audioaufzeichnungen nicht grundsätzlich, legt aber Vorgaben fest, die bei der Erfassung oder Aufzeichnung von Audioinhalten zu beachten sind. Wenn Sie ein vorhandenes System zur Videoüberwachung um Audiofunktionen ergänzen, ist zu prüfen, ob die Rechtsgrundlage, auf der Sie personenbezogene Daten gemäß DSGVO verarbeiten, weiterhin gilt.

### 3.2 Möglichkeiten prüfen

Ein gängiges Missverständnis lautet, dass Überwachungslösungen mit Audiofunktionen grundsätzlich nicht erlaubt seien. Dieses Missverständnis ist so verbreitet, dass die Möglichkeit, ein Überwachungssystem durch Audiofunktionen noch leistungsstärker zu machen, gar nicht erst in Betracht gezogen wird.

Tatsächlich ließen sich aber viele Installationen vollkommen rechtskonform realisieren, z. B. wenn die betroffene Zielgruppe informiert oder das Einverständnis eingeholt wird. In jedem Fall ist zu prüfen, welche Gesetze und Vorschriften für die jeweilige Region und den konkreten Anwendungsfall gelten. Auch wenn das Szenario *Aufzeichnen und speichern* nicht erlaubt ist, lassen sich viele Sicherheitslösungen datenschutzkonform anpassen – z. B. im Rahmen der Szenarien *Hören und handeln*, *Hören und bezeugen* oder *Ereignisse mittels Audioanalyse erkennen*.

## 4 Die Installation

Die Position eines Mikrofons entscheidet darüber, wofür sich die von ihm erfassten Daten verwenden lassen. Vor der Installation von Geräten zur Audioerfassung empfehlen sich sorgfältige Überlegungen zu ihrer Position und Konfiguration. Dies kann einige Planung erfordern, stellt aber eine einfache Maßnahme dar, um die Nutzbarkeit der erfassten Audiodaten zu erhöhen.

Je vorteilhafter die Mikrofonposition und Akustik der ausgewählten Stelle sind, umso wahrscheinlicher werden die jeweiligen Überwachungsziele erreicht. Zunächst muss ein Mikrofon natürlich so positioniert sein, dass es relevante Geräusche gut erfasst. Typische Standorte sind mitten im Raum, in Verbindung mit einer Kamera oder in der Nähe möglicher relevanter Vorfälle. Ein Mikrofon sollte nicht in der Nähe einer Geräuschquelle montiert werden (z. B. Ventilator oder Maschine), die schwächere oder weiter entfernte Geräusche übertönen kann.



### *Typische Mikrofonstandorte*

- 1 *Am Ort relevanter Geschehnisse*
- 2 *In einer Kamera*
- 3 *Mitten im Raum*

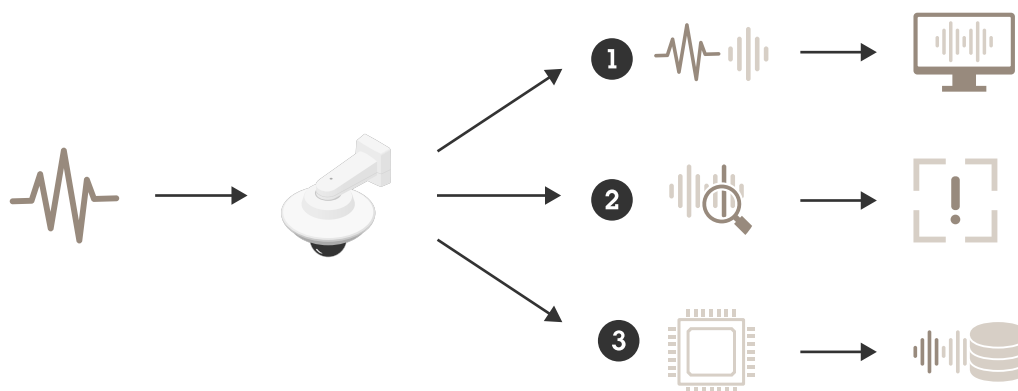
Die schallabsorbierenden Eigenschaften von Wänden, Decke und Boden oder bauliche Besonderheiten wie sehr lange Flure können verschiedene Nachhall- und Echoeffekte erzeugen, die sich je nach Standort stark auf das Schallfeld auswirken können. So hört sich ein Geräusch in einem stark gedämmten Raum (z. B. Konferenzraum mit spezieller Akustikausstattung) völlig anders an als in einer Kirche oder einem vollständig gekachelten Badezimmer. In akustisch anspruchsvollen Situationen kann die Mikrofonposition den entscheidenden Unterschied machen.

Neben der Installation und Konfiguration der Geräte (z. B. Verstärkungseinstellungen) ist auch die Einbindung der Audioausstattung in das Überwachungssystem von wesentlicher Bedeutung. Systeminstallateure und -integratoren können Empfehlungen für bestimmte Situationen geben.

Dabei können die Empfehlungen zur Audioanalyse manchmal von denen zur allgemeinen Audioaufzeichnung abweichen. Lesen Sie stets die mitgelieferte Dokumentation, um ggf. geltende Bedingungen zu erfüllen.

## 5 Audiodaten vorbereiten

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



- 1 *Umwandlung*
- 2 *Edge-Analyse in Echtzeit*
- 3 *Verarbeitung und Kodierung*

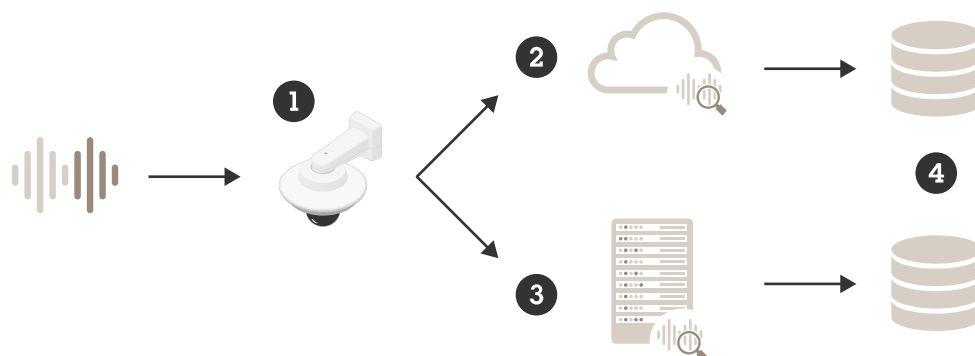
- **Umwandlung.** Die akustische Information wird abstrahiert und z. B. als spektrales Klangdiagramm visualisiert. Dieser Schritt lässt sich nicht rückgängig machen: Sie können den Originalton aus einem spektralen Klangdiagramm nicht wiederherstellen.
- **Edge-Analyse in Echtzeit.**  
Bei einer Analyse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

## 6 Topologie der 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.

- **Edge-Analyse.** Wenn die Analyse direkt im Gerät erfolgt, brauchen keine digitalen Audiodaten die Kamera zu verlassen. Bei einer reinen Erfassung von Audiodaten ohne Speicherung wird stattdessen nur das Ergebnis der durchgeführten Analyse gesendet, also die Metadaten oder Auslöser.
- **Server-Analyse.** Wenn die Analyse auf einem Server erfolgt, müssen digitale Audiodaten die Kamera verlassen. Wenn im Gerät (Kamera) eine Vorverarbeitung erfolgt, kann es sich bei diesen Daten um abstrahierte bzw. anonymisierte Daten handeln. Da ein Server in der Regel Teil eines geschlossenen Systems ist (Betreiber des Systems als Verantwortlicher), lässt sich der Datenschutz der übermittelten Audiodaten gewährleisten. Dennoch ist unbedingt sicherzustellen, dass die geltenden Gesetze und Vorschriften eingehalten werden.
- **Cloud-Analyse.** Digitale Audiodaten lassen sich auch an einen Cloud-Server senden. 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.



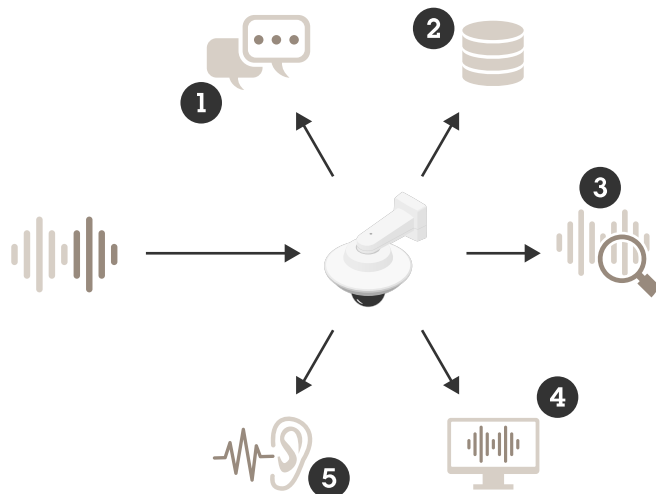
- 1 Edge
- 2 Cloud
- 3 Server
- 4 Speicherung

## 7 Anwendungsfälle und Beispiele

In vielen Videokameras ist die Funktion zur Audioaufzeichnung bereits integriert. Die verantwortliche und professionelle Aufzeichnung von Audiodaten ist in vielen Fällen von Vorteil und kann entscheidende Hinweise liefern. Häufig liefert sie z. B. das fehlende Teil in einem Ermittlungsverfahren, oder sie ermöglicht die Echtzeiterkennung kritischer Situationen, die das sofortige Eingreifen von Sicherheitskräften



oder medizinischem Personal erfordern. Außerdem kann bereits das Wissen darüber, dass eine Audioaufzeichnung stattfindet, eine abschreckende Wirkung haben.



*Typische Zwecke der Audioaufzeichnung:*

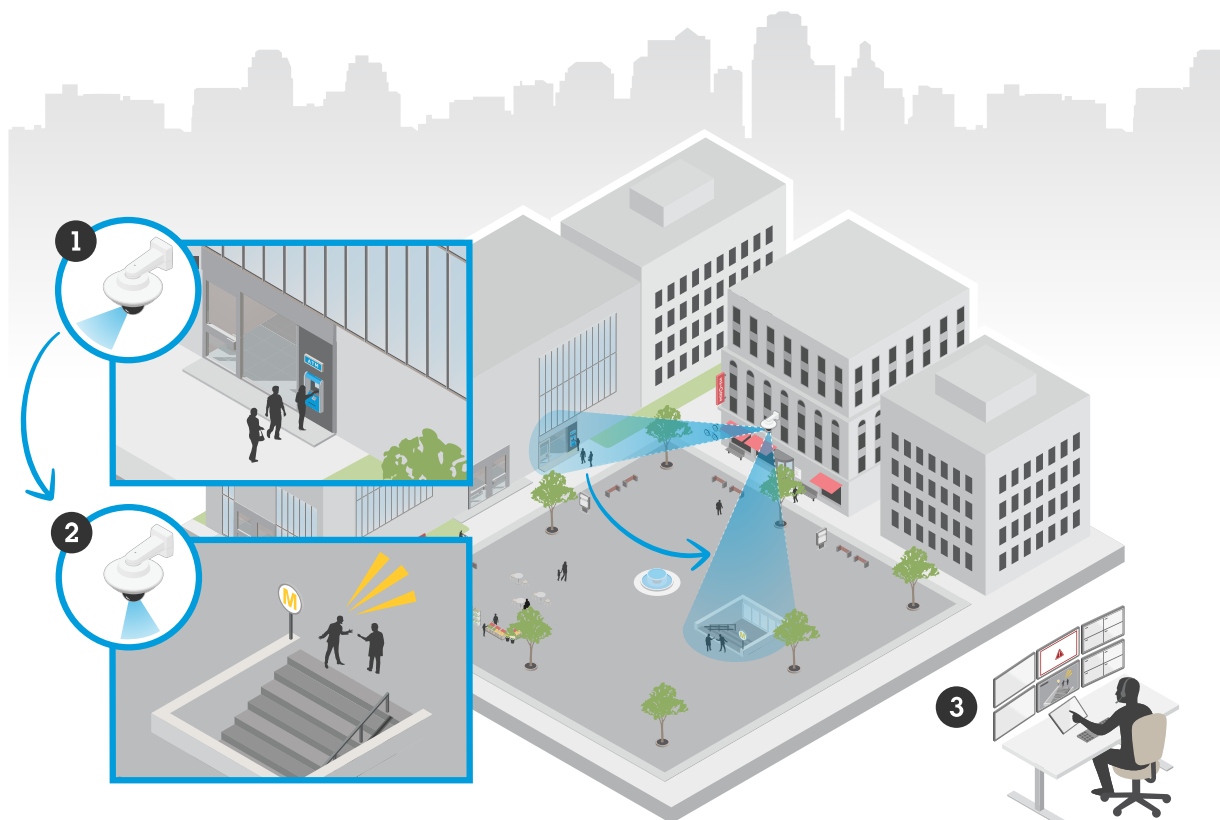
- 1 *Kommunizieren*
- 2 *Aufzeichnen*
- 3 *Analysieren*
- 4 *Visualisieren*
- 5 *Hören*

## 7.1 Ereignisse mittels Audioanalyse erkennen

Audioanalyse-Anwendungen sind Softwareprogramme, die erfasste Audiodaten verarbeiten, um bestimmte Informationen zu finden und herauszufiltern. Sie dienen dazu, Ereignisse wie Schüsse, Glasbruch oder aggressives Verhalten zu erkennen. Die Verarbeitung erfasster Audiodaten durch solche Anwendungen kann z. B. darin bestehen, die Ja-Nein-Frage „Ist eine Fensterscheibe zerbrochen?“ zu beantworten. Je nach Antwort benachrichtigt das System dann meist automatisch das Personal, indem es eine visuelle Warnmeldung anzeigt oder einen Alarm auslöst. Dies ermöglicht frühe Warnungen und damit schnelles Reagieren und Eingreifen.

### 7.1.1 Kameraausrichtung anpassen

Ein weiteres Beispiel für die Nutzung der Audioanalyse ist die angepasste Ausrichtung einer PTZ-Kamera. Hierbei wird durch die Verknüpfung von Audio- und Videofunktion erkannt, woher ein Geräusch kommt, und die Kamera automatisch in Richtung der Geräuschquelle geschwenkt.



- 1 Eine PTZ-Kamera überwacht einen Geldautomaten.
- 2 Das Kameramikrofon erfasst ein lautes, plötzliches Geräusch und schwenkt die Kamera sofort in dessen Richtung.
- 3 Das Personal am Überwachungsmonitor erhält einen Alarm und überprüft den Vorfall.

### 7.2 Klang in Videos visualisieren

Der in einem Video erfasste Klang lässt sich als spektrales Klangdiagramm auf einem Monitor visualisieren. Bei Überschreitung eines definierten Grenzwerts gibt das Diagramm einen Alarm aus.

Eine solche Visualisierung kann sinnvoll sein, wenn die gleichzeitige Audioüberwachung mehrerer Orte gewünscht ist (z. B. mehrere Patientenzimmer in einem Krankenhaus). Da das gleichzeitige Abhören mehrerer Audioquellen (Ton aus mehreren Zimmern) schwierig sein kann, ist die visuelle Darstellung auf

einem Monitor oft die bessere Lösung. Wenn ein Videostream aus den Zimmern gesendet wird, lässt sich das Klangdiagramm als Overlay über das Videobild legen.



*Klangdiagramm als Overlay eines Videostreams im Krankenhaus.*

## 7.3 Hören und handeln

Die vielleicht grundlegendste und intuitivste Nutzung der Audioüberwachung besteht darin, dass das Bedienpersonal bestimmte Ereignisse mitbekommt und direkt reagieren kann. Ein typisches Beispiel ist das Mithören eines verdächtigen Gesprächs und die Entsendung von Sicherheitskräften zu den beteiligten Personen. In einem Krankenhaus oder Pflegeheim kann das Bedienpersonal hören, wenn Patienten Hilfe brauchen und Pflegekräfte in das betreffende Zimmer schicken. Mithilfe von Audiodaten lässt sich auch feststellen, ob ein „merkwürdiges“ Geräusch von links oder rechts kommt, woraufhin eine PTZ-Kamera in Richtung der Geräuschquelle ausgerichtet werden kann.

In solchen Szenarien haben eine oder mehrere Personen in einem Kontrollraum oder über ein Sicherheitsprogramm auf einem mobilen Endgerät Zugriff auf das Audiosystem. Der Mensch ist Teil des Systems: Das Ohr erfasst das Geräusch, und das Gehirn entscheidet, worauf es in der jeweiligen Situation ankommt. In Verbindung mit einer Videoüberwachung liefern Audiodaten eine zusätzliche Informationsebene, die dann ebenfalls in die Entscheidungsfindung einfließt. Manchmal sind Audiodaten auch die einzige Ebene, wenn z. B. die Geräuschquelle außerhalb des Kamerasichtfelds liegt oder schwierige Lichtverhältnisse herrschen.

### 7.3.1 Hören und bezeugen

Die Ergebnisse einer Audioüberwachung lassen sich auch als Zeugenaussage verwenden, wenn ein Ereignis unmittelbar mitgehört wurde. Dieser Fall unterscheidet sich vom Szenario *Hören und handeln*, da die Erfassung hier keiner Entscheidungsfindung dient, wobei sich die beiden Anwendungsfälle häufig

überlappen. Wenn z. B. der/die AnwenderIn einen Streit mit belastenden Aussagen mithört, kann er/sie nicht nur Sicherheitskräfte an den betreffenden Ort schicken, sondern später auch die mitgehörten Inhalte belegen.

## 7.4 Aufzeichnen und speichern

Gegebenenfalls kann die Erfassung und Aufzeichnung von Audiodaten einen wichtigen zusätzlichen Beweis liefern. Nehmen wir z. B. belastende Aussagen oder abgegebene Schüsse. Aufgezeichnete Audiodaten können beweisen, wer was gesagt hat, wie viele Schüsse abgefeuert wurden oder andere forensisch relevante Sachverhalte belegen.

Bei der Aufzeichnung von ggf. forensisch relevanten Audiodaten ist zu beachten, dass die Originaldaten erhalten bleiben und nicht verarbeitet werden (was in anderen Zusammenhängen erforderlich oder förderlich sein kann). Bei forensisch relevanten Aufzeichnungen kann jegliche Verarbeitung als Manipulation von Beweisen gelten. Algorithmen für die optimierte Wiedergabe gesprochener Inhalte können die forensische Nutzbarkeit zwar erhöhen. Solche Algorithmen sind aber anschließend und auf eine Kopie des aufgezeichneten Originalmaterials anzuwenden. Die Aufzeichnung sollte so wenig Verarbeitung wie möglich beinhalten, um die Optionen einer späteren Verwendung nicht unnötig einzuschränken.

## 7.5 Mehr aus Ihrem Überwachungssystem machen

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. Manchmal werden nicht-visuelle Sensoren als autonome Lösung verwendet, weil eine Videoüberwachung nicht in Frage kommt. In vielen Fällen ergänzen nicht-visuelle Sensoren aber das Kamerasystem, indem sie andere Arten von Informationen beisteuern.

Die allermeisten Nutzungsszenarien profitieren davon, wenn eine Überwachungslösung um Geräuschsensoren (Mikrofone) erweitert wird. Die zusätzliche Ausstattung eines Systems mit Audiofunktionen ermöglicht die Interaktion mehrerer Sensoren durch Analysefunktionen oder menschliches Eingreifen.

*Hören und handeln* ist ein einfaches Beispiel dafür, dass das Personal Situationen sehr viel besser einschätzen kann, wenn neben Bildern auch Töne ankommen. Wer Menschen nur sieht, erkennt aggressives Verhalten nicht immer – sie auch zu hören, macht es deutlich einfacher.

Ein weiteres Anwendungsbeispiel ist die Bewegungserkennung mittels Videoanalyse. Wenn die Analyse der Kamera z. B. bei schwachem Licht nicht ihr volles Potenzial entfaltet, kann Audioanalyse die Erkennungsgenauigkeit erhöhen.

# 8 Überwachung und Erkennung

Audiodaten liefern verschiedene Arten von Informationen, die sich zur Überwachung und Analyse nutzen lassen. Mithilfe verschiedener Verarbeitungs- und Bewertungsschritte werden diese Informationen extrahiert und verfeinert, wodurch sie sich besser nutzen und in das Gesamtsystem einbinden lassen.

## 8.1 Klangeigenschaften

Im Zusammenhang mit einer Überwachung können Klangeigenschaften wie Lautstärke oder Tonhöhe wichtige Informationen liefern. Wie lange ein Ton hörbar ist, ob er sich bewegt, wie weit er entfernt ist – all diese Puzzleteile entscheiden darüber, wie wir den Ton deuten. Hard- und Software zur Audioüberwachung und -erkennung verarbeiten die gleiche Art von Informationen: Sie „hören“ komplexe Kombinationen bestimmter Merkmale heraus, wie z. B. den Schallpegel oder die Energie unterschiedlicher Frequenzen im zeitlichen Verlauf.

- **Informationen zum Raum.** Diese betreffen die physische Welt, die uns umgibt, wie z. B. Ort, Richtung oder Abstand. Mithilfe räumlicher Informationen lässt sich die Audioerfassung anders ausrichten oder verstärken, um eine höhere Aufzeichnungsqualität zu erreichen. Analysefunktionen können mit ihrer Hilfe bestimmen, aus welcher Richtung ein Geräusch kommt oder wie weit seine Quelle entfernt ist.
- **Informationen zur Zeit.** Informationen zur Zeit sind sowohl im relativen (Änderung im zeitlichen Verlauf) als auch im absoluten Sinn (wann ist was passiert?) von Bedeutung und werden häufig zu den Informationen anderer Sensoren in Bezug gesetzt (z. B. aus der Videoerfassung). Zeitliche Informationen zeigen, was sich wann und über welchen Zeitraum ereignet hat und spielen damit eine wichtige Rolle für die Analyse von Verhalten.
- **Informationen zum Spektrum.** Diese betreffen Frequenzen und geben z. B. die Höhe eines Tons oder bei komplexeren Klängen die vorkommenden Frequenzkombinationen an. In der Audioüberwachung eingesetzte Mikrofone haben einen flachen Frequenzgang, sind also so konstruiert, dass sie möglichst alle Frequenzen innerhalb des hörbaren Bereichs (20 Hz bis 20 kHz) gleichermaßen erfassen. Dies unterscheidet sich vom menschlichen Hören: Wir erkennen die Frequenzen, die typischerweise beim Sprechen auftreten, leichter als andere Frequenzen.
- **Informationen zur Amplitude.** Diese betreffen die Intensität oder Lautstärke eines Geräuschs. Informationen zur Amplitude können Informationen zum Spektrum ergänzen und in Verbindung mit diesen darstellen, wie das erfasste Geräusch strukturiert ist.

## 8.2 Signalverarbeitung

Im Bereich der Audioüberwachung werden Signale in der Regel verarbeitet, um die Übertragung, Speichereffizienz oder subjektive Qualität zu verbessern oder relevante Bestandteile hervorzuheben oder zu erkennen. Dies geschieht mithilfe von Software-Algorithmen, die akustische Daten auf unterschiedliche Weise ändern oder analysieren.

### 8.2.1 Signale bearbeiten

Mit Algorithmen lassen sich Signale für einen bestimmten Zweck bearbeiten, wie z. B.:

- Verbesserung des Signals, um z. B. mittels automatischer Verstärkungsregelung die Verständlichkeit zu erhöhen.
- Umwandlung des Signals, z. B. um mithilfe eines Equalizers den relativen Frequenzgehalt zu ändern.
- Eingrenzung des Signals, indem bestimmte Frequenzen oder Amplituden entfernt werden. Dies kann z. B. eingesetzt werden, um mittels Komprimierung die Datenmenge zu begrenzen oder mittels Stimmenverschlüsselung den Datenschutz zu wahren.

## 8.2.2 Signale analysieren

Funktionen zur Audioanalyse werten die relevanten Eigenschaften erfasster (aber in der Regel nicht aufgezeichneter) Audiodaten aus und erzeugen daraus Ergebnisse, die ihrerseits keine Audiodaten darstellen. Im Wesentlichen wandeln solche Anwendungen die Audiodaten in ein anderes Format um, das besser geeignet ist, um bestimmte Aktionen einzuleiten. Manche Analyseanwendungen wurden z. B. speziell dafür entwickelt, aggressives Verhalten, Schüsse, Glasbruch oder Fahrzeugalarme zu erkennen.

Algorithmen des maschinellen Lernens können aus großen Datenmengen Vorhersagen ableiten, ohne speziell hierfür programmiert zu werden. Ein Beispiel hierfür ist ein Algorithmus, der das Geräusch einer sich schließenden Tür zuverlässig erkennt, nachdem er mit Tausenden solcher Geräusche gefüttert wurde.

## 8.3 Menschliches Hören

Das menschliche Ohr gehört zu den leistungsstärksten Instrumenten überhaupt, wenn es um die Erkennung und Analyse von Geräuschen geht. Auch in sehr lauten Umgebungen sind das menschliche Ohr und Gehirn noch in der Lage, gesprochene Sprache zu erkennen und zu deuten – anders als die meisten Algorithmen.

Mithilfe unserer Ohren können wir auch räumliche Informationen ableiten und z. B. eine Geräuschquelle lokalisieren oder erkennen, ob sich diese bewegt. Da wir zwei Ohren haben, hören wir, ob ein Geräusch von links oder rechts kommt oder ob sich die Geräuschquelle irgendwo zwischen diesen beiden Polen befindet. Außerdem erfassen Ohren und Gehirn, ob uns ein Geräusch von oben oder unten sowie von vorn oder hinten erreicht. Mehrere „Filterstufen“ im Gehirn verarbeiten zeitliche Abweichungen zwischen den Informationen, die wir über die Ohren aufnehmen, und lassen uns schon anhand von Sekundenbruchteilen bestimmte Arten von Vorfällen erkennen. Wir sind also Spezialisten in Sachen Audioanalyse – insbesondere bei menschlichen Stimmen, aber auch in Bezug auf Geräusche, die mit erlernten Gefahren einhergehen.

Unter den richtigen Bedingungen (gute Tonqualität, Stereoklang, keine zu große Verzögerung) können Menschen die zur Erfassung eingesetzte Hard- oder Software als leistungsstarkes „Analysetool“ ergänzen. Durch eine Audioüberwachung mit gerade einmal zwei Mikrofonen kann das Bedienpersonal räumliche Informationen zu einer Situation ableiten und z. B. bestimmen, woher ein Geräusch kommt und ob dieses sich bewegt.

## 9 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 Garantie 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 DIE WIRKUNG DIESES DOKUMENTS ÜBERNIMMT DER NUTZER DES DOKUMENTS, UND AXIS SCHLIESST IM GESETZLICH MAXIMAL ERLAUBTEN RAHMEN SÄMTLICHE GARANTIEEN AUS, OB GESETZLICH, AUSDRÜCKLICH ODER STILLSCHWEIGEND, UNTER ANDEREM SÄMTLICHE STILLSCHWEIGENDE GARANTIEEN DER VERMARKTBARKEIT, EIGNUNG FÜR EINEN BESTIMMTEN ZWECK, RECHTSANSPRUCH UND NICHTVERLETZUNG SOWIE PRODUKTHAFTUNG, ODER GARANTIEEN, DIE SICH IN BEZUG AUF DIESES DOKUMENT AUS EINEM VORSCHLAG, EINER SPEZIFIKATION ODER EINEM MUSTER ERGEBEN.

# Anhang 1 Begriffe zur Audioqualität

## Digitalaudio

Der Begriff Digitalaudio bezeichnet analoge Audiodaten (häufig mit einem Mikrofon erfasste akustische Signale), die in digitalisierter Form aufgezeichnet werden. Normalerweise wird beim Digitalaudio die Schallwelle des Audiosignals numerisch als kontinuierliche Abtastfolge kodiert. Die Genauigkeit hängt von der Anzahl signifikanter Stellen ab, die der Encoder erfasst. Für Audio-CDs werden z. B. 44.100 Samples (Abtastvorgänge) pro Sekunde mit einer Abtasttiefe von je 16 Bit durchgeführt.

## Rauschen:

Der Begriff Rauschen bezeichnet unerwünschte (und manchmal unvermeidliche) Geräusche, die das lautlose Ende des Lautstärkebereichs definieren oder begrenzen. Es wird von allen Teilen des Audiosystems erzeugt: von der Quelle des aufgezeichneten Tons (z. B. Ventilator im Raum) über das Mikrofon (z. B. Eigenrauschen, Vibrationen, Wind) und die Kabel (z. B. Interferenzen, Überlagerungen) bis hin zum Gerät für die Audioerfassung (z. B. Eigenrauschen, digitales Abtausrauschen). Die Summe all dieser Geräusche wird in der Regel als Grundrauschen bezeichnet.

Gemessen wird das Rauschen meist als SRV (Signal-Rausch-Verhältnis), das den gesamten Bereich von einem bestimmten Pegel (manchmal die höchste Lautstärke, die das System verarbeiten kann) bis zum Grundrauschen umfasst.

*Das Pendant bei der Videoverarbeitung ist das Bildrauschen („Schnee“) – ein (meist) zufälliges Muster aus statischen Pixeln, die das Sichtbare in einem dunklen Bild begrenzen (entsprechend der Begrenzung des Hörbaren bei Lautlos-Signalen).*

## Verzerrung:

Als Verzerrung werden alle unerwünschten Abweichungen von einem Signal bezeichnet, die dessen ursprünglichen, also wahren Gehalt ändern. Wenn von Verzerrung gesprochen wird, ist das oben erläuterte Rauschen in der Regel nicht enthalten. Eine Verzerrung mindert die subjektive Qualität (wobei häufig eine als angenehm empfundene Verzerrung vorliegt) und mindert durch die weniger gute Hörbarkeit, insbesondere bei der Inhaltsanalyse, den objektiven Informationsgehalt und die Leistungsstärke der Analysefunktion.

Angegeben wird die Verzerrung meist mit den Parametern THD (harmonische Gesamtverzerrung) und IMD (intermodulare Verzerrung).

*Bei Videos mindert eine Verzerrung in Form von Artefakten (z. B. chromatische Aberration, Vignettierung oder Unschärfe) Bildqualität und Detailgenauigkeit.*

## Abtastrate und Frequenzantwort:

In einem digitalen System werden Audiodaten in einer bestimmten Häufigkeit pro Sekunde abgetastet. Hierbei handelt es sich um die Abtastrate (meist zwischen 8.000 und 48.000 Mal pro Sekunde oder Hz). Gemäß der Signaltheorie (und insbesondere dem Nyquist-Shannon-Abtasttheorem) muss die Abtastrate mindestens doppelt so hoch sein wie die maximal erforderliche oder gewünschten Frequenz des analogen Signals, um einen Ton mit angemessener Genauigkeit zu erfassen.

Das menschliche Ohr ist je nach Alter und anderen Faktoren normalerweise in der Lage, Frequenzen zwischen 20 Hz und etwa 15–20 kHz wahrzunehmen. Vereinfacht ausgedrückt bilden tiefe Frequenzen ab einigen Hundert Hz häufig die Basis bestimmter Klänge (z. B. die Klangfarbe einer Stimme), während höhere Frequenzen über einigen Tausend Hz mehr Details liefern.



*Der Frequenzbereich bei Klängen entspricht der Auflösung und Bildrate bei Videos: Je tiefer diese eingestellt ist, desto weniger Details sind zu sehen.*

### **Bittiefe:**

Bei jedem Abtasten eines Audiosignals wird ein analoger Wert erfasst und in einen digitalen umgewandelt. Da die digitale Ebene das Konzept der Unendlichkeit nicht vorsieht, ist die Detailmenge auf eine festgelegte Bittiefe beschränkt. Jedes Bit steht für einen von zwei Werten (0 oder 1, niedrig oder hoch usw.), der in Verbindung mit einem definierten Amplitudenbereich (z. B. Spannungswert oder Schalldruckpegel), Teilstücke dieses Bereichs erstellt. Zwei Bits erzeugen vier Teilstücke, drei Bits acht usw. Vereinfacht ausgedrückt wird ein mit drei Bits abgetastetes 1-Volt-Signal in 1/8-Volt-Schritte aufgeteilt und dargestellt.

Zumindest für das menschliche Ohr reichen 16 Bits (bzw. 65.536 Schritte) in der Regel aus, um eine ausreichende Audioqualität zu erzielen. Diese Bittiefe wird auch für Audio-CDs verwendet. Für Analysezwecke oder besonders anspruchsvolle Anwendungen kommen eher 24 Bit in Frage.

*Auf den Videobereich übertragen entspricht die Bittiefe dem Kontrast, also dem Luminanz- oder Chrominanzbereich, den ein Pixel darstellen kann.*

# 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