

WHITE PAPER

# Audio analytics for security and safety

February 2024

## Summary

Audio analytics for security and safety can detect sound patterns and highlight unexpected sounds in live audio. The analytics can, for instance, identify verbal aggression to prevent escalation and assault, detect glass breakage to prevent break-ins, or provide early warnings of people in distress.

Using different types of sensors, such as video and audio (a camera and a microphone) increases confidence in detection results and contributes to more actionable insights.

AXIS Audio Analytics is a software platform for edge-based analytics that are integrated in Axis device software. These analytics capture and detect sounds without saving the original audio stream. This is a way to safeguard privacy, and it works because the analytics are edge-based and provide audio metadata.

# Table of Contents

1	Introduction	4	
2	Technology	4	
	2.1	Detecting sound events	4
	2.2	Capturing and processing audio	4
	2.3	Edge-based or server-based analytics	5
	2.4	Metadata	6
3	AXIS Audio Analytics	7	
	3.1	Privacy	7
	3.2	Use cases	8
4	Legal restrictions	12	
5	Disclaimer	12	

# 1 Introduction

Audio analytics for security and safety can detect sound patterns and highlight unexpected sounds in live audio. The analytics can, for instance, identify sounds associated with aggression, detect glass breakage, or provide early warnings of people in distress.

Audio analytics combined with video surveillance can alert operators to ongoing potential incidents, and guide them to the relevant camera views. This can enable early detection, swift intervention, and in many cases, prevention of further escalation.

This white paper presents how audio analytics can be used for security and safety. We discuss the technology of capturing and processing audio, with a focus on real-time edge analytics such as AI-based sound classification directly in the camera or microphone. We also show how audio analytics at the edge enables several options for maintaining privacy through the use of audio metadata.

This paper does not provide legal advice. Before installing any surveillance system you need to investigate which laws and regulations apply in your region and to your use case. It is up to the owner of the system to ensure that it complies with regional laws, regulations, and recommendations.

## 2 Technology

### 2.1 Detecting sound events

A sound event is an audio segment that humans typically would identify as a distinctive concept, for instance the concept of screaming or glass breaking. These kinds of conceptual sounds can be detected and labeled in a similar manner as object classes are detected and labeled in video analytics.

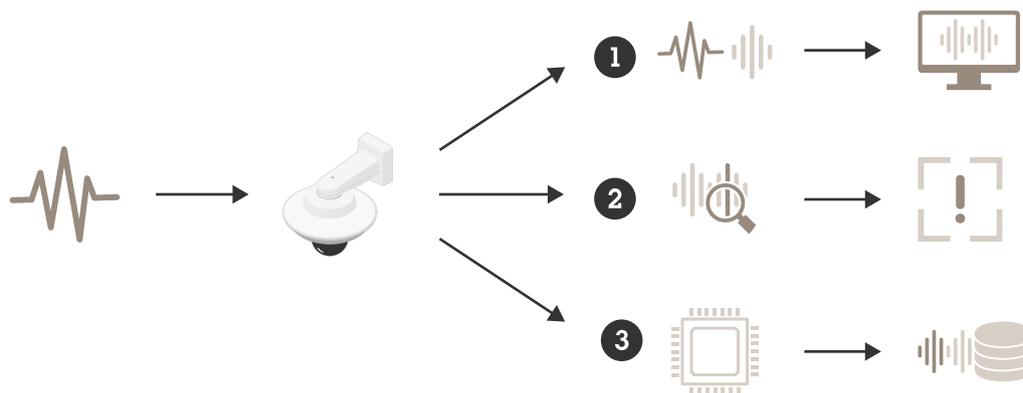
Analytics that are trained to recognize sound patterns typically listen for a combination of characteristics ranging from decibel level to the energy in different frequencies over time. When a specific sound pattern is detected, the system can send an automatic notification to staff through a visual alert or by triggering an alarm.

If AI-based algorithms are used, they can be trained from large amounts of data. For example, an algorithm can reliably detect human screams after having been trained with thousands of such sounds.

### 2.2 Capturing and processing audio

Audio analytics use captured audio data and analyze the relevant sound characteristics to generate non-audio output. Capturing audio basically means digitizing it and making it available for use in software. This is done through picking up the sound vibrations in the air using a microphone, converting these analog signals to digital signals and passing them on to a processing unit. If the captured audio is not placed on any permanent medium such as a flash memory or a hard drive, it is not recorded. In Axis cameras, audio streaming and recording are off by default.

After the initial audio capturing, the captured information is prepared for the next processing steps. The different preparations can be made in parallel or exclusive.



#### Audio preparation

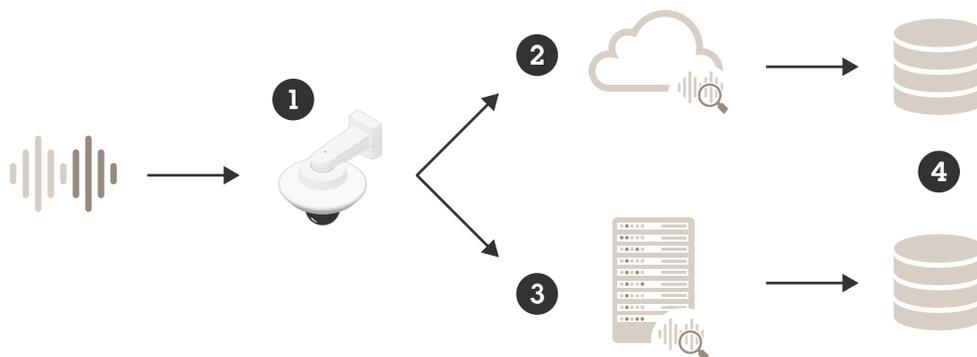
- 1 Transformation
- 2 Realtime edge analytics
- 3 Processing and encoding for streaming or storage — if you use an Axis camera, audio is neither streamed nor stored unless you actively turn on audio streaming.

- **Transformation.** The sound is made abstract and converted into, for example, visual information as a graph showing the sound spectrum. This process cannot be reversed: you cannot retrieve the original sound from the spectrum graph.
- **Realtime edge analytics.**  
  
A **sound classifier** can be used if the sound is processed on the edge. The outcome will be metadata describing the sound's characteristics. The original sound cannot be recreated from its metadata.  
  
A **sound detector** can be used to recognize patterns, levels, or frequency, and provide status information. Again, the original sound is not restorable.
- **Processing and encoding.** For cases where the original audio will be used (not transformed or analyzed), some processing and encoding is normally performed to prepare the audio data for the intended use cases. These use cases can involve storing audio data on the edge, streaming to external clients for additional processing (on server or cloud), or external storage. With an Axis camera you first need to actively turn on audio streaming, which is off by default for privacy reasons (audio privacy control).

## 2.3 Edge-based or server-based analytics

The location of the analytics engine in the system is important for many reasons. Especially for managing privacy concerns and complying with regulations regarding personal data, it matters where the software algorithm analyzes the audio data. There are situations where audio data cannot be sent over the network and it is critical that captured (but not stored) audio data can be analyzed locally. If very

computation-intensive algorithms are needed, such that cannot run on the edge, it might be required to send digital audio data to the cloud or a server.



*Possible locations for running the analytics*

- 1 Edge
- 2 Cloud
- 3 Server
- 4 Storage – only if audio streaming is activated the original audio can be stored.

- **Edge analytics.** When analytics run on the edge, no audio stream needs to leave the camera. Only the output from the performed analytics, that is, metadata or triggers, does. The analytics within AXIS Audio Analytics are edge-based.
- **Server analytics.** When analytics run on a server, audio data needs to be transmitted from the camera. If preprocessed in the camera, this can be limited to abstracted or depersonalized metadata. A server is normally a part of a closed system (a system owner is in control), so it is possible to manage privacy concerns of transmitted audio. Nonetheless, it must be ensured that applicable rules and regulations are followed.
- **Cloud analytics.** Audio data can also be transmitted to a server in a cloud context. As in the server analytics case, the audio information can be preprocessed into metadata. Cloud usage is often decentralized, so it is even more important to address privacy issues and ensure that regulations are complied with.

## 2.4 Metadata

Audio analytics generate a constant metadata stream of audio level data. The analytics also generate metadata based on events detected by classifiers and adaptive audio detection. Analytics that run at the edge analyze the audio information within the camera. They do not have to transmit the actual audio stream anywhere – it is possible to transmit just the metadata, which provides insight into what is happening in the scene.

The metadata stream enables visual inspection of the audio envelope, for example represented in a dashboard where audio events and video events can be aligned and viewed together. This way, metadata makes it possible to search efficiently and quickly for specific events and unexpected sounds. This can save investigators hours and hours of time in searches on vast amounts of footage.

## 3 AXIS Audio Analytics

AXIS Audio Analytics is a software platform that will continuously introduce new and upgraded analytics and features. The analytics are integrated in the device software AXIS OS, and are included for free with cameras and other Axis devices.

Analytics within the platform are edge-based, with the algorithms running directly on the camera. This provides optimal scalability, low data traffic, and privacy. Only the output from the analytics (metadata or triggers) is saved. No audio is recorded or streamed from the camera, and the original sounds cannot be recreated from the metadata.

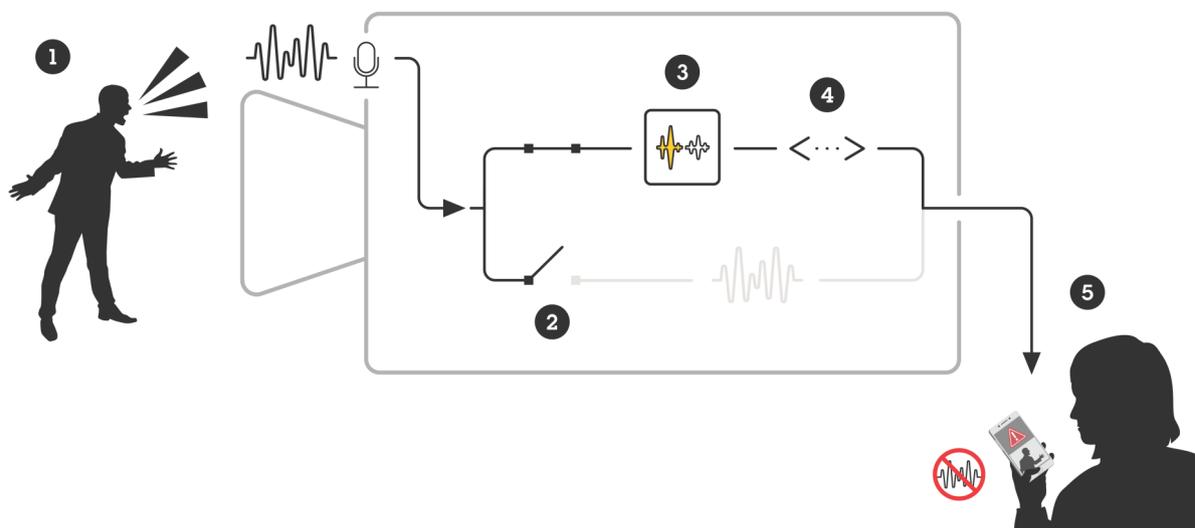
- **Audio classification (scream and shout).** This is an AI-based sound classifier that detects and analyzes specific sounds such as screams and shouts. More audio classifications, such as glass break, will be added in the future. The outcome of classification analytics is metadata that describes the sound's characteristics.
- **Adaptive audio detection.** This is a sound detector that creates an event when there is a sudden change in the audio level. It detects sound peaks of any kind, with the advantage of adapting to the ambient noise even when the noise level varies.

### 3.1 Privacy

Audio analytics in general do not record incoming audio, nor transmit it from the camera. They just process sounds to enable the search for specific events, patterns, or sound levels in a receiving system, such as a dashboard for further investigation or a video management software for alerting operators. No audio data can be reconstructed, and no private conversation can be recorded. This is because these analytics are edge-based and provide audio metadata.

AXIS Audio Analytics neither record nor stream audio, but transmit only metadata. For privacy reasons, all audio streaming is also off by default in Axis devices (audio privacy control), meaning that audio is neither streamed, recorded, nor possible to recreate. You can turn audio streaming on if you need it, but you will be notified when something relevant happens even if audio streaming is off.

The illustration shows an overview of how AXIS Audio Analytics works together with audio privacy control in picking up sounds and using metadata to create an alert.



*Audio privacy control and AXIS Audio Analytics in Axis cameras, from a scream (1) to an alert (5).*

- 1 Sounds are picked up by the microphone.*
- 2 By default, audio streaming is off.*
- 3 The audio classifier of AXIS Audio Analytics detects screaming or shouting in the incoming audio.*
- 4 Metadata, including an event notification, is generated by the audio classifier.*
- 5 Stakeholders receive an alert based on the event notification and metadata. They can verify by checking the video stream. No audio stream is available.*

## 3.2 Use cases

Although AI-based analytics have a high potential for filtering out irrelevant noise, they may provide false alerts when there is a lot of background noise. Rain against windows, thunder, sirens, music, or busy scenes with people talking can trigger false alerts. Thus, typical use cases include quiet areas such as banks and reception desks but also many types of indoor spaces after business hours, such as stores, restaurants, staircases, or offices.

### 3.2.1 Sound detection with alerts

In a bank or reception desk, sound classification analytics can monitor the area and detect sounds associated with aggression or glass breakage. Upon detection, the system's event data and audio metadata

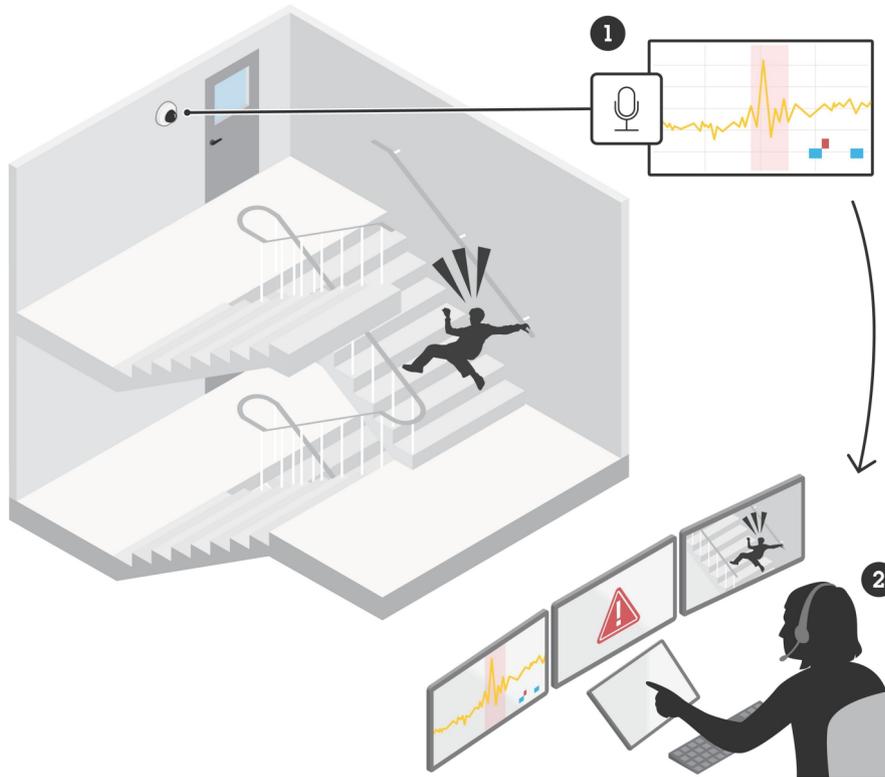
sends automatic notifications to staff through a visual alert or by triggering an alarm. This provides early warning that enables quick responses and intervention.



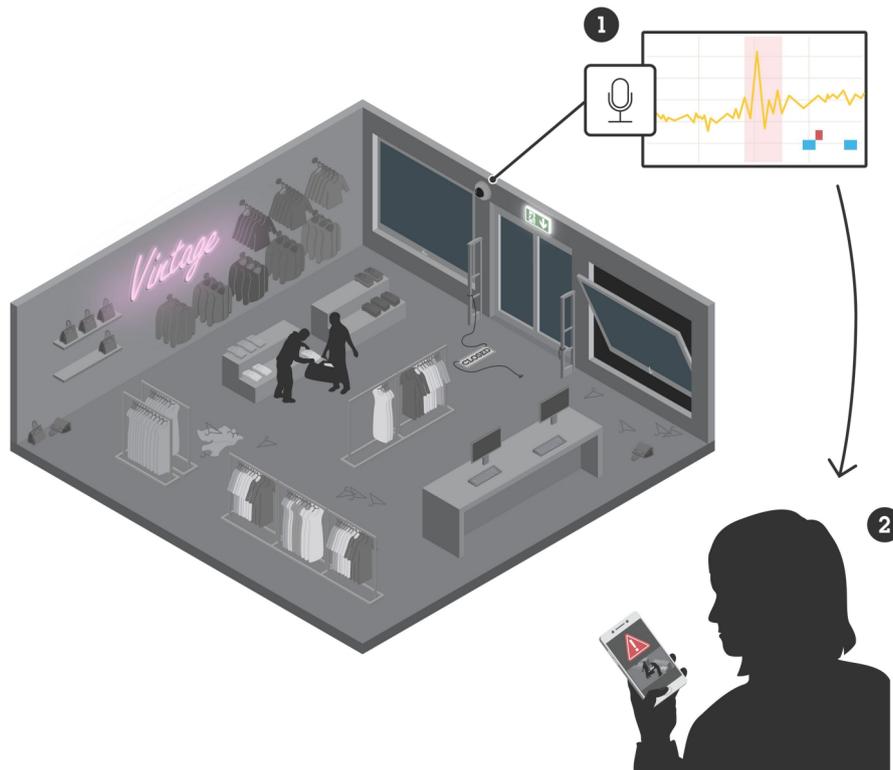
- 1 A camera with sound classification analytics detects screaming or shouting at the reception desk.
- 2 An operator receives an alert and can verify by checking the video stream before taking further action.

Adaptive audio detection analytics can be used to detect unexpected sounds outside business hours. The analytics analyze ambient sounds and respond when it detects voices, breaking windows, or other

sudden, short-lived noises. When events are detected, the analytics forward the metadata to notify operators accordingly.



- 1 A camera with sound classification analytics detects unexpected sounds after business hours.
- 2 An operator receives an alert and can verify by checking the video stream before taking further action.



- 1 A camera with sound classification analytics detects sounds in a shop after business hours.
- 2 The shop owner receives an alert and can verify by checking the video stream before taking further action.

### 3.2.2 Combining sensors to get more out of your surveillance system

Surveillance systems often incorporate several types of sensors. The camera's image sensor is one, of course, registering the visual aspect of a scene. Non-visual sensors are also commonly used, such as motion detectors based on radar technology or infrared radiation emissions. Non-visual sensors complement the camera installation by adding other types of information input.

By also employing audio sensors (microphones) in a surveillance installation, the great majority of all possible use cases are reinforced. Adding audio capability and audio analytics to a non-audio system enables multisensor interaction. If you are using video analytics, adding audio analytics can increase the detection confidence. This is the case especially if the video analytics is challenged by, for example, low-light conditions.

You can set up the system, for example in video management software, so that it triggers actions only when both video analytics and audio analytics have reacted. For example, the audio analytics detects a scream, and the video analytics detects an individual in the camera's field of view. In some environments, this combination provides just the right level of security.

### **3.2.3 Providing input to dashboards**

The audio metadata can be input to analytics dashboards, such as VSaaS (video surveillance as a service), IoT (internet of things), or business intelligence platforms, that gather and present the metadata visually. These analyze real-time and historical trends to generate instant overview and actionable insights. Statistical analyses based on customer flow or customer experience enable data-driven decision making to improve operations.

With dashboards you can see results without ever listening to the actual audio, or ever being able to retrieve the original sound. Instead you can get actionable insights from, for example, counting events, and there should be no doubts regarding data privacy. Note that there may be differences in legal restrictions depending on whether audio is recorded or just captured.

## **4 Legal restrictions**

Many people have concerns regarding the use of microphones in video surveillance. These concerns are typically linked to the recording of plain speech along with the video material. But with audio analytics you typically neither record nor stream any audio. Laws that regulate surveillance vary by region and by country, so be sure to know what is permitted before using audio in your surveillance system.

Audio capturing and audio recording can be prohibited or require special consideration for several reasons, by national legislature or various types of local rules and regulations. While one region or environment might allow audio capture, it can still prohibit audio recordings. Companies can also prohibit audio surveillance within their premises.

## **5 Disclaimer**

This document and its content is provided courtesy of Axis and all rights to the document or any intellectual property rights relating thereto (including but not limited to trademarks, trade names, logotypes and similar marks therein) are protected by law and all rights, title and/or interest in and to the document or any intellectual property rights related thereto are and shall remain vested in Axis Communications AB.

Please be advised that this document is provided "as is" without warranty of any kind for information purposes only. The information provided in this document does not, and is not intended to, constitute legal advice. This document is not intended to, and shall not, create any legal obligation for Axis Communications AB and/or any of its affiliates. Axis Communications AB's and/or any of its affiliates' obligations in relation to any Axis products are subject exclusively to terms and conditions of agreement between Axis and the entity that purchased such products directly from Axis.

FOR THE AVOIDANCE OF DOUBT, THE ENTIRE RISK AS TO THE USE, RESULTS AND PERFORMANCE OF THIS DOCUMENT IS ASSUMED BY THE USER OF THE DOCUMENT AND AXIS DISCLAIMS AND EXCLUDES, TO THE MAXIMUM EXTENT PERMITTED BY LAW, ALL WARRANTIES, WHETHER STATUTORY, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, TITLE AND NON-INFRINGEMENT AND PRODUCT LIABILITY, OR ANY WARRANTY ARISING OUT OF ANY PROPOSAL, SPECIFICATION OR SAMPLE WITH RESPECT TO THIS DOCUMENT.



# About Axis Communications

Axis enables a smarter and safer world by creating solutions for improving security and business performance. As a network technology company and industry leader, Axis offers solutions in video surveillance, access control, intercom, and audio systems. They are enhanced by intelligent analytics applications and supported by high-quality training.

Axis has around 4,000 dedicated employees in over 50 countries and collaborates with technology and system integration partners worldwide to deliver customer solutions. Axis was founded in 1984, and the headquarters are in Lund, Sweden