

Coughing fit and stressed voice detection

Audio analytics in selected cameras with integrated microphone

May 2024

Summary

Coughing fit and stressed voice detection is an audio analytics application that detects incidents by listening to the surrounding audio 24/7.

The application comes preinstalled in selected Axis cameras with integrated microphone. It consists of two separate detection algorithms, and you can choose to use one of them, or both.

The cough detector detects single coughs or coughing fits, allowing for personnel to respond quickly to persons in need. The stressed voice detector identifies sound patterns associated with duress, anger, or fear, making it an effective tool for crime deterrence, assault reduction, or indicating persons in need of help.

Privacy is protected because *coughing fit and stressed voice detection* does not need to store any audio data to work properly. Audio is not recorded unless you have explicitly enabled audio recording.

You can configure several settings to make the analytics work optimally for your use case. *Coughing fit and stressed voice detection* also performs continuous health checks to verify proper operation.

Table of Contents

1	Introduction	4
2	Cough detector and stressed voice detector	4
3	Enabling optimal audio detection	5
4	Multisensor awareness	5
5	Overlays	6
6	Event types and health status	6
7	Privacy	7

1 Introduction

Audio detection analytics in a camera are powerful complements to video surveillance. They enable early detection and alerting on potential incidents, possibly before they are discovered in the video.

Coughing fit and stressed voice detection is an audio analytics application that listens to the surroundings 24/7 to classify and filters out sounds. When a coughing fit or stressed voice is detected, the application generates an alert.

This white paper presents *coughing fit and stressed voice detection* and how it is configured for optimal detection.

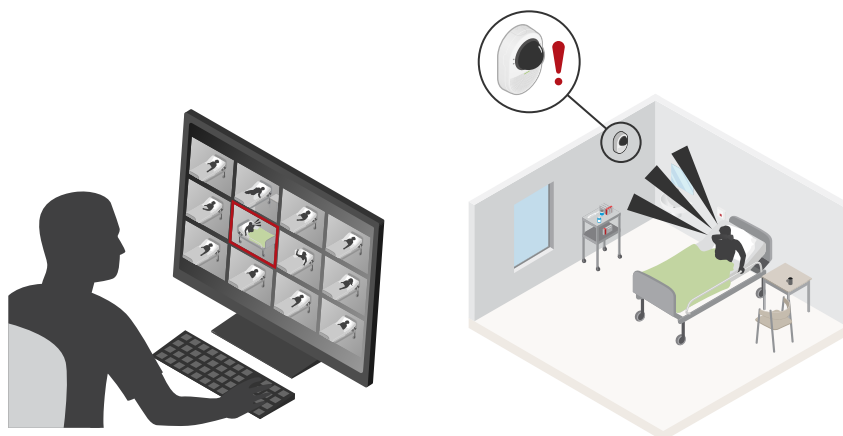
2 Cough detector and stressed voice detector

Coughing fit and stressed voice detection comes preinstalled in selected Axis cameras with integrated microphone. The detectors catch audible indicators of incidents in real time, directly in the camera. You can choose to use one of the detectors, or both.

The cough detector works by detecting coughs and counting them within a time frame. It allows for personnel to respond quickly when somebody is coughing or is having a coughing fit. The detector can detect coughing fits or a single cough, depending on how you set it up.

To reduce multiple event notifications within a very short time frame, a 5 second block time starts as soon as the first cough is counted. If the analytic is set to 3 coughs within 30 seconds, it will only count the next cough if there is at least 5 seconds after the previous cough. Intermediate coughs within the 5 second block time will not be counted. This means that, with these settings, an alert will only be sent after 3 coughs have been counted with at least 5 seconds between each cough.

The stressed voice detector identifies sound patterns associated with duress, anger, or fear in a person's voice. Upon recognition, the system sends an automatic notification to staff through a visual alert or by triggering an alarm. The early warning allows for personnel to respond quickly. They can offer help to a person in need or prevent escalation that could otherwise lead to physical aggression.



Coughing fit and stressed voice detection used in a healthcare environment.

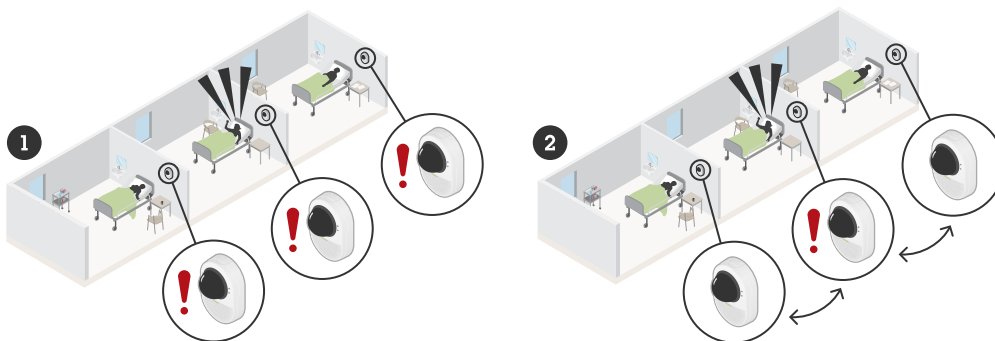
3 Enabling optimal audio detection

- **Camera placement.** The camera with the analytics should be placed at least 1.5 meters (5 feet) away from interfering noise sources, such as HVAC systems, PA systems or speakers, and slamming doors. Also, the camera should preferably be placed in line of sight of the area where you want to detect audio. While line of sight is not a strict requirement, it may enable more accurate detection. This is because sounds can be affected when they bend around corners or obstacles. For example, not all frequencies bend to the same degree.
- **Sensitivity.** The detection system can be fine-tuned through the sensitivity settings. A higher sensitivity will yield more detections. It increases the risk of unwanted detections (false alarms), but may be required when it is crucial to never miss a detection. With a lower sensitivity, detections will be reported only when it is very certain that the sound is correctly classified. This increases the risk of missing potential incidents, but a low sensitivity may be needed when there would otherwise be many false alarms.
- **Data gathering mode.** You can use the data gathering mode for a period of time after installation in order to gain insight into what types of audio are detected. The results and analysis can provide information about which level of sensitivity is optimal for the particular installation.
- **Cough detection threshold.** You can set the threshold for how many coughs should be required. An alarm will be triggered only when the number of coughs reaches the threshold in the assigned period of time.
- **Advanced settings.** Advanced settings are for expert users only. Changes may lead to incorrect detections or no detections at all. For specific scenarios, however, you may need to change these settings. This should be done only when advised by, or in consultation with, a system expert.

4 Multisensor awareness

When cameras are placed near each other, for instance in adjacent rooms, the same audio incident may be detected by multiple cameras. This can make it harder to pinpoint where the incident is taking place.

For stressed voice detections, the multisensor awareness feature can be helpful in these cases. When it is activated and multiple cameras pick up the same stressed voice, only the camera that picked it up first will trigger a notification. This way, nearby cameras work together to reduce false events and reduce duplicate notifications for the same event.



1 No multisensor awareness: cameras in adjacent rooms detect the same stressed voice incident and create multiple alarms.

2 With multisensor awareness, only the closest camera reports a detection.

With multisensor awareness you create peer groups to group together nearby cameras that are within audio pickup range of each other. Some restrictions apply:

- All peers should be configured to use NTP time synchronization.
- All peers should be running the same version of *coughing fit and stressed voice detection*.
- All peers should be able to reach each other over the network.

If any of the above fails, the peer will fall back to standalone mode and mark itself degraded.

5 Overlays

A live spectrogram and application notifications can be overlaid on top of the video feed. You can customize the overlays in size and drag them to the desired position. You can adjust the overlay opacity using a slider.

Application notifications will show events detected by the camera and what the application status is.

The spectrogram provides a visual representation of the audio. Hearing the audio and simultaneously seeing its visual representation can help you quickly determine the severity of an incident.

6 Event types and health status

Events generated by the cough detector and the stressed voice detector are *stateless*. They are momentary occurrences that are triggered by a detection. After the event block time (five seconds, configurable) has expired, a detection will generate a new event.

The health status of *coughing fit and stressed voice detection* is reflected by use of *stateful events*. With stateful events, the event state stays active as long as the condition occurs, and toggles only when the condition is resolved.

Health checks are built in to verify proper operation and alert when something is off. Three states of health can be distinguished for *coughing fit and stressed voice detection*:

- Healthy state: normal operation. Detections are possible.
- Degraded state: operation is running in degraded mode. This is typically caused by temporary factors, such as the loss of a peer camera, audio clipping due to very loud sounds, or audio buffer overrun. In degraded state, detections are possible, but there may be more false detections or missed detections. The degraded state typically resolves itself.
- Malfunction state: no operation. No detections are possible. This is typically caused by factors that do not resolve themselves, such as audio support being disabled in the device settings, or audio input gain being muted.

Degraded and malfunction state will show in the info panel and also on the text overlay (if enabled), so that the operator knows that the application is running with degraded health or detected malfunction.

A heartbeat event is triggered every 60 seconds (when enabled, configurable time setting). This can be used on the receiving end to verify if the analytics is up and running and alert if no heartbeats are received. Heartbeat events are not sent while the malfunction state is active.

7 Privacy

Audio data is processed and analyzed in the camera, and no storage of audio data is needed for *coughing fit and stressed voice detection* to work properly. Only when enabled explicitly, recordings of audio during events will be made. This can be helpful for forensics when investigating incidents, for troubleshooting when false positives are reported, or for listening back to incidents in case this is not supported by the video management system.

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