

Dźwięk z potencjałem

Przechwytywanie i przetwarzanie dźwięków na potrzeby zwiększania świadomości sytuacyjnej i pozyskiwania materiału dowodowego

Czerwiec 2021

Spis treści

1	Streszczenie	3
2	Wprowadzenie	4
	2.1 Przechwytywanie bez rejestrowania	4
3	W obliczu przeszkód	4
	3.1 Co mówią przepisy ustawowe i wykonawcze?	5
	3.2 Ocena możliwości	5
4	Zagadnienia związane z instalacją	6
5	Przygotowanie dźwięku	7
6	Topologia analiz	8
7	Zastosowania i przykłady	8
	7.1 Detekcja incydentów przy użyciu analizy audio	9
	7.2 Wizualizacja dźwięku w materiale wizyjnym	10
	7.3 Słuchanie i działanie	11
	7.4 Rejestrowanie i przechowywanie	12
	7.5 Pełniejsze wykorzystanie systemu dozoru	12
8	Monitorowanie i detekcja	12
	8.1 Cechy dźwięku	12
	8.2 Przetwarzanie sygnału	13
	8.3 Ludzki zmysł słuchu	14
9	Wyłączenie odpowiedzialności	15
Załącznik 1	Terminologia dotycząca jakości dźwięku	16

1 Streszczenie

Funkcje przechwytywania dźwięku, czy to zintegrowane i gotowe do użycia w kamerze czy udostępniane przez mikrofon zewnętrzny, otwierają drogę do wielu ważnych zastosowań. Odpowiedzialne, profesjonalne wykorzystanie odbieranego dźwięku może zapewnić szereg istotnych korzyści w systemie dozoru. Informacje dźwiękowe mogą na przykład uzupełnić materiał dowodowy zebrany podczas prac dochodzeniowych albo ułatwić bieżącą detekcję zdarzeń, które wymagają natychmiastowej reakcji pracowników ochrony lub personelu szpitalnego. Ponadto sama obecność dozoru dźwiękowego może działać odstraszająco na potencjalnych przestępców.

Przechwytywanie dźwięku (często w połączeniu z jego natychmiastową analizą) można wdrożyć jako odrębną technologię, umożliwiając kilka rodzajów zastosowań w obszarach zapobiegania przestępczości, ochrony i prowadzenia prac wyjaśniających.

Jednak w połączeniu z systemem wizyjnym funkcje przechwytywania dźwięku mogą wzbogacić zdecydowaną większość dotychczasowych zastosowań dozoru. Przykładowo operator systemu bezpieczeństwa może zyskać znacznie lepszy ogląd zdarzeń, jeśli odbierany strumień wideo zostanie uzupełniony strumieniem audio.

Tak jak stosuje się kilka rodzajów analizy *wideo* na potrzeby automatycznego wykrywania zdarzeń i przekazywania alarmów na podstawie detekcji wizualnej, tak również funkcje analizy *audio* mogą monitorować strumienie audio i wyzwać reakcję w przypadku wykrycia czegoś nietypowego.

Oprogramowanie do analizy audio można skonfigurować tak, aby automatycznie uruchamiało alarm lub inną czynność, gdy mikrofon wychwyci odgłosy związane z krzykami, tłuczeniem szkła lub strzałami z broni palnej. Jest to wczesne ostrzeżenie, które umożliwia szybką reakcję i interwencję.

Analiza audio pozwala także stwierdzić, czy nieoczekiwany odgłos dobiegł z lewej czy z prawej strony, oraz automatycznie przekierować kamerę PTZ na źródło dźwięku. W szpitalu lub domu opieki funkcje analizy audio mogą wykrywać głośne dźwięki wskazujące na cierpienie pacjentów albo podopiecznych i wysłać automatyczne powiadomienia do pielęgniarek lub opiekunów. W tym zastosowaniu analizy przydaje się funkcja wizualizacji dźwięku, która ułatwia jednoczesne monitorowanie dźwięku z wielu miejsc.

Istnieje różnica między przechwytywaniem dźwięków a ich rejestrowaniem. W wielu zastosowaniach do osiągnięcia zamierzonego celu nie jest wymagane rejestrowanie dźwięku, co zmniejsza obawy związane z prywatnością i pomaga w przestrzeganiu przepisów dotyczących danych osobowych. Zasadniczo aplikacje do analizy audio nie rejestrują dźwięku w sposób ciągły. Zazwyczaj jedynie przetwarzają one odbierany sygnał audio, szukając określonych wzorców, poziomów lub częstotliwości. Gdy funkcje analizy działają na brzegu sieci (w kamerze), żadne cyfrowe dane audio nie opuszczają kamery – przekazuje ona jedynie wyniki wykonanych analiz, czyli metadane lub wyzwalacze.

Axis nie udostępnia porad prawnych. Przepisy dotyczące dozoru różnią się w zależności od regionu i kraju, a obowiązek prowadzenia dozoru w sposób zgodny z obowiązującym prawem spoczywa na użytkownikach produktów (którym zazwyczaj jest klient końcowy). Podobnie jak w przypadku rozwiązania dozoru wizyjnego instalację systemu dozoru dźwiękowego należy poprzedzić dokładnym zbadaniem odnośnych aspektów prawnych.

Po wdrożeniu niezbędnych środków zapewniających zgodność z przepisami należy dokładnie rozważyć takie aspekty instalacji jak rozmieszczenie i konfiguracja sprzętu. Może to wymagać pewnego planowania, ale zazwyczaj nie jest trudne, a proste czynności mogą zdecydować o dużej użyteczności materiału dźwiękowego.

2 Wprowadzenie

Informacje dźwiękowe mogą być cennym zasobem w obszarach zapobiegania przestępczości, ochrony i prowadzenia prac wyjaśniających. Przechwycony materiał dźwiękowy może być też przetwarzany w czasie rzeczywistym przez oprogramowanie analityczne, co pozwala na bardzo efektywny monitoring audio na potrzeby detekcji działań, zachowań lub zdarzeń.

W tym dokumencie omówiono potencjał dźwięku w systemach bezpieczeństwa i przedstawiono przykłady typowych zastosowań. Zaprezentowano także różne rodzaje analiz audio wraz z krótkim omówieniem sposobu ich działania.

Dokument nie zawiera porad prawnych, lecz przedstawia różne rozwiązania techniczne, które mogą się przydać na etapie instalacji i konfiguracji. W zależności od wybranego sposobu wdrożenia funkcji analizy audio istnieje możliwość zachowania zgodności z regionalnymi przepisami i zaleceniami oraz skutecznego wykorzystania tego wydajnego narzędzia.

Tematyka tego dokumentu ogranicza się do *przechwytywania i ewentualnego rejestrowania* dźwięku, czyli do *wejściowego* materiału dźwiękowego/audio. Inne typowe zastosowanie dźwięku w rozwiązaniach bezpieczeństwa dotyczy jego *emitowania*, czyli *wyjściowego* materiału dźwiękowego/audio – zazwyczaj w zakresie odtwarzania komunikatów głosowych lub alarmów w celu odstraszenia intruzów czy złodziei sklepowych. Więcej informacji na temat emitowania dźwięku w systemach bezpieczeństwa jest dostępnych na stronie www.axis.com/products/audio.

2.1 Przechwytywanie bez rejestrowania

Istnieje możliwość przechwytywania i wykorzystywania materiału dźwiękowego bez jego rejestrowania. Zasadniczo przechwytywanie dźwięku oznacza przekształcanie go do postaci cyfrowej i udostępnianie go do użycia w oprogramowaniu. W tym celu mikrofon wychwytuje drgania dźwięku występujące w powietrzu, a następnie taki sygnał analogowy jest konwertowany na postać cyfrową (przy użyciu odpowiednich przetworników) i przekazywany do modułu przetwarzającego.

Jeśli przechwycony materiał dźwiękowy nie zostanie umieszczony na żadnym trwałym nośniku danych, na przykład w pamięci flash lub na dysku twardym, nie zostanie zarejestrowany. W niektórych zastosowaniach, na przykład gdy operator słucha przechwytywanego materiału dźwiękowego w czasie rzeczywistym, nagrania są niepotrzebne. W pewnych sytuacjach rejestrowanie materiału audio może być wręcz *niewskazane*. Obowiązujące ograniczenia prawne mogą być różne w zależności od tego, czy materiał dźwiękowy jest rejestrowany czy tylko przechwytywany (monitorowany).

Narzędzia do analizy audio zazwyczaj nie rejestrują dźwięku w sposób ciągły. Aby działać prawidłowo, buforują dźwięk przez pewien czas. Wiele systemów można skonfigurować tak, aby rejestrowały materiał zbuforowany bezpośrednio przed detekcją i bezpośrednio po niej. Dzięki temu pracownicy ochrony mogą zweryfikować detekcję i ewentualnie zachować materiał dźwiękowy jako dowód do prac wyjaśniających lub dochodzeniowych.

3 W obliczu przeszkód

Wiele osób ma zastrzeżenia związane z używaniem mikrofonów w kontekście dozoru wizyjnego. Zastrzeżenia te zwykle dotyczą rejestrowania głosu razem z materiałem wizyjnym.

Tę pierwszą przeszkodę można pokonać przez uświadomienie sobie, że z odbieranymi danymi audio wiąże się wiele innych możliwości niż tylko ich rejestrowanie. W wielu zastosowaniach nie jest wymagane rejestrowanie jakiegokolwiek materiału dźwiękowego.

Przepisy regulujące kwestie związane z dozorem różnią się w zależności od kraju i regionu, więc przed rozszerzeniem systemu dozoru o funkcje audio należy koniecznie zapoznać się z obowiązującymi przepisami.

3.1 Co mówią przepisy ustawowe i wykonawcze?

Podobnie jak w przypadku rozwiązania dozoru wizyjnego instalację systemu dozoru dźwiękowego należy poprzedzić dokładnym zbadaniem odnośnych aspektów prawnych. Jeśli to konieczne, należy złożyć niezbędne wnioski i uzyskać potrzebne zezwolenia. Tam, gdzie to wymagane, należy umieścić odpowiednie oznaczenia i informacje.

Wykorzystywanie i rejestrowanie materiału dźwiękowego może być zabronione lub objęte szczególnymi obostrzeniami wynikającymi z ustawodawstwa krajowego bądź różnego rodzaju przepisów i regulacji lokalnych. W pewnych regionach lub środowiskach może być dopuszczalne przechwytywanie dźwięku przy jednoczesnym zakazie wykonywania nagrań audio. Również firmy mogą zabraniać prowadzenia dozoru dźwiękowego na swoim terenie.

3.1.1 Przykłady ze Stanów Zjednoczonych

Przepisy ustawowe i wykonawcze obowiązujące w Stanach Zjednoczonych różnią się w zależności od stanu.

Niektóre stany wymagają jednostronnej zgody na rejestrowanie dźwięku. Oznacza to, że wystarczy zgoda jednej osoby uczestniczącej w rozmowie, aby dozór był legalny.

W innych stanach wymagana jest zgoda wszystkich stron (lub zgoda dwustronna), co oznacza, że nagranie można wykonać tylko w przypadku, gdy wszystkie strony rozmowy zgodzą się na jej rejestrowanie.

Wyjątki od zasady zgody wszystkich stron mogą obowiązywać w miejscach publicznych, w których nie można oczekiwać zachowania prywatności.

W niektórych regionach ocena sytuacji prawnej może także doprowadzić do innego wniosku, jeśli używana aplikacja do analizy audio nie rejestruje dźwięku. Dlatego należy sprawdzić, jakie przepisy ustawowe i wykonawcze obowiązują w danym stanie.

3.1.2 Przykłady europejskie

W Europie kwestie związane z dozorem dźwiękowym są uregulowane w przepisach krajowych. Dlatego należy sprawdzić, jakie przepisy ustawowe i wykonawcze obowiązują w danym kraju.

Nagrania audio mogą zawierać dane osobowe, które podlegają Ogólnemu rozporządzeniu o ochronie danych (RODO). RODO niekoniecznie zabrania wykonywania nagrań audio, ale określa szczególne czynniki, które należy uwzględnić w przypadku przechwytywania czy rejestrowania dźwięku. W przypadku poszerzenia dotychczasowego systemu dozoru wizyjnego o funkcje audio należy się zastanowić, czy podstawa prawna do przetwarzania danych osobowych zgodnie z RODO zachowuje swą ważność.

3.2 Ocena możliwości

Istnieje błędne przekonanie, że w systemach dozoru w ogóle nie wolno korzystać z funkcji audio. Przekonanie to jest tak powszechne, że w wielu przypadkach klient nawet nie rozważa możliwości rozszerzenia systemu dozoru o rozwiązania audio.

Jednak wiele rodzajów instalacji jest dopuszczalnych, jeśli przykładowo odpowiednie osoby zostaną poinformowane, wprost wyrażą zgodę itp. Dlatego warto zbadać przepisy ustawowe i wykonawcze obowiązujące w określonym kraju lub regionie w odniesieniu do danego rodzaju zastosowań. Nawet jeśli w danym systemie dozoru nie będzie dozwolone zastosowanie typu „rejestrowanie i przechowywanie”,

wiele innych zastosowań, takich jak „słuchanie i działanie”, „słuchanie i zaświadczenie” oraz „detekcja incydentów przy użyciu analizy audio”, można zmodyfikować tak, aby nie naruszać prawa do prywatności.

4 Zagadnienia związane z instalacją

Umieszczenie mikrofonu w obrębie danej sceny określa zakres możliwych zastosowań. Przed zainstalowaniem sprzętu audio należy szczegółowo rozważyć jego rozmieszczenie i konfigurację. Może to wymagać pewnego planowania, ale zazwyczaj nie jest trudne, a proste czynności mogą znacznie zwiększyć użyteczność materiału dźwiękowego.

Rozważenie właściwej lokalizacji mikrofonu i wybór miejsca o dobrych cechach akustycznych zwiększy prawdopodobieństwo realizacji celów związanych z dozorem. To oczywiste, że mikrofon należy umieścić tak, aby mógł łatwo rejestrować interesujące nas dźwięki. Często będzie to środek pomieszczenia, bezpośrednia bliskość kamery lub miejsce, w którym może dochodzić do określonych zdarzeń. Mikrofonu nie należy instalować w pobliżu źródeł hałasu, na przykład wentylatorów lub maszyn, które mogą zagłuszyć dźwięki o niższym natężeniu lub dochodzące z większej odległości.



Typowe umiejscowienie mikrofonu

- 1 W miejscu, w którym dochodzi do określonych czynności
- 2 W kamerze
- 3 Na środku pomieszczenia

Akustyczne właściwości otoczenia, na przykład różny stopień pochłaniania dźwięku przez ściany, sufit i podłogę, oraz cechy architektoniczne, na przykład bardzo długie korytarze, prowadzą do powstawania pogłosu i echa, które mogą negatywnie wpływać na pole dźwiękowe w niektórych miejscach. Przykładowo

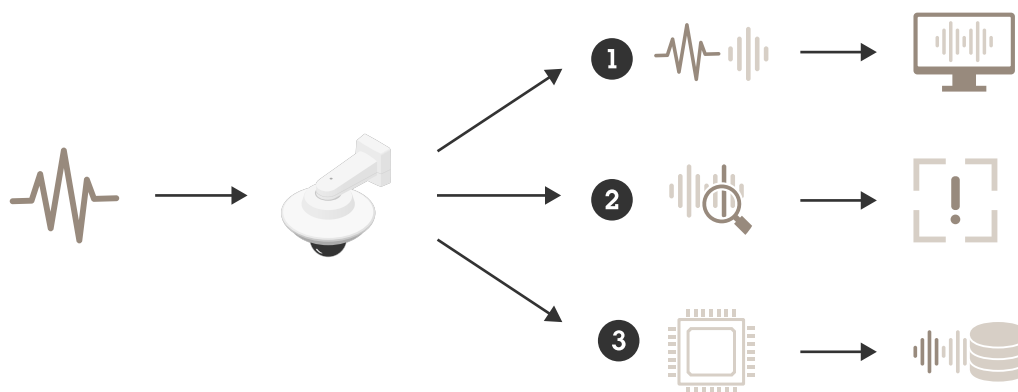
głos zabrzmie inaczej w mocno wytłumionym pomieszczeniu (na przykład w specjalnie przystosowanej sali konferencyjnej), a inaczej w kościele lub wyłożonej płytkami łazience. W trudnych warunkach akustycznych umiejscowienie mikrofonu może mieć krytyczne znaczenie.

Duże znaczenie ma właściwa instalacja i konfiguracja sprzętu (na przykład ustawienie wzmocnienia dźwięku), a także integracja urządzeń audio z systemem dozoru. Instalatorzy i integratorzy systemów mogą przedstawić zalecenia dostosowane do konkretnej sytuacji.

W przypadku analizy audio mogą istnieć pewne wskazania, które różnią się od zaleceń dotyczących typowych nagrań dźwiękowych. Aby poznać obowiązujące wymagania wstępne, należy zapoznać się z dokumentacją użytkownika.

5 Przygotowanie dźwięku

Po pierwotnym przechwyceniu dźwięku pozyskane w ten sposób informacje są przygotowywane pod kątem kolejnych etapów przetwarzania. Poszczególne czynności przygotowawcze mogą być wykonywane równoległe lub pojedynczo.



- 1 Przekształcenie
- 2 Analizy brzegowe w czasie rzeczywistym
- 3 Przetwarzanie i kodowanie

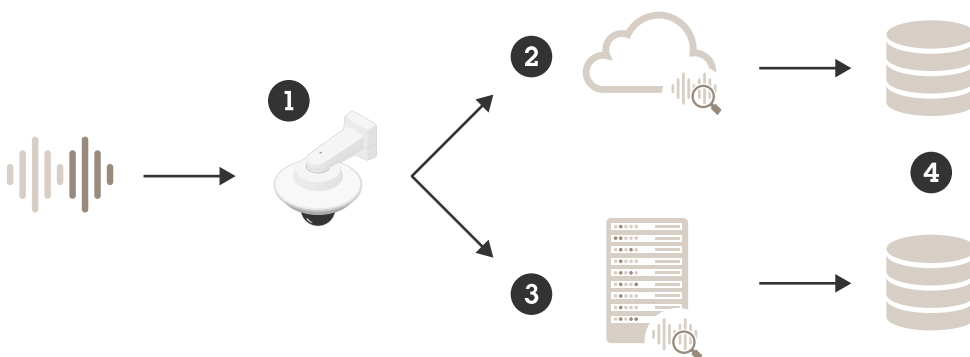
- **Przekształcenie.** Dźwięk zyskuje postać abstrakcyjną i jest konwertowany na informacje wizualne, np. na wykres przedstawiający spektrum dźwięku. Jest to proces nieodwracalny: z wykresu spektrum nie można odtworzyć pierwotnego dźwięku.
- **Analizy brzegowe w czasie rzeczywistym.** Jeśli dźwięk jest przetwarzany na brzegu sieci, można użyć **klasyfikatora dźwięku**. Wynikiem jego zastosowania są metadane opisujące cechy dźwięku. Z metadanych nie można odtworzyć pierwotnego dźwięku. Można także użyć **detektora dźwięku** w celu rozpoznawania wzorców, poziomów lub częstotliwości i uzyskiwania informacji o stanie. Również w tym przypadku nie jest możliwe przywrócenie pierwotnego dźwięku.
- **Przetwarzanie i kodowanie.** W przypadkach, gdy używane są pierwotne dane audio (bez przekształcenia czy analizy), zazwyczaj wykonywane są określone czynności z zakresu przetwarzania i kodowania, które mają na celu przygotowanie danych dźwiękowych do zamierzonych zastosowań. Zastosowania te mogą obejmować przechowywanie danych audio na brzegu sieci, ich strumieniowe przesyłanie

do zewnętrznych urządzeń klienckich na potrzeby dodatkowego przetwarzania (na serwerze lub w chmurze) albo przechowywanie w lokalizacji zewnętrznej.

6 Topologia analiz

Umieszczenie mechanizmu analiz w systemie jest ważne z wielu powodów. Zwłaszcza z perspektywy obaw o ochronę prywatności i przestrzegania przepisów dotyczących danych osobowych duże znaczenie ma to, *gdzie* algorytm programowy analizuje dane audio. Są sytuacje, gdy danych dźwiękowych nie można przesłać przez sieć i przechwycone (ale niezachowane) dane audio muszą być analizowane lokalnie. Jeśli stosowane algorytmy wymagają dużej mocy obliczeniowej i nie mogą być wykonywane na brzegu sieci, może być konieczne wysłanie cyfrowych danych audio do chmury lub na serwer.

- **Analizy brzegowe.** Gdy funkcje analiz działają na brzegu sieci, cyfrowe dane audio nie muszą opuszczać kamery. W przypadku przechwytywania dźwięku bez jego rejestrowania wysyłane są tylko wyniki wykonanych analiz, czyli metadane lub wyzwalacze.
- **Analizy serwerowe.** Gdy analizy są wykonywane na serwerze, cyfrowe dane audio muszą opuścić kamerę. Jeśli wstępne przetwarzanie odbywa się w kamerze (na brzegu sieci), urządzenie mogą opuszczać metadane w postaci abstrakcyjnej lub zdepersonalizowanej. Serwer zazwyczaj jest elementem zamkniętego systemu (kontrolę sprawuje jego właściciel), co zmniejsza obawy dotyczące prywatności przesyłanych danych audio. Niemniej jednak należy zadbać o przestrzeganie obowiązujących przepisów i regulacji.
- **Analizy w chmurze.** Cyfrowe dane audio mogą być także przekazywane na serwer w środowisku chmurowym. Tak jak w przypadku analiz serwerowych informacje audio mogą być wstępnie przetwarzane do postaci metadanych. Chmura ma często charakter zdecentralizowany, więc tym bardziej należy wyjaśnić kwestie dotyczące prywatności i zapewnić przestrzeganie przepisów.

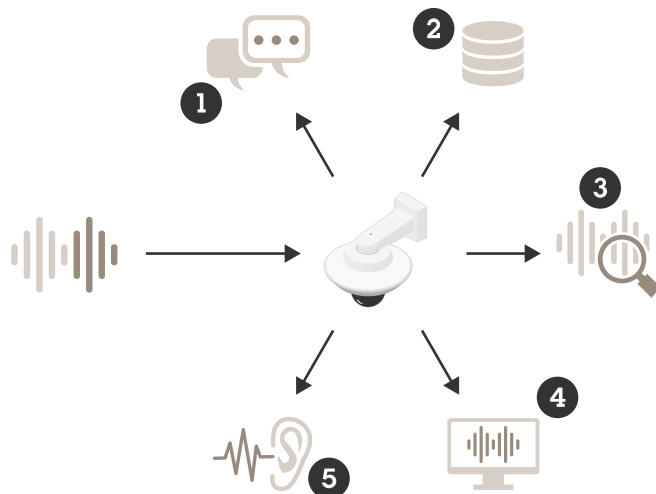


- 1 Brzeg sieci
- 2 Chmura
- 3 Serwer
- 4 Przechowywanie

7 Zastosowania i przykłady

Funkcje audio często są zintegrowane z kamerami i gotowe do użycia. Istnieją różne rodzaje zastosowań, w których odpowiedzialne i profesjonalne wykorzystanie odbieranego materiału dźwiękowego może zapewnić szereg istotnych korzyści. Informacje dźwiękowe mogą na przykład uzupełnić materiał dowodowy zebrany podczas prac dochodzeniowych albo ułatwić bieżącą detekcję zdarzeń, które wymagają natychmiastowej

reakcji pracowników ochrony lub personelu szpitalnego. Ponadto sama obecność dozoru dźwiękowego może działać odstrasżająco na potencjalnych przestępców.



Typowe cele przechwytywania dźwięku:

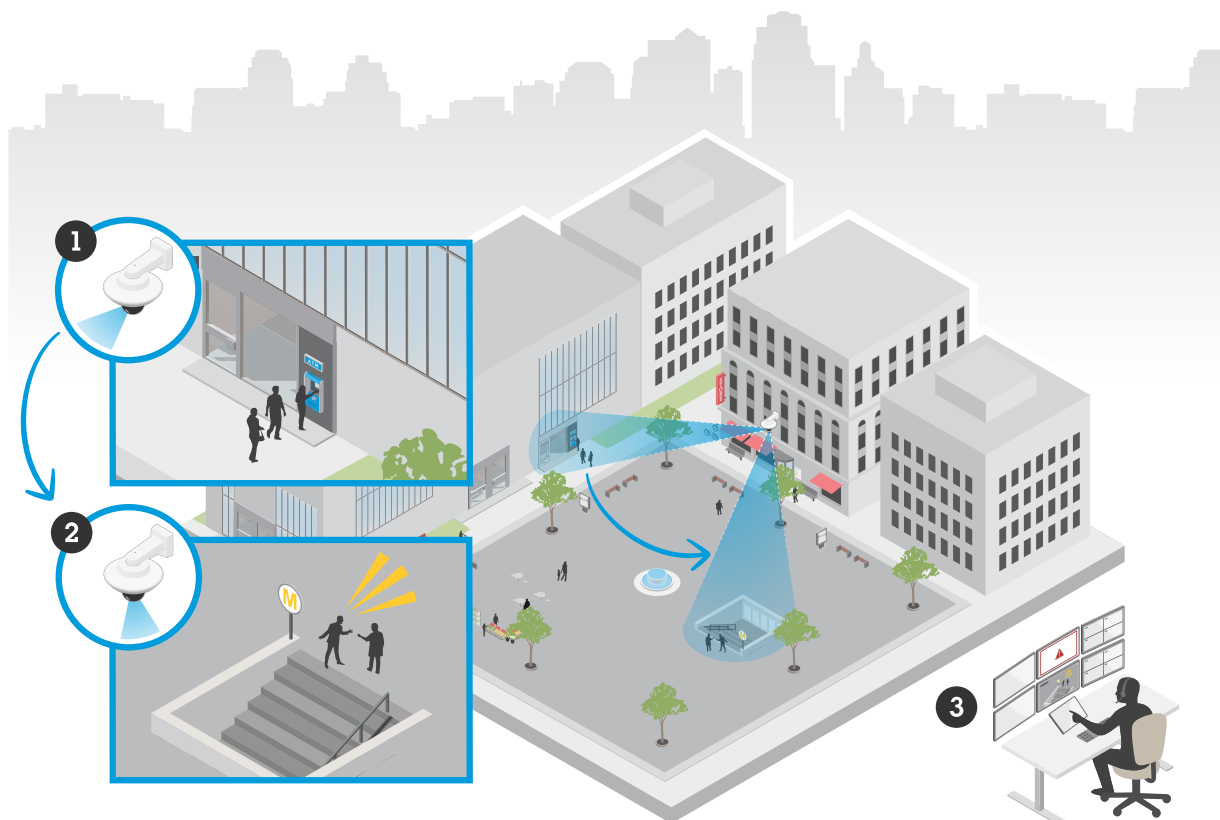
- 1 *Komunikacja*
- 2 *Rejestrowanie*
- 3 *Analiza*
- 4 *Wizualizacja*
- 5 *Słuchanie*

7.1 Detekcja incydentów przy użyciu analizy audio

Aplikacje do analizy audio to rodzaj oprogramowania, które przetwarza przechwycony materiał dźwiękowy, aby znaleźć w nim i wyodrębnić określone informacje. Aplikacje te służą do detekcji takich zdarzeń i stanów jak strzały z broni palnej, tłuczenie szkła czy agresja. Przykładowo taka aplikacja po przetworzeniu wejściowego materiału audio może dać jasną odpowiedź na pytanie, czy doszło do stłuczenia okna. Po pomyślnej detekcji system zazwyczaj przekazuje automatyczne powiadomienie pracownikom – w postaci alertu wzrokowego lub przez wyzwolenie alarmu. Jest to wczesne ostrzeżenie, które umożliwia szybką reakcję i interwencję.

7.1.1 Przekierowanie kamery

Kolejnym przykładem zastosowania analizy audio jest przekierowanie kamery PTZ. W tym przypadku mamy do czynienia z połączeniem funkcji audio i wideo: po wykryciu, skąd dochodzi dźwięk, kamera jest automatycznie przekierowywana na jego źródło.



- 1 Kamera PTZ monitoruje bankomat.
- 2 Mikrofon kamery wychwytuje nagły, głośny hałas i kamera natychmiast przekierowuje się na zdarzenie.
- 3 Operator otrzymuje alarm i weryfikuje zdarzenie.

7.2 Wizualizacja dźwięku w materiale wizyjnym

Przechwycony dźwięk związany z materiałem wizyjnym może być wyświetlany w postaci diagramu spektrum dźwięku na monitorze. W przypadku przekroczenia zadanej wartości progowej na diagramie pojawi się wskazanie alarmu.

Taka wizualizacja dźwięku może się przydać w sytuacji, gdy trzeba jednocześnie monitorować dźwięki pochodzące z wielu źródeł, na przykład z kilku sal w szpitalu. Równoczesne słuchanie dźwięku z wielu źródeł (w tym przypadku sal) byłoby trudne – rozwiązanie dające pielęgniarkom wzrokowy dostęp do wielu

wizualizacji na jednym monitorze jest znacznie łatwiejsze w obsłudze. Jeśli są dostępne przekazy wizyjne z sal, wizualizacje można dodać w postaci nakładki na obraz wideo.



Wizualizacja dźwięku dodana w postaci nakładek na przekazy wizyjne w szpitalu.

7.3 Słuchanie i działanie

Chyba najbardziej podstawowe i intuicyjne zastosowanie polega na dozorcze dźwiękowym z bezpośrednim działaniem operatora w celu zwiększenia świadomości sytuacyjnej. Przykładem może być sytuacja, gdy operator słyszy podejrzaną rozmowę i wysyła pracownika ochrony w celu szczegółowego zbadania zdarzenia. W szpitalu lub domu opieki może chodzić o możliwość usłyszenia pacjenta potrzebującego pomocy i wezwania pielęgniarki. Czasem może chodzić o możliwość stwierdzenia, czy „dziwny” odgłos dobiegł z lewej czy z prawej strony, oraz skierowania kamery PTZ na źródło dźwięku.

W tego rodzaju zastosowaniach jeden lub kilku operatorów ma dostęp do środowiska audio w pomieszczeniu kontrolnym lub za pośrednictwem specjalnej aplikacji na urządzeniu mobilnym. Ucho ludzkie rejestruje dźwięki, a mózg wydobywa z nich informacje odnoszące się do danej sceny lub sytuacji. W połączeniu z dozorem wizyjnym dźwięk zapewnia dodatkowy wymiar informacji na potrzeby podejmowania decyzji. W niektórych przypadkach – na przykład gdy źródło dźwięku znajduje się poza polem widzenia kamery lub panują trudne warunki oświetleniowe – dźwięk stanowi wręcz jedyny dostępny wymiar informacji.

7.3.1 Słuchanie i zaświadczenie

Dozór dźwiękowy może także służyć do celów bezpośredniego składania zeznań na podstawie usłyszanych zdarzeń. Ten przypadek różni się od zastosowania typu „*słuchanie i działanie*”, ponieważ jego celem nie jest podejmowanie decyzji, ale oba te rodzaje zastosowań często współistnieją ze sobą. Przykładowo operator, który usłyszy nasilającą się kłótnię zawierającą groźby karalne, może nie tylko wysłać ochroniarzy, ale później również złożyć zeznanie na temat tego, co usłyszał.

7.4 Rejestrowanie i przechowywanie

W odpowiednich przypadkach zastosowanie polegające na przechwytywaniu i rejestrowaniu danych audio może zapewnić dodatkowy materiał dowodowy o dużej wartości. Może przykładowo chodzić o obciążające wypowiedzi lub strzały z broni palnej. Zarejestrowany materiał dźwiękowy może pomóc w wyjaśnieniu, kto wypowiedział określone słowa, ile strzałów padło z broni palnej czy zapewnić inne przydatne informacje na etapie prac dochodzeniowych.

Jeśli dźwięk jest rejestrowany w kontekście prac wyjaśniających lub dochodzeniowych, należy dopilnować, by zostały zachowane oryginalne dane, i nie poddawać ich przetwarzaniu (które w innych kontekstach może być wymagane lub korzystne). W przypadku nagrań przeznaczonych do prac wyjaśniających wszelkie przetwarzanie może zostać uznane za manipulowanie materiałem dowodowym. Można zastosować algorytmy wzmacniające głos, które zwiększą słyszalność nagranej mowy i podniosą wartość materiału na etapie dochodzenia. Jednak czynność tę należy wykonać na odrębnej kopii zarejestrowanego materiału. Pozostawienie nagrania w nieprzetworzonej postaci daje szerokie możliwości późniejszego wykorzystania materiału.

7.5 Pełniejsze wykorzystanie systemu dozoru

Systemy dozoru często obejmują kilka rodzajów czujników. Jednym z nich jest przetwornik obrazu w kamerze, który rejestruje dane wizualne na temat obserwowanej sceny. Często stosowane są też czujniki niewizualne, na przykład detektory ruchu oparte na technologii radarowej lub emisji promieniowania podczerwonego. Zdarza się, że dozór wizyjny jest nieodpowiedni, a wówczas czujniki niewizualne są używane jako urządzenia autonomiczne. Jednak w wielu przypadkach czujniki niewizualne uzupełniają system kamer, dostarczając innych rodzajów informacji.

Przez poszerzenie systemu dozoru o czujniki dźwiękowe (mikrofony) zyskuje się nowe możliwości w ogromnej większości zastosowań. Funkcje audio dodane do pozbawionego ich wcześniej systemu otwierają drogę do interakcji wieloczujnikowej – przy użyciu analiz lub za pośrednictwem operatora.

Dobrym przykładem jest proste zastosowanie typu „*słuchanie i działanie*”, w którym operator zyskuje znacznie lepszy ogląd zdarzeń dzięki odbieranemu równolegle strumieniowi audio. Czasem trudno jest stwierdzić agresywne zachowanie przez wyłącznie wzrokową obserwację uczestników zajścia, natomiast możliwość ich usłyszenia znacznie ułatwia to zadanie.

Kolejnym typowym przykładem jest korzystanie z analizy wideo, na przykład wizyjnej detekcji ruchu. Jeśli aplikacja analityczna ma trudności z detekcją np. z powodu słabego oświetlenia, obecność funkcji analiz audio może zwiększyć skuteczność detekcji.

8 Monitorowanie i detekcja

Dźwięk zawiera kilka rodzajów informacji, których można używać do monitorowania i analizy audio. Różne rodzaje przetwarzania pomagają w wydobyciu i oczyszczeniu tych informacji na potrzeby dalszego wykorzystania i interakcji z otoczeniem.

8.1 Cechy dźwięku

W kontekście dozoru duże znaczenie mogą mieć takie cechy dźwięku jak głośność i wysokość tonu. Czas słyszalności, ewentualne przemieszczanie się źródła dźwięku, jego odległość od odbiorcy – to przykładowe elementy układanki, które trzeba uwzględnić, zanim wyciągnie się wnioski na temat słyszanego dźwięku. Sprzęt oraz oprogramowanie do monitorowania i detekcji audio korzystają z tych samych rodzajów

informacji, „nasłuchując” w nich złożonych połączeń cech, od poziomu natężenia dźwięku po rozkład energii na różnych częstotliwościach w czasie.

- **Informacje przestrzenne.** Dotyczą otoczenia fizycznego, w tym takich cech jak lokalizacja, kierunek i odległość. Informacji przestrzennych można użyć, by zogniskować przechwytywanie dźwięku w określonym kierunku w celu uzyskania lepszych nagrań. Ponadto za pomocą tych informacji aplikacja analityczna może określić, skąd dochodzi dźwięk lub w jakiej odległości znajduje się jego źródło.
- **Informacje czasowe.** Informacje czasowe są ważne zarówno w ujęciu dynamicznym (zmiana w czasie), jak i w ujęciu bezwzględnym (kiedy coś się wydarzyło). Często rozpatruje się je przez odniesienie do informacji pochodzących z innych urządzeń, na przykład materiału wizyjnego. Informacje czasowe odgrywają ważną rolę w analizie zachowań, ponieważ pomagają ustalić, kiedy doszło do określonego zdarzenia i jak długo ono trwało.
- **Informacje spektralne.** Dotyczą częstotliwości, na przykład wysokości tonu dźwięku lub połączenia wysokości tonu w przypadku bardziej złożonych dźwięków. Mikrofony używane w dozorcze dźwiękowym są projektowane z płaską charakterystyką przenoszenia, co oznacza, że powinny równie skutecznie wychwytywać wszystkie częstotliwości z pasma słyszalnego (20 Hz – 20 kHz). Różni się to od sposobu działania ludzkiego zmysłu słuchu, ponieważ człowiek łatwiej wychwytuje częstotliwości typowe dla ludzkiej mowy.
- **Informacje o amplitudzie.** Dotyczą intensywności czyli głośności dźwięku. Informacje o amplitudzie w połączeniu z informacjami spektralnymi pozwalają uzyskać pełniejszy obraz struktury odbieranego dźwięku.

8.2 Przetwarzanie sygnału

W obszarze dozoru dźwiękowego przetwarzanie sygnału najczęściej ma na celu polepszenie transmisji, wydajności przechowywania lub subiektywnej jakości albo wzmocnienie lub detekcję interesujących elementów składowych. W tym celu algorytmy programowe na różne sposoby analizują lub modyfikują sygnał audio.

8.2.1 Modyfikacja sygnału

Za pomocą odpowiednich algorytmów można zmodyfikować sygnał na potrzeby określonego celu. Zazwyczaj obejmuje to:

- Poprawę jakości sygnału, na przykład zwiększenie słyszalności przez automatyczną kontrolę wzmocnienia.
- Modyfikację sygnału, na przykład przez zmianę względnego rozkładu pasm częstotliwości przy użyciu korektora.
- Ograniczenie sygnału przez usunięcie określonych częstotliwości lub amplitud. Może tu chodzić o zmniejszenie ilości danych przez kompresję lub zapewnienie ochrony prywatności przez szyfrowanie głosu.

8.2.2 Analiza sygnału

Narzędzia do analizy audio wykorzystują przechwycone (ale zazwyczaj niezarejestrowane) dane audio i przez analizę ich charakterystyki dźwiękowej generują wyniki w postaci niedźwiękowej. Zasadniczo aplikacje te przekształcają dane dźwiękowe w bardziej użyteczny zasób w innym formacie. Istnieją

aplikacje analityczne stworzone specjalnie do wykrywania np. agresji, strzałów z broni palnej, odgłosów tłuczenia szkła lub alarmów samochodowych.

Jeśli używane są algorytmy uczenia maszynowego, można je wyszkolić pod kątem przewidywania za pomocą dużej ilości danych, nie programując ich bezpośrednio do tego celu. Przykładem w kontekście funkcji audio może być algorytm, który niezawodnie wykrywa dźwięk zamykanych drzwi, gdy zostanie przeszkolony z użyciem tysięcy takich dźwięków.

8.3 Ludzki zmysł słuchu

Ludzkie ucho należy do najlepszych narzędzi detekcji i analizy dźwięku. Nawet w warunkach dużego hałasu ludzkie ucho i mózg są w stanie wykrywać i interpretować mowę, chociaż większość algorytmów wówczas zawodzi.

Używając zmysłu słuchu, jesteśmy w stanie stworzyć charakterystykę przestrzenną danego miejsca, określając na przykład, skąd dochodzi dźwięk i czy jego źródło się porusza. Ponieważ mamy dwoje uszu, słyszymy, czy dźwięk dochodzi z lewej czy z prawej strony, a może z punktu znajdującego się między nimi. Uszy i głowa są też zbudowane tak, że słyszymy, czy dźwięk dochodzi z góry czy z dołu oraz z przodu czy z tyłu. Kilka „stopni filtrujących” w mózgu rekompensuje czasowe różnice między uszami, błyskawicznie wykrywając odchylenia rzędu mikrosekund i uwrażliwiając nas na określone rodzaje zdarzeń. Mamy dobrze wykształconą zdolność do analizy sygnałów dźwiękowych, zwłaszcza tych związanych z ludzkim głosem, ale także tych kojarzących się z historycznymi zagrożeniami.

W odpowiednich okolicznościach (dźwięk stereofoniczny o wysokiej jakości, niezbyt duże opóźnienie) człowiek pełniący funkcję operatora może być skutecznym „narzędziem analitycznym” uzupełniającym sprzęt lub oprogramowanie do detekcji. Używając produktu do dozoru dźwiękowego wyposażonego w zaledwie dwa mikrofony, operator może pozyskiwać z obserwowanej sceny informacje przestrzenne, na przykład na temat źródła dźwięku i jego przemieszczania się.

9 Wyłączenie odpowiedzialności

Ten dokument i jego treść są udostępniane dzięki uprzejmości firmy Axis, a wszelkie prawa do dokumentu oraz związane z nim prawa własności intelektualnej (w tym, między innymi, znaki towarowe, nazwy handlowe, logotypy i podobne zawarte w nim znaki) są chronione na mocy prawa oraz wszelkie prawa, tytuł własności i/lub udziały w dokumencie i wszelkie związane z nim prawa własności intelektualnej należą i będą należeć do Axis Communications AB.

Zwracamy uwagę, że ten dokument jest udostępniany w swojej aktualnej postaci bez jakichkolwiek gwarancji wyłącznie do celów informacyjnych. Informacje zawarte w tym dokumencie nie stanowią i nie mają stanowić porady prawnej. Ten dokument nie ma tworzyć i nie tworzy żadnego zobowiązania prawnego po stronie firmy Axis Communications AB i/lub jakichkolwiek podmiotów z nią powiązanych. Zobowiązania firmy Axis Communications AB i/lub jakichkolwiek podmiotów z nią powiązanych w odniesieniu do dowolnych produktów Axis są w całości uregulowane warunkami umowy zawartej między firmą Axis i podmiotem, który kupił takie produkty bezpośrednio od firmy Axis.

DLA UNIKNIĘCIA WĄTPLIWOŚCI WYJAŚNIA SIĘ, ŻE CAŁE RYZYKO ZWIĄZANE Z UŻYCIEM TEGO DOKUMENTU I JEGO REZULTATAMI PONOSI UŻYTKOWNIK DOKUMENTU, A AXIS W MAKSYMALNYM ZAKRESIE DOZWOLONYM PRZEZ PRAWO WYKLUCZA I WYŁĄCZA WSZELKIE USTAWOWE, WYRAŻNE LUB DOROZUMIANE GWARANCJE, W TYM, MIĘDZY INNYMI, WSZELKIE DOROZUMIANE GWARANCJE WARTOŚCI HANDLOWEJ, PRZYDATNOŚCI DO OKREŚLONEGO CELU, TYTUŁU I NIENARUSZENIA PRAW ORAZ ODPOWIEDZIALNOŚCI ZA PRODUKT, A TAKŻE WSZELKIE GWARANCJE WYNIKAJĄCE Z JAKIEJKOLWIEK OFERTY, SPECYFIKACJI LUB PRÓBK, KTÓRE MOŻNA BY ODNIEŚĆ DO TEGO DOKUMENTU.

Załącznik 1 Terminologia dotycząca jakości dźwięku

Dźwięk cyfrowy:

Dźwięk cyfrowy (audio cyfrowe) to odwzorowanie dźwięku analogowego (często mającego postać sygnału akustycznego zarejestrowanego przez mikrofon) zapisane w formie cyfrowej. W dźwięku cyfrowym fala dźwiękowa sygnału audio najczęściej zakodowana jest jako nieprzerwany ciąg cyfrowych próbek. Dokładność kodowania zależy od liczby cyfr znaczących rejestrowanych przez koder. Przykładowo w standardzie CD audio próbki są pobierane 44 100 razy na sekundę, a każda ma 16-bitową głębię.

Szum:

Szum (nazywany też hałasem) to niepożądany (i czasem nieunikniony) dźwięk, który określa lub przesuwają granicę ciszy w zakresie głośności. Generują go wszystkie elementy łańcucha audio, począwszy od rejestrowanego źródła (np. wentylator włączony w pomieszczeniu) przez mikrofon (np. szum własny, drgania, wiatr) i okablowanie (np. zakłócenia, przesłuch) po urządzenie przechwytyjące (np. szum własny, szum próbkowania cyfrowego), których połączenie tworzy tzw. podłogę szumową.

Szum zazwyczaj określa się parametrem SNR (signal-to-noise ratio – stosunek sygnału do szumu), oznaczającym cały zakres od określonego poziomu (czasem jest to najgłośniejszy dźwięk możliwy do uzyskania w systemie) do podłogi szumowej.

Odpowiednikiem szumu w systemach wizyjnych jest szum wizyjny, obserwowany jako „śnieżenie” złożone z losowych (zazwyczaj) statycznych pikseli. Ogranicza on widoczność ciemnych obrazów (podobnie jak szum dźwiękowy ogranicza słyszalność cichych dźwięków).

Zniekształcenia:

Wszelkie niepożądane zmiany sygnału w stosunku do jego pierwotnej „prawdziwej” postaci są nazywane zniekształceniami. (Omówionego powyżej szumu zazwyczaj nie zalicza się do zniekształceń). Zniekształcenia pogarszają subiektywną jakość (choć bywają zniekształcenia o „ładnym” brzmieniu) i przysłaniają obiektywną zawartość informacyjną, powodując gorszą słyszalność sygnału, zwłaszcza podczas analizy treści, oraz ograniczają funkcjonalność analizy.

Do parametrów często używanych w ilościowej analizie zniekształceń należą THD (total harmonic distortion – współczynnik zawartości harmonicznym) i IMD (inter-modulation distortion – zniekształcenia intermodulacyjne).

Odpowiednikiem zniekształceń w systemach wizyjnych są artefakty, takie jak aberracja chromatyczna, winietowanie, rozmycie itp., które sprawiają, że obraz wygląda „źle” i jest mniej szczegółowy.

Częstotliwość próbkowania i charakterystyka przenoszenia:

W systemie cyfrowym dźwięk jest próbkowany określoną liczbą razy w ciągu sekundy. Jest to częstotliwość próbkowania, która zazwyczaj wynosi od 8000 do 48 000 razy na sekundę (czyli herców). Teoria sygnałów (a konkretnie twierdzenie Nyquista-Shannona) mówi, że aby odwzorować dźwięk analogowy z dostateczną szczegółowością, częstotliwość próbkowania musi być co najmniej dwukrotnie wyższa od najwyższej pożądanej lub wymaganej częstotliwości sygnału analogowego.

W zależności od wieku i innych czynników normalne ludzkie ucho słyszy częstotliwości od 20 Hz do około 15–20 kHz. Z grubsza biorąc, pasmo niskich częstotliwości, powyżej kilkuset herców, często stanowi podstawę określonych dźwięków (np. zasadniczą składową głosu ludzkiego), natomiast pasmo wyższych częstotliwości, powyżej kilku tysięcy herców, zawiera więcej „szczęgółów”.

Zakres częstotliwości dźwięku odpowiada rozdzielczości i poklatkowości w systemach wizyjnych: im niższa wartość tego ustawienia, tym mniejsza ilość uzyskiwanych szczegółów.

Głębina bitowa:

Podczas każdego próbkowania dźwięku określona wartość analogowa zostaje przechwycona i przekształcona na swój cyfrowy odpowiednik. W świecie cyfrowym nie ma wielkości nieskończonych, więc ilość szczegółów jest ograniczona do określonej głębi bitowej. Każdy bit reprezentuje współczynnik równy dwa (0 lub 1, wartość niska lub wysoka itp.), który w połączeniu z określonym zakresem amplitud (*np.* wybrane napięcie lub poziom ciśnienia akustycznego) pozwala uzyskać ułamki tego zakresu. Dwa bity dają cztery ułamki, trzy bity – osiem ułamków itp. Mówiąc w uproszczeniu, sygnał o napięciu jednego wolta poddany próbkowaniu przy użyciu trzech bitów zostanie podzielony i przedstawiony w postaci stopni o wielkości 1/8 wolta.

Do uzyskania dostatecznej jakości dźwięku zazwyczaj wystarcza 16 bitów (reprezentujących 65 536 stopni), przynajmniej w przypadku ludzkiego ucha. Tak jest w standardzie CD audio. W analizie i bardziej wymagających zastosowaniach lepiej sprawdzają się 24 bity.

Odpowiednikiem głębi bitowej w systemach wizyjnych jest kontrast, czyli zakres luminancji i chrominancji możliwy do odwzorowania w każdym pikselu.

O firmie Axis Communications

Axis umożliwia tworzenie mądrzejszego i bezpieczniejszego świata, tworząc rozwiązania zwiększające bezpieczeństwo i wydajność biznesową. Jako firma z branży technologicznej będąca liderem na rynku, Axis oferuje systemy dozoru wizyjnego, kontroli dostępu, domofonowe i rozwiązania audio. Rozwiązania te są wzbogacone o inteligentne aplikacje analityczne i wysokiej jakości szkolenia

Firma Axis zatrudnia około 4000 zaangażowanych pracowników w ponad 50 krajach i współpracuje z partnerami z sektora technologii oraz integracji systemów na całym świecie, aby dostarczać rozwiązania dla klientów. Firma Axis powstała w 1984 roku, a jej siedziba znajduje się w Lund w Szwecji