

# Le potentiel de l'intégration de l'audio

Enregistrer et traiter les sons pour percevoir correctement  
une scène et récolter des preuves

Juin 2021

# Table des matières

1	Avant-propos	3
2	Introduction	4
	2.1 Capturer sans enregistrer	4
3	Surmonter les obstacles	4
	3.1 Que disent les lois et la réglementation ?	5
	3.2 Étudier la possibilité	5
4	Importance de l'installation	6
5	Préparation de l'audio	7
6	Topologie des analyses	8
7	Cas d'usage et exemples	9
	7.1 Détecter les incidents à l'aide des analyses audio	9
	7.2 Visualiser le son dans la vidéo	10
	7.3 Écouter et interagir	11
	7.4 Enregistrer et stocker	12
	7.5 Profitez davantage de votre système de surveillance	12
8	Contrôle et détection	12
	8.1 Caractéristiques du son	12
	8.2 Traitement du signal	13
	8.3 L'ouïe humaine	14
9	Clause de non-responsabilité	15
Annexe 1	Terminologie de la qualité audio	16

# 1 Avant-propos

Les fonctionnalités de capture audio, qu'elles soient intégrées et prêtes à l'emploi dans une caméra vidéo ou fournies par un microphone externe, permettent divers cas d'usage importants. Une utilisation professionnelle et responsable de l'audio intégré peut ajouter une valeur et des avantages cruciaux à une installation de sécurité. Il peut, par exemple, fournir une preuve manquante lors d'une enquête criminelle ou permettre la détection en temps réel d'événements nécessitant une attention immédiate des agents de sécurité ou du personnel hospitalier. La simple présence de la surveillance audio peut également avoir des effets de dissuasion et éviter des crimes et délits.

La capture audio (souvent combinée à l'action des analyses instantanées) peut être déployée en tant que technologie autonome, ce qui permet plusieurs cas d'usage dans les domaines de la prévention des crimes et délits, de la protection et des enquêtes judiciaires.

Mais combinée à la vidéo, la capture audio a également le potentiel de renforcer la grande majorité des cas d'usage de surveillance existants. Par exemple, les opérateurs de sécurité peuvent avoir un bien meilleur aperçu des événements d'une scène si leur flux vidéo est complété par un flux audio.

Tout comme vous pouvez employer plusieurs types d'analyses *vidéo* pour la détection automatique d'événements et le lancement d'alarme basé sur la détection visuelle, les analyses *audio* peuvent surveiller les flux audio et réagir lorsqu'un événement se démarque.

Les logiciels d'analyse audio peuvent être configurés pour déclencher des alarmes automatiques et d'autres actions lorsqu'un microphone détecte des sons associés au cri de personnes, à du bris de verre ou à des coups de feu. Cet avertissement rapide permet des réponses et une intervention rapides.

En matière d'analyses audio, il peut également s'agir de détecter si un son inattendu vient de la droite ou de la gauche et rediriger automatiquement une caméra PTZ vers la source sonore. Dans un hôpital ou un centre de soins, les analyses audio peuvent être utilisées pour détecter des niveaux sonores élevés indiquant qu'un patient est en détresse et envoyer une notification automatique à une infirmière. Ce cas d'usage peut également tirer profit des analyses de visualisation audio qui facilitent la surveillance simultanée des sons depuis plusieurs emplacements.

Il existe une différence entre la capture des sons et leur enregistrement. Pour de nombreux types d'utilisation, il n'est pas nécessaire d'enregistrer l'audio pour atteindre l'objectif fixé, et cela peut permettre de gérer les problèmes de confidentialité et de se conformer à la réglementation relative aux données à caractère personnel. En général, les applications d'analyse audio n'enregistrent pas le son en continu. Généralement, elles traitent uniquement l'audio entrant afin de rechercher des modèles, niveaux ou fréquences spécifiques. Lorsque les analyses sont réalisées en périphérie de réseau (dans la caméra), aucune donnée audio numérique n'a besoin de quitter la caméra, seuls les résultats provenant des analyses réalisées, c'est-à-dire les métadonnées ou les déclencheurs, sont envoyés.

Axis ne fournit aucun conseil de nature légale. Les lois qui régulent la surveillance varient d'une région, d'un état et d'un pays à l'autre et il appartient à l'utilisateur des produits (en général le client final) de s'assurer que la surveillance est réalisée conformément à la réglementation en vigueur. Tout comme pour la vidéosurveillance, l'installation de la surveillance audio doit être précédée par un examen et la bonne compréhension des aspects légaux d'une telle installation.

Une fois que les mesures nécessaires ont été prises pour respecter les conditions légales, il faut soigneusement étudier l'installation en termes de disposition et de configuration de l'équipement. Cela peut nécessiter un peu d'organisation mais cela ne s'avère pas très compliqué et de simples mesures peuvent en faire beaucoup pour obtenir la facilité d'utilisation de l'audio.

## 2 Introduction

Les informations audio peuvent représenter un atout précieux pour la prévention des crimes et délits, pour la protection et pour un usage judiciaire. L'audio capturé peut également être traité en temps réel par un logiciel d'analyse qui permet un contrôle audio très efficace pour la détection d'activités, de comportement ou d'événements.

Ce livre blanc décrit le potentiel de l'audio dans la sécurité en s'appuyant sur des exemples de cas d'usage typiques. De nombreux types d'analyses audio sont présentés, ainsi que de brefs aperçus de leur fonctionnement.

Ce livre ne fournit aucun conseil de nature juridique, mais présente les différentes solutions techniques qui peuvent s'avérer utiles pour la configuration d'une installation. En fonction de la façon dont vous choisissez de mettre en place les analyses audio, il peut être possible de se frayer un chemin dans la législation et les recommandations régionales et utiliser cet outil puissant où vous en avez besoin.

La portée du présent livre est limitée à la *capture et l'enregistrement éventuel* de l'audio, c'est-à-dire *l'entrée* audio. Une autre utilisation habituelle de l'audio dans les solutions de sécurité concerne la *diffusion* d'audio, c'est-à-dire *la sortie* audio, généralement pour la lecture de messages vocaux ou d'alarmes pour dissuader les intrus et les voleurs. Vous trouverez de plus amples informations sur la diffusion d'audio dans le domaine de la sécurité à l'adresse [www.axis.com/products/audio](http://www.axis.com/products/audio).

### 2.1 Capturer sans enregistrer

Il est possible de capturer et d'utiliser l'audio sans l'enregistrer. Capturer l'audio signifie simplement le numériser et le rendre disponible dans un logiciel pour utilisation. Cette action est réalisée par l'enregistrement des vibrations sonores dans l'air à l'aide d'un microphone, puis la conversion de ces signaux analogiques en signaux numériques (à l'aide d'un équipement de conversion analogique/numérique) et le passage dans une unité de traitement.

Si l'audio capturé n'est pas placé sur un support permanent tel qu'une clé USB ou un disque dur, il n'est pas enregistré. Les enregistrements ne sont pas forcément nécessaires dans certains cas d'usage, tels que lorsque l'opérateur humain écoute l'audio capturé en temps réel. Dans certaines situations, il existe même des raisons précises de ne *pas* enregistrer l'audio. Il peut exister des différences de restrictions juridiques en fonction de la nature de l'audio, qu'il soit enregistré ou simplement capturé.

En général, les analyses audio n'enregistrent pas le son en continu. Pour fonctionner correctement, elles doivent temporairement mettre le son en mémoire tampon. De nombreux systèmes peuvent être configurés pour enregistrer ce qui a été mis en mémoire tampon juste avant ou après une détection afin de permettre à la sécurité de vérifier la détection et éventuellement, conserver le son dans le cadre d'une enquête judiciaire.

## 3 Surmonter les obstacles

Nombreux sont ceux qui sont inquiets de l'utilisation de microphones dans les situations de vidéosurveillance. Ces préoccupations sont généralement liées à l'enregistrement de conversations avec la vidéo.

Il est possible de dépasser cet obstacle initial si on comprend que l'intégration de l'audio offre bien plus de possibilités que le simple enregistrement. Il existe de nombreux cas d'usage pour lesquels il n'est pas nécessaire d'enregistrer des informations sonores.

Les lois qui régulent la surveillance varient d'une région à l'autre et d'un pays à l'autre, il faut donc s'assurer de bien savoir ce qui est autorisé avant d'ajouter l'audio à votre système de surveillance.

### **3.1 Que disent les lois et la réglementation ?**

Tout comme pour la vidéosurveillance, l'installation de la surveillance audio doit être précédée par un examen et la bonne compréhension des aspects légaux d'une telle installation. Le cas échéant, les documents de demande appropriés doivent être envoyés et les permis obtenus. Des panneaux ou des déclarations publiques peuvent être utilisés lorsqu'ils sont exigés.

L'utilisation et/ou l'enregistrement de l'audio peuvent être interdits par la législation nationale ou divers types de réglementation locale ou nécessiter un examen spécial pour plusieurs raisons au regard de ces lois et de cette réglementation. Tandis qu'une région ou un environnement pourrait autoriser la capture audio, elle/il pourrait interdire les enregistrements audio. Les entreprises peuvent également interdire l'utilisation de la surveillance audio au sein de leurs locaux.

#### **3.1.1 Exemples américains**

Les lois et la réglementation aux États-Unis varient en fonction des états.

Certains états exigent le consentement unilatéral pour l'enregistrement audio. Cela signifie qu'une seule partie impliquée dans une conversation doit donner son consentement à la surveillance pour que cela soit légal.

Les autres états exigent le consentement de toutes les parties (ou deux parties), ce qui signifie que toutes les parties doivent donner leur consentement à être enregistrées avant que l'enregistrement puisse avoir lieu. Certaines exceptions à ce consentement de toutes les parties peuvent s'appliquer dans certains espaces publics lorsqu'une conversation ne peut être complètement privée.

L'examen légal d'une situation d'utilisation d'une application d'analyses audio qui n'enregistre pas l'audio peut également aboutir à un autre résultat dans certaines régions. Vous devez donc rechercher les lois et la réglementation qui s'appliquent dans votre état en particulier.

#### **3.1.2 Exemples européens**

La surveillance audio est régulée par des lois nationales dans les pays européens. Vous devez donc rechercher les lois et la réglementation qui s'appliquent dans votre pays en particulier.

Les enregistrements audio peuvent contenir des données à caractère personnel qui sont soumises au RGPD (Règlement général sur la protection des données). Le RGPD n'interdit pas forcément les enregistrements audio mais la capture ou l'enregistrement d'audio doit faire l'objet d'un examen particulier. Lorsque vous ajoutez l'audio à votre vidéosurveillance existante, vous devez tenir compte du fondement juridique de votre pays en matière de traitement des données à caractère personnel dans le respect du RGPD.

### **3.2 Étudier la possibilité**

Il existe une idée fausse généralement répandue que l'audio n'est jamais autorisé en surveillance. Cette idée fausse est si répandue que dans de nombreux cas, la possibilité de renforcer une installation de surveillance par l'audio n'est même jamais envisagée.

Mais de nombreux types d'installations peuvent être autorisés, par exemple, si les personnes sont informées, si vous disposez de leur consentement, etc. Vous devez vous renseigner sur les lois et réglementations qui s'appliquent dans votre région ou pour votre cas d'usage. Même si le cas d'usage *Enregistrer et stocker*

n'était pas autorisé dans votre installation de sécurité, de nombreux cas d'usage pourraient être adaptés afin de ne pas enfreindre les droits relatifs à la confidentialité, comme *Écouter et interagir*, *Écouter et témoigner* et *détecter les incidents à l'aide des analyses audio*.

## 4 Importance de l'installation

Le positionnement du microphone dans une scène définit les éventuelles applications. Avant d'installer un équipement audio, sa disposition et sa configuration doivent être soigneusement étudiées. Cela peut nécessiter un peu d'organisation mais cela ne s'avère pas très compliqué et de simples mesures peuvent en faire beaucoup afin d'améliorer la facilité d'emploi de l'audio.

En réfléchissant à la bonne disposition d'un microphone et en choisissant un bon emplacement d'un point de vue acoustique, vous augmenterez la probabilité d'atteindre vos objectifs de surveillance. Un microphone doit, bien sûr, être placé de telle sorte qu'il puisse capturer facilement les sons appropriés. Les dispositions habituelles d'un microphone se trouvent au milieu d'une pièce, associé à une caméra, ou à proximité d'un point où des événements spécifiques pourraient avoir lieu. Un microphone ne doit pas être placé à proximité d'une source de bruit, telle qu'une ventilation ou des machines. Ces dernières pourraient prendre le dessus sur les sons plus faibles ou provenant d'une source plus éloignée.



*Disposition habituelle d'un microphone*

- 1 *Emplacement où les actions dignes d'intérêt peuvent se produire*
- 2 *Dans une caméra*
- 3 *Au centre de la pièce*

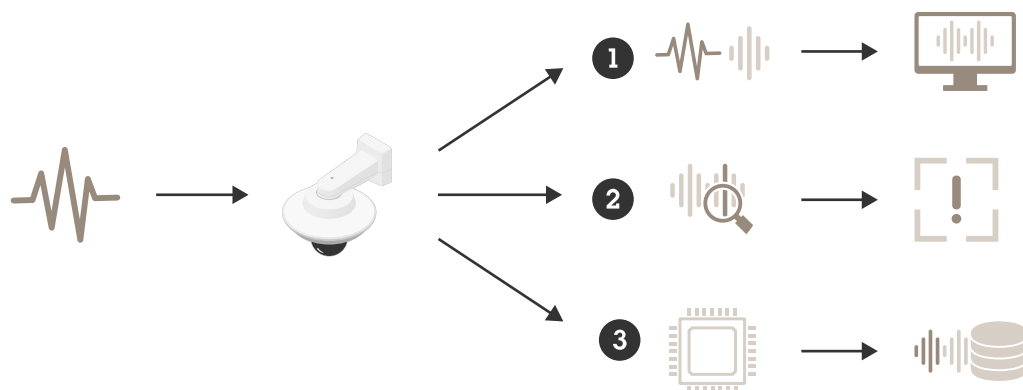
L'environnement acoustique, tel que les propriétés d'absorption du son des murs ou des plafonds/sols et les complexités dimensionnelles (comme de très longs couloirs), produira différentes réverbérations et différents échos qui peuvent grandement affecter le champ sonore à certains emplacements. Par exemple, une voix produira un son très différent dans un espace très atténué (tel qu'une salle de conférence traitée acoustiquement) par rapport au son produit dans une église ou une salle de bains entièrement carrelée. Dans les situations difficiles d'un point de vue acoustique, la disposition du microphone peut devenir cruciale.

L'installation et la configuration (par exemple le réglage du gain basse fréquence) de l'équipement sont importants, ainsi que l'intégration de l'équipement audio dans le système de surveillance. Les installateurs et intégrateurs de systèmes peuvent fournir des recommandations pour les situations particulières.

Pour les analyses audio, des recommandations particulières s'appliquent parfois, elles peuvent différer des recommandations pour les enregistrements audio en général. Toujours étudier le manuel de l'utilisateur pour connaître les prérequis qui s'appliquent.

## 5 Préparation de l'audio

Après la capture initiale de l'audio, les informations capturées sont préparées pour les étapes suivantes de traitement. Les différentes préparations peuvent être réalisées en parallèle ou en exclusivité.



1 *Transformation*

2 *Analyses en périphérie de réseau en temps réel*

3 *Traitement et cryptage*

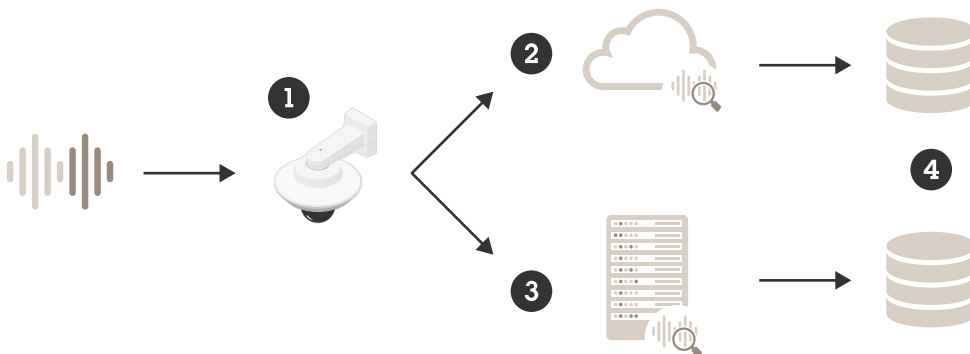
- **Transformation.** Le son est rendu abstrait et converti, par exemple, en informations visuelles telles qu'un graphique montrant le spectre sonore. Ce procédé ne peut être annulé : vous ne pouvez pas récupérer le son d'origine à partir du graphique du spectre.
- **Analyses en périphérie de réseau en temps réel.**  
Un **classificateur de son** peut être utilisé si le son est traité en périphérie de réseau. Il en résultera des métadonnées qui décrivent les caractéristiques du son. Le son d'origine ne peut être recréé à partir de ses métadonnées.  
Un **détecteur de son** peut être utilisé pour reconnaître les modèles, les niveaux ou les fréquences et fournir des informations relatives au statut. À nouveau, le son d'origine ne peut être restauré.
- **Traitement et cryptage.** Pour les cas dans lesquels l'audio d'origine sera utilisé (non transformé ou analysé), un traitement et un cryptage sont normalement réalisés pour préparer les données audio en vue des cas d'usage prévus. Ces cas d'usage peuvent impliquer le stockage de données audio en

périphérie de réseau, la diffusion à des clients extérieurs pour traitement complémentaire (sur serveur ou cloud) ou le stockage externe.

## 6 Topologie des analyses

L'emplacement du moteur d'analyse dans le système est important pour de nombreuses raisons. En particulier pour la gestion des questions de confidentialité et la conformité avec les réglementations relatives aux données à caractère personnel, il est important de savoir *où* le logiciel d'algorithmes analyse les données audio. Il existe des situations dans lesquelles les données audio ne peuvent être envoyées sur le réseau et il est essentiel que les données audio capturées (mais non stockées) puissent être analysées localement. Si des algorithmes à très forts besoins en calcul sont nécessaires, de telle sorte qu'ils ne peuvent être exécutés en périphérie de réseau, il pourrait être nécessaire d'envoyer les données audio numériques sur le cloud ou un serveur.

- **Analyses en périphérie de réseau.** Lorsque les analyses s'exécutent en périphérie de réseau, aucune donnée audio numérique n'a besoin de quitter la caméra. Dans le cas d'une capture audio sans enregistrement, seuls les résultats de l'analyse exécutée, c'est-à-dire les métadonnées ou les déclencheurs, seront envoyés.
- **Analyses sur serveur.** Lorsque les analyses sont exécutées sur un serveur, l'audio numérique doit quitter la caméra. Si elles sont prétraitées sur la caméra (en périphérie de réseau), ces données pourraient être des métadonnées dissociées ou dépersonnalisées. Un serveur est normalement une partie d'un système fermé (contrôlé par un propriétaire), aussi les questions de confidentialité de l'audio transféré peuvent être gérées. Néanmoins, il faut s'assurer que les lois et réglementations en vigueur sont respectées.
- **Analyses sur le cloud.** L'audio numérique peut également être transféré sur un serveur dans un contexte de cloud. Comme dans le cas des analyses sur serveur, les informations audio peuvent être prétraitées en métadonnées. L'utilisation du cloud est souvent décentralisée, il est par conséquent d'autant plus important de s'occuper des problèmes de confidentialité et de veiller à la conformité avec la réglementation.

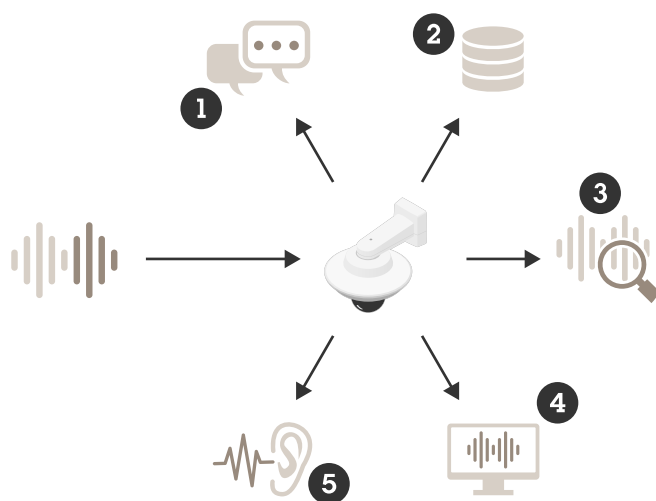


- 1 Périphérie de réseau
- 2 Cloud
- 3 Serveur
- 4 Stockage



## 7 Cas d'usage et exemples

La fonctionnalité audio est souvent intégrée et prête à l'emploi dans les caméras vidéo. Il existe de nombreux cas d'usage pour lesquels une utilisation responsable et professionnelle des fonctionnalités audio intégrées peut offrir une plus-value essentielle et de nombreux avantages potentiels. Il peut, par exemple, être utilisé pour présenter une preuve manquante lors d'une enquête criminelle ou permettre la détection en temps réel d'événements nécessitant une attention immédiate des agents de sécurité ou du personnel hospitalier. La présence de la surveillance audio peut également avoir des effets de dissuasion et éviter des crimes et délits.



Objectifs type de la capture audio :

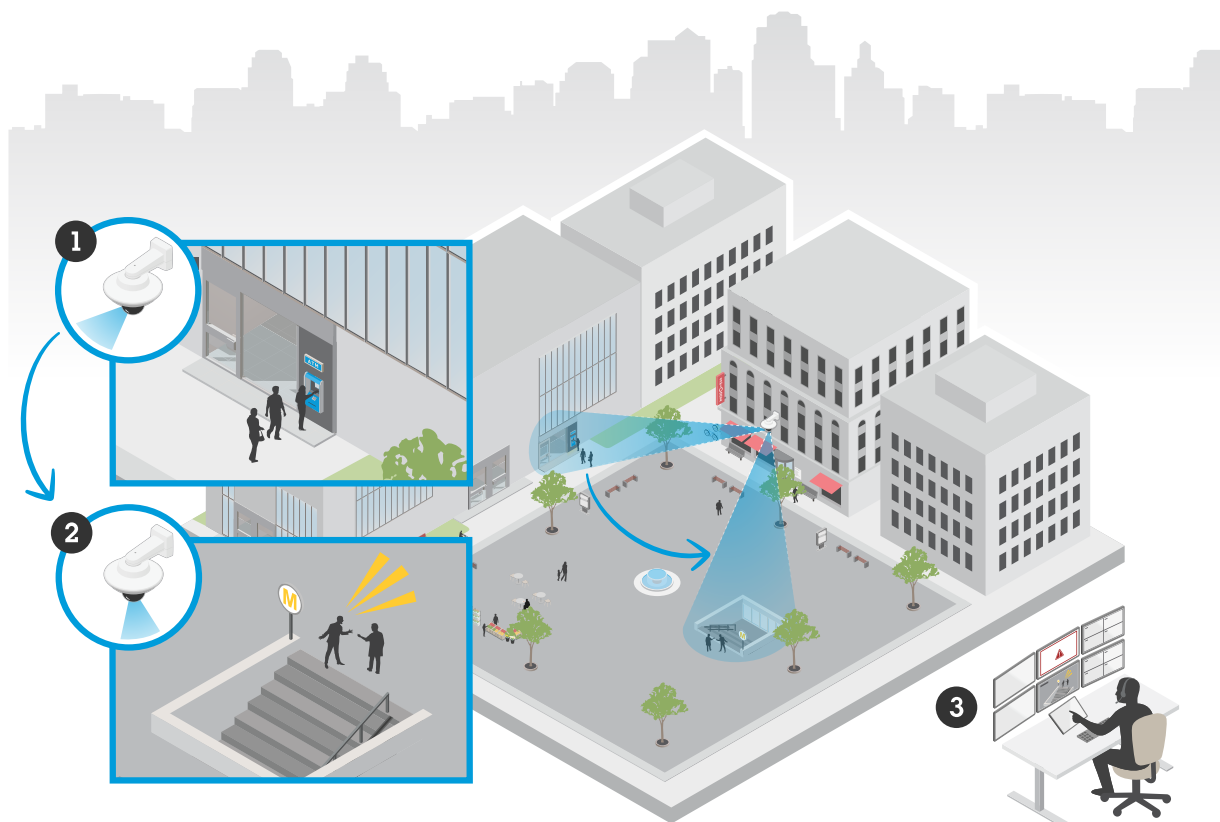
- 1 Communiquer
- 2 Enregistrer
- 3 Analyser
- 4 Visualiser
- 5 Écouter

### 7.1 Détecter les incidents à l'aide des analyses audio

Les applications d'analyses audio sont des programmes logiciels qui traitent l'audio capturé afin de trouver et d'extraire des informations spécifiques. Elles sont utilisées pour détecter des événements tels que des coups de feu, le bris de glace ou les agressions. Elles peuvent, par exemple, traiter les entrées audio pour répondre par oui ou par non à la question « Une fenêtre a-t-elle été cassée ? ». En cas de détection, le système envoie généralement une notification automatique au personnel par une alerte visuelle ou le déclenchement d'une alarme. Cet avertissement rapide permet des réponses et une intervention rapides.

### 7.1.1 Rediriger une caméra

Un autre exemple d'application de l'analyse audio est la redirection de caméra PTZ. Elle combine les fonctionnalités audio et vidéo en détectant l'origine de l'audio et en redirigeant automatiquement la caméra vers la source du son.



- 1 La caméra PTZ contrôle un distributeur automatique.
- 2 Le microphone de la caméra perçoit un fort bruit soudain et la caméra se redirige instantanément vers l'incident.
- 3 L'opérateur reçoit une alarme et vérifie l'incident.

### 7.2 Visualiser le son dans la vidéo

Le son capturé dans une vidéo peut être visualisé et affiché en tant que diagramme du spectre sonore sur un moniteur. Si un seuil défini est dépassé, le diagramme commencera à indiquer une alarme.

Une telle visualisation du son peut être profitable dans les situations où vous devez contrôler en même temps les sons provenant de plusieurs sources, par exemple, de plusieurs chambres de patients à l'hôpital. Alors qu'il serait trop difficile d'écouter simultanément beaucoup de sources audio (le son provenant de plusieurs pièces), il serait bien plus facile d'afficher de nombreuses visualisations sur un moniteur dans le

bureau des infirmières. Si les flux vidéo provenant des pièces sont disponibles, les visualisations peuvent être ajoutées en incrustation sur l'image vidéo.



Visualisation sonore ajoutée en incrustation aux flux vidéo à l'hôpital.

## 7.3 Écouter et interagir

La surveillance audio avec interaction directe d'un opérateur pour une meilleure perception de la scène représente le cas d'usage probablement le plus basique et intuitif. Parmi les exemples usuels, on trouve la perception d'une conversation suspecte et l'envoi un agent de sécurité sur place afin de chercher à en savoir plus. Ou bien, dans un hôpital ou un centre de soins, entendre qu'un patient est en détresse et faire appel à une infirmière. Il pourrait également s'agir de détecter si un son « étrange » vient de la droite ou de la gauche afin de pointer une caméra PTZ vers la source sonore.

Ces cas d'usage impliquent un ou plusieurs opérateurs disposant d'un accès à l'environnement audio depuis une salle de contrôle ou via une application de sécurité sur un téléphone portable. L'oreille humaine capture les sons et le cerveau extrait les informations pertinentes de la scène ou de la situation. Associé à la vidéosurveillance, l'audio ajoute une autre dimension d'informations pour la prise de décisions. Dans certains cas, l'audio s'avèrera être la seule dimension, si la source audio est extérieure au champ de vision de la caméra ou si les conditions d'éclairage sont difficiles.

### 7.3.1 Écouter et témoigner

La surveillance audio peut également être utilisée aux fins de témoignage direct relatif à des événements auxquels le témoin a assisté (qu'il a entendu). Ce cas d'usage diffère de *Écouter et interagir* car l'objectif ici n'est pas la prise de décision, mais ces cas d'usage coexistent souvent. Par exemple, lorsqu'il entend une dispute qui dégénère avec des paroles compromettantes, un opérateur peut non seulement envoyer des gardes mais également par la suite, témoigner de ce qu'il a entendu.

## 7.4 Enregistrer et stocker

Le cas échéant, le cas d'usage Capter et enregistrer des données audio peut fournir d'importantes preuves supplémentaires. Il peut s'agir de paroles compromettantes ou d'un coup de feu. L'audio enregistré peut fournir des preuves sur l'auteur de certaines paroles, sur le nombre de coups de feu tirés ou sur des événements similaires d'intérêt sur le plan judiciaire.

Lorsque l'audio est enregistré dans un contexte judiciaire, le plus grand soin doit être apporté à conserver les données d'origine et éviter tout traitement (qui, dans d'autres contextes, pourrait être nécessaire ou bénéfique). Dans le cadre des enregistrements judiciaires, tout type de traitement pourrait être considéré comme une falsification de preuves. Des algorithmes d'amélioration de la voix peuvent être utilisés pour augmenter l'audibilité de paroles enregistrées, cela pourrait améliorer la valeur judiciaire. Mais ces algorithmes doivent être appliqués a posteriori, sur une copie de l'enregistrement. En conservant l'enregistrement aussi peu traité que possible, on garde toutes les options ouvertes pour l'utilisation ultérieure des données enregistrées.

## 7.5 Profitez davantage de votre système de surveillance

Les systèmes de surveillance intègrent souvent plusieurs types de capteurs. Le capteur d'image de la caméra est l'un de ces capteurs. Bien entendu, il enregistre l'aspect visuel d'une scène. Des capteurs non visuels sont également généralement utilisés, tels que les détecteurs de mouvement basés sur la technologie du radar ou les émissions de radiation infrarouge. Parfois, la vidéosurveillance n'est pas appropriée, les capteurs non visuels sont alors utilisés comme des appareils autonomes. En revanche, dans de nombreux cas, les capteurs non visuels sont utilisés pour compléter l'installation de caméra en ajoutant d'autres types d'informations.

En employant également des capteurs audio (microphones) dans une installation de surveillance, la grande majorité de tous les cas d'usage possibles sont renforcés. L'ajout de la fonctionnalité audio à un système sans audio permet l'interaction multi-capteur, soit via les analyses ou via l'interaction de l'opérateur.

Le cas d'usage *Écouter et interagir* est un exemple simple, dans lequel l'opérateur dispose d'un bien meilleur aperçu des événements d'une scène lorsqu'il reçoit également un flux audio. Il peut être difficile de détecter un comportement agressif en observant simplement les gens, mais il sera bien plus facile à déterminer si vous pouvez également entendre ce qu'il se passe.

Un autre exemple typique est l'utilisation des analyses vidéo, telles que la détection de mouvements vidéo. Si l'application d'analyse est remise en question par une faible luminosité par exemple, la présence des analyses audio peut augmenter l'indice de confiance de la détection.

# 8 Contrôle et détection

L'audio contient plusieurs types d'informations qui peuvent être utilisés pour le contrôle et les analyses audio. Divers types de traitement et de caractérisation aident à l'extraction et au peaufinage de ces informations pour une utilisation et une interaction plus aisées avec le système environnant.

## 8.1 Caractéristiques du son

Des caractéristiques telles que le volume et le ton peuvent contenir d'importantes informations dans un contexte de surveillance. La durée pendant laquelle il est audible, s'il se déplace ou s'il vient de loin ou de près, sont des exemples de pièces à ajouter au puzzle lorsque nous tirons des conclusions sur le son que nous entendons. Le matériel et le logiciel pour la détection et la surveillance de l'audio sont conçus

pour fonctionner avec les mêmes types d'informations, « écouter » des combinaisons complexes de caractéristiques allant du niveau de décibels à l'énergie à différentes fréquences au fil du temps.

- **Informations spatiales.** Il s'agit du monde physique qui nous entoure, y compris les concepts tels que l'emplacement, la direction et la distance. Les informations spatiales peuvent être utilisées pour se concentrer ou zoomer sur la capture audio dans différentes directions afin d'améliorer les enregistrements. Elles peuvent également être utilisées par les analyses pour déterminer de quelle direction provient un son ou la distance de la source de celui-ci.
- **Informations temporelles.** Les informations temporelles sont importantes à la fois pour le sens dynamique (elles changent avec le temps) et le sens absolu (quand un événement s'est-il produit ?), souvent observées en regard des informations provenant d'autres capteurs comme la vidéo. Les informations temporelles jouent un rôle important dans l'analyse comportementale : connaître la nature d'un événement, le moment de son apparition et sa durée.
- **Informations spectrales.** Il s'agit des fréquences, telles que la hauteur du ton d'un son ou la combinaison de tons dans les sons plus complexes. Les microphones utilisés en surveillance audio sont conçus pour avoir une réponse en fréquence uniforme, c'est-à-dire qu'ils essaient de capturer toutes les fréquences de la même manière dans une gamme audible (20 Hz – 20 kHz). Ce fonctionnement est différent de celui du système auditif humain, car nous pouvons détecter plus facilement ces fréquences que l'on trouve habituellement dans les paroles humaines, que les autres fréquences.
- **Informations relatives à l'amplitude.** Il s'agit de l'intensité et du volume d'un son. Les informations d'amplitude peuvent compléter les informations spectrales et être utilisées conjointement pour dresser un portrait de la structure de l'audio entrant.

## 8.2 Traitement du signal

Au sein de la surveillance audio, en termes de traitement du signal il s'agit généralement d'améliorer la transmission, l'efficacité de stockage, la qualité subjective, ou pour accentuer ou détecter les composants dignes d'intérêt. Cette opération est réalisée grâce à des logiciels d'algorithmes qui modifient ou analysent l'audio de diverses façons.

### 8.2.1 Modifier des signaux

Les algorithmes peuvent être utilisés pour modifier le signal dans un but précis, généralement pour :

- améliorer le signal, par exemple, augmenter l'audibilité par le contrôle automatique du gain.
- modifier le signal, par exemple, en changeant le contenu de fréquence relative à l'aide d'un égaliseur.
- limiter le signal en supprimant des fréquences ou amplitudes spécifiques. Et ce afin de réguler le volume de données grâce à la compression ou assurer la confidentialité par le brouillage de la voix.

### 8.2.2 Analyser des signaux

Les analyses audio utilisent les données audio capturées (mais normalement non enregistrées) et analysent les caractéristiques sonores pertinentes afin de générer des résultats non sonores. Les applications convertissent principalement les données audio en élément plus concret dans un autre format. Certaines applications d'analyse sont spécialement développées pour détecter, par exemple, les agressions, les coups de feu, le bris de glace ou des alarmes de voiture.

Si des algorithmes de machine learning sont utilisés, ils peuvent être formés à partir d'un grand nombre de données pour apprendre à faire des prédictions sans y avoir été particulièrement programmés. Un exemple dans un contexte audio pourrait être un algorithme qui peut détecter de façon fiable le son d'une porte qui se ferme après avoir été formé par des milliers de sons identiques.

### 8.3 L'ouïe humaine

L'oreille humaine est l'un des meilleurs outils disponibles pour détecter et analyser l'audio. Dans des environnements très bruyants, l'oreille et le cerveau humains peuvent détecter et interpréter des paroles lorsque la plupart des algorithmes ne pourraient pas réussir.

Grâce à nos oreilles, nous pouvons déduire des informations spatiales d'une scène, telles que l'origine d'un son et si la source de l'audio se déplace. Puisque nous avons deux oreilles, nous pouvons entendre si un son provient de la gauche ou de la droite ou d'un point entre les deux. Les oreilles et la tête sont aussi conçues de telle sorte que nous entendons si un son provient du dessus ou du dessous, de l'avant ou de l'arrière. Plusieurs « étapes de filtre » dans le cerveau fonctionnent avec des différences temporelles entre les oreilles, détectant instantanément des écarts aussi faibles que des microsecondes afin de détecter des types d'événements spécifiques. Nous disposons d'une capacité bien développée d'analyse du signal audio, particulièrement en matière de voix humaine mais également de sons associés à des dangers historiques.

Dans de bonnes circonstances (telles qu'une bonne qualité sonore, un son stéréophonique, un délai peu important), un opérateur humain peut être un puissant « outil d'analyse » et compléter le matériel ou le logiciel de détection. À l'aide d'un produit de surveillance audio équipé de seulement deux microphones, un opérateur peut déduire des informations spatiales d'une scène telle que l'origine d'un son et le mouvement de celui-ci.

## 9 Clause de non-responsabilité

Le présent document et son contenu sont offerts par Axis et tous les droits sur le document ou tout droit de la propriété intellectuelle s'y afférant (y compris mais sans s'y limiter les marques déposées, les noms commerciaux, les logos et les symboles similaires stipulés aux présentes) sont protégés par la loi et tous les droits, titre et/ou intérêt au présent document ou tout droit de la propriété intellectuelle s'y afférant sont et resteront acquis à Axis Communications AB.

Veuillez noter que le présent document est fourni « en l'état » sans garantie d'aucune nature à des fins d'informations uniquement. Les informations fournies dans le présent document ne constituent pas des conseils de nature légale et ne sont pas destinées à constituer des conseils de nature légale. Le présent document n'est pas destiné et ne devra pas être destiné à créer d'obligation légale de quelque nature que ce soit pour Axis Communications AB et/ou ses filiales. Les obligations d'Axis Communications AB et/ou de ses filiales relatives aux produits Axis sont soumises exclusivement aux conditions générales de l'accord entre Axis et l'entité qui s'est portée acquéreur desdits produits directement auprès d'Axis.

POUR ÉVITER TOUT DOUTE, TOUS LES RISQUES RELATIFS À L'UTILISATION, LES RÉSULTATS ET L'EXÉCUTION DU PRÉSENT CONTRAT SONT ASSUMÉS PAR L'UTILISATEUR DU DOCUMENT ET AXIS REJETTE ET EXCLUT, DANS LES LIMITES PRÉVUES PAR LA LOI, TOUTES LES GARANTIES, QUELLES SOIENT STATUTAIRES, EXPRESSES OU TACITES, Y COMPRIS MAIS SANS S'Y LIMITER TOUTE GARANTIE TACITE DE QUALITÉ MARCHANDE, D'ADÉQUATION À UN USAGE PARTICULIER, DE TITRE ET DE NON VIOLATION ET DE RESPONSABILITÉ ASSOCIÉE AUX PRODUITS, OU TOUTE GARANTIE RÉSULTANT DE TOUTE PROPOSITION, SPÉCIFICATION OU ÉCHANTILLON EN CE QUI CONCERNE LE PRÉSENT DOCUMENT.

# Annexe 1 Terminologie de la qualité audio

## Audio numérique :

L'audio numérique est une représentation de l'audio analogique (souvent un signal acoustique capturé par un microphone) enregistré sous forme numérique. En audio numérique, l'onde sonore du signal audio est généralement cryptée en une séquence continue d'échantillons numériques. La précision dépend du nombre de chiffres significatifs enregistrés par l'encodeur. Par exemple, dans un CD audio, les échantillons sont pris 44 100 fois par seconde, chacun avec une profondeur de 16 bit.

## Bruit:

Le bruit est un son indésirable (et parfois inévitable) qui définira ou limitera la fin silencieuse d'une gamme acoustique. Il est généré par toutes les pièces d'une chaîne audio, de la source enregistrée (par exemple, un ventilateur dans une pièce), en passant par le microphone (*par ex.* bruit propre, vibrations, vent) et le câblage (*par ex.* interférences, diaphonie) jusqu'au dispositif de capture (*par ex.* bruit propre, bruit d'échantillonnage numérique), tous ces éléments créant ce qu'on appelle généralement le plancher de bruit.

Le bruit est normalement défini par le SNR (rapport signal sur bruit), la gamme entière d'un niveau défini (parfois le son le plus fort qu'un système puisse gérer) au plancher de bruit.

*L'équivalent vidéo est le bruit vidéo, vu comme un modèle aléatoire de pixels (normalement) statiques, la « neige ». Il limite ce que vous pouvez voir dans les images sombres (tout comme il limite ce que vous entendez pour les signaux sonores).*

## Distorsion :

Toute modification indésirable d'un signal par rapport à la « vérité » d'origine est appelée distorsion (le bruit, comme expliqué ci-dessus, est normalement exclu des spécifications de la distorsion). La distorsion réduit la qualité subjective (normalement, il existe des distorsions qui sonnent « bien ») et dissimule les informations objectives, rendant le signal plus difficile à entendre, en particulier pour l'analyse de contenu, et réduit la fonctionnalité d'analyse.

Le taux de distorsion harmonique (THD, total harmonic distortion) et la distorsion d'intermodulation (IMD, inter-modulation distortion) sont deux propriétés normalement utilisées pour quantifier la distorsion.

*La distorsion est liée à la vidéo en tant qu'objets, tels que l'aberration chromatique, le vignettage, le flou, etc. Elle provoque une mauvaise qualité de l'image et limite le nombre de détails perceptibles.*

## Taux d'échantillonnage et réponse en fréquence :

Dans un système numérique, l'audio est échantillonné un nombre défini de fois par seconde. On appelle cela le taux d'échantillonnage (normalement compris entre 8 000 et 48 000 fois par seconde, ou Hz). Pour capturer correctement un son, la théorie du signal (plus particulièrement le théorème de Nyquist-Shannon) énonce que le taux d'échantillonnage doit être égal à au moins deux fois la fréquence requise ou souhaitée la plus élevée dans un signal analogique.

L'oreille d'un être normal perçoit des fréquences comprises entre 20 Hz et environ 15-20 kHz en fonction de l'âge et d'autres facteurs. Grosso modo, la gamme de basse fréquence, supérieure à des centaines de Hz, définit souvent la base de sons spécifiques (comme la base des voix), tandis que la gamme de haute fréquence, au-dessus de quelques milliers de Hz, contient plus de « détails ».

*La gamme de fréquence en audio est liée à la résolution et à la fréquence d'images en vidéo. Plus elle est faible, moins vous percevez de détails.*

## Profondeur de bits:



À chaque fois qu'un audio est échantillonné, une valeur analogique est capturée et traduite en représentation numérique. Dans le domaine numérique, l'infini n'existe pas, aussi le nombre de détails est limité à une profondeur de bits définie. Chaque bit représente un facteur de deux (0 ou 1, bas ou haut, etc.) qui, associé à une gamme d'amplitude définie (*par ex.* une tension ou un niveau de pression sonore choisi) crée des fractions de cette gamme. Deux bits produisent quatre fractions, trois bits en produisent huit etc. Pour simplifier, un signal d'un volt, échantillonné avec trois bits, sera fractionné et représenté en pas de 1/8 volt.

Pour une qualité audio suffisante, 16 bits sont normalement suffisants (ce qui représente 65 536 pas), du moins pour l'oreille humaine. C'est ce qu'utilise le CD audio. Pour les analyses ou les utilisations plus exigeantes, 24 bits sont plus pertinents.

*La profondeur de bits est liée au contraste en vidéo, la gamme de luminance ou chrominance que chaque pixel peut reproduire.*

# À propos d'Axis Communications

En concevant des solutions qui améliorent la sécurité et les performances de l'entreprise, Axis crée un monde plus clairvoyant et plus sûr. En tant qu'entreprise de technologie de réseau et leader de l'industrie, Axis propose des solutions de vidéosurveillance, de contrôle d'accès, d'interphonie et de systèmes audio. Les performances de ces solutions sont améliorées grâce à des applications d'analyse intelligentes et une formation de haute qualité.

Axis emploie près de 4 000 personnes dans plus de 50 pays et collabore avec des partenaires technologiques et d'intégration de systèmes dans le monde entier pour fournir des solutions clients adaptées. Axis a été fondée en 1984 et le siège social se trouve à Lund, en Suède.