DOCUMENTO TÉCNICO

El potencial de la captura de audio

Captura y procesamiento de sonidos para saber qué está pasando y obtener pruebas

Junio 2021

AXIS COMMUNICATIONS

Índice

1 **Resumen** 3

2 **Introducción** 4

2.1 Capturar sin grabar 4

3 **Sortear los obstáculos** 4

3.1 ¿Qué dicen las leyes y las regulaciones? 5

3.2 Investigue qué es posible 5

4 **La instalación es importante** 6

5 **Preparación del audio** 7

6 **Topología de la analítica** 7

7 **Casos de uso y ejemplos** 8

7.1 Detectar incidentes usando la analítica de audio 9

7.2 Visualizar el sonido en un vídeo 10

7.3 Escuchar e interactuar 11

7.4 Grabar y guardar 12

7.5 Saque más partido a su sistema de vigilancia 12

8 **Supervisión y detección** 12

8.1 Características del sonido 12

8.2 Procesamiento de la señal 13

8.3 La audición humana 14

9 **Nota legal** 15

Apéndice Terminología sobre la calidad de audio 16

1 Resumen

La captura de audio, ya sea como función integrada y lista para usar en una videocámara o facilitada por un micrófono externo, abre la puerta a varios casos de uso importantes. Una utilización responsable y profesional del audio capturado puede ser clave y aportar numerosas ventajas a una instalación de seguridad. Por ejemplo, puede ayudar a obtener esa prueba que falta en una investigación forense o permitir la detección en tiempo real de eventos que requieren una atención inmediata por parte de guardias de seguridad o del personal del hospital. Además, la vigilancia de audio puede tener un efecto disuasorio y evitar actos delictivos.

La captura de audio (a menudo combinada con una analítica instantánea) puede utilizarse de forma independiente y con finalidades muy concretas en el ámbito de la prevención de delitos, la protección y la investigación forense.

Sin embargo, si la combinamos con el vídeo, la captura de audio tiene también el potencial de reforzar la inmensa mayoría de casos de uso del sector de la vigilancia. Por ejemplo, los operadores de seguridad pueden hacerse una idea mucho más exacta de lo que ocurre en una escena si la transmisión de vídeo se complementa con información de audio.

Del mismo modo que utilizamos diferentes tipos de analítica de *vídeo* para la detección automática de eventos y la generación de alarmas basadas en la detección visual, la analítica de *audio* puede supervisar las transmisiones de audio y reaccionar cuando algo se sale de lo normal.

El software de analítica de audio puede configurarse para activar alarmas automáticas y otras acciones cuando un micrófono capta un sonido que pueda relacionarse con gente gritando, cristales rotos o disparos. Este sistema facilita una detección precoz y, por lo tanto, permite responder e intervenir con mayor rapidez.

La analítica de audio puede servir también para detectar si un sonido singular viene de la derecha o de la izquierda y redirigir una cámara PTZ automáticamente hacia el lugar de donde procede el sonido. En un hospital o residencia, la analítica de audio puede utilizarse para detectar niveles de sonido altos. Este tipo de sonido puede implicar que un paciente se encuentra mal, por lo que se envía una notificación automática al personal de enfermería. En este caso de uso concreto, también es útil la analítica de visualización del sonido, que permite supervisar el sonido desde varios sitios simultáneamente.

Existe una diferencia entre capturar sonidos y grabarlos. En muchos casos de uso, no es necesario grabar el audio para llegar al objetivo, lo que contribuye a disipar las inquietudes en materia de privacidad y cumplir con las regulaciones relativas a la información personal. En general, las aplicaciones de analítica de audio no graban de forma continua. Normalmente lo que hacen es procesar el audio recibido en busca de patrones, frecuencias o niveles concretos. Además, al utilizar la analítica en el extremo (en la cámara), no sale de la cámara ningún dato de audio digital, solo los resultados obtenidos, es decir, metadatos o activadores.

Axis no ofrece asesoramiento legal. Las leyes que regulan la vigilancia varían según la región, el estado y el país, y el usuario de los productos (normalmente el cliente final) es el responsable de comprobar que la vigilancia cumple con las disposiciones legales. Como ocurre en el caso de videovigilancia, antes de empezar con una instalación de vigilancia de audio es necesario investigar y entender las implicaciones legales.

Una vez aplicadas las consideraciones necesarias para cumplir con las condiciones legales, es necesario analizar meticulosamente la instalación para determinar el mejor sitio para colocar el equipo y su configuración. Este proceso requiere una cierta planificación pero en general no es complicado, y unas medidas sencillas pueden tener una gran repercusión en la usabilidad del audio.

2 Introducción

La información de audio puede resultar muy útil para objetivos como la prevención de delitos, la protección o las investigaciones. El audio capturado puede procesarse en tiempo real usando software de analítica, una opción muy práctica para detectar actividades, comportamientos o eventos mediante la supervisión por audio.

Este documento técnico explica el potencial del audio en el sector de la seguridad con ejemplos de los casos de uso más habituales. Además, se presentan distintos tipos de aplicaciones de analítica de vídeo, con una breve descripción sobre su funcionamiento.

Este documento no ofrece asesoramiento legal, pero presenta diferentes soluciones técnicas que pueden ser útiles a la hora de diseñar una instalación. Según los objetivos que persiga con la analítica de vídeo, quizás sea posible encontrar su encaje en las leyes y recomendaciones regionales para utilizar esta potente herramienta cuando sea necesario.

El tema que nos ocupa en este documento es la *captura y posible grabación de audio*, es decir, la *entrada* de audio. En el ámbito de las soluciones de seguridad, otra de las aplicaciones habituales del audio es la *emisión*, es decir, la *salida* de audio, normalmente mediante la reproducción de mensajes de voz o alarmas para disuadir a ladrones o personas que cruzan un perímetro delimitado. Encontrará más información sobre la emisión de audio en el sector de la seguridad en www.axis.com/products/audio.

2.1 Capturar sin grabar

Es posible capturar y usar el audio sin grabarlo. Capturar el audio significa digitalizarlo y ponerlo a disposición del software. Este proceso consiste en registrar las vibraciones de sonido en el aire usando un micrófono, convertir estas señales analógicas en señales digitales (usando un equipo de conversión A/D) y transferirlas a una unidad de procesamiento.

Si el audio capturado no se guarda en un soporte permanente, como una memoria flash o un disco duro, no se graba. En determinados casos de uso, las grabaciones son innecesarias, por ejemplo cuando un operador humano está escuchando en tiempo real el audio capturado. En situaciones concretas, puede haber incluso argumentos para *no* grabar el audio. Existen diferencias en las restricciones legales en función de si el audio se graba o solo se captura.

En general, las soluciones de analítica no graban el sonido de forma continua. Lo que hacen es guardarlo temporalmente para su correcto funcionamiento. Muchos sistemas pueden configurarse para grabar solo lo que se ha guardado antes y después de una detección para que el equipo de seguridad pueda verificarlo y, si es necesario, guardar el sonido como prueba.

3 Sortear los obstáculos

A mucha gente le preocupa el uso del micrófono en situaciones de videovigilancia. Normalmente lo que les preocupa es que se graben conversaciones junto con el material de vídeo.

Sin embargo, es fácil sortear este obstáculo inicial si entendemos que las posibilidades de la captura del audio van más allá de las grabaciones. Son muchos los casos de uso en los que no es necesario grabar ningún tipo de sonido.

Las leyes que regulan la vigilancia son diferentes en cada región y en cada país, por lo que es importante investigar qué está permitido antes de añadir el audio a un sistema de vigilancia.

3.1 ¿Qué dicen las leyes y las regulaciones?

Como ocurre en el caso de videovigilancia, antes de empezar con una instalación de vigilancia de audio es necesario investigar y entender las implicaciones legales. A veces hay que enviar las solicitudes pertinentes y obtener los permisos exigidos. También hay que colocar los avisos o carteles cuando sea necesario.

La utilización y/o la grabación del audio puede estar prohibida o sujeta a consideraciones especiales por varios motivos, según lo establecido por la legislación nacional o por varios tipos de normativas y reglamentos locales. Una misma región o espacio puede permitir la captura de audio pero prohibir las grabaciones. Las empresas pueden prohibir también la vigilancia de audio en sus instalaciones.

3.1.1 Ejemplos estadounidenses

En Estados Unidos, las leyes y las regulaciones son diferentes en cada estado.

En algunos se exige solo el consentimiento de una parte. Por lo tanto, la vigilancia es legal si una de las partes da su consentimiento.

Otros estados, en cambio, requieren el consentimiento de todas las partes (o de dos), es decir, todas las partes deben dar su consentimiento antes de que se produzca la grabación. Pueden existir excepciones al consentimiento de todas las partes en espacios públicos donde, por definición, no hay privacidad.

Desde un punto de vista legal, las conclusiones pueden ser diferentes en algunas regiones si se utiliza una aplicación de analítica de audio que no grabe el audio. Por lo tanto, tendrá qué investigar qué leyes y regulaciones se aplican en cada estado.

3.1.2 Ejemplos europeos

En Europa, la vigilancia a través del audio está regulada por leyes nacionales. Por lo tanto, tendrá qué investigar qué leyes y regulaciones se aplican en su país.

Las grabaciones de audio pueden incluir datos personales que están sujetos al RGPD (Reglamento General de Protección de Datos). El RGPD no necesariamente prohíbe las grabaciones de audio, aunque existen consideraciones especiales para la captura o la grabación de audio. Al incorporar el audio en su solución de videovigilancia actual, tiene que analizar si se mantienen los fundamentos legales para el procesamiento de los datos personales de conformidad con el RGPD.

3.2 Investigue qué es posible

Existe una falsa creencia y es que el audio nunca está permitido en el ámbito de la vigilancia. Está tan extendida que en muchos casos ni tan siquiera se estudia la posibilidad de reforzar una instalación de vigilancia incorporando el audio.

Sin embargo, la realidad es que en muchas instalaciones sí pueden tener su lugar, por ejemplo, si se informa o si se obtiene el consentimiento. Es necesario investigar qué legislaciones y regulaciones se aplican en su región y a su caso de uso particular. Aunque el caso de uso *grabar y guardar* no esté permitido en su instalación de seguridad, muchos casos de uso pueden adaptarse para no infringir los derechos de privacidad, como *escuchar e interactuar*, *escuchar y ser testigo* y *detectar incidentes usando la analítica de audio*.

4 La instalación es importante

El sitio donde está colocado el micrófono en una escena determina sus aplicaciones potenciales. Antes de instalar un equipo de audio, es importante estudiar meticulosamente su colocación y configuración. Este proceso requiere una cierta planificación pero en general no es complicado, y unas medidas sencillas pueden tener una enorme repercusión en la usabilidad del audio.

Si analiza cuál es el mejor sitio para colocar un micrófono y elige un lugar adecuado desde el punto de vista acústico, será más probable que alcance sus objetivos de vigilancia. Evidentemente, los micrófonos deben colocarse en lugares desde los que sea fácil capturar los sonidos relevantes. Entre los sitios elegidos con mayor frecuencia, encontramos el centro de una habitación, junto con una cámara, o cerca de lugares donde suelen producirse eventos concretos. Hay que evitar colocar los micrófonos cerca de una fuente de ruido, como equipos de ventilación o maquinaria, porque podrían ahogar sonidos débiles o lejanos.

Colocación habitual de los micrófonos

1. *Lugares donde se producen acciones de interés*
2. *En una cámara*
3. *En el centro de una habitación*

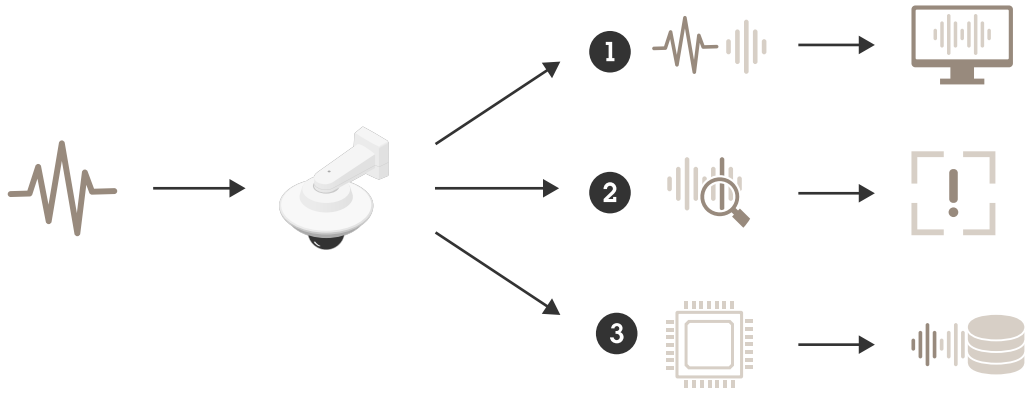
El entorno acústico, como las propiedades insonorizantes de paredes o techos/suelos y las complejidades relacionadas con las dimensiones (como pasillos muy largos), provoca unas reverberaciones y unos ecos concretos, que tienen un extraordinario impacto sobre el campo acústico en determinados sitios. Una voz, por ejemplo, suena muy diferente en una zona con un gran nivel de atenuación (como una sala de conferencias con tratamiento acústico) que en una iglesia o un baño con azulejos en todas las superficies. En situaciones complicadas desde un punto de vista acústico, la colocación del micrófono puede ser crucial.

Son importantes tanto la instalación como la configuración del equipo (por ejemplo, el ajuste de la ganancia de audio), al igual que su integración en el sistema de vigilancia. Los instaladores e integradores del sistema pueden dar recomendaciones según cada situación concreta.

En el caso de la analítica de audio, a veces habrá recomendaciones específicas, que serán diferentes de las directrices para las grabaciones de audio generales. Estudie siempre la documentación de cada proyecto para conocer los requisitos concretos.

5 Preparación del audio

Una vez capturado el audio, la información capturada se prepara para los siguientes pasos de procesamiento. Los diferentes preparativos pueden realizarse en paralelo o de forma independiente.



1 *Transformación*
2 *Analítica en el extremo en tiempo real*
3 *Procesamiento y codificación*

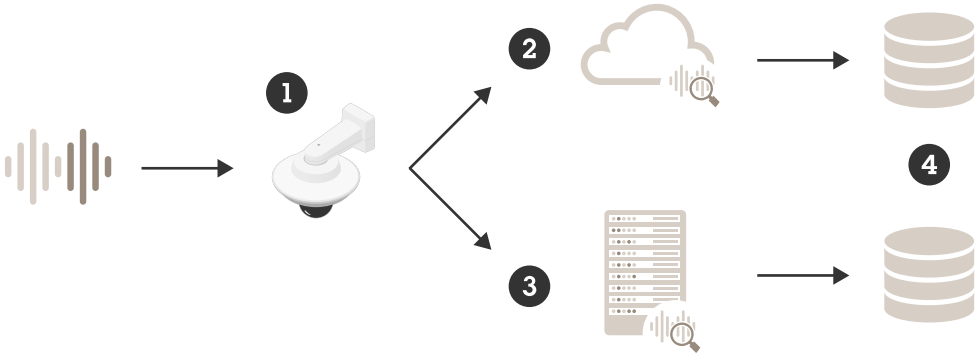
- **Transformación.** El sonido pasa a ser abstracto y se convierte, por ejemplo, en información visual, como un gráfico que muestra el espectro acústico. No es posible volver al punto de origen: es imposible recuperar el sonido original a través del gráfico del espectro.
- **Analítica en el extremo en tiempo real.**
  Se puede utilizar un **clasificador de sonido** si el audio se procesa en el extremo. El resultado son metadatos que describen las características del sonido. No es posible recrear el sonido original a partir de sus metadatos.
  Se puede utilizar un **detector de sonido** para reconocer patrones, niveles o frecuencias y proporcionar información sobre el estado. De nuevo, es imposible recuperar el sonido original.
- **Procesamiento y codificación.** En aquellos casos en los que se utilizará el audio original (sin transformar ni analizar), suele incluirse una parte de procesamiento y codificación para preparar los datos de audio para una finalidad concreta. Algunas opciones serían guardar los datos de audio en el extremo, transmitirlos a clientes externos para su procesamiento (en un servidor o en la nube) o almacenarlos en servicios externos.

6 Topología de la analítica

El lugar del sistema donde se encuentra el motor de la analítica es importante por muchas razones. Sobre todo para las cuestiones relacionadas con la privacidad y el cumplimiento de los reglamentos sobre

información personal, es determinante *dónde* analiza los datos de audio el algoritmo de software. Hay situaciones en las que los datos de audio no pueden transmitirse por la red y es vital que los datos de audio capturados (pero no guardados) puedan analizarse en local. Sin embargo, cuando se necesitan algoritmos con una gran potencia de computación, seguramente no podrán ejecutarse en el extremo y deberán enviarse los datos de audio digital a la nube o a un servidor.

- **Analítica en el extremo.** Cuando la analítica se ejecuta en el extremo, los datos de audio digital no salen de la cámara. En el caso del audio capturado pero no grabado, solo se envían los resultados de la analítica, es decir, los metadatos o activadores.
- **Analítica en servidores.** Cuando se ejecuta en un servidor, el audio digital tiene que salir de la cámara. Si se procesa previamente en la cámara (extremo), estos datos pueden convertirse en metadatos abstractos o anonimizados. Un servidor normalmente forma parte de un sistema cerrado (controlado por el responsable del mismo), lo que permite gestionar las cuestiones de privacidad del audio transportado. De todos modos, es importante respetar las regulaciones y normativas aplicables.
- **Analítica en la nube.** El audio digital también puede transportarse a un servidor en una nube. Al igual que sucede con la analítica en un servidor, la información de audio puede procesarse antes y convertirse en metadatos. Como la utilización de la nube suele estar descentralizada, es todavía más importante abordar las cuestiones de privacidad y garantizar el cumplimiento de las normativas.

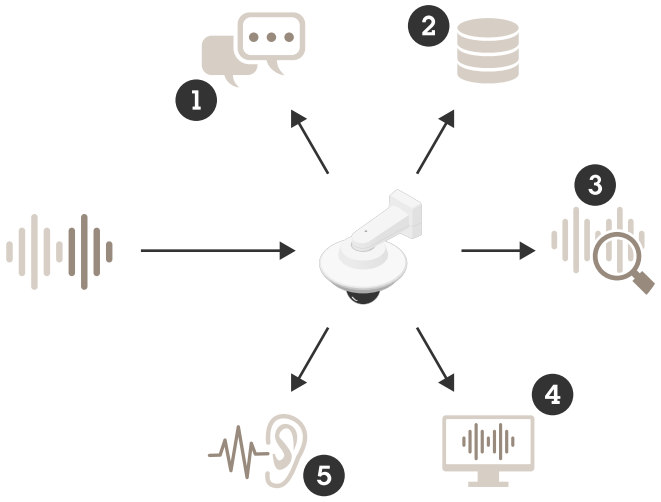


1 Extremo
2 Nube
3 Servidor
4 Almacenamiento

7 Casos de uso y ejemplos

El audio es a menudo una función integrada y lista para usar en las videocámaras. En distintos casos de usos, una utilización responsable y profesional del audio capturado puede ser clave y aportar numerosas ventajas. Por ejemplo, puede servir para obtener esa prueba que falta en una investigación forense o permitir la detección en tiempo real de eventos que requieren una atención inmediata por parte de

guardias de seguridad o del personal del hospital. Además, la vigilancia de audio puede tener un efecto disuasorio y evitar actos delictivos.



Aplicaciones habituales de la captura de audio:

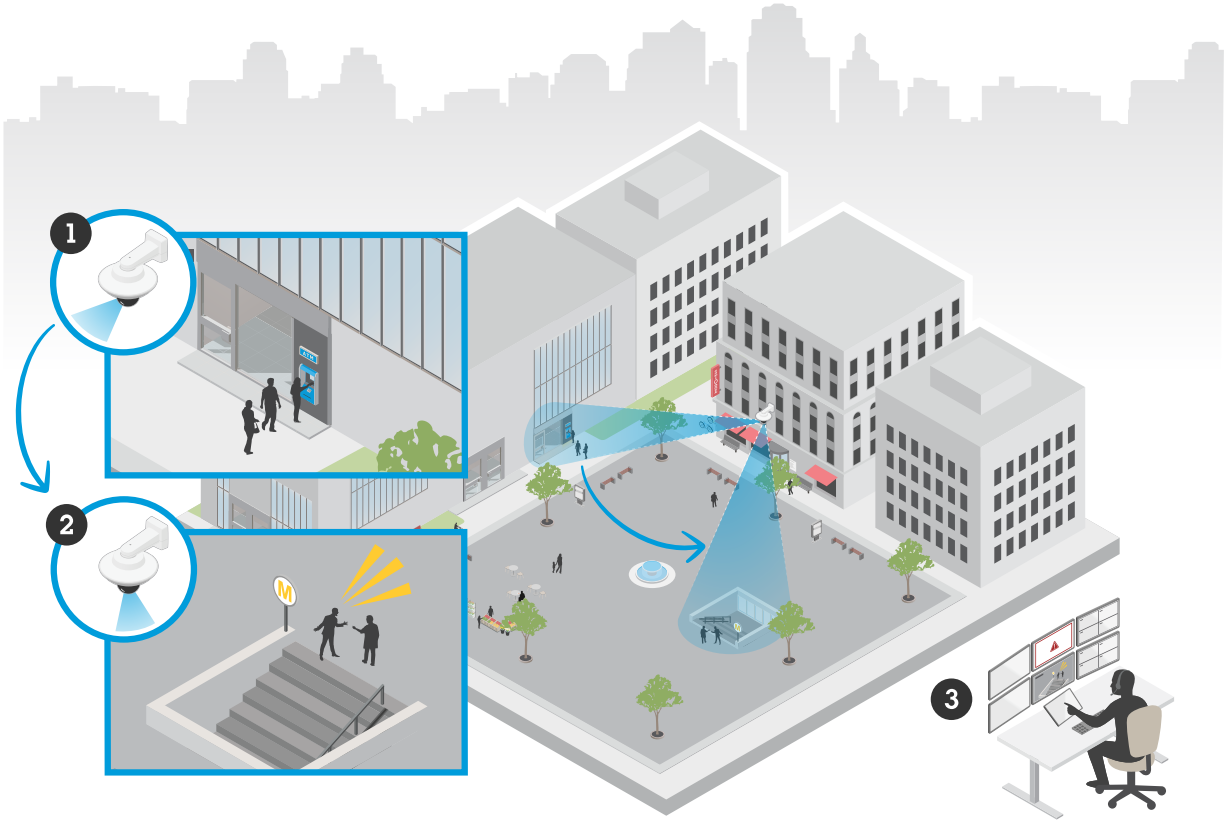
1. *Comunicar*
2. *Grabar*
3. *Analizar*
4. *Visualizar*
5. *Escuchar*

7.1 Detectar incidentes usando la analítica de audio

Las aplicaciones de analítica de audio son programas de software que procesan el audio capturado para encontrar y extraer información específica. Se utilizan para detectar eventos como disparos, cristales rotos o agresiones. Por ejemplo, pueden procesar el audio recibido para responder con sí/no a la pregunta "¿Se ha roto una ventana?". Si la respuesta es afirmativa, el sistema suele enviar una notificación automática al personal a través de una alerta visual o activando una alarma. Este sistema facilita una detección precoz y, por lo tanto, permite responder e intervenir con mayor rapidez.

7.1.1 Redirigir una cámara

Otro ejemplo de analítica de audio es el redireccionamiento de una cámara PTZ. En este caso, se combinan las funciones de audio y de vídeo: se detecta de donde procede el sonido y se redirige automáticamente la cámara hacia allí.



1 *Se utiliza una cámara PTZ para la supervisión de un cajero automático.*
2 *El micrófono de la cámara captura un ruido repentino muy alto e inmediatamente se redirige la cámara hacia el lugar del incidente.*
3 *El operador recibe una alarma y verifica el incidente.*

7.2 Visualizar el sonido en un vídeo

El sonido capturado en un vídeo puede visualizarse y mostrarse en un monitor como un diagrama del espectro acústico. Si se supera el umbral fijado, aparecerá una alarma en el diagrama.

La visualización del sonido puede ser interesante en situaciones en las que sea necesario supervisar el sonido de varias fuentes al mismo tiempo, por ejemplo, de varias habitaciones de un hospital. Si bien sería demasiado complicado escuchar tantas fuentes simultáneamente, resulta mucho más factible ver varias

visualizaciones en un monitor instalado en el mostrador de enfermería. Si se reciben vídeos de varias habitaciones, las visualizaciones pueden aparecer como superposición en la imagen.

Visualizaciones del sonido mostradas como superposiciones en el vídeo de un hospital.

7.3 Escuchar e interactuar

Seguramente el caso de uso más básico e intuitivo es la vigilancia de audio con la interacción directa del operador para saber mejor qué está ocurriendo en una escena. Algunos ejemplos prácticos serían detectar una conversación sospechosa y poder enviar a un guardia de seguridad para investigar qué sucede. O, en un hospital o residencia, escuchar si un paciente se encuentra mal y llamar al personal de enfermería. Puede servir también para detectar si ese sonido "extraño" viene de la derecha o de la izquierda y apuntar una cámara PTZ hacia el lugar de donde procede el sonido.

En estos casos de uso, uno o varios operadores tienen acceso al entorno de audio desde una sala de control o a través de una aplicación de seguridad en un dispositivo móvil. El oído humano captura los sonidos y el cerebro extrae la información relevante de la escena o de la situación. Combinado con la videovigilancia, el audio aporta información de otra dimensión, que resulta muy útil para tomar decisiones. En algunos casos, el audio será la única dimensión, por ejemplo si la fuente de audio está fuera del campo de visión de la cámara o si las condiciones de iluminación son complicadas.

7.3.1 Escuchar y ser testigo

La vigilancia de audio puede utilizarse también para disponer de un testigo directo de los eventos presenciados (escuchados). Este caso de uso es diferente de la aplicación *escuchar e interactuar* porque aquí el objetivo no es tomar decisiones, aunque a menudo estos dos casos de uso coexisten. Por ejemplo, al escuchar una discusión que va subiendo de tono con palabras incriminatorias, el operador no solo puede enviar guardias de seguridad sino también ser testigo de lo que se ha dicho.

7.4 Grabar y guardar

En determinadas situaciones, el caso de uso capturar y grabar datos de audio puede ser una vía fantástica para conseguir más pruebas, desde palabras incriminatorias hasta disparos. El audio grabado puede revelar quién dijo qué, cuántos disparos hubo o eventos similares con un interés para la investigación.

En las grabaciones de audio utilizadas para una investigación, es importante conservar los datos originales y evitar su procesamiento (que, en otros contextos, puede ser necesario o beneficioso). En las grabaciones forenses, cualquier tipo de procesamiento puede considerarse una manipulación de las pruebas. Pueden utilizarse algoritmos de optimización de la voz para que se entiendan mejor las palabras grabadas. De hecho, este tipo de procesamiento puede aumentar su valor en una investigación. Sin embargo, estos algoritmos deben aplicarse después, en una copia del material grabado. Al guardar la grabación con el mínimo nivel de procesamiento posible, el material puede utilizarse para lo que sea necesario. Todas las opciones están abiertas.

7.5 Saque más partido a su sistema de vigilancia

Los sistemas de vigilancia a menudo incorporan varios tipos de sensores. Uno de ellos es el sensor de imagen de la cámara, responsable de capturar el componente visual de una escena. Son también habituales los sensores no visuales, como los detectores de movimiento basados en la tecnología de radar o en emisiones de radiación infrarroja. A veces, la videovigilancia no es la opción más adecuada y los sensores no visuales se utilizan solos. Sin embargo, en la mayoría de casos sirven para complementar una instalación con cámaras, puesto que aportan otros tipos de información.

Además, cuando se añaden sensores de audio (micrófonos) a una instalación de vigilancia, hay ventajas para la gran mayoría de casos de uso potenciales. La incorporación del audio a un sistema sin audio abre la puerta a la interacción multisensor, ya sea a través de la analítica o del contacto con un operador.

El caso de uso *escuchar e interactuar* es un ejemplo sencillo: el operador puede hacerse una idea mucho más completa de lo que ocurre en una escena si también recibe el audio. Puede ser complicado detectar comportamientos agresivos solo con la imagen de una persona; es mucho más fácil si también podemos oír qué dice.

Otro buen ejemplo es la analítica de vídeo, como la detección de movimiento por vídeo. Si la aplicación de analítica tiene que lidiar con unas condiciones complicadas como poca luz, la presencia de la analítica de audio puede aumentar la fiabilidad de la detección.

8 Supervisión y detección

El audio contiene diferentes tipos de información, que puede utilizarse tanto para la supervisión como para la analítica de audio. Distintos tipos de acciones de procesamiento y caracterización contribuyen a extraer y pulir esta información para facilitar su uso y la interacción con el sistema.

8.1 Características del sonido

Características como el volumen o el tono pueden darnos una información importante en un contexto de vigilancia. Al extraer conclusiones sobre un sonido, pequeños detalles como el tiempo durante el que se oye, si se mueve o si viene de lejos o está cerca, tienen su transcendencia. Los equipos y las aplicaciones de software para la supervisión y la detección de audio están diseñados para usar también estos tipos de información, para "percibir" combinaciones complejas de características, desde el nivel de decibelios hasta la energía en diferentes frecuencias a lo largo del tiempo.

- **Información espacial.** Hace referencia al mundo físico que nos rodea e incluye conceptos como la ubicación, la dirección y la distancia. La información espacial puede utilizarse para capturar el audio en diferentes direcciones para optimizar las grabaciones. Puede resultar útil también en la analítica para determinar de qué dirección procede un sonido o a qué distancia se encuentra su fuente.
- **Información temporal.** La información temporal es importante tanto desde una perspectiva dinámica (cambio a largo del tiempo) como absoluta (¿cuándo ocurrió un incidente?), a menudo en relación con la información proporcionada por otros sensores, como el vídeo. La información temporal tiene un peso decisivo en el análisis del comportamiento, para saber qué ocurrió, cuándo y durante cuánto tiempo.
- **Información sobre el espectro.** Este concepto tiene que ver con las frecuencias, por ejemplo si un sonido tiene un tono alto o la combinación de tonos en sonidos más complejos. Los micrófonos que se utilizan en la vigilancia de audio están diseñados para ofrecer una respuesta de frecuencia plana, es decir, tratan de capturar todas las frecuencias del campo auditivo (20 Hz – 20 kHz) por igual. Su funcionamiento difiere del sistema de audición humano, porque las personas podemos detectar más fácilmente las frecuencias más habituales en el habla.
- **Información sobre la amplitud.** Este componente tiene que ver con la intensidad o el volumen de un sonido. La información sobre la amplitud puede complementar la información sobre el espectro y combinarse para dibujar una imagen sobre cómo está estructurado el audio recibido.

8.2 Procesamiento de la señal

En el mundo de la vigilancia de audio, el procesamiento de la señal consiste generalmente en mejorar la transmisión, la eficiencia del almacenamiento o la calidad subjetiva, o enfatizar o detectar componentes de interés. Y eso se consigue usando algoritmos de software que modifican o analizan el audio de distintas formas.

8.2.1 Modificación de las señales

Los algoritmos pueden utilizarse para modificar la señal para una finalidad concreta, normalmente para:

- mejorar la señal, por ejemplo, aumentando la audibilidad a través del control automático de la ganancia.
- alterar la señal, por ejemplo, modificando el contenido de la frecuencia relativa con un ecualizador.
- limitar la señal eliminando frecuencias o amplitudes específicas. Algunas aplicaciones concretas podrían ser mantener el volumen bajo a través de la compresión o garantizar la privacidad mediante la distorsión de las voces.

8.2.2 Análisis de las señales

La analítica de audio utiliza datos de audio capturados (pero normalmente no grabados) y analiza las características relevantes del sonido para generar resultados en un formato distinto al audio. En otras palabras, las aplicaciones convierten los datos de audio a un formato que permita tomar decisiones o iniciar acciones más fácilmente. Existen aplicaciones de analítica desarrolladas específicamente para detectar, por ejemplo, agresiones, disparos, cristales rotos o alarmas de coches.

Si se utilizan algoritmos de aprendizaje automático, pueden entrenarse usando grandes cantidades de datos para que aprendan a hacer predicciones sin estar explícitamente programados para ello. Un buen ejemplo en el contexto del audio sería el de un algoritmo capaz de detectar el sonido de una puerta cerrándose tras haberlo entrenado con miles de sonidos de este tipo.

8.3 La audición humana

El oído humano es una de las mejores herramientas que existen para detectar y analizar el audio. En entornos con mucho ruido, el oído y el cerebro humanos son capaces de detectar e interpretar las palabras en situaciones en las que fallan la mayoría de los algoritmos.

A través del oído, podemos sacar información espacial de una escena, como descubrir de dónde procede un sonido y también si la fuente de audio se está moviendo. Además, como tenemos dos oídos, somos capaces de distinguir si el sonido viene de la derecha, de la izquierda o de cualquier otro sitio a medio camino entre los dos. Además, los oídos y la cabeza están diseñados para que podamos identificar si un sonido viene de arriba o de abajo, de delante o de atrás. En el cerebro existen diferentes "filtros" para detectar diferencias temporales entre los oídos, que permiten identificar desviaciones minúsculas, de tan solo microsegundos, y nos proporcionan información sobre diferentes tipos de eventos. La capacidad de análisis de las señales de audio está extraordinariamente desarrollada en las personas, sobre todo en lo que respecta a voces humanas pero también a sonidos vinculados a peligros relevantes históricamente.

Con las condiciones adecuadas (como una buena calidad de sonido, sonido estereofónico, poco retardo), un operador humano puede ser una potente "herramienta de análisis" y complementar el hardware o software de detección. Usando un producto de vigilancia de audio con solo dos micrófonos, un operador puede extraer información espacial de una escena, como el lugar del que procede el sonido y el movimiento del sonido.

9 Nota legal

Este documento y su contenido son cortesía de Axis, y todos los derechos sobre el documento y los derechos de propiedad intelectual derivados del mismo (lo que incluye, a título meramente ilustrativo, marcas comerciales, nombres de marcas, logotipos y marcas similares) están protegidos por la ley. Asimismo, todos los derechos y/o intereses vinculados al documento o derechos de propiedad intelectual relacionados con el mismo son y serán propiedad de Axis Communications AB.

Este documento se proporciona "tal cual", sin garantía de ningún tipo y únicamente con una finalidad informativa. La información facilitada en este documento no pretende ofrecer ningún tipo de asesoramiento legal. Asimismo, no pretende crear ningún tipo de obligación legal para Axis Communications AB ni para sus filiales. Las obligaciones de Axis Communications AB y/o sus filiales en relación con cualquier producto Axis están sujetas exclusivamente a los términos y condiciones del contrato entre Axis y la persona física o jurídica que compra dicho producto a Axis directamente.

PARA DESPEJAR CUALQUIER TIPO DE DUDA, EL USUARIO DEL DOCUMENTO ASUME LA INTEGRIDAD DEL RIESGO VINCULADO A LA UTILIZACIÓN, LOS RESULTADOS Y LA APLICACIÓN DE ESTE DOCUMENTO, Y AXIS RENUNCIA A TODAS LAS GARANTÍAS EN LA MEDIDA MÁXIMA QUE PERMITA LA LEGISLACIÓN, YA SEAN LEGALES, EXPLÍCITAS O IMPLÍCITAS, LO QUE INCLUYE, A TÍTULO MERAMENTE ILUSTRATIVO, GARANTÍAS DE COMERCIABILIDAD, IDONEIDAD PARA UN FIN PARTICULAR, TITULARIDAD Y AUSENCIAS DE INFRACCIÓN Y RESPONSABILIDAD SOBRE EL PRODUCTO, O CUALQUIER GARANTÍA DERIVADA DE UNA PROPUESTA, ESPECIFICACIÓN O MUESTRA VINCULADA A ESTE DOCUMENTO.

Apéndice Terminología sobre la calidad de audio

Audio digital:

El audio digital es una representación del audio analógico (a menudo una señal acústica capturada por un micrófono) grabado en formato digital. En el audio digital, la onda acústica de la señal de audio suele codificarse como una secuencia continua de muestras numéricas. La precisión depende del número de dígitos relevantes grabados por el codificador. Por ejemplo, en el audio de un CD, las muestras se toman 44.100 veces por segundo, cada una con una profundidad de muestreo de 16 bits.

Ruido:

El ruido es un sonido no deseado (y a veces inevitable) que define o limita el extremo silencioso del intervalo de volumen sonoro. Es generado por todos los componentes de la cadena de audio, desde la fuente grabada (por ejemplo, un ventilador en una habitación), el micrófono (*por ejemplo,* ruidos propios del dispositivo, vibraciones, viento) o los cables (*por ejemplo,* interferencias, superposición de sonidos) hasta el dispositivo usado para capturar el sonido (*por ejemplo,* ruidos propios del dispositivo, ruido de muestreo digital). Juntos forman lo que se conoce como umbral mínimo de ruido.

El ruido viene definido normalmente por la relación señal/ruido, todo el intervalo entre el nivel definido (a veces el sonido más alto que el sistema puede procesar) y el umbral mínimo de ruido.

En vídeo, el equivalente es el ruido del vídeo, entendido como un patrón de píxeles aleatorios (normalmente) estáticos, popularmente conocido como "nieve"; el ruido limita lo que se puede ver en imágenes oscuras (en audio limita lo que se puede oír con las señales silenciosas).

Distorsión:

Todas las alteraciones no deseadas de una señal restan fidelidad al original y se consideran distorsiones (el ruido, entendido según la definición de la entrada anterior, normalmente se excluye de este apartado). La distorsión reduce la calidad subjetiva (normalmente hay distorsiones que suenan "bien") y empaña el contenido objetivo, por lo que resulta más complicado procesar la señal, sobre todo para fines de análisis, y disminuye su utilidad desde el punto de vista analítico.

A la hora de cuantificar la distorsión, normalmente se utilizan dos propiedades: la distorsión armónica total (THD) y la distorsión de intermodulación (IMD).

En el vídeo, el equivalente de la distorsión serían los artefactos, como la aberración cromática, la degradación, el desenfoque, etc. El resultado es una "mala" imagen y una pérdida de los detalles que se pueden ver.

Frecuencia de muestreo y respuesta de frecuencia:

En un sistema digital, el audio se muestrea un determinado número de veces por segundo. Es lo que se conoce como frecuencia de muestreo (normalmente entre 8.000 y 48.000 veces por segundo o Hz). Para una captura adecuada del sonido, la teoría de las señales (concretamente el teorema de muestreo de Nyquist Shannon) nos dice que la frecuencia de muestreo tiene que ser como mínimo el doble de la frecuencia más alta deseada o imprescindible en la señal analógica.

El oído humano puede oír en general frecuencias de entre 20 Hz y 15-20 kHz aproximadamente, en función de la edad y otros factores. Por regla general, el intervalo de frecuencia baja, por encima de varios cientos de Hz, define la base de sonidos concretos (como los componentes esenciales de las voces), mientras que el intervalo de frecuencia alta, por encima de varios miles de Hz, contiene más "detalles".

El intervalo de frecuencia en el ámbito del audio equivale a la resolución y la velocidad de fotogramas en el universo del vídeo; cuanto más bajo, menos detalles.

Profundidad de bits:

Cada vez que se muestrea el audio, se captura un valor analógico y se traduce a una representación digital. En el universo digital no hay infinitos, por lo que el nivel de detalle está limitado por la profundidad de bits definida. Cada bit representa un factor de dos (0 o 1, bajo o alto, etc.) que, combinado con el intervalo de amplitud definido (*por ejemplo,* el nivel de presión del sonido o de tensión definido), crea fracciones de este intervalo. Con dos bits se obtienen cuatro fracciones, con tres bits se obtienen ocho y así sucesivamente. Simplificando, podemos decir que una señal de un voltio, muestreada con tres bits, se dividiría y representaría en fracciones de 1/8 voltios.

Para obtener una calidad de audio suficiente, suelen bastar 16 bits (que representan 65.536 fracciones), al menos para el oído humano. Esta cantidad es la que se usa en el audio de los CD. En aplicaciones de analítica y para fines más exigentes, se suele recurrir a los 24 bits.

El equivalente de la profundidad de bits en el mundo del vídeo sería el contraste, el intervalo de luminancia o crominancia que puede reproducir cada píxel.

T10167735/ES/M2.2/2106

Acerca de Axis Communications

Axis contribuye a crear un mundo más inteligente y seguro a través de soluciones en red que mejoran la seguridad y suponen una nueva manera de hacer negocios. Como líder de la industria del vídeo en red, Axis pone a su disposición productos y servicios de videovigilancia y analítica, control de accesos y sistemas de audio e intercomunicación. Axis cuenta con más de 3800 empleados especializados en más de 50 países, y proporciona soluciones a sus clientes en colaboración con empresas asociadas de todo el mundo. Fundada en 1984, su sede central se encuentra en Lund, Suecia.

Para más información sobre Axis, visite nuestro sitio web *axis.com*.

©2021 Axis Communications AB. AXIS COMMUNICATIONS, AXIS, ARTPEC y VAPIX son marcas comerciales registradas de Axis AB en distintas jurisdicciones. El resto de marcas comerciales son propiedad de sus respectivos dueños. Nos reservamos el derecho a introducir modificaciones sin previo aviso.