

O potencial da "entrada de áudio"

Captura e processamento de sons para percepção e evidência da cena

Junho 2021

Sumário

1	Resumo	3
2	Introdução	4
	2.1 Captura sem gravação	4
3	Navegação pelos obstáculos	4
	3.1 O que as leis e regulamentos dizem?	5
	3.2 Investigue a possibilidade	5
4	A instalação importa	6
5	Preparação de áudio	7
6	Topologia da análise	7
7	Casos de uso e exemplos	8
	7.1 Detecte incidentes usando análise de áudio	9
	7.2 Visualização do som em vídeo	10
	7.3 Escutar e interagir	11
	7.4 Gravação e armazenamento	12
	7.5 Tire mais proveito do seu sistema de monitoramento	12
8	Monitoramento e detecção	12
	8.1 Características do som	12
	8.2 Processamento do sinal	13
	8.3 A audição humana	14
9	Aviso de isenção de responsabilidade	15
	Apêndice 1 Terminologia de qualidade de áudio	16

1 Resumo

A capacidade de captura de áudio, integrada e pronta para uso em uma câmera de vídeo ou fornecida por um microfone externo, permite vários casos de uso importantes. O uso responsável e profissional da "entrada de áudio" pode agregar valor crítico e benefícios a uma instalação de segurança. Ele poderia, por exemplo, fornecer a peça de evidência que está faltando em uma investigação forense ou permitir a detecção de eventos em tempo real que requerem atenção imediata dos agentes de segurança ou da equipe do hospital. O mero fato de que o monitoramento de áudio está acontecendo também pode ter um efeito de dissuasão e impedir o crime.

A captura de áudio (geralmente combinada com ação instantânea de análise) pode ser implantada como uma tecnologia autônoma, permitindo vários casos de uso na prevenção, proteção e perícia forense de crimes.

Mas combinada com vídeo, a captura de vídeo também tem o potencial de reforçar a grande maioria dos casos de uso de monitoramento existentes. Por exemplo, os operadores de segurança podem ter uma visão geral significativamente melhor dos eventos da cena se o stream de vídeo for complementado com um stream de áudio.

Assim como você pode empregar vários tipos de análise de vídeo para detecção automática de eventos e alarmes com base na detecção visual, a análise de áudio pode monitorar os streams de áudio e reagir quando algo se destaca.

O software de análise de áudio pode ser configurado para disparar alarmes automáticos e outras ações quando um microfone capta sons associados a gritos de pessoas, vidros quebrados ou tiros. Isso fornece um aviso prévio que permite respostas e intervenção rápidas.

A análise de áudio também pode ser sobre como detectar se um som inesperado veio da esquerda ou da direita e redirecionar automaticamente uma câmera PTZ para a origem do som. Em um hospital ou unidade de saúde, a análise de áudio pode ser usada para detectar altos níveis sonoros, implicando que um paciente está em perigo e enviar uma notificação automática para uma enfermeira. Este caso de uso também pode se beneficiar da análise de visualização sonora, o que torna mais fácil monitorar simultaneamente o som de muitos locais.

Há uma diferença entre capturar sons e gravá-los. Para muitos tipos de uso, não há necessidade de gravar áudio para alcançar o objetivo, e isso pode ajudar a gerenciar questões com privacidade e cumprir com os regulamentos relacionados a dados pessoais. Os aplicativos de análise de áudio, em geral, não gravam o som continuamente. Eles normalmente apenas processam o áudio de entrada para pesquisar padrões, níveis ou frequências específicos. Quando as análises são executadas na borda (na câmera), nenhum dado de áudio digital precisa sair da câmera - apenas os resultados das análises realizadas, ou seja, metadados ou gatilhos, deixam.

A Axis não fornece consultoria jurídica. As leis que regulam o monitoramento variam por região, estado e país, e é o usuário dos produtos (normalmente o cliente final) que é responsável por garantir que qualquer monitoramento seja conduzido de maneira compatível. Como o caso do videomonitoramento, a instalação do monitoramento em áudio deve ser precedida por uma investigação e entendimento dos aspectos legais de tal instalação.

Uma vez que as medidas necessárias tenham sido tomadas para atender às condições legais, a instalação deve ser cuidadosamente considerada quanto à colocação e configuração do equipamento. Isso pode exigir algum planejamento, mas geralmente não é difícil e medidas simples ajudam muito a alcançar a usabilidade do áudio.

2 Introdução

As informações de áudio podem ser um ativo valioso para a prevenção ao crime, proteção ou uso forense. O áudio capturado pode também ser processado em tempo real pelo software de análise que permite um monitoramento de áudio eficiente para detecção de atividades, comportamento ou eventos.

Este white paper descreve o potencial do áudio em segurança com exemplos de casos de uso típicos. Vários tipos de análise de áudio são apresentados junto com breves visões gerais de como trabalham.

Este documento não fornece nenhum aconselhamento legal, mas apresenta diferentes soluções técnicas que podem ser úteis para configurar uma instalação. Dependendo de como você escolhe implementar a análise de áudio, pode ser possível navegar por leis regionais e recomendações e empregar esta ferramenta poderosa onde necessário.

O escopo deste documento está limitado à *captura e possível gravação* de áudio, isto é, *entrada* de áudio. Outro uso comum do áudio em soluções de segurança diz respeito à *transmissão* de áudio, isto é, *saída* de áudio, tipicamente para reproduzir mensagens de voz ou alarmes para deter invasores ou ladrões. Mais informações sobre transmissão de áudio em segurança estão disponíveis em www.axis.com/products/audio.

2.1 Captura sem gravação

É possível capturar e usar o áudio sem gravá-lo. Capturar áudio basicamente significa digitalizá-lo e torná-lo disponível para uso em software. Isso é feito através do registro de vibrações sonoras no ar usando um microfone, convertendo esses sinais analógicos em digitais (usando equipamento de conversão A/D) e passando-os para uma unidade de processamento.

Se o áudio capturado não for colocado em uma mídia permanente, como uma memória flash ou um disco rígido, não é gravado. As gravações podem não ser necessárias para alguns casos de uso, tais como quando um operador humano está escutando em tempo real ao áudio capturado. Em algumas situações, há ainda razões específicas para *não* gravar o áudio. Pode haver diferenças em restrições legais, dependendo se o áudio é gravado ou apenas capturado.

Em geral, a análise de áudio não grava sons continuamente. Para funcionar apropriadamente, eles temporariamente armazenam o som. Muitos sistemas poderiam ser configurados para gravar o que foi armazenado em buffer antes e depois de uma detecção para permitir que a segurança verifique a detecção e, possivelmente, preserve o som para evidência forense.

3 Navegação pelos obstáculos

Muitas pessoas têm preocupações em relação ao uso de microfones em situações de videomonitoramento. Essas preocupações normalmente são vinculadas à gravação de fala simples juntamente com o material de vídeo.

Podemos superar esse obstáculo inicial se entendermos que há muito mais possibilidades com "entrada de áudio" do que apenas gravá-lo. Existem muitos casos de uso em que não há necessidade de gravar nenhuma informação sonora.

As leis que regulam o monitoramento variam por região e país, portanto, certifique-se de saber o que é permitido antes de adicionar áudio ao seu sistema de monitoramento.

3.1 O que as leis e regulamentos dizem?

Como o caso do videomonitoramento, a instalação do monitoramento em áudio deve ser precedida por uma investigação e entendimento dos aspectos legais de tal instalação. Se aplicável, os documentos apropriados de aplicação devem ser enviados e as licenças adquiridas. Sinais ou declarações públicas devem ser usados quando necessário.

O uso e/ou gravação de áudio pode ser proibido ou exigir consideração especial por diversas razões, por legislatura nacional ou vários tipos de regras e regulamentos locais. Enquanto uma região ou ambiente pode permitir a captura de áudio, ela ainda pode proibir gravações de áudio. As empresas também podem proibir o uso de monitoramento de áudio dentro de suas instalações.

3.1.1 Exemplos norte-americanos

As leis e regulamentos nos Estados Unidos variam entre os diferentes estados.

Alguns estados requerem consentimento unilateral para a gravação de áudio. Isso significa que somente uma parte em uma conversa precisa consentir para que o monitoramento seja legal.

Outros estados requerem consentimento de todas as partes (ou duas partes), o que significa que todas as partes devem consentir serem gravados antes que a gravação aconteça. Exceções ao consentimento de todas as partes podem se aplicar a locais públicos onde uma pessoa não pode ter privacidade.

Sua avaliação legal pode também levar a outro resultado em algumas regiões ao usar um aplicativo de análise de áudio que não grava áudio. Portanto, você precisa investigar que leis e regulamentos se aplicam no seu estado específico.

3.1.2 Exemplos europeus

O monitoramento em áudio é regulado por leis nacionais nos países europeus. Portanto, você precisa investigar que leis e regulamentos se aplicam no seu país específico.

As gravações de áudio podem conter dados pessoais que são sujeitos ao RGPD (Regulamento Geral de Proteção de Dados). O GDPR não necessariamente proíbe gravações de áudio, mas a captura ou gravação do áudio requer considerações especiais. Ao adicionar áudio ao videomonitoramento existente, você precisa considerar se sua base legal para processamento dos dados pessoais de acordo com o GDPR ainda se aplica.

3.2 Investigue a possibilidade

Existe um equívoco geral de que o áudio nunca é permitido no monitoramento. Este equívoco é tão difundido que, em muitos casos, a possibilidade de reforçar uma instalação de monitoramento com áudio nunca é considerada.

Mas muitos tipos de instalações podem ser permitidos, por exemplo, se as pessoas forem informadas, se você tiver consentimento e assim por diante. Você precisa investigar quais leis e regulamentos se aplicam à sua região e ao seu caso de uso. Mesmo se o caso de uso de *gravação e armazenamento* não fosse permitido na sua instalação de segurança, muitos casos de uso podem ser adaptados para não infringir direitos de privacidade, tais como *escuta e interação*, *escuta e testemunho* e *detecção de incidentes usando análise de áudio*.

4 A instalação importa

O posicionamento do microfone em uma cena define os aplicativos potenciais. Antes de instalar um equipamento de áudio, sua localização e configuração devem ser cuidadosamente pensados. Isso pode exigir algum planejamento, mas geralmente não é difícil e medidas simples ajudam muito a aumentar a usabilidade do áudio.

Refletir sobre o posicionamento correto de um microfone e escolher um local acusticamente bom aumentará a probabilidade de atingir seus objetivos de monitoramento. É claro que um microfone deve ser posicionado de forma a facilmente capturar os sons que são relevantes. Os posicionamentos típicos são no meio de uma sala, em conjunto com uma câmera, ou perto de onde eventos específicos de interesse podem ocorrer. Um microfone não deve ser colocado perto de uma fonte de ruídos, como ventilação ou maquinário, que poderia ofuscar sons mais fracos ou mais distantes.



Posicionamento típico de um microfone

- 1 Onde ocorrem as ações de interesse
- 2 Em uma câmera
- 3 No meio de uma sala

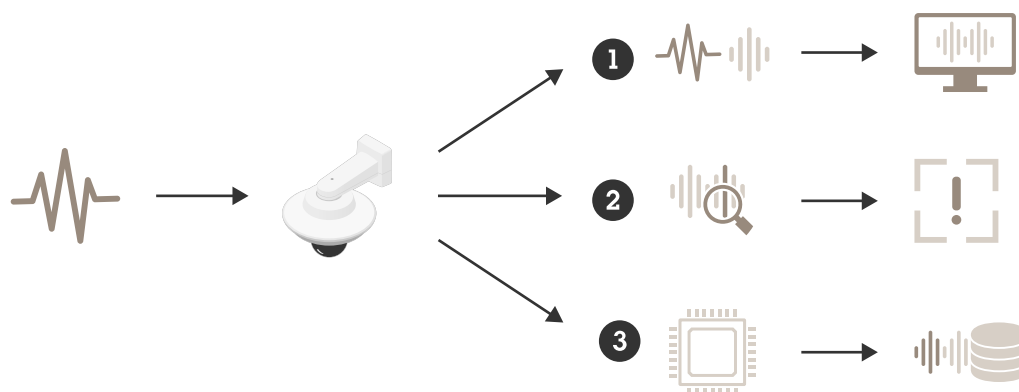
O ambiente acústico, como propriedades de absorção de som de paredes ou teto/piso e complexidades dimensionais (como corredores muito longos), produzirá reverberações e ecos diferentes que podem afetar gravemente o campo sonoro em determinados locais. Por exemplo, uma voz soará muito diferente em uma área fortemente atenuada (como uma sala de conferências tratada acusticamente), em comparação com uma igreja ou em um banheiro totalmente revestido com azulejos. Em situações acusticamente desafiadoras, o posicionamento do microfone pode se tornar crítico.

Tanto a instalação quanto a configuração (por exemplo, a configuração de ganho de áudio) do equipamento são importantes, bem como a integração do equipamento de áudio com o sistema de monitoramento. Os instaladores e integradores do sistema podem fornecer recomendações para situações específicas.

Para análise de áudio, às vezes se aplicam recomendações específicas, que podem ser diferentes das recomendações para gravações de áudio em geral. Sempre estude a documentação do usuário para estar ciente dos pré-requisitos aplicáveis.

5 Preparação de áudio

Após a captura de áudio inicial, as informações capturadas são preparadas para as próximas etapas de processamento. As diferentes preparações podem ser feitas em paralelo ou de forma exclusiva.



- 1 *Transformação*
- 2 *Análise da borda em tempo real*
- 3 *Processamento e codificação*

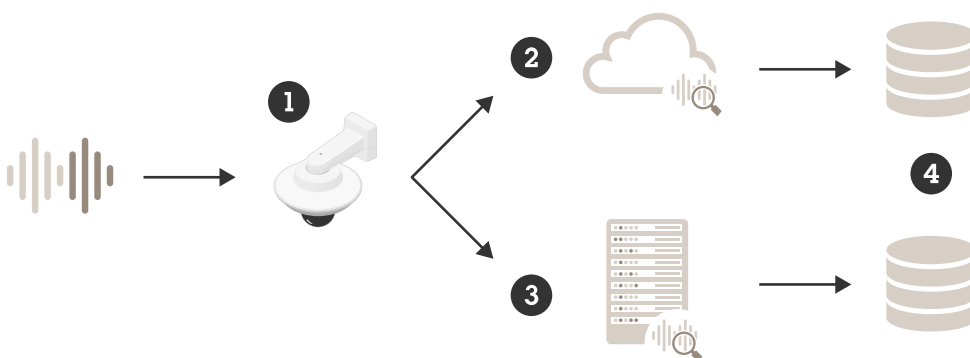
- **Transformação** O som é sintetizado e convertido em, por exemplo, informações visuais como um gráfico, mostrando o espectro do som. Este processo não pode ser revertido. Você não pode recuperar o som original do gráfico do espectro.
- **Análise da borda em tempo real.**
Um **classificador de som** pode ser usado se o som for processado na borda. O resultado serão metadados que descrevem as características do som. O som original não pode ser recriado a partir de seus metadados.
Um **detector de som** pode ser usado para reconhecer padrões, níveis ou frequência e fornecer informações de status. De novo, o som original não é restaurável.
- **Processamento e codificação.** Para os casos em que o áudio original será usado (não transformado ou analisado), algum processamento e codificação é normalmente executado para preparar os dados de áudio para os casos de uso pretendidos. Estes casos de uso podem envolver o armazenamento dos dados de áudio na borda, transmissão para clientes externos para processamento adicional (no servidor ou na nuvem) ou armazenamento externo.

6 Topologia da análise

A localização do mecanismo de análise no sistema é importante por muitos motivos. Especialmente para gerenciar questões de privacidade e cumprir com os regulamentos relacionados aos dados pessoais,

é importante *onde* o algoritmo do software analisa os dados de áudio. Há situações em que os dados de áudio não podem ser enviados pela rede e é fundamental que os dados do áudio capturado (mas não armazenados) possam ser analisados localmente. Se forem necessários algoritmos de computação intensiva, tais como os que não podem executados na borda, pode ser necessário enviar dados de áudio digital para a nuvem ou para um servidor.

- **Análise da borda.** Quando a análise é executada na borda, não são necessários dados de áudio digital para deixar a câmera. No caso de captura de áudio sem gravação, apenas o resultado das análises realizadas, ou seja, metadados ou acionadores, será enviado.
- **Análise do servidor.** Quando executado em um servidor, o áudio digital precisa deixar a câmera. Se pré-processados na câmera (borda), estes dados podem ser metadados extraídos ou despersonalizados. Um servidor normalmente faz uma parte de um sistema fechado (o proprietário do sistema está no controle), portanto, as questões de privacidade do áudio transportado podem ser gerenciadas. No entanto, deve-se garantir que as regras e regulamentos aplicáveis sejam seguidos.
- **Análise da nuvem.** O áudio digital também pode ser transportado para um servidor em um contexto de nuvem. Como no caso da análise do servidor, as informações de áudio podem ser pré-processadas em metadados. O uso da nuvem geralmente é descentralizado, por isso é ainda mais importante abordar as questões de privacidade e assegurar que os regulamentos sejam cumpridos.

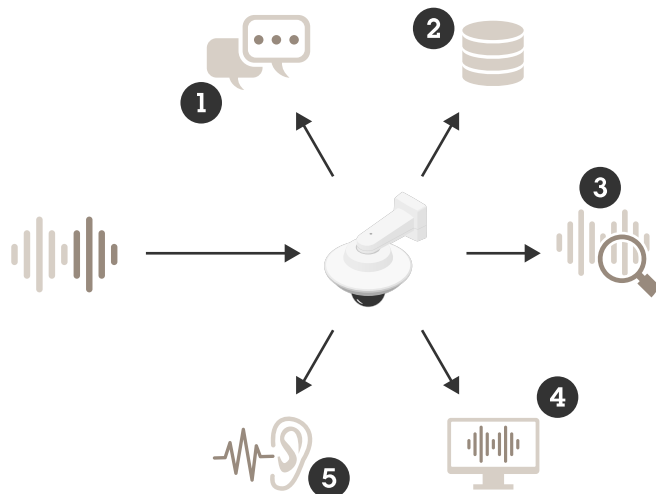


- 1 Edge
- 2 Nuvem
- 3 Servidor
- 4 Armazenamento

7 Casos de uso e exemplos

A capacidade de áudio costuma estar integrada e pronta para uso em câmeras de vídeo. Há vários casos de uso onde o uso responsável e profissional da "entrada de áudio" pode fornecer valor críticos e vários benefícios potenciais. Ele pode, por exemplo, ser usado para apresentar a peça de evidência que está faltando em uma investigação forense ou permitir a detecção de eventos em tempo real que requerem a

atenção imediata dos agentes de segurança ou da equipe do hospital. O fato de que o monitoramento de áudio está acontecendo também pode ter um efeito de dissuasão e impedir o crime.



Finalidades típicas da captura de áudio:

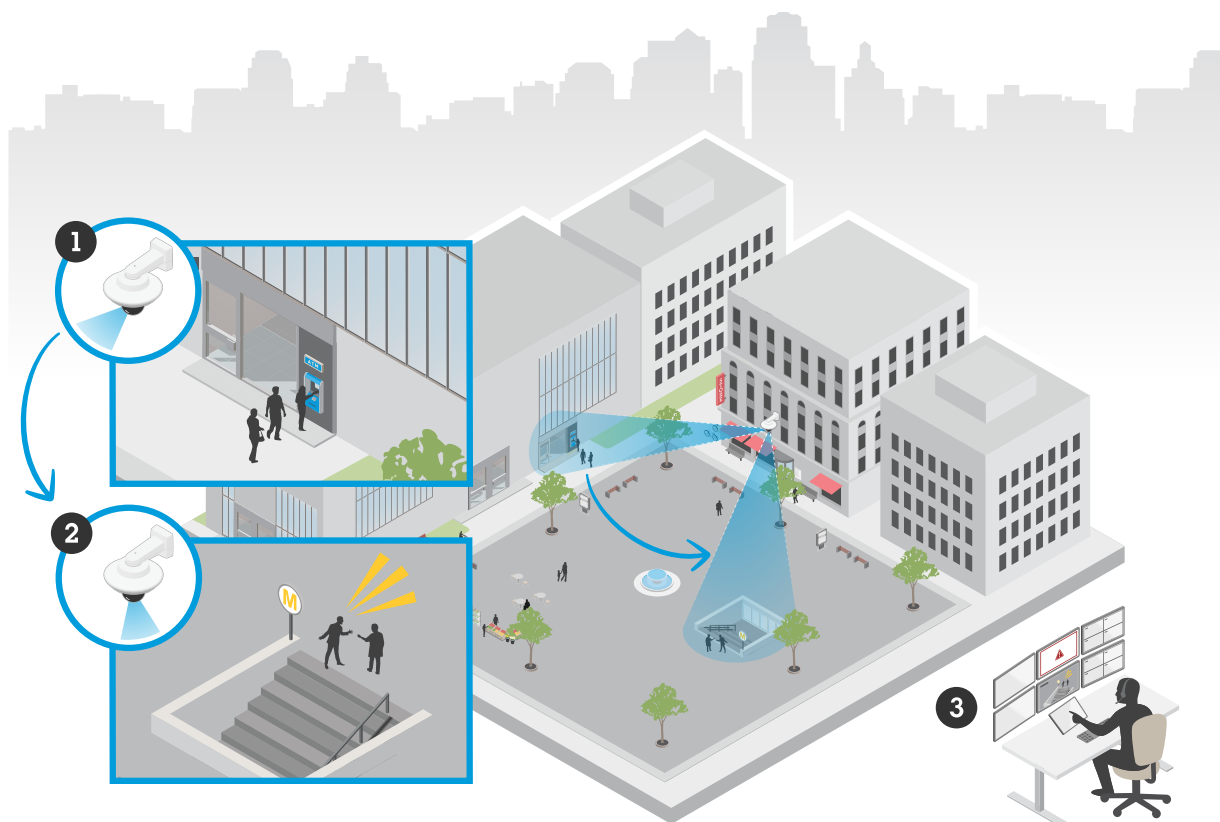
- 1 *Comunicar*
- 2 *Gravar*
- 3 *Analisar*
- 4 *Visualizar*
- 5 *Escutar*

7.1 Detecte incidentes usando análise de áudio

Os aplicativos de análise de áudio são programas de software que processam áudio capturado para encontrar e extrair informações específicas. São usados para detectar eventos, como tiros, quebra de vidros ou agressão. Eles podem, por exemplo, processar áudio de entrada para oferecer uma resposta sim-ou-não para a pergunta "uma janela quebrou?" Após a detecção, o sistema normalmente envia uma notificação automática para a equipe por meio de um alerta visual ou pelo acionamento de um alarme. Isso fornece um aviso prévio que permite respostas e intervenção rápidas.

7.1.1 Redirecionamento de uma câmera

Um outro exemplo de análise de áudio é um aplicativo de redirecionamento de câmera PTZ. Isso combina as funcionalidades de áudio e vídeo ao detectar de onde vem o áudio e direcionar automaticamente a câmera em direção à fonte do áudio.



- 1 *Uma câmera PTZ está monitorando um caixa eletrônico.*
- 2 *O microfone da câmera capta um ruído alto e repentino e a câmera instantaneamente direciona para o incidente.*
- 3 *O operador recebe um alarme e verifica o incidente.*

7.2 Visualização do som em vídeo

O som capturado em um vídeo pode ser visualizado e exibido como um diagrama de espectro de som em um monitor. Se um limite definido for excedido, o diagrama começará a indicar um alarme.

Essa visualização de som pode ser valiosa em situações em que você precisa monitorar sons de várias fontes ao mesmo tempo, por exemplo, vários quartos de pacientes em um hospital. Embora seja muito difícil ouvir muitas fontes de áudio (som de muitas salas) simultaneamente, seria muito mais fácil ver muitas

visualizações em um monitor no posto de enfermagem. Se estiverem disponíveis os feeds de vídeo das salas, as visualizações podem ser adicionadas como sobreposição à imagem de vídeo.



Visualização de som adicionada como sobreposições aos feeds de vídeo em um hospital.

7.3 Escutar e interagir

Talvez o caso de uso mais básico e intuitivo seja o monitoramento de áudio com interação direta com o operador para aumentar a percepção da cena. Os exemplos são normalmente perceber uma conversa suspeita e ser capaz de enviar um agente de segurança para investigar mais profundamente. Ou em um hospital ou centro de saúde, para saber se um paciente está em perigo e chamar uma enfermeira. Também pode ser sobre como detectar se um som "estranho" veio da esquerda ou da direita e apontar uma câmera PTZ para a origem do som.

Esses casos de uso envolvem um ou vários operadores tendo acesso ao ambiente de áudio de uma sala de controle ou por meio de um aplicativo de segurança em um dispositivo móvel. O ouvido humano captura sons e o cérebro extrai o que é relevante na cena ou na situação. Se usado em conjunto com videomonitoramento, o áudio adiciona uma outra dimensão de informação para a tomada de decisão. Em alguns casos, o áudio será realmente a única dimensão e sua origem estiver fora do campo de visão da câmera ou se as condições de luz forem desafiadoras.

7.3.1 Escute e testemunhe

O monitoramento por áudio também pode ser usado para fins de testemunho direto, com base em eventos testemunhados (ouvidos). Este caso de uso difere do uso de *ouvir e interagir* porque o propósito não é a tomada de decisão, mas os casos de uso costumam coexistir. Por exemplo, ao escutar uma discussão crescente com discurso incriminatório, um operador não pode apenas enviar seguranças, porém, mais tarde, também testemunhar sobre o que foi ouvido.

7.4 Gravação e armazenamento

Se apropriado, o caso de uso de captura e gravação de dados em áudio pode fornecer grande evidência adicional. Isso pode envolver discurso incriminatório ou tiroteio. A gravação de áudio pode fornecer provas de quem disse o quê, quantos tiros foram disparados ou eventos similares de interesse forense.

Quando o áudio é gravado em um contexto forense, deve-se tomar cuidado para conservar os dados originais e evitar o processamento (o que, em outros contextos, pode ser necessário ou benéfico). Para gravações forenses, todos os tipos de processamento podem ser considerados violação de evidência. Algoritmos de aprimoramento de voz podem ser usados para aumentar a audibilidade da fala gravada e podem melhorar o valor forense. Mas tais algoritmos devem ser aplicados posteriormente, em uma cópia do material gravado. Ao manter a gravação o menos processada possível, as opções são mantidas abertas sobre como o material pode ser usado posteriormente.

7.5 Tire mais proveito do seu sistema de monitoramento

Os sistemas de monitoramento frequentemente incorporam vários tipos de sensores. O sensor de imagem da câmera é um deles, é claro, registrando o aspecto visual de uma cena. Sensores não visuais também são comumente usados, como detectores de movimento com base em tecnologia de radar ou emissões de radiação infravermelha. Algumas vezes, o videomonitoramento não é apropriado e os sensores não visuais são consequentemente usados como dispositivos autônomos. Mas em muitos casos, os sensores não visuais são usados para completar a instalação da câmera ao adicionar outros tipos de informação.

Também empregando sensores de áudio (microfones) em uma instalação de monitoramento, a grande maioria de todos os casos de uso possíveis são reforçados. A adição de capacidade de áudio a um sistema sem áudio permite a interação multisensor, tanto pela análise ou pela interação do operador.

O caso de uso *ouvir e interagir* é um exemplo simples, em que o operador obtém uma visão geral muito melhor dos eventos da cena ao também receber um stream de áudio. Pode ser difícil detectar comportamento agressivo apenas olhando as pessoas, mas é muito mais fácil se você também pode escutá-las.

Outro exemplo típico é usar a análise de vídeo, como detecção de movimento por vídeo. Se o aplicativo de análise for desafiado por, por exemplo, condições de pouca luz, a presença de análise de áudio pode aumentar a confiança da detecção.

8 Monitoramento e detecção

O áudio contém vários tipos de informações que podem ser usados para monitoramento e análise de áudio. Vários tipos de processamento e caracterização ajudam a extrair e refinar essas informações para facilitar o uso e a interação com o sistema em volta.

8.1 Características do som

Características, como volume e tom, podem incluir informações importantes em um contexto de monitoramento. Por quanto tempo é audível, se está se movendo, ou se vem de perto ou de longe, são todos exemplos de peças que se somam ao quebra-cabeça quando tiramos conclusões sobre um som que ouvimos. Os hardware e software para monitoramento e detecção de áudio são projetados para trabalhar com os mesmos tipos de informações, "ouvindo" combinações complexas de características desde o nível de decibéis até a energia em diferentes frequências ao longo do tempo.

- **Informações espaciais.** Isso se refere ao mundo físico ao nosso redor, incluindo conceitos como localização, direção e distância. As informações espaciais podem ser usadas para focar ou ampliar a captura de áudio em diferentes direções para permitir melhores gravações. Ele também pode ser usado por análises para determinar de qual direção um som está vindo ou a que distância sua fonte está localizada.
- **Informações temporais.** As informações temporais (tempo) são importantes tanto no sentido dinâmico (mudança ao longo do tempo) quanto no sentido absoluto (quando algo aconteceu?), muitas vezes vistas em relação às informações de outros sensores, como o vídeo. As informações temporais desempenham um papel importante na análise comportamental - saber o que aconteceu, quando e por quanto tempo.
- **Informações espectrais.** Isso se refere a frequências, tais como o nível do tom que um som tem ou a combinação de tons em sons mais complexos. Os microfones usados em monitoramento de áudio são projetados para ter uma resposta de frequência plana, ou seja, eles tentam capturar todas as frequências dentro da faixa audível (20 Hz – 20 kHz) de forma igual. Isso difere de como o sistema auditivo humano funciona, porque podemos detectar mais facilmente essas frequências que ocorrem normalmente na fala humana do que as outras frequências.
- **Informações de amplitude.** Trata-se de quão intenso ou alto é um som. As informações de amplitude podem complementar as informações espectrais e ser usadas juntas para construir uma imagem de como o áudio de entrada está estruturado.

8.2 Processamento do sinal

No monitoramento de áudio, o processamento do sinal normalmente visa melhorar a transmissão, a eficiência do armazenamento ou a qualidade subjetiva, ou para enfatizar ou detectar componentes de interesse. Isso é feito por meio de algoritmos de software que modificam ou analisam o áudio de várias maneiras.

8.2.1 Modificando sinais

Algoritmos podem ser usados para alterar o sinal para uma finalidade específica, normalmente para:

- melhorar o sinal, por exemplo, aumentar a audibilidade por meio do controle automático de ganho.
- alterar o sinal, por exemplo, mudando o conteúdo da frequência relativa com um equalizador.
- limitar o sinal, removendo frequências ou amplitudes específicas. Isso pode ser sobre como manter o volume de dados baixo por meio da compactação ou sobre como garantir a privacidade por meio de codificação de voz.

8.2.2 Analisando sinais

A análise de áudio usa dados de áudio capturado (mas normalmente não gravado) e analisa características relevantes do som para gerar resultados que não são de áudio. Os aplicativos essencialmente convertem os dados de áudio em um ativo mais acionável em outro formato. Existem aplicativos de análise desenvolvidos especialmente para detectar, por exemplo, agressões, tiros, vidros quebrados ou alarmes de carros.

Se forem usados algoritmos de aprendizado de máquina, eles podem ser treinados com grandes quantidades de dados para aprender a fazer previsões sem serem explicitamente programados para isso. Um exemplo em um contexto de áudio pode ser um algoritmo que pode detectar com segurança o som de uma porta se fechando depois de ter sido treinado com milhares desses sons.

8.3 A audição humana

O ouvido humano é uma das melhores ferramentas disponíveis para detectar e analisar áudio. Em ambientes muito barulhentos, o ouvido e o cérebro humanos ainda podem detectar e interpretar a fala onde a maioria dos algoritmos pode não funcionar.

Usando nossos ouvidos, podemos derivar informações espaciais de uma cena, como de onde um som está vindo e se a fonte de áudio está se movendo. Como temos dois ouvidos, podemos ouvir se um som está vindo da esquerda ou da direita ou de algum lugar intermediário. Os ouvidos e a cabeça também são projetados para que possamos ouvir se um som vem de cima ou de baixo, da frente ou de trás. Várias "etapas de filtro" no cérebro trabalham com diferenças temporais entre as orelhas, detectando instantaneamente desvios tão pequenos quanto microssegundos para nos tornar cientes de tipos específicos de eventos. Temos uma capacidade bem desenvolvida de análise de sinais de áudio, especialmente em relação a vozes humanas, mas também sons associados a perigos históricos.

Sob as circunstâncias certas (como boa qualidade de som, som estereofônico, sem muito atraso), um operador humano pode ser uma "ferramenta de análise" poderosa e complementar o hardware ou software de detecção. Usando um produto de monitoramento de áudio com apenas dois microfones, um operador pode derivar informações espaciais de uma cena, como de onde um som está vindo e o movimento desse som.

9 Aviso de isenção de responsabilidade

Este documento e seu conteúdo são fornecidos como cortesia da Axis e todos os direitos ao documento ou quaisquer direitos de propriedade intelectual relacionados a ele (incluindo, mas não se limitando a marcas registradas, nomes comerciais, logotipos e marcas semelhantes neles) são protegidos por lei e todos os direitos, títulos e/ou interesse no e para o documento ou quaisquer direitos de propriedade intelectual relacionados a ele são e devem permanecer propriedade da Axis Communications AB.

Informamos que este documento é fornecido como se encontra ("as is"), sem garantia de qualquer tipo apenas para fins informativos. As informações fornecidas neste documento não constituem, nem pretendem constituir, aconselhamento jurídico. Este documento não se destina a, e nem deve, criar qualquer obrigação legal para a Axis Communications AB e/ou qualquer uma de suas afiliadas. As obrigações da Axis Communications AB e/ou de qualquer uma de suas afiliadas em relação aos produtos Axis estão sujeitas exclusivamente aos termos e condições do contrato entre a Axis e a entidade que adquiriu tais produtos diretamente com a Axis.

PARA EVITAR DÚVIDAS, TODO O RISCO QUANTO AO USO, RESULTADOS E DESEMPENHO DESTES DOCUMENTOS É ASSUMIDO PELO USUÁRIO DO DOCUMENTO E A AXIS REJEITA E EXCLUI, ATÉ O LIMITE MÁXIMO PERMITIDO > POR LEI, TODAS AS GARANTIAS, SEJAM ESTATUTÁRIAS, EXPRESSAS OU IMPLÍCITAS, INCLUINDO, MAS NÃO SE LIMITANDO A QUAISQUER GARANTIAS IMPLÍCITAS DE COMERCIALIZAÇÃO, ADEQUAÇÃO A UMA DETERMINADA FINALIDADE, TÍTULO E NÃO VIOLAÇÃO E RESPONSABILIDADE PELO PRODUTO, OU QUALQUER GARANTIA DECORRENTE DE QUALQUER PROPOSTA, ESPECIFICAÇÃO OU AMOSTRA COM RESPEITO A ESTES DOCUMENTOS.

Apêndice 1 Terminologia de qualidade de áudio

Áudio digital:

O áudio digital é uma representação do áudio analógico (geralmente um sinal acústico capturado com um microfone) gravado em formato digital. No áudio digital, a onda sonora do sinal de áudio é geralmente codificada como uma sequência contínua das amostras numéricas. A precisão depende do número de dígitos significativos que o codificador registra. Por exemplo, no áudio de CD, as amostras são obtidas a 44.100 vezes por segundo, cada uma com uma profundidade de amostra de 16 bits.

Ruídos:

O ruído é um som indesejado (e, às vezes, inevitável) que definirá ou limitará o fim do silêncio da faixa de volume. É gerado por todas as partes de uma cadeia de áudio, a partir da fonte gravada (por exemplo, um ventilador na sala), por meio do microfone (*por exemplo*, ruído próprio, vibrações, vento) e cabeamento (*por exemplo*, interferência, diafonia), para o dispositivo de captura (*por exemplo*, ruído próprio, ruído de amostragem digital), todos combinados criando o que é normalmente chamado de piso de ruído.

O ruído é normalmente definido por SNR (relação sinal-ruído), toda a faixa de um nível definido (às vezes, o som mais alto que o sistema pode suportar) até o piso de ruído.

O equivalente de vídeo é o ruído de vídeo, visto como padrão de pixel estático aleatório (normalmente), "neve"; limita o que você pode ver em imagens escuras (assim como limita o que você ouve para sinais silenciosos).

Distorção:

Todas as alterações indesejadas de um sinal subtraem de uma "verdade" original e isso é chamado de distorção (ruído, como explicado acima, é normalmente excluído da especificação de distorção). A distorção reduz a qualidade subjetiva (normalmente, há distorção que soa "legal") e obscurece o conteúdo da informação objetiva, tornando o sinal mais difícil de ser escutado, especialmente para análise de conteúdo, e reduz a funcionalidade da análise.

A THD (distorção harmônica total) e a IMD (distorção de intermodulação) são duas propriedades normalmente usadas para quantificar a distorção.

A distorção se correlaciona com o vídeo como artefatos, tais como aberração cromática, vinhetas, desfocado, etc.; faz uma imagem parecer "ruim" e limita a quantidade de detalhes que você pode ver.

Taxa de amostragem e resposta de frequência:

Em um sistema digital, o áudio é tomado como amostra por um número determinado de vezes por segundo. Essa é a taxa de amostragem (normalmente de 8.000 a 48.000 vezes por segundo, ou Hz). Para capturar um som adequadamente, a teoria do sinal (especificamente o teorema de amostragem de Nyquist Shannon) nos diz que a taxa de amostragem precisa ser pelo menos duas vezes maior que a frequência mais alta desejada ou exigida no sinal analógico.

Um ouvido humano normalmente escuta frequências de 20 Hz a aproximadamente 15-20 kHz, dependendo da idade e outros fatores. Grosso modo, a faixa de baixa frequência, acima de centenas de Hz, geralmente define a base de sons específicos (como bases em vozes), enquanto a faixa de alta frequência, acima de algumas centenas de Hz, contém mais "detalhes".

A faixa de frequência no áudio se correlaciona à resolução e à taxa de quadros do vídeo; quanto mais baixo for definido, menos detalhes obterá.

Profundidade de bits:

Cada vez que se obtém uma amostra de áudio, um valor analógico é capturado e traduzido para uma representação digital. No domínio digital não há infinitos, então a quantidade de detalhes é limitada a uma profundidade de bits definida. Cada bit representa um fator de dois (0 ou 1, baixo ou alto, etc.) que, combinado com uma faixa de amplitude definida (*por exemplo*, uma tensão escolhida ou nível de pressão sonora), cria frações desta faixa. Dois bits geram quatro frações, três bits geram oito e assim por diante. Simplificando, um sinal de um volt, obtido como amostra com três bits, seria dividido e representado em etapas de 1/8 de volt.

Para uma qualidade de áudio suficiente, os 16 bits normalmente são suficientes (representando 65.536 etapas), ao menos para o ouvido humano. Isso é o que o áudio de CD está usando. Para análise ou uso mais exigente, 24 bits são mais relevantes.

A profundidade de bits se correlaciona ao contraste no vídeo, a faixa de luminância ou crominância que cada pixel pode reproduzir.

Sobre a Axis Communications

A Axis torna possível um mundo mais inteligente e seguro criando soluções para melhorar a segurança e o desempenho dos negócios. Como empresa de tecnologia de rede e líder do setor, a Axis oferece soluções em vigilância por vídeo, controle de acesso, intercomunicação e áudio. Nossas soluções são aprimoradas por aplicativos de análise inteligentes e apoiados por treinamento de alta qualidade.

A Axis tem cerca de 4.000 funcionários dedicados em mais de 50 países e colabora com parceiros de tecnologia e integração de sistemas em todo o mundo para fornecer soluções aos clientes. A Axis foi fundada em 1984 e tem sede em Lund, Suécia