ホワイトペーパー

セキュリティと安全のための音声分析

7月 2025

概要

セキュリティと安全性に関する音声分析機能を活用することで、音声のパターンを検知し、ライブ音声の中に潜んでいる重要な音声に焦点を当てることができます。分析機能は、悲鳴や怒鳴り声を識別してオペレーターに警告できるため、オペレーターは事態の悪化や暴行などを避けるために追加のスタッフの派遣が必要かどうかを確認できます。ガラスの割れる音も検知できるため、不法な侵入を防ぐこともできます。

ビデオや音声（カメラやマイク）といったさまざまなタイプのセンサーを使用することで、検知結果の信頼性が高まり、より実用的な洞察が得られます。

AXIS Audio Analyticsは対応するAxisデバイスに組み込まれています。元の音声ストリームを保存することなく、音声をキャプチャーして検知します。このようにしてプライバシーを保護します。AXIS Audio Analyticsがエッジベースであり、音声メタデータを提供するためこれが可能になっています。

目次

1 はじめに 4
2 テクノロジー 4
2.1 音声イベントの検出 4
2.2 音声のキャプチャーと処理 4
2.3 エッジベースまたはサーバーベースの分析機能 5
2.4 メタデータ 6
3 AXIS Audio Analytics 6
3.1 プライバシー 7
3.2 ユースケース 8
3.2.1 アラートを伴う音声検知 8
3.2.2 センサーを組み合わせることで監視システムを最大限に活用 10
3.2.3 ダッシュボードへの入力 10
4 法的な制限 11
5 免責事項 11

1 はじめに

セキュリティと安全性に関する音声分析機能を活用することで、音声のパターンを検知し、ライブ音声の中に潜んでいる重要な音声に焦点を当てることができます。分析機能は、悲鳴、怒鳴り声、話し声を識別し、ガラスが割れる音を検知し、オペレーターに通知して早い段階で警告を与えることができます。

音声分析と映像監視を組み合わせることで、進行している潜在的なインシデントをオペレーターに警告し、関連するカメラビューを指示することができます。これにより、早期発見と迅速な介入が実現し、多くの場合、事態の悪化を防止することが可能となります。

本ホワイトペーパーでは、音声分析機能をセキュリティと安全性の向上に活かす方法をご紹介します。カメラやマイクで直接実行されるAIベースの音声分類など、リアルタイムのエッジ分析に焦点を当てながら、音声をキャプチャーして処理するテクノロジーについて考察します。また、エッジでの音声分析機能があれば、音声メタデータを使用することでプライバシーを保護することができます。このいくつかのオプションについての説明も含まれています。

本ホワイトペーパーは、法的なアドバイスを提供することを意図するものではありません。監視システムを設置する前に、所在する地域とそのユースケースに適用される法規制を確認する必要があります。地域の法規制および推奨事項に準拠してシステムを使用する責任はシステムの所有者が負います。

2 テクノロジー

2.1 音声イベントの検出

音声イベントは、悲鳴やガラスの割れる音など、人間が通常、特徴的な概念として認識する音声セグメントです。こうしたタイプの概念的な音は、ビデオ分析でオブジェクトクラスが検知およびラベル付けされるのと同様の方法で検知してラベル付けすることができます。

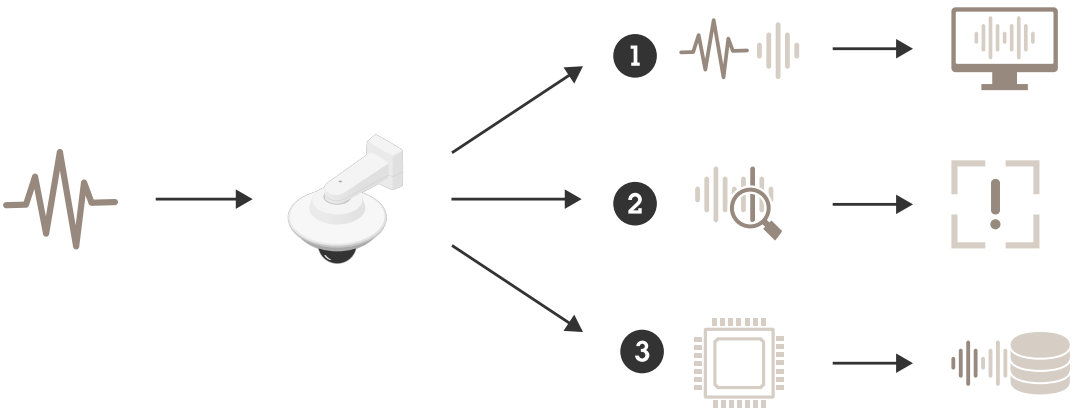
音声パターンを認識できるようにトレーニングされている分析機能は、一般的に音量レベルからさまざまな周波数のエネルギーに至るまで、その特性の組み合わせを経時的に判断できるように構成されています。特定の音声パターンが検知されると、視覚的なアラートまたはアラームのトリガーという手段で、システムからスタッフに自動的に通知が送信されるように設定することができます。

AIベースのアルゴリズムを使用すれば、大量のデータを用いて機能をトレーニングすることが可能となります。たとえば、アルゴリズムを用いて、数千に上る音声でトレーニングされた機能は、人間の悲鳴を確実に検知できるようになります。

2.2 音声のキャプチャーと処理

音声分析機能により、キャプチャーされた音声データを使用して関連性の高い音声特性が分析され、音声以外の出力が生成されます。音声のキャプチャーとは、基本的に音声をデジタル化してソフトウェアで使用できるようにすることを意味します。これは、マイクにより空気中の音の振動を拾い、そのアナログ信号がデジタル信号に変換されてから、処理ユニットに転送されることで実現します。キャプチャーされた音声は、フラッシュメモリやハードドライブなどの常設メディアに取り込まなければ、録音されません。Axisデバイスでは、音声のストリーミングと録音がデフォルトでオフに設定されています。

最初の音声キャプチャーの後、キャプチャーされた情報は次の処理段階用に準備されます。さまざまな準備を並行して行うことも、個々に行うこともできます。



音声の準備

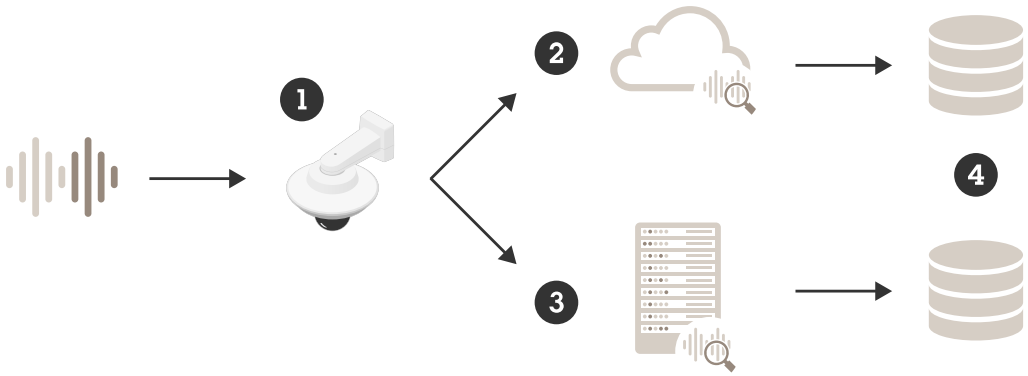
1. 変換
2. リアルタイムでのエッジ分析
3. ストリーミングまたはストレージのための処理とエンコード – *Axis*デバイスを使用する場合、音声ストリーミングを意図的にオンにしない限り、音声のストリーミングや保存は行われません。

- **変換。** 音が抽象化され、音響スペクトルを示すグラフなどの視覚情報に変換されます。このプロセスを逆にすることはできません。つまり、スペクトルグラフから元の音声を取り出すことはできません。
- **リアルタイムでのエッジ分析。**

  音声をエッジで処理する場合は、**音声分類機能**を使用することができます。これにより、音の特性を説明するメタデータが生成されます。そのメタデータから、元の音声を復元することはできません。
  音声検知機能を使用すれば、パターン、音量、周波数を検知し、ステータス情報を取得することができます。この場合も、元の音声を復元することはできません。
- **処理およびエンコード。** 通常、元の（変換や分析が行われていない）音声が使用される場合は、何らかの処理とエンコードが実行され、意図されているユースケースに合わせて音声データが準備されます。こうしたユースケースでは、エッジでの音声データの保存、追加処理のための外部クライアント（サーバーまたはクラウド）へのストリーミング、または外部ストレージなどが必要になる場合があります。Axisデバイスでは、まず音声ストリーミングを意図的にオンにする必要があります。プライバシー保護の観点から、音声ストリーミングはデフォルトでオフになっています (音声プライバシー制御)。

2.3 エッジベースまたはサーバーベースの分析機能

システム内の分析エンジンの場所は、多くの理由で重要です。特にプライバシーの問題を管理し、個人データに関する規制を遵守するためには、ソフトウェアアルゴリズムがどこで音声データを分析するかが重要です。音声データをネットワーク経由で送信できない状況があり、この場合、キャプチャーした (保存されていない) 音声データをローカルで分析できる必要があります。エッジで実行できないような、計算負荷が非常に高いアルゴリズムが必要な場合は、デジタル音声データをクラウドまたはサーバーに送信する必要がある場合があります。



分析を実行できるところ

1. エッジ
2. クラウド
3. サーバー
4. ストレージ – 音声ストリーミングが有効になっていれば、元の音声を保存することができます。

- **エッジ分析。** 分析機能をエッジで実行する場合は、音声ストリームをデバイスから送信する必要はありません。実行された分析から得られた出力（メタデータまたはトリガー）のみが送信されます。AXIS Audio Analyticsはエッジベースです。
- **サーバー分析。** 分析機能をサーバーで実行する場合は、音声データをデバイスからサーバーに送信する必要があります。音声データがデバイス上で前処理される場合、送信する必要があるのは抽象化または非個人化されたメタデータのみです。通常、サーバーはシステム所有者が管理するクローズドシステム (外部から閉鎖されたシステム) の一部であるため、送信される音声のプライバシーを管理できます。この場合でも、適用される規則や規制に準拠していることを確認する必要があります。
- **クラウド分析。** 音声データをクラウド経由でサーバーに送信することができます。サーバー分析の場合と同様に、音声情報を前処理してメタデータにすることが可能です。クラウドの使用は分散化されていることが多いため、プライバシーの問題に対処し、規制に準拠していることを確認することがさらに重要です。

2.4 メタデータ

音声分析では、音声レベルデータのメタデータストリームが継続的に生成されます。また、分析機能は、音声分類、SPL (音圧レベル)、適応音声検知によって検知されたイベントに基づいてメタデータを生成します。エッジで実行する分析機能の場合は、デバイス内で音声情報が分析されます。つまり、実際の音声のストリームを外部に送信する必要がないということです。シーンで発生した事態の洞察が得られるメタデータのみを送信することが可能となります。

メタデータのストリームにより、オーディオエンベロープの視覚的な検査が可能となります。たとえば、ダッシュボードに音声イベントとビデオイベントを並べて表示することができます。メタデータを活用することで、特定のイベントや不測の音声を効率的かつ迅速に検索できるようになるのです。これにより、調査担当者は膨大な量の映像を検索する時間を何時間も節約することが可能となります。

3 AXIS Audio Analytics

AXIS Audio AnalyticsはデバイスソフトウェアのAXIS OSに組み込まれており、対応するカメラやその他のAxisデバイスに無料で含まれています。

AXIS Audio Analyticsはエッジベースで、アルゴリズムがデバイス上で直接実行されます。これにより、最適なスケーラビリティ、低データトラフィック、プライバシー保護が実現します。分析機能による出力（メタデータまたはトリガー）のみが保存されます。音声は録音されたりデバイスからストリーミングされることはなく、メタデータから元の音声を再作成することもできません。

- **音声分類。** 悲鳴、怒鳴り声、話し声、ガラスが割れる音などの特定の音声を検知および分析するAIベースの音声分類機能です。分類分析により、結果として音声の特性を説明するメタデータが生成されます。
- **SPL (音圧レベル)。** 音の大きさをデシベル (dB) で測定します。SPLの測定は、音声の品質から聴覚に対する安全性といった各側面の評価に役立ちます。
- **適応音声検知。** 音声レベルが突然変化するとイベントを作成する音声検知機能です。あらゆる種類の音声のピークを検知でき、ノイズレベルが変化しても周囲のノイズに適応できるというメリットがあります。

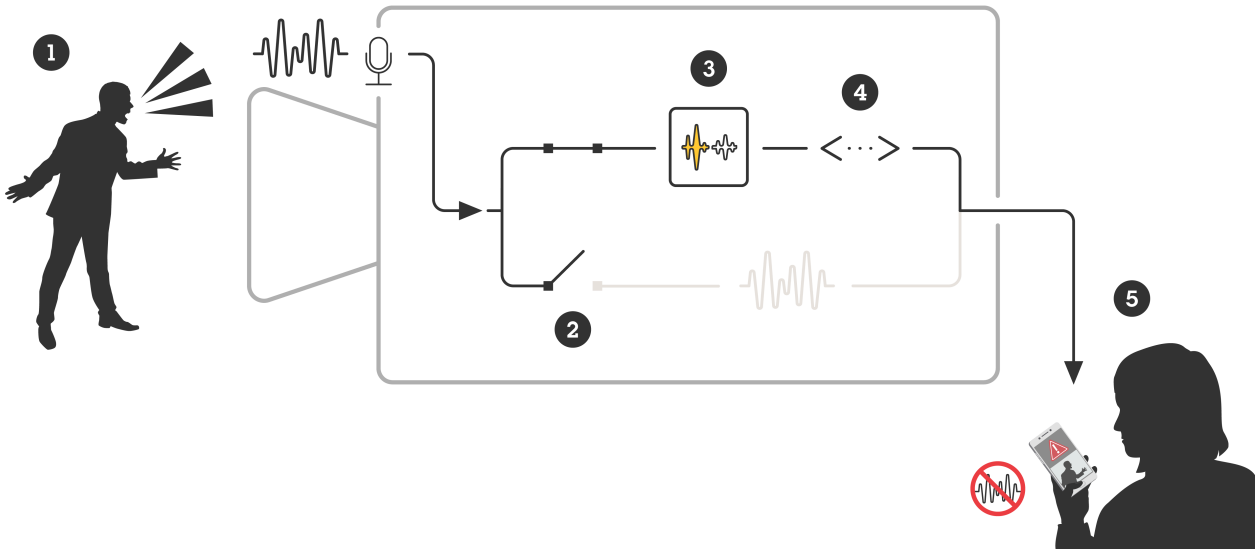
AXIS Audio Analyticsは、新機能やアップグレード機能を継続的に導入します。

3.1 プライバシー

通常、音声分析機能によって、受信音声が録音されることやデバイスから送信されることはありません。追加調査を実行するダッシュボードやオペレーターへの警告を発信するビデオ管理ソフトウェアなど、受信システムで特定のイベント、パターン、音声レベルの検索を行えるように、単に音声が処理されるだけです。音声データが再構築されることも、個人的な会話が録音されることもありません。これは、分析機能がエッジベースであり、音声メタデータが生成されるためです。

AXIS Audio Analyticsは音声の録音やストリーミングを行わず、メタデータの送信のみを行うようにデフォルトで設定されています。プライバシー保護という観点から、Axisデバイスでは音声ストリーミングがすべてデフォルトでオフになっています（音声プライバシー制御）。つまり、音声はストリーミング、録音はされず、再作成もできないということです。必要に応じて、意図的に音声ストリーミングをオンにすることはできますが、関連性の高い何らかの事象が発生した場合は通知が発信されます。これは、音声ストリーミングがオフになっている場合も同様です。

下図には、AXIS Audio Analyticsで、音声プライバシー制御と連携して音声がキャプチャーされ、メタデータを用いてアラートが作成される仕組みが示されています。



Axisデバイスの音声プライバシー制御とAXIS Audio Analytics：悲鳴 (1) からアラート (5) までの流れ

1. マイクで音声がキャプチャーされます。
2. 音声ストリーミングはデフォルトでオフになっています。
3. *AXIS Audio Analytics*の音声分類機能により、受信音声の中に含まれている悲鳴や怒鳴り声が検知されます。
4. 音声分類機能により、イベント通知などのメタデータが生成されます。

5 イベント通知とメタデータに基づいて、関係者にアラートが発信されます。関係者はビデオストリームをチェックすることで、状況を確認することができます。音声ストリーミングはありません。

Axisはマイクの代わりに音響センサーを搭載したデバイスも提供しています。音響センサーにより、デバイスは音声ストリーミングの可能性を完全に排除しながら、AXIS Audio Analyticsを使用できます。これらのデバイスは、音声のストリーミングや録音を行わず、音声メタデータのみを生成するように設計されています。

3.2 ユースケース

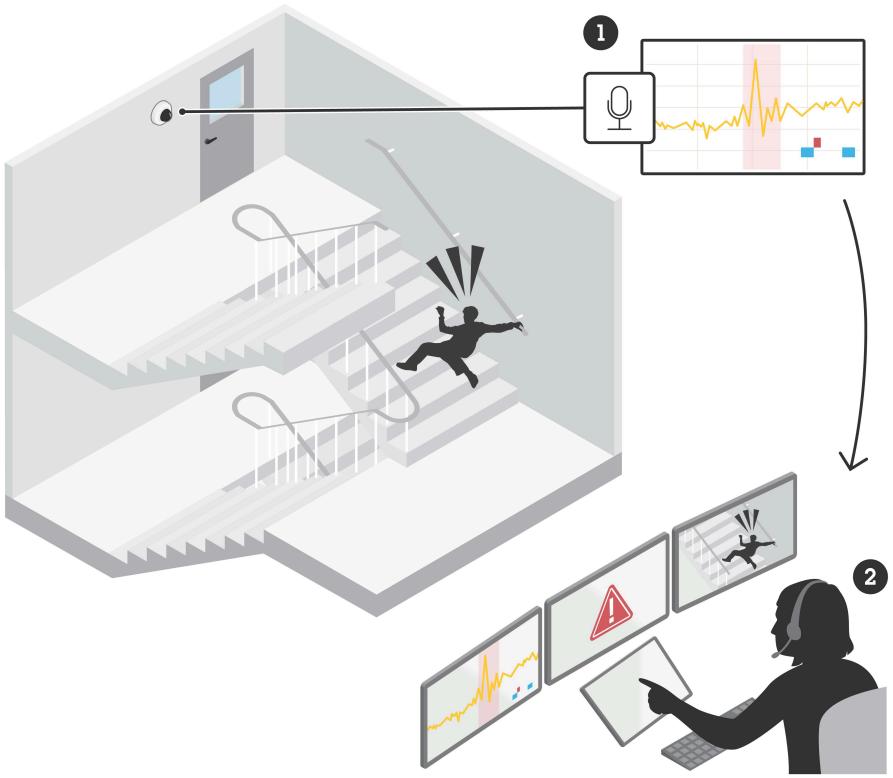
無関係なノイズの除去という点でAIベースの分析機能は高い可能性を秘めていますが、背景ノイズが多いと誤報が発生する場合があります。窓ガラスに当たる雨、雷、サイレン、音楽、賑やかな場所での人々の話し声は、誤報発生の要因となり得ます。そのため、一般的な使用場所として、銀行や受付デスクといった静かなエリア、および営業時間外の店舗やレストラン、階段、オフィスなどのさまざまなタイプの屋内空間が挙げられます。

3.2.1 アラートを伴う音声検知

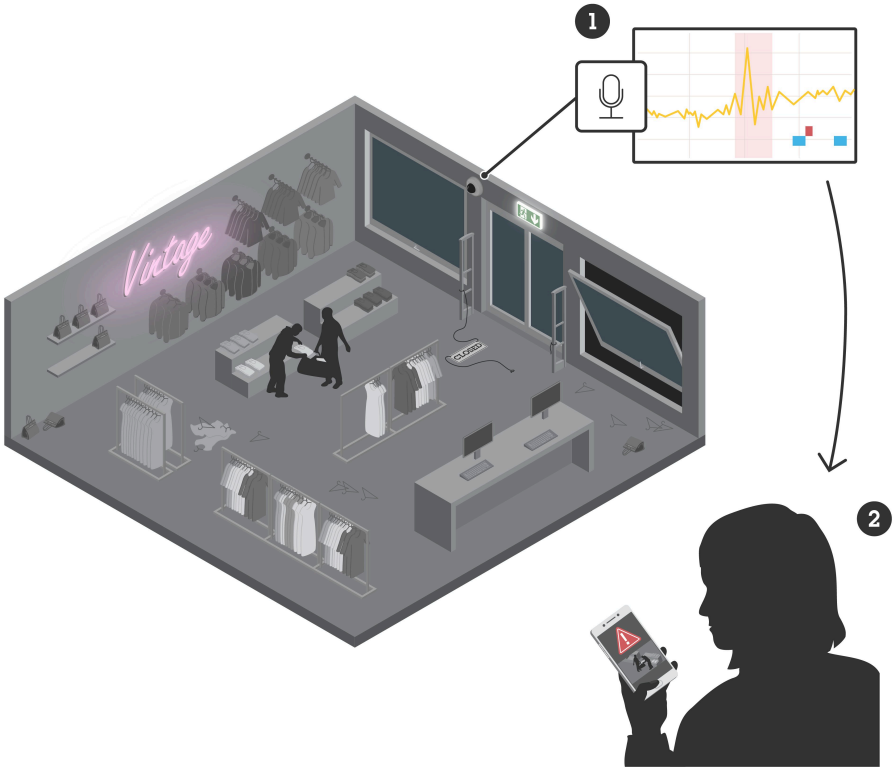
銀行や受付デスクでは、音声分類機能を使用してエリアを監視し、悲鳴、怒鳴り声、話し声、ガラスが割れる音などを検知することができます。こうした音声が検知されると、視覚的なアラートまたはアラームのトリガーという手段で、システムのイベントデータと音声メタデータによってスタッフに自動的に通知が送信されます。このように早期に警告が発信されることで、迅速な対応と介入が可能となります。

1 音声分類機能を搭載したデバイスが受付デスクで悲鳴や怒鳴り声を検知します。
2 オペレーターがアラートを受信し、措置を講じる前にビデオストリームをチェックして状況を確認することができます。

適応音声検知機能を使用して、営業時間外に異常な音声を検知することができます。分析機能によって環境音が分析され、人の声や窓ガラスの割れる音、またその他の突発的な異常音が検知された場合はそれに応じた操作が実行されます。イベントが検知された場合は、分析機能によってメタデータが転送され、それに応じてオペレーターに通知が発信されます。



1 音声分類機能を搭載したデバイスが営業時間外に異常な音声を検知します。
2 オペレーターがアラートを受信し、措置を講じる前にビデオストリームをチェックして状況を確認することができます。



1 音声分類機能を搭載したデバイスが営業時間外に店内の音声を検知します。
2 店舗経営者がアラートを受信し、措置を講じる前にビデオストリームをチェックして状況を確認することができます。

3.2.2 センサーを組み合わせることで監視システムを最大限に活用

監視システムには、多くの場合、複数のタイプのセンサーが組み込まれています。カメラのイメージセンサーは、シーンの視覚的側面を記録するものです。レーダー技術に基づく動体検知器や赤外線放射などの非視覚センサーもよく使用されます。他のタイプの情報が得られる非視覚センサーを活用することで、カメラの機能を補完することができます。

また、監視システムに音声センサー (マイクまたは音響センサー) を採用することで、想定されるすべてのユースケースの大部分が強化されます。音声機能と音声分析機能を非音声システムに追加することで、マルチセンサーによる相互作用が実現します。ビデオ分析機能に音声分析機能を追加することで、検知の信頼性が向上します。特にこれは、低照度条件下や、ビデオキャプチャーが許可されていない場所や不可能な場所でビデオ分析が困難な場合に当てはまります。

また、ビデオ管理ソフトウェアを用いて、ビデオ分析機能と音声分析機能の両方が反応した場合にのみアクションがトリガーされるようにシステムを設定することもできます。例として、音声分析機能によって悲鳴などを検知し、ビデオ分析機能によってカメラ視野内の個人を検知するといった活用方法が挙げられます。環境によっては、この組み合わせによって適切なレベルのセキュリティを確保することができます。

3.2.3 ダッシュボードへの入力

音声メタデータは、メタデータを収集して視覚的に表示する分析ダッシュボードやビジネスインテリジェンスプラットフォームに入力することができます。これにより、リアルタイム分析と過去の傾向の分析を行い、速やかに全体像と実用的な洞察を取得することができます。顧客の流れや顧客体験に基づいた統計分析を取得すれば、データに基づいた意思決定と業務の改善が実現します。

ダッシュボードを使用することで、音声を実際に聞かなくても、また元の音声を取得しなくても、結果を確認することができます。これなら、たとえばイベントを計量することで実用的な洞察を得られるだけでなく、データのプライバシー保護を徹底することが可能となります。音声を録音するか、キャプチャーのみを行うかによって、法的な制限が異なる場合があることに注意してください。

4　法的な制限

映像監視にマイクを使用することに対して懸念を抱く人は少なくありません。通常、こうした懸念は、映像素材と共にプライベートな会話が記録されることに関連しています。その点、音声分析機能を活用すれば、一般的に音声を録音またはストリーミングする必要がなくなります。監視を規制する法律は地域や国によって異なるため、監視システムに音声機能を使用する前に、法律で許可および禁止されている事項を確認する必要があります。

音声のキャプチャーや録音については、国の立法機関や地域の種々の規則や規制によって、複数の理由で禁止されている場合、または特別な配慮が必要な場合があります。地域や環境によっては、音声キャプチャーが許可されていても、音声録音が禁止されているところがあります。また、企業が自社の敷地内での音声監視を禁止している場合もあります。

5　免責事項

本書とそのコンテンツはAxisの厚意により提供されており、本書に対するすべての権利またはそれに関連する知的財産権 (商標、商品名、ロゴタイプ、および同様のマークを含むがこれらに限定されない) は法律によって保護され、本書に対するすべての権利、権原、および/または利益、またはそれに関連する知的財産権は、Axis Communications ABに帰属し、今後も継続して帰属するものとします。

本書は情報提供のみを目的として「現状有姿」で提供されており、いかなる保証も伴いません。本書に記載されている情報は、法的助言を意図したものではありません。本書は、Axis Communications ABおよび/またはその関連会社に、いかなる法的義務も発生させるものではありません。Axis製品に関連するAxis Communications ABおよび/またはその関連会社の義務は、Axisとかかる製品をAxisから直接購入した事業体との間の契約条件のみに従うものとします。

解釈上の疑義を回避するため、本書の使用、結果、および性能に関するすべてのリスクは、本書のユーザーが負うものとします。またAXISは、法律で認められる最大限の範囲で、法定、明示、黙示を問わず、商品性、製造物責任、特定の目的への適合性、権原の黙示の保証、または本書に関する提案、仕様、またはサンプルから生じる保証を含むがこれらに限定されないすべての保証を否認および除外します。

T10202823_JA/JA/M3.3/2507

Axis Communicationsについて

Axisは、セキュリティ、安全性、運用効率、ビジネスインテリジェンスを向上させることで、よりスマートでより安全な世界の実現を目指しています。ネットワークテクノロジー企業として、また業界をけん引するリーダーとして、Axisは映像監視、アクセスコントロール、インターコム、音声ソリューションを提供しています。これらのソリューションは、インテリジェントアプリケーションによって強化され、質の高いトレーニングによってサポートされています。

Axisは50ヶ国以上に5,000人を超える熱意にあふれた従業員を擁し、世界中のテクノロジーパートナーやシステムインテグレーションパートナーと連携することで、カスタマーソリューションをお届けしています。Axisは1984年に創業し、本社はスウェーデン・ルンドにあります。

© 2024 — 2025 Axis Communications AB. AXIS COMMUNICATIONS、AXIS、ARTPEC、VAPIXは、スウェーデンおよびその他の国で登録されたAxis ABの登録商標です。その他のすべての商標は、それぞれの所有者に帰属します。記載内容は予告なく変更されることがあります。