Technical guide to network audio

Technologies and considerations for a world of audio possibilities

Introduction

This *Technical guide to network audio* was created to support you in your work with Axis network audio systems and help you stay up-to-date with the changing technological landscape. It provides a complete overview of network audio, from explaining the very basics of audio and acoustics, via the technologies which enable audio over IP, all the way to Axis network audio products and complete systems.

Table of Contents

1. Network audio: overview, purposes, and benefits

Network audio uses standardized IP-based networks for transporting audio, as opposed to the dedicated audio cabling used in traditional audio systems. Sending digital audio streams over an IP network enables individual speaker control from remote, including zone configuration, content management, scheduling, and health monitoring, all managed from a single user interface.

1.1 Overview of a network audio system

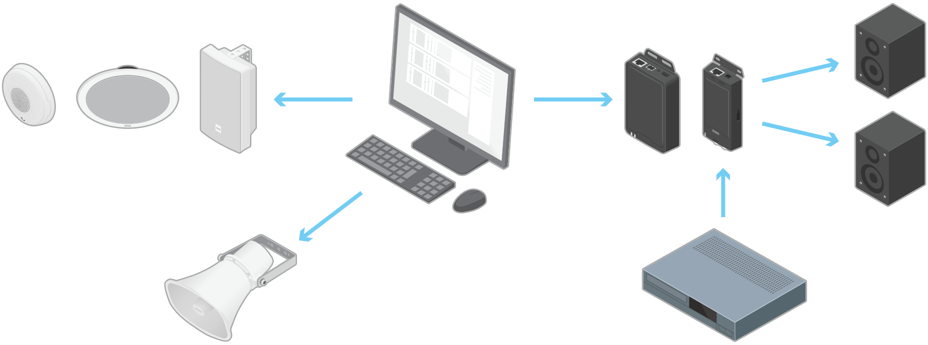


Figure 1.1a *A network audio system can be controlled from a single point via an intuitive interface. System devices make it possible to also control analog audio sources and analog speakers as part of the network audio system.*

Network audio uses an IP network and standard IT equipment as the backbone for transporting and managing audio. The core components of a network audio system are network speakers, network microphone consoles, the network itself, a server, and audio management software. A system from Axis can also include system devices with the purpose to provide a smooth migration from a legacy (analog) audio system to network audio.

The network, the server and storage components are all common off-the-shelf equipment. As the speakers and microphone consoles are computer-based, they have capabilities that cannot be matched by analog audio systems, and network audio is digital all the way.

1.2 Purposes

Network audio and voice messages can play an important part in a security system, often integrated with network video surveillance, but network audio is also ideal for safety systems, public address announcements, and background music.

> **Improve security.** A network audio system is the perfect addition to a video-based security installation. Perimeter protection is just one example: imagine a potential intruder climbing a fence. A network camera registers the event and sends an alert to a security guard. He then uses the network audio system to communicate a warning to the intruder, "We can see you, you're trespassing." More often than not, this type of warning is sufficient, preventing the need for additional security measures.

> **Increase safety.** With network audio, urgent safety information can be quickly and efficiently broadcast to a very large number of people. In cities, schools, or in areas that are under threat of earthquakes or other natural disasters, network audio (public address) can have a role in a larger safety system. Network speakers can be used for emergency callouts regarding evacuation or other warnings. With their integrated microphones, Axis network audio speakers can even be used as active security sensors. If you run audio analytics that can detect, for example, glass breaking or gunshots, it adds a completely new dimension to an audio system.

> **Improve operational efficiency.** In public address systems, network audio provides a flexible way to manage and play different kinds of informative messages and updates in schools, retail stores, hotels, and public buildings. You can make live announcements calling someone to come to a specific area (such as a colleague to the clothing department), make scheduled announcements (for instance, about the start of the school day), or issue live or triggered announcements during an emergency. You can broadcast to single or multiple zones to make callouts to specific rooms. Network audio can also create ambiance with background music. It is easily managed with both central and remote control, and preset volume and music choices. Background music can be combined with scheduled and live announcements for the best customer experience.

1.3 Applications in key industry segments

Network audio can be used in an almost unlimited number of applications, both for security or safety purposes and for operational efficiency. Typical use in key industry segments include:

- **Smart cities.** Network audio can be a useful, crime-deterring complement to network video surveillance in cities. It can also increase safety by guiding people in cases of emergency.
- **Retail.** Network audio can provide retail stores with a flexible and easily managed system for making announcements and security callouts to staff or customers. It can be used to deter shoplifters and loiterers through the right voice messages, but also to play commercial announcements.
- **Critical infrastructure.** In power plants and heavy industries, network audio systems can provide announcements and warnings regarding safety hazards and incidents such as gas leaks, oil spill, or chemical leaks.
- **Industrial.** In warehouses, datacenters, and logistics centers, network audio can provide triggered or live announcements about security concerns and incidents.
- **Education.** In a school or on a campus, network audio can provide a flexible public address system for information callouts, as well as increasing security and safety through live or triggered instruction announcements and warnings in case of incidents.
- **Transportation.** In train stations, subways, and tunnels, network audio can be used for security callouts to prevent vandalism and theft, and warnings and safety instructions in case of incidents. It can also be used to communicate valuable information about delays or schedule changes.

Figure 1.3a *Network audio can be installed for security or safety purposes, or for operational efficiency.*

1.4 Benefits of network audio

A fully digital network audio system provides many benefits and advanced functionalities, such as remote accessibility, easy integration possibilities, central remote health monitoring, and scalability, flexibility, and cost-effectiveness. And because a network audio system consists of true

IT devices that connect to the standard network, no dedicated audio cabling is needed. If you prefer to keep using legacy audio cabling or analog speakers, however, there are specific system devices that enable you to do that while still enjoying some of the digital benefits.

> **Digital all the way.** In network audio, the audio is digital all the way from the source to the speaker. The sound is digitally stored, processed, and transmitted, without any analog-to-digital or digital-to-analog conversions. The integrated digital signal processor makes sure that the sound is optimized right in the speakers, which means that no technician has to go onsite to control the audio quality, and there is no risk of equipment interfering with the audio signal. The sound signal remains strong no matter how long the cables are, and audio announcements are clear and easy to understand.
>
> By comparison, in analog audio the sound waves are converted into electrical voltage. This electric signal can then be transported, stored, and eventually converted back into sound waves. When the electric signal gets passed through many steps, or over long hauls of wire, its voltage level will drop and will need to be amplified. This process can increase the noise and lower the quality, and the output audio will sound less good. In digital audio, however, instead of using varying voltage, numerical values are used to describe the waveform. These values can be stored and transmitted without changing even the slightest, even in areas of electromagnetic disturbance. They will not be corrupted by noise, and when converted back into sound waves they will sound exactly the same as the original audio.

> **Central accessibility and health monitoring from remote**. Network audio devices can be configured and accessed remotely, enabling authorized users to control them from virtually any networked location in the world. Network speakers can also have self-test functionality which works from remote. This is a way to check that the speaker is in working order by having it provide audio feedback to the system.

> **Easy, future-proof integration**. The use of open-standard IP technology simplifies integration with other systems as speakers can be integrated directly into a VMS (video management system) or a standard Voice over IP (VoIP) phone system using Session Initiation Protocol (SIP). With one, integrated system, it is easy to comply with changing needs.

> **Scalability and flexibility**. IP offers flexibility and broad ability for future reconfiguration, in case of any unforeseen changes in system requirements. Speaker groups and zones can be reconfigured without the need for new cabling. The audio system can be expanded or changed when needed and this is easily done in the software.

> **Cost-effectiveness**. An IP audio system typically has a low total cost of ownership because it uses an IP network infrastructure which is often already in place and used for other applications. There is no need for a dedicated audio cabling network. A network audio system

can be health monitored and maintained from remote, which is also cost efficient with less staff hours and less equipment needed.

> **Everything built in**. A traditional audio system generally needs a dedicated technical room to place the rack cabinets with all audio components, such as power amplifiers, digital signal processors and more. With network audio, these tasks are done through edge processing in the speakers, which means that network speakers are complete, high-quality audio systems in themselves and there is no need for a technical room. Even audio management software for handling zoning, content, and scheduling is built in.

> **Power over Ethernet (PoE)**. Network audio products support PoE technology. PoE enables networked devices to receive power from a PoE-enabled switch or midspan through the same Ethernet cable that transports the data (audio). There is, therefore, no need for a power outlet at the speaker's location, nor any extra audio cables or technicians to calculate cable dimensions and capacity. PoE provides substantial savings in installation costs and can increase the reliability of the system.

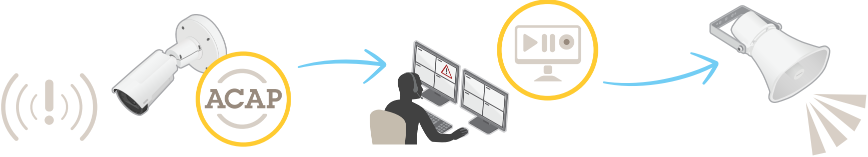
1.5 Video surveillance with audio

A security solution with audio and video can be monitored or unmonitored, with different levels of camera integration.

> **Prerecorded messages triggered by motion detection.** A camera detects motion and sends a command to a speaker to play a prerecorded message. The speaker plays the message.

> **Prerecorded messages manually triggered.** The camera detects motion and sends a notification to an administrator. Upon getting that notification, the administrator triggers a prerecorded audio message. The speaker plays the message.

> **Live voice messages.** A camera detects motion and sends a notification to an administrator. The administrator then speaks a live message using his microphone, and the speaker plays that live message.



Network audio can also be used in a number of ways as a standalone audio system for deterrence, information, safety messages, or background music.

2. Audio fundamentals

As digital as an audio system may be, the audible sound itself consists of physical vibrations. This chapter provides background information about the physics concepts of audio, such as frequencies, sound pressure, human sound perception, and the basic units of sound measurement.

The content of this chapter is not essential for most network audio use cases, but it serves as background information which may be of interest for more complex installations and solutions.

2.1 What is sound?

Sound is an audible pressure wave. Vibrations from your vocal chords or the diaphragm of a speaker disturb parts of the air. These pressure differences are what we hear. The qualities of a sound — its loudness and pitch — are determined by the properties of the sound wave.

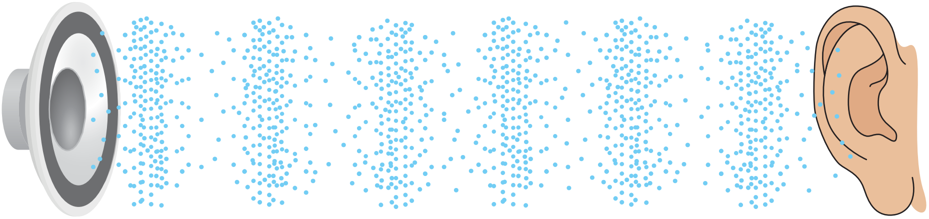


Figure 2.1a *Pressure differences in the air are interpreted as sound.*

2.2 Longitudinal waves

A sound wave is a longitudinal wave, which means that the air vibrates parallel to the direction of propagation of the wave. Air molecules vibrate back and forth in the same direction as the sound travels. You can visualize it as the movement that occurs when you stretch and compress a coil (typically a Slinky toy), where the distance between coils increases and decreases.

However, longitudinal waves are not as easily visualized as the other type of wave, a transverse wave. In a transverse wave, the vibration back and forth takes place in directions perpendicularly to the direction in which the wave travels. You can visualize this wave, too, with a coil, by moving the ends of the coil in a direction perpendicularly to the length of the coil. One example of a transverse wave is an electromagnetic wave (such as light), where the electric and magnetic fields vary perpendicularly to the direction of propagation.

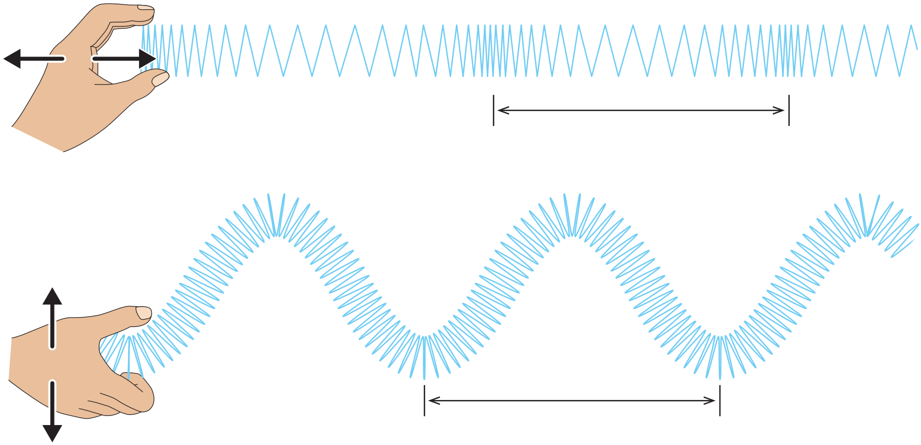


Figure 2.2a *Using a coil to visualize a longitudinal wave (top), such as a sound wave, and a transverse wave (bottom). The wavelength is marked as the distance between two minima.*

Since it is easier to visualize a transverse wave than a longitudinal wave, a sound wave is usually graphically represented by a transverse wave. Despite not being completely true to the nature of the wave, this graphical waveform provides a good understanding of a sound wave as it moves through the air over time, and it makes it easier to see the wave's fundamental characteristics: wavelength, amplitude, and frequency. Both longitudinal and transverse waves have these characteristics.

Do not be confused by a graphical representation of a sound wave, such as that on an oscilloscope which displays a transverse wave when displaying sounds. The real sound is always longitudinal.

2.3 Pitch

When we talk about pitch – how high or low a sound is – we are talking about the frequency of the pressure wave, that is, how many times per second the sound wave vibrates. Frequency is measured in cycles per second, unit Hz (Hertz).

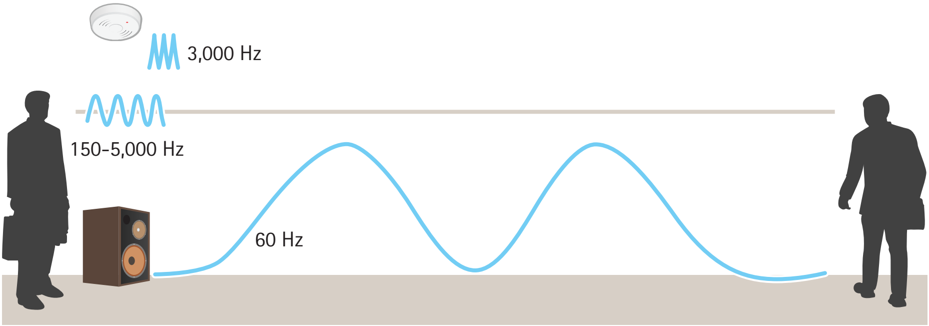


Figure 2.3a *Sounds, including human voices, can vary in pitch — vibrations per second (also known as frequency).*

2.4 Sound pressure

A noise can be barely audible, or extremely loud — but with the same pitch. These noises have the same frequency, but their waveforms differ instead in amplitude. The sound pressure is higher for a sound with a higher amplitude, and the changes in pressure is measured in Pascal, Pa. We can perceive sounds with as low sound pressure as 20 µPa (could be a mosquito ten feet away), and as high as 20 billion µPa (typically the launch of a space shuttle). When dealing with such a large range of numbers, it is easier to use a logarithmic scale. This is one of the reasons why sound pressure levels are more often measured in dB SPL.

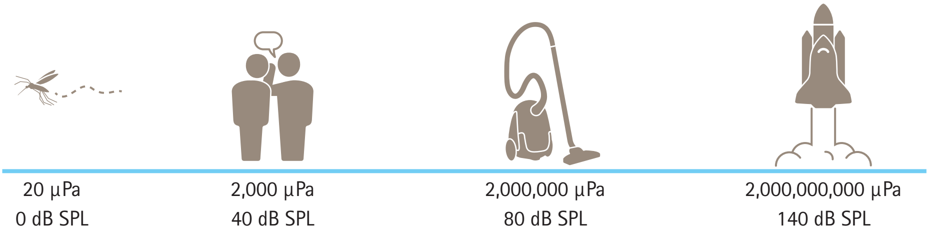


Figure 2.4a *Typical sound pressures from familiar sources, measured in both Pascal and decibel.*

2.5 SPL values

Sound pressure level (SPL) is the RMS (root mean square) value of the instantaneous sound pressures measured, in dB, over a specified period of time. SPL is not a constant average value of loudness but rather an average of the short peak values. An SPL value given for a speaker is assumed to be measured for a 1 kHz tone at a distance of 1 m, if nothing else is stated.

The sound pressure level of an audio source decreases with the distance from the source. Defined to start at 0 dB at 1 m from the source, the SPL decreases by 6 dB with each doubling of the distance from the source.

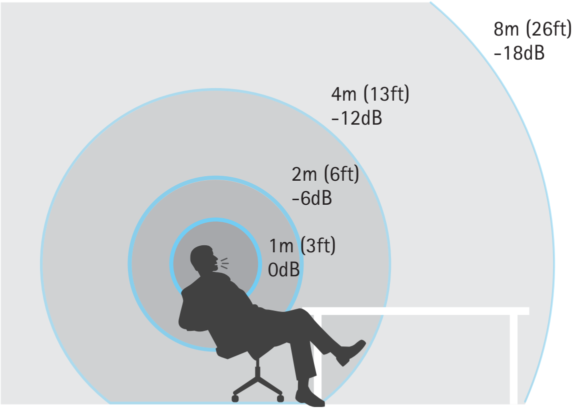


Figure 2.5a *The sound pressure level from an audio source decreases by 6 dB with each doubling of the distance from the source.*

2.6 Sound power

The unit of power, watt (W), is familiar from various electrical components, such as light bulbs, laptop chargers, and speakers. The unit can, however, be used in different ways, and in audio terminology we come across varieties like instantaneous power, average power, RMS (root mean square) power, and peak power.

An amplifier might be constructed to be able to deliver 300 W over a very short period of time, such as when a drum, explosion, or any other audio with a short and loud transient, will be heard. This means that the instantaneous power will increase really fast from very low to very high. The same amplifier might, however, only be rated for 50 W continuous use, since continuous use will produce a lot more heat, which impacts both the electrical components and the amplifier's performance.

2.7 Decibels

Because sound is perceived non-linearly, it is best measured and described using the non-linear unit decibel (dB). A doubling (measured in W) of the sound power equals to a 3 dB increase, and a doubling of the loudness equals a 10 dB increase.

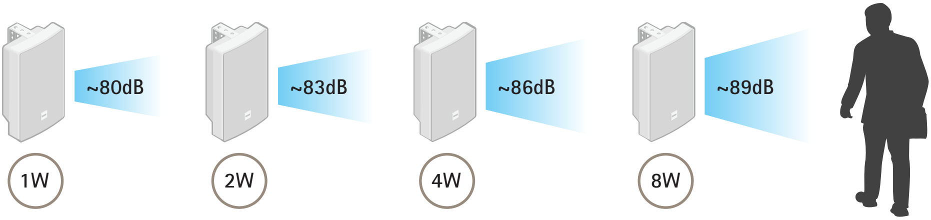


Figure 2.7a *A doubling of sound power, as measured in Watts, corresponds to a 3 dB increase.*

A sound pressure level given in the weighted dBA scale has been compensated for the human ear's frequency-dependent perception of sound. Using the unweighted dB scale, a 100 dB level at 100 Hz will, for example, be perceived to have a loudness equal to only 80 dB at 1 kHz, while 100 dBA will be perceived as equally loud at all frequencies.

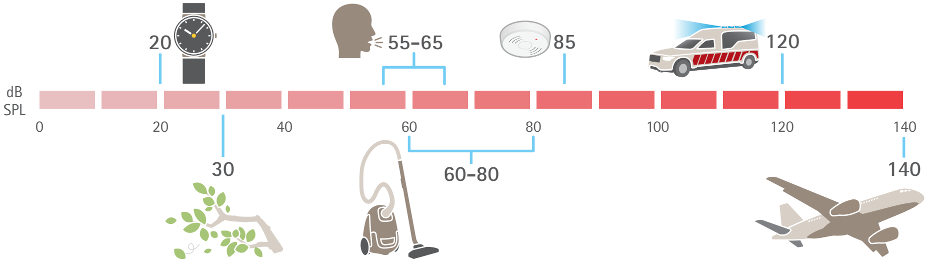


Figure 2.7b *Approximate sound levels, in decibel, from familiar audio sources.*

The decibel unit is often referring to a relative change in loudness. For expressing an absolute value, dB SPL should be used. A value of 0 dB SPL is the softest sound that the human ear can perceive.

2.8 Perceived loudness

The human ear is, in theory, able to perceive frequencies from 20 Hz to 20 kHz. The upper limit of 20 kHz is lowered with age but the high frequencies can still add character through overtones to audio with lower frequencies. Human speech, being complex with lots of harmonies, is scattered over frequencies from around 85 Hz (lowest for human male) to around 8 kHz (overtones for human female). In telephony, only the range of 300 Hz to 3.4 kHz is commonly used, and while it makes the voice audible, the audio will not be as clear as a full frequency recorded voice.

Even though the ear is sensitive to all frequencies between 20 Hz and 20 kHz, the sensitivity varies with the frequency. Sounds of a specific power will thus be perceived as having different loudness at different frequencies. The loudness unit *phon* takes this sensitivity into account and, for example, a sinusoidal tone of 50 phons is perceived as equally loud at all frequencies.

The difference in sensitivity can be visualized in equal-loudness curves. One line represents the sound level that must be used, in order for the sound to be perceived at the same volume for all frequencies. The different lines represent different phon values. It is evident from the curves that the sound level must be substantially higher at the lower frequencies in order to be perceived as equally loud as higher frequencies. This is because the human ear is less sensitive to lower frequencies. The minimum of the curves is placed around 2 kHz – 5 kHz, meaning that this is the frequency range to which a human ear is most sensitive, and in which the ear can best decipher a conversation. It is also the frequency range of human speech.

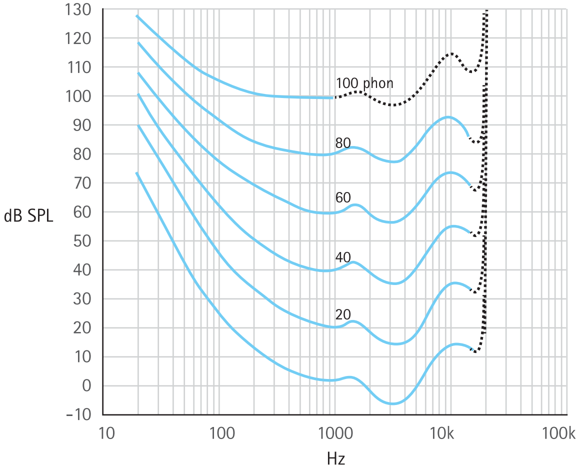


Figure 2.8a *Equal-loudness curves showing the sound pressure levels needed at different frequencies in order to make a sound perceived as equally loud over all frequencies. The curves are originally from the ISO standard ISO 226:2003.*

2.9 Sampling frequency

The sampling frequency is the number of audio “snapshots” taken per second of the analog input audio stream, in order to digitally reconstruct it. Lower sampling frequencies sample data less frequently. With a low sampling frequency, parts of the audio are not captured and the overall sound quality will be lower. With a higher sampling rate the audio stream can be more correctly recreated, delivering higher quality.

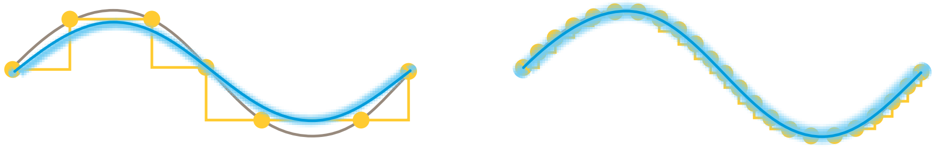


Figure 2.9a *Visualization of an analog sound wave and how it can be digitally represented. Left: with a too low sampling frequency (sampling at every yellow dot), the original wave (thin black line) will be wrongly represented (as the blue line) which results in lower sound quality. Right: with a high enough sampling frequency the analog wave can be accurately reconstructed.*

The sampling frequency must be at least twice as high as the highest input audio frequency that should be reconstructed, but even higher for good quality. In audio files and CDs, 44.1 kHz is a commonly used sampling frequency, thus using 44,100 samples per second.

2.10 Dynamic range compression

Audio that has large differences between the quietest and the loudest parts is said to have a “large dynamic range”.

Dynamic range compression is an audio signal processing operation that makes the quietest parts louder, while the loud parts either stay the same or become less loud. The operation decreases the dynamic range, and we perceive the recording as louder.

Compression of dynamic range is often applied in audio systems for restaurants, retail, and similar public environments that play background music at a relatively low volume. Apart from making the volume more constant, the compression also makes the quieter parts of the audio more audible over ambient noise.

2.11 Speaker sensitivity

A speaker’s sensitivity is its ability to reproduce sound when fed a certain power. Determining the sensitivity is usually done by feeding an audio signal of 1 W (typically at 1 kHz) and then measuring the sound pressure level in dB at a 1 m distance. Common values for speakers are around 85 - 92 dB. The higher the sensitivity, the louder the sound will be from the speaker when fed a certain power.

The sensitivity of the speaker is usually an indicator of the quality of the speaker. Lower sensitivity indicates a less powerful magnet and/or a smaller and cheaper coil. Therefore, in regards to audio quality, a larger speaker is not necessarily better than a smaller speaker. The size of a speaker is, in a way, what megapixels are to a camera: unless we also have a good camera lens (or speaker sensitivity), a higher resolution (or increased speaker size) is worth nothing.

2.12 Polar response

Some speakers have a very narrow direction of sound in order to achieve a high sound pressure in one direction. Others are made to have as wide spread of the sound as possible. A speaker's ability to reconstruct audio is dependent on the audio frequency. Generally, lower frequencies have a wider spread while higher frequencies are more directional.

A polar diagram can visualize how frequencies spread out differently from a speaker.

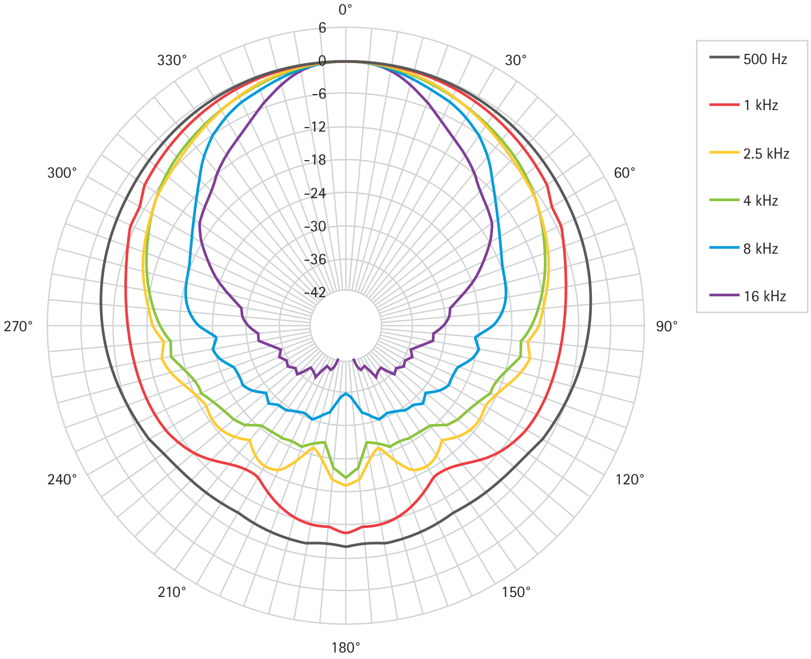


Figure 2.12a *Polar diagram showing the spread from a generic example speaker (which was located in the center of the diagram). The lower frequencies have a wider spread (even behind the speaker, at 180 degrees) while higher frequencies are more directional.*

3. Acoustics

The surroundings can have a great impact on the sound from an audio system. While calculating the effects can be a complex task, some understanding of acoustics goes a long way.

The content of this chapter is not essential for most network audio use cases, but it serves as background information which may be of interest for more complex installations and solutions.

3.1 Echoes

In a room that is completely empty, there will be reverb and/or delay in the sound. This is because all the flat surfaces are perfect for the audio waves to reflect against. If fabrics and uneven surfaces are added, such as sofas, curtains, and carpets, there will be less reverb, but the sound will also be perceived slightly less loud because of the absorption.

Sound waves are often reflected multiple times before reaching our ears. Knowing that the speed of sound in air is around 340 m/s (1020 feet/s), we can calculate the distance that an echo has travelled. If we hear the echo 0.25 s after the initial sound, for example, the sound has travelled around 85 m (0.25 s x 340 m/s), or 255 feet. For each reflection, the audio fades a little bit until we cannot hear it anymore.

3.2 The impact of room dimensions

The size of the room has a large effect on the audio experience. This is because the sound waves are reflected against walls, ceilings, and furniture. To better understand the impact of these reflections, it may be helpful to talk about the wavelength of an audio wave.

The wavelength, denoted by the Greek letter lambda (λ) is related to the speed of sound (v=340 m/s in air) and the frequency (or pitch, denoted by f, measured in Hz), according to:

> $\lambda=v/f$

A frequency of 20 kHz (20,000 vibrations per second) corresponds to a wavelength of about 1.7 cm (0.7 inches) while a lower frequency of 20 Hz (20 vibrations per second) corresponds to a longer wavelength of about 17 m (56 feet).

With wavelengths up to 17 m (56 feet) for the lowest bass, audible sound waves in a small room will be reflected against the walls before the waves have properly developed. This results in resonances and associated standing waves, causing some frequencies to be amplified (higher volume), and others to be attenuated (lower volume). We need a rather large room to hear the bass without distortion.

The impact of resonances on the experienced audio quality increases with the sound volume. With higher volume, the reflections will interfere more with the sound from the source.

3.3 Neutralizing room acoustics

In order to reduce annoying echoes in large or empty rooms, acoustic panels can be installed in the ceiling, on the walls, or both. The panels are made from sound-absorbing materials and create more neutral acoustics in spaces such as shopping malls, auditoriums, offices, and conference rooms. A similar effect can, however, be achieved by using curtains or other interior fabrics. Acoustic panels are usually quite effective for frequencies above 300 Hz, while the absorption capabilities gradually decrease for lower frequencies.

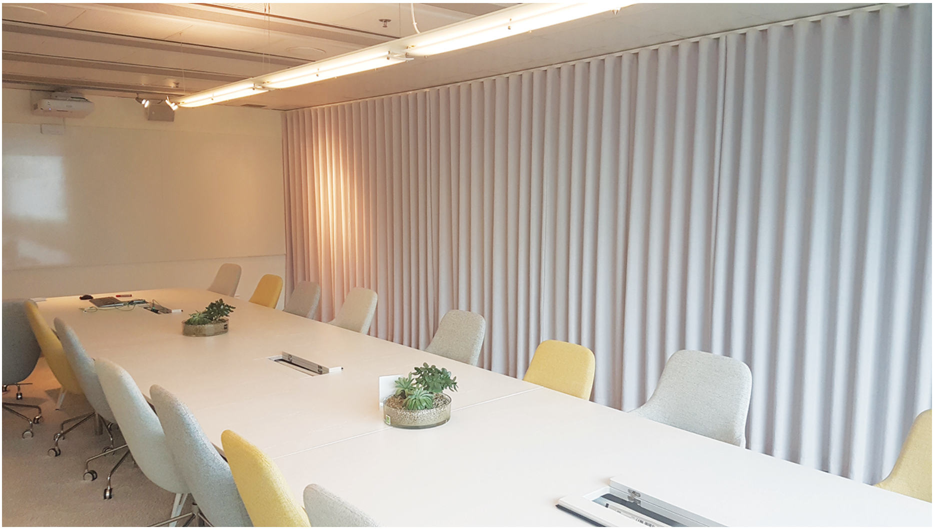


Figure 3.3a *Curtains and other pieces of fabric can significantly improve room acoustics.*

4. Audio technologies

Various audio technologies are involved in enabling network audio devices to provide high-quality sound. In security and safety installations, high quality typically means clear and intelligible speech, and this is ensured by use of several sound optimization techniques which are built into the speakers. The quality is maintained throughout the system by means of techniques for compressing, transmitting, and synchronizing audio streams.

4.1 High quality for network audio

In the entertainment industry, the term “high-quality audio” may refer to hi-fi speakers with stereophonic sound. Such speakers are designed to reproduce audio very accurately at high loudness, and a speaker system may consist of several types of speaker elements to manage as many audible frequencies as possible. There may be a bass element that reproduces sound up to 500 Hz, a mid-range element for frequencies between 500 Hz and 9 kHz, and a treble element for frequencies above 9 kHz, for example. Entertainment systems can also create very powerful listening experiences through the use of stereophonic, or stereo, sound, where two or more independent audio channels are used in separate speakers in order to create the impression of sound coming from various directions.

Network audio from Axis, however, is not about high loudness, an extensive frequency range, or stereo sound. For security or safety purposes, our solutions are all about maximizing the loudness of the rather narrow frequency range where human speech is most discernible. Even when network audio is used for background music, the loudness is pretty low. And there is no need for stereo sound — when the intended audience consists of people moving around in a school, a hospital, or a store, there is no left or right. For these reasons, network audio uses mono speakers with relatively low loudness.

4.2 Sound optimization techniques

Axis network speakers come equipped with digital signal processing (DSP) capabilities. DSP for public announcement is about analyzing and manipulating sound to improve speech intelligibility. In Axis network speakers, several sound optimization techniques – such as frequency optimization, loudness compensation, and dynamic range control – are built into the speakers to deliver excellent audio quality in any environment.

> **Frequency optimization**. The edge processing in Axis network speakers means they are frequency optimized, which gives the same characteristics to every speaker. As a result, they can combine without the need for manual tuning or configuration, and the system can be easily expanded just by connecting more Axis speakers.

> **Dynamic range control**, or dynamic range compression. An audio signal will often have peaks and troughs in volume, and dynamic range control can balance these to make sure that sound is broadcast at the ideal volume for listeners.

> **Loudness compensation**. If you have a 'loudness' button on your stereo, you might be familiar with the basic concept. At low volumes, some frequencies are less perceptible to the human ear (see section 2.8 about perceived loudness). Loudness compensation boosts those frequencies so that the listener does not miss anything. This happens automatically in Axis speakers, which makes them suitable both for important audio messages and for background music.

4.3 Audio communication modes

Depending on the application, there may be a need to send audio in only one direction or both directions. This relates to three basic modes of audio communication:

> **Simplex** means that audio can be sent in one direction only. In network audio, audio is usually sent from the user to a speaker, for example for communicating warnings or announcements through the speaker. But audio in simplex mode could instead be sent from the speaker to the user or operator, for example in remote monitoring applications where live audio from a monitored site is sent over a network.

> **Half duplex** means that audio can be sent and received in both directions, but only in one direction at a time. This type of communication is similar to a walkie-talkie. To speak, an operator must press and hold down a push-to-talk button. Releasing the button enables the operator to receive audio. With half duplex, there is no risk of echo problems.

> **Full duplex** means that users can send and receive audio (talk and listen) at the same time. This mode of communication is similar to a telephone conversation. Full duplex requires both the client (PC, SIP microphone, or VoIP phone) and the speaker to be able to handle full-duplex

audio. While full duplex has the advantage of simultaneous audio in both directions, it also increases the demands on available bandwidth.

4.4 Audio codecs

An audio codec (coder-encoder) is a software system that can digitize and compress data for transmission and decompress the received data. Axis network audio products and video products support three audio codecs:

AAC-LC (Advanced audio coding - low complexity) is a licensed standard, also known as MPEG-4 AAC. AAC-LC, particularly at a sampling rate of 16 kHz or higher and at a bit rate of 64 kbit/s or more, is the recommended codec to use when the best possible audio quality is required.

G.711 and **G.726**, which are non-licensed ITU-T standards. They have lower delay and requires less computing power than AAC-LC. G.711 and G.726 are speech codecs that are primarily used in telephony and have low audio quality. Both have a sampling rate of 8 kHz. G.711 has a bit rate of 64 kbit/s. Axis G.726 implementation supports 24 and 32 kbit/s. With G.711, Axis products support only µ-law, which is one of two sound compression algorithms in the G.711 standard. When using G.711, it is important that the client also uses the µ-law compression. In addition, the Axis products supporting SIP can also use the following codecs: opus, L16/16000, L16/8000, speex/8000, speex/16000, G.726-32.

4.5 Synchronization of audio and video

Synchronization of audio and video data is handled by a media player, or by a multimedia framework such as Microsoft DirectX®.

In combined audio/video products, audio and video are sent over a network as two separate packet streams. For the client or player to perfectly synchronize the audio and video streams, the audio and video packets must be time-stamped. The timestamping of video packets using Motion JPEG compression may not always be supported in a network camera. If this is the case and if it is important to have synchronized video and audio, the video format to choose is MPEG-4, H.264, or H.265 since such video streams, along with the audio stream, are sent using RTP (Real-time transport protocol), which timestamps the video and audio packets.

4.6 Synchronization between multiple network audio devices

Network audio from Axis supports inter-destination media synchronization. This means that synchronized playback is possible between multiple devices over any network, since synchronization between their clocks is maintained.

5. Network audio hardware

Network speakers form the physical foundation for network audio. They can be of various types depending on the purpose of the system. Audio system devices provide possibilities to connect analog audio and network audio, and microphone consoles may be used to complete the system with live public address functionality. Axis serves as a one-stop shop for network audio, with everything that is needed for a complete audio system.

5.1 Network speakers

Axis network speakers are complete, high-quality audio systems in themselves, with integrated amplifier and digital signal processor. They are powered by Power over Ethernet (PoE) technology and connect to standard networks.

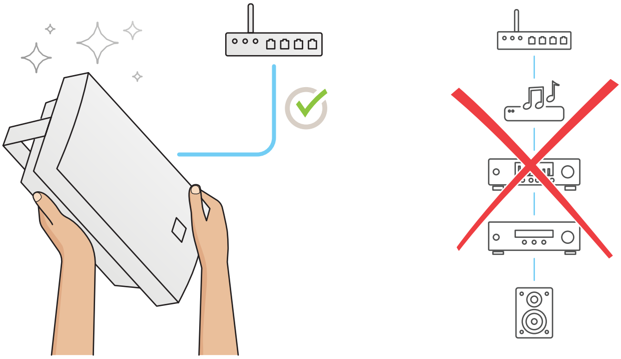


Figure 5.1a *Network audio speakers are complete audio systems.*

Every speaker has built-in audio management software which provides support for live or prerecorded announcements, background music, audio content scheduling, zoning, and priority of audio sources. They come with preconfigured sound and onboard memory for storing audio clips.

Each speaker also has an integrated microphone. This can be used for a built-in test function with test tones to verify the functionality from remote.

5.1.1 Speaker types

Form factors, sound pressures, and mounting possibilities vary – some speaker types are optimal for conveying clear and audible announcements in noisy outdoor areas, while others work better in small spaces.

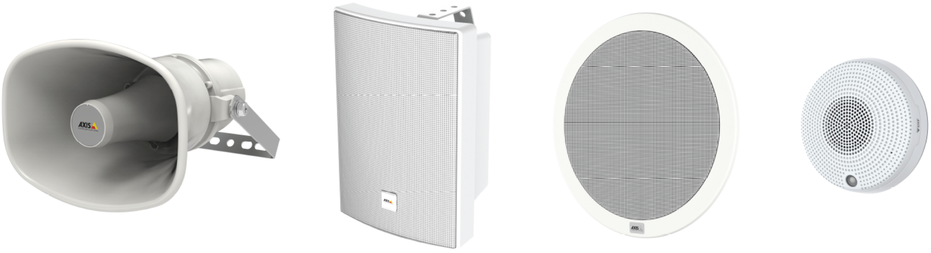


Figure 5.1b *Network speakers: horn speaker, cabinet speaker, ceiling speaker, and a mini speaker.*

Horn speaker. An Axis network horn speaker has a high sound pressure level and maximizes the loudness of those frequencies to which the human ear is the most sensitive. This means that a message can be conveyed as clearly as possible. Due to its shape, the speaker directs all sound in one direction, which further enhances the sound pressure. A horn speaker can be used in noisy indoor areas like warehouses and plants, or in outdoor installations. It can be mounted on a pole or a wall.

Cabinet speaker. An Axis network cabinet speaker provides a medium sound pressure level and should be used in less noisy areas, such as hospitals, schools, retail shops, or office buildings. It can be used both indoors and semi-outdoors, which means it can be mounted below a roof that protects it from heavy rain. It can be mounted horizontally or vertically, on a wall, in a ceiling, or with a pendant kit.

Ceiling speaker. An Axis network ceiling speaker provides a medium sound pressure level and should be used in less noisy, indoor areas, such as hospitals, schools, retail shops, or office buildings. It can be mounted in a drop ceiling where it will be very discreet and physically well-integrated.

Mini speaker. An Axis network mini speaker provides a low sound pressure level and should be used in quieter indoor areas, such as hospitals, schools, retail shops, or office buildings. It is small and discreet and fits into small spaces. It also has a wide audio coverage which means that you need fewer speakers. The mini speaker has a built in PIR sensor for motion detection, which can be set up so that the speaker automatically plays an audio message when someone is approaching.

5.2 Audio system devices

Audio system devices enable a smooth migration from analog to network audio. These devices make it possible to combine legacy equipment, such as analog speaker systems with or without amplifiers, with network audio equipment and gain the benefits of network audio without having to replace all the equipment at once. Axis offers both a network audio amplifier and a network audio bridge for this end.

5.2.1 Network audio amplifier

A network audio amplifier from Axis is a device for connecting one or multiple analog speakers. It has a built-in amplifier and digital signal processor (DSP) that delivers a total power output of 15 W. The amplifier is powered by PoE.

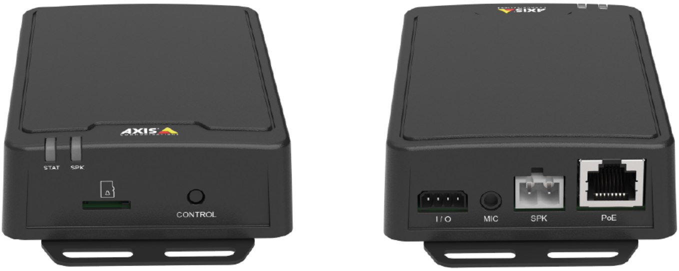


Figure 5.2a *A network audio amplifier, front and back, from Axis.*

This is a product for those who want to keep their passive speakers. This could be for design reasons – typically in environments that are designed to provide a very specific visual experience, such as a hotel lobby – or because large investments have been made in a passive speaker system.

With the audio amplifier passive speakers and network audio speakers can be mixed and matched in one installation and they will work together seamlessly. This is a good solution for existing installations that you want to migrate into network audio, as well as new installations with specific speaker requirements.

When the amplifier is connected to one or several speakers, the amplifier and the speaker together will, in all relevant aspects, act as a network speaker. Passive speakers that are connected this way can be managed through an audio management system – both network speakers and passive speakers can be controlled and managed from one location. The addition of a network audio amplifier also enables system health tests to be performed on passive speakers.

Ordinary speaker cables are used to connect the speaker to the amplifier, and the amplifier is then connected to the IP network in the building. The amplifier should be placed in the same room as the passive speakers. If connected to only one speaker, the amplifier can be mounted right at the back of that speaker. If connected to speakers in a drop ceiling, the amplifier can also be discreetly mounted above that ceiling, away from sight. Depending on the use case and the surroundings, the amplifier is recommended to be used with up to eight speakers.

5.2.2 Network audio bridge

The network audio bridge from Axis is a very versatile device that connects and combines analog and network audio systems.

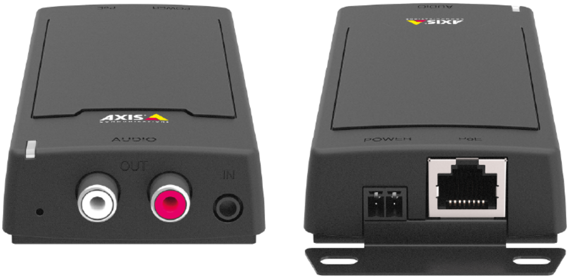


Figure 5.2b *A network audio bridge, front and back, from Axis.*

It has ports for both analog and digital connections and enables network speakers to be used in an analog audio system and analog audio sources to be used in an Axis network audio system. It will all function as one system. One single network audio bridge can be used for hundreds of speakers.

The audio bridge can be used to connect any digital audio source to an analog speaker system or vice versa. With the bridge you can also trigger announcements or make live callouts via your IP telephone system.

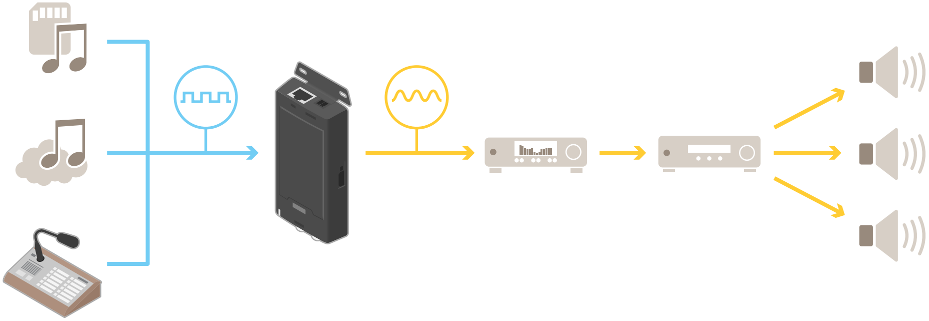


Figure 5.2c *How to use network audio bridge to connect digital audio sources to an analog audio system.*

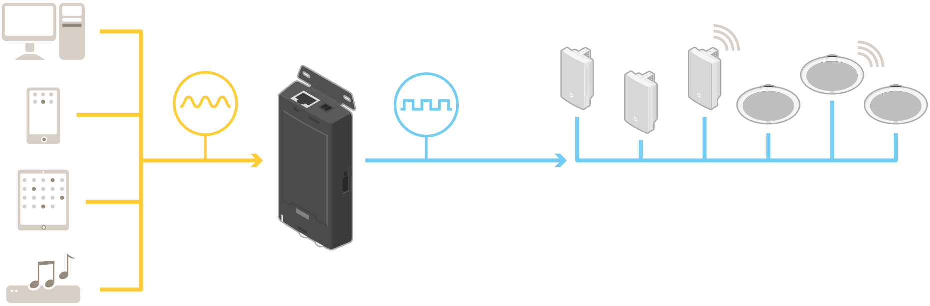


Figure 5.2d *How to use network audio bridge to combine analog audio sources with network speakers.*

Network audio bridge also enables a customer to continue to use an existing, analog audio system while expanding it with network audio for added functionality and future proofing. Network audio bridge can combine both systems and provide improved network functionality even for analog systems, while protecting the possibly very large investment of the existing analog system.

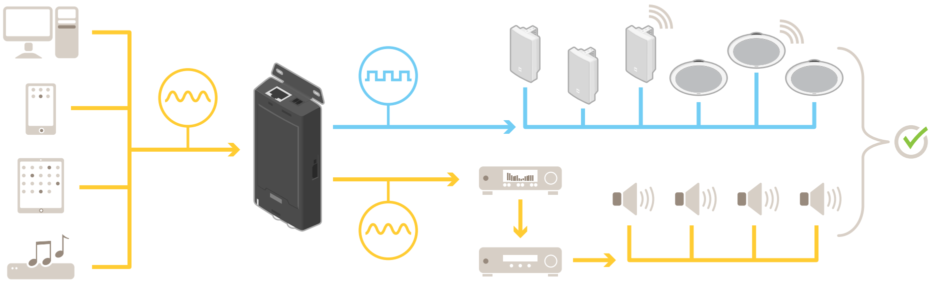


Figure 5.2e *How to use network audio bridge to combine an analog system with a network audio system.*

The audio bridge can be powered by PoE but can also use an ordinary power supply.

5.3 Network microphone consoles

In our portfolio we have a microphone console that includes a microphone, several configurable buttons, and a built-in audio management server. It can be connected to a standard network using PoE and integrated with network speakers to create a complete public address system for both live and prerecorded announcements. Using the web user interface of the microphone's built-in software, you can configure the buttons to the actions of your choice. Each button can be associated with a single zone or a combination of zones, and changed as often as you like.

6. Network audio management software

An important aspect of an audio system is managing the audio content, as well as managing the devices. With the right audio management software, it is easy to update scheduling, zoning, and audio content, and a network audio system can be efficiently managed and controlled regardless of its size and complexity.

6.1 Management features

Sophisticated network audio management software is central to the flexibility of Axis network audio. With features such as central control, easy zone and content management from remote, and centralized health monitoring, a network audio system can be fully controlled through the software. Changes can be applied without any new cabling or anyone needing to physically go onsite.

> **Central control.** Guidelines, rules, and protocols can be implemented at once throughout the system without anyone having to physically visit the sites.

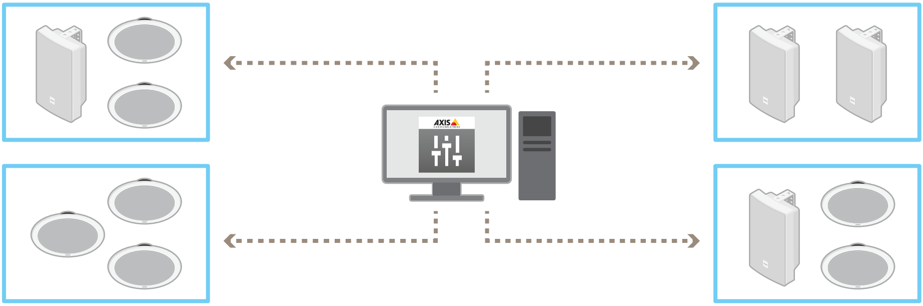


Figure 6.1a *Central control of multiple audio devices.*

> **User management.** Users can be assigned access with varying levels of privileges that fit their role in the organization.

- **Zone management.** Large spaces can be subdivided into zones, permitting announcements (prerecorded or live) and music to specific areas. Zoning the system also allows volume to be tuned for the different content types separately for each zone.
- **Content management.** Local files for announcements, advertisements, music, and configuration for streaming content can be managed for all sites and zones through one single user interface. It should be easy to set up and combine audio sources, configurations, and destinations.
- **Scheduling.** Recurring announcements can be played by use of scheduling, which makes sure that the announcements are played on specific times and days, and in specific parts of the system. An example could be that a specific voice announcement is played at noon on all weekdays in zones 1 and 2. Scheduling enables long-term planning for announcements and music, but it also provides flexibility and a possibility to tailor audio well in advance. In addition to calendar-based scheduling, content can be set up to play based on external events, such as when motion is detected by a video camera or by PIR sensors.
- **Prioritization.** Announcements or audio clips can be set to automatically overrule any background music played so that no announcements are missed. When the site receives a scheduled announcement, this will overrule the background music, which is muted during the announcement. If then a security announcement is triggered, that announcement will overrule both the scheduled announcement and the background music.

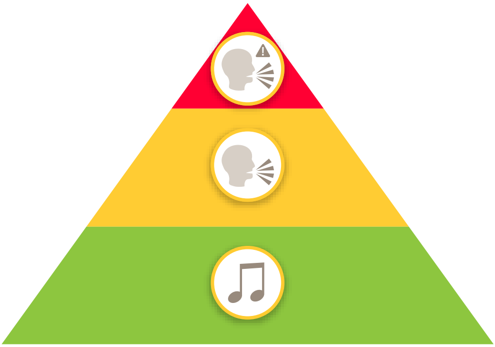


Figure 6.1b *Priority can be set so that some types of audio content will automatically overrule less important content.*

- **Health monitoring.** A network audio management system can detect malfunctioning speakers and send feedback about it in real time. Through the software interface you can see connection status of all speakers at once.

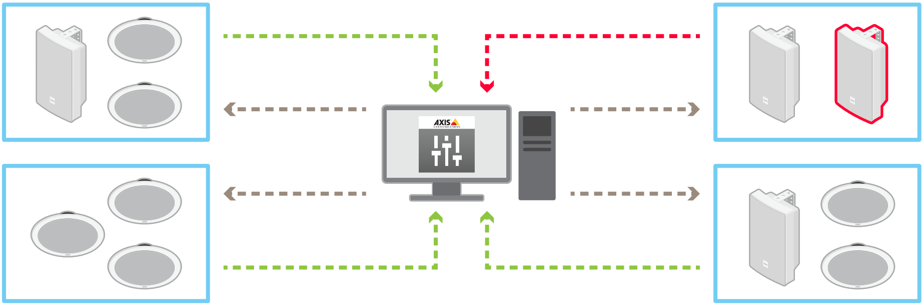


Figure 6.1c *Central system health monitoring.*

6.2 Audio management systems

Each network audio speaker from Axis comes with a built-in management software, **AXIS Audio Manager Edge.** It makes every speaker a complete, all-in-one sound system with no need for a separate software management server. This software is intended for low-complexity use cases on small to medium-sized sites, where it can be used to manage up to 200 speakers in up to 20 zones.

For larger and more advanced use cases, there is **AXIS Audio Manager Pro** which can handle a large number of zones and thousands of speakers in a single interface. It facilitates long-term scheduling and advanced priority settings.

7. Integration with other systems

When audio is integrated with other systems, such as, video surveillance, VoIP, and access control, information from those systems can be used to trigger functions in the network audio system and vice versa. Users also benefit from having one common interface for managing the different systems.

7.1 Open standards enable integration

Axis network audio is a truly open system. It uses open standards which enable easy integration with other open systems. For Axis, open standards mean that we provide full access to the application programming interfaces (APIs) and we work with protocols that are commonly used and available for free. Integration possibilities are virtually endless, and users can develop their own software for use in Axis products. Open standards allow users to make the most of their network audio installation by enabling extended use cases.

The following subsections show some examples of how audio can be integrated with other systems.

7.1.1 Video surveillance integration

There are numerous examples of situations where a security system benefits from having both video surveillance and audio capabilities. Both video analytics and audio analytics can enhance this system integration. When trespassers are caught on video, the audio system can be triggered to send out a deterring voice message, live or prerecorded, to let the intruders know that they are being watched. It could also be the other way around, so that audio analytics applications such as glass break detection or aggression detection trigger event-based recording in the network video system.

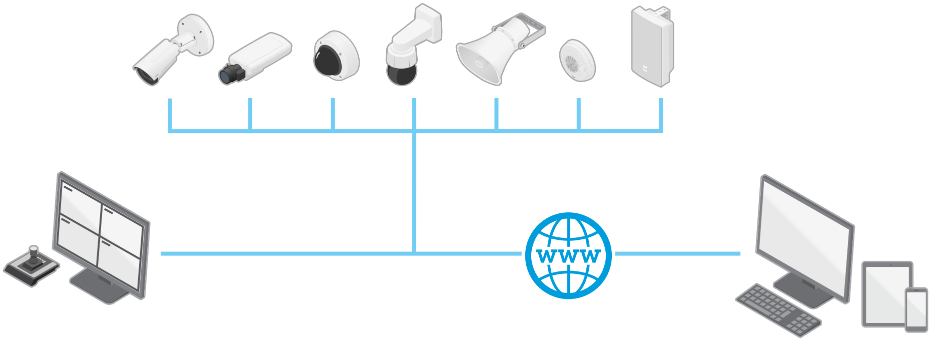


Figure 7.1a *Seamless integration of audio and video products in an IP-based security system.*

Axis network audio can be easily integrated with video management systems such as AXIS Camera Station and AXIS Companion, but also with Milestone (through a plugin for audio), Genetec, and with a range of customized software solutions developed by Axis application development partners.

7.1.2 VoIP integration

Voice over IP (VoIP) is a way to communicate over IP networks. VoIP applications can include voice and video elements and are used for video conferencing, call control, and instant messaging. It is enabled through the standard SIP protocol, which constitutes a way to connect, integrate, and control audio products over an IP network.

Using VoIP you can integrate your network audio system with a company PBX phone system to make callouts possible from a traditional desk phone.

7.1.3 Edge-to-edge integration

One type of integration is enabled by so called edge-to-edge technology. This is a way to make IP devices communicate directly with each other. Using edge-to-edge, you can extend the functionality of an Axis network camera with audio, even if it has no built-in speaker or line-out capability.

8. Network technologies

A network audio system employs many network technologies. Local area networks, in particular Ethernet, provide the foundation and also enables network audio products to be powered by Power over Ethernet (PoE). IP addressing and data transport protocols enable the products to be accessed over the Internet, and technologies that provide Quality of Service allows different network applications to co-exist without consuming each other's bandwidth.

8.1 Local area networks and Ethernet

A local area network (LAN) is a group of computers that are connected in a localized area to communicate with one another and share resources such as printers. Data is sent in the form of packets, and to regulate the transmission of the packets, different technologies can be used. The most widely used LAN technology is Ethernet, which is specified in a standard called IEEE 802.3. Other types of LAN networking technologies include token ring and FDDI (Fiber Distributed Data Interface).

Today Ethernet uses a star topology in which the individual nodes (devices) are networked with one another via active networking equipment such as switches. The number of networked devices in a LAN can range from two to several thousand.

One good rule of thumb is to always build a network with greater capacity than that currently required. To future-proof a network, it is a good idea to design a network so that it only uses 30% of the total capacity when first used. As more and more applications run over networks, more and more network performance is required. While network switches are easy to upgrade after a few years, cabling is normally much more difficult to replace.

8.1.1 Types of Ethernet networks

The following are the most common types of Ethernet networks used today. They can be based on twisted pair or fiber optic cables.

Fast Ethernet. Can transfer data at a rate of 100 Mbit/s, which is enough for most network audio applications. The older 10 Mbit/s Ethernet is still installed and used, but such networks may not provide the necessary bandwidth for some modern applications. Most devices are equipped with a 10BASE-T/100BASE-TX Ethernet interface, most commonly called a 10/100 interface, which supports both 10 Mbit/s and Fast Ethernet. The type of twisted pair cable that supports Fast Ethernet is called a Cat-5 cable.

Gigabit Ethernet. Supports a data rate of 1,000 Mbit/s (1 Gbit/s) and is now more commonly used than Fast Ethernet. 1 or 10 Gbit/s Ethernet may be necessary for the backbone network that connects many network devices. The type of twisted pair cable that supports Gigabit Ethernet is a Cat-5e cable, where all four pairs of twisted wires in the cable are used to achieve the high data rates. Most interfaces are backwards compatible with 10 and 100 Mbit/s Ethernet and are commonly called 10/100/1000 interfaces. For transmission over greater distances, fiber optic cables such as 1000BASE-SX (up to 550 m/1804 ft.) and 1000BASE-LX (up to 550 m with multimode optical fibers and 5 km/3 miles with single-mode fibers) can be used.

10 Gigabit Ethernet. Supports a data rate of 10 Gbit/s (10,000 Mbit/s). 10GBASE-LX4, 10GBASE-ER and 10GBASE-SR based on an optical fiber cable can be used to cover distances up to 10 km/ 6 miles). With a 10GBASE-T twisted pair solution, a very high quality cable (Cat-6a or Cat-7) is required. 10 Gbit/s Ethernet is mainly used for backbones in applications that require high data rates.

8.1.2 Connecting network devices and network switch

To network multiple devices in a LAN, network equipment, such as a network switch, is required. Its main function is to forward data from one device to another on the same network. The switch does this efficiently by directing data from one device directly to the target device, without affecting other devices on the same network.

A network switch works by registering the MAC (Media Access Control) address of each device that connects to it. Each and every networking device has a unique MAC address, made up of a series of figures and letters in hexadecimal notation, as set by the manufacturer. The address is often found on the product label. When a network switch receives data, it forwards it only to the port that is connected to the device with the appropriate destination MAC address.

Network switches typically indicate their performance in per port rates, and in backplane or internal rates (both in bitrates and in packets per second). The port rates indicate the maximum rates on specific ports. This means that the speed of a switch, for example 100 Mbit/s, is often the performance of each port.

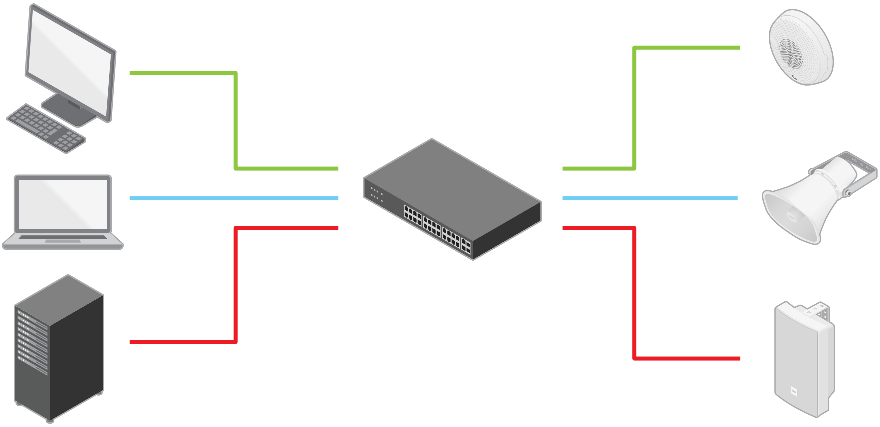


Figure 8.1a *In a network switch, data transfer is managed very efficiently as data traffic can be directed from one device to another without affecting any other ports on the switch.*

Network switches often have 10/100/1000 interfaces, thus supporting 10 Mbit/s, Fast Ethernet, and Gigabit Ethernet simultaneously. The transfer rate and mode between a port on a switch and a connected device are normally determined through auto-negotiation, whereby the highest common data rate and best transfer mode are used. A network switch also allows a connected device to function in full-duplex mode, that is, send and receive data at the same time, resulting in increased performance.

Network switches may come with different features or functions, for example, some may include router functionality. A switch may also support Power over Ethernet or Quality of Service, which controls how much bandwidth is used by different applications.

8.1.3 Power over Ethernet (PoE)

Power over Ethernet (PoE) is used to supply power to devices connected to an Ethernet network over the data-communication cable. Apart from its use in network audio, PoE is widely used to power IP phones, wireless access points, and network cameras.

The main benefit of PoE is the inherent cost savings. Hiring a certified electrician to install a separate power line is not required when running PoE. This is advantageous, particularly in difficult-to-reach areas. The fact that power cabling is not required can reduce costs and it also makes it easier to move a device to a new location.

Additionally, PoE makes it easier to make an audio system more secure. A system with PoE can be powered from the server room, which is often backed up by a UPS (uninterruptible power supply). This means that the system can stay operational even during a power outage.

Due to the benefits of PoE, it is recommended for use with as many devices as possible. The power available from the PoE-enabled switch or midspan should be sufficient for the connected devices and the devices should support power classification.

8.1.3.1 PoE standards

Most PoE devices today conform to the IEEE 802.3af standard. It uses Cat-5 or higher cables, and ensures that data transfer is not affected. In the standard, the device that supplies the power is referred to as the power sourcing equipment (PSE). This can be a PoE-enabled switch or midspan. The device that receives the power is referred to as a powered device (PD). The functionality is normally built into the network device, or it can be provided from a standalone splitter.

Backward compatibility to non-PoE compatible network devices is guaranteed. The standard includes a method for automatically identifying if a device supports PoE, and only when this is confirmed will power be supplied to the device. This also means that the Ethernet cable connected to a PoE switch will not supply any power if not connected to a PoE-enabled device. This eliminates the risk of electrical shock when installing or rewiring a network.

In a twisted pair cable, there are four pairs of twisted wires. PoE can use either the two 'spare' wire pairs, or it can overlay the current on the pairs used for data transmission. Switches with built-in PoE often supply power through the two pairs of wires used for transferring data, while midspans normally use the two spare pairs. A PD supports both options.

According to IEEE 802.3af, a PSE provides a voltage of 48 V DC with a maximum power of 15.4 W per port. Considering that there will be some power loss over a twisted pair cable, only 12.95 W is guaranteed as available for the PD. The standard specifies various performance categories for PDs.

PSE such as PoE-enabled switches and midspans normally supply a certain amount of power, typically 300-500 W. On a 48-port switch, this would mean 6-10 W per port, if all ports are connected to devices that use PoE. Unless the PDs support power classification, a full 15.4 W must be reserved for each port that uses PoE, which means a switch with 300 W can only supply power on 20 of the 48 ports. However, if all devices let the switch know that they are Class 1 devices, then 300 W will be enough to supply power to all 48 ports.

Other PoE standards are IEEE 802.3at (also known as PoE+) and IEEE 802.3bt. Using PoE+, the power limit is raised to at least 30 W via two pairs of wires from a PSE. For power requirements that are higher than the PoE+ standard, Axis uses the term High PoE, which raises the power limit to at least 60 W via four pairs of wires, and 51 W is guaranteed for the device (PD).

Class	Type	Minimum power level at PSE	Maximum power level used by PD
0	Type 1, 802.3af	15.4 W	0.44 W - 12.95 W
1	Type 1, 802.3af	4.0 W	0.44 W - 3.84 W
2	Type 1, 802.3af	7.0 W	3.84 W - 6.49 W
3	Type 1, 802.3af	15.4 W	6.49 W - 12.95 W
4	Type 2, 802.3at	30 W	12.95 W - 25.5 W
6	Type 3, 802.3bt	60 W	51 W
8	Type 4, 802.3bt	100 W	71.3 W

Table 8.1a *Power classifications according to IEEE 802.3af, IEEE 802.3at, and IEEE 802.3bt.*

Most network audio devices can receive power via PoE using the IEEE 802.3af standard and are normally identified as Class 3 devices.

8.1.3.2 Midspans

Midspans are devices that enable an existing network to support PoE. The midspan, which injects power to an Ethernet cable, is placed between the network switch and the powered devices. To ensure that data transfer is not affected, the maximum distance between the source of the data (e. g., switch) and the network audio product must not exceed 100 m (330 ft.). The midspan is not a repeater and does not amplify the Ethernet data signal.

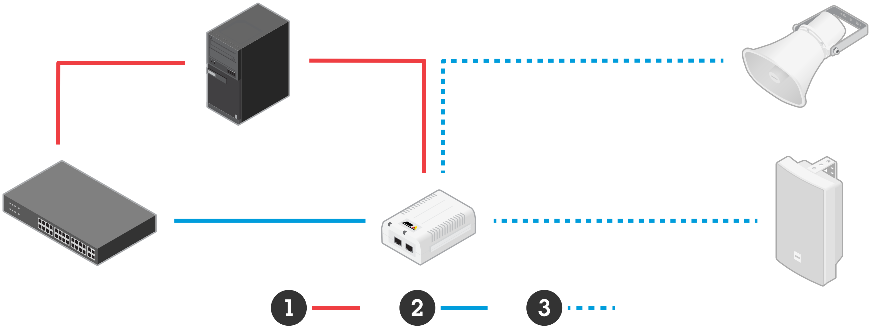


Figure 8.1b *An existing system with separate power (1) and Ethernet (2) cables can be upgraded with PoE functionality (3) using a midspan.*

8.2 Sending data over the Internet

The basic elements of Internet communication include:

Routers. To forward data packages from one LAN to another via the Internet, a network router must be used. This device routes information from one network to another (hence the name) based on IP addresses. A router only forwards data packages destined for another network, and is most commonly used for connecting a local network to the Internet. Routers are sometimes referred to as gateways.

Firewalls. A firewall is designed to prevent unauthorized access to or from a private network. Firewalls can be implemented in both hardware and software, or in a combination of both. Firewalls are frequently used to prevent unauthorized Internet users from accessing private networks connected to the Internet. Messages entering or leaving the Internet pass through the firewall, which examines each message, and blocks those that do not meet the specified security criteria.

Sending data from a device on one LAN to a device on another LAN requires a standard method of communication, since local area networks may use different technologies. This requirement led to the development of IP addressing and the many IP-based protocols for communicating over the Internet.

8.2.1 IP addressing

Devices wishing to communicate via the Internet must have unique and appropriate IP addresses, which identify the sending and receiving devices. There are currently two IP versions: IP version 4

(IPv4) and IP version 6 (IPv6). The main difference between the two is that an IPv6 address is longer (128 bits compared with 32 bits for an IPv4 address). IPv4 addresses are the most commonly used today.

8.2.2 IPv4 addresses

IPv4 addresses are grouped into four blocks, with each block separated by a dot. Each block represents a number between 0 and 255; for example, 192.168.12.23.

Certain blocks of IPv4 addresses have been reserved exclusively for private use. These private IP addresses are 10.0.0.0 to 10.255.255.255, 172.16.0.0 to 172.31.255.255 and 192.168.0.0 to 192.168.255.255. These addresses can only be used on private networks and are not allowed to be forwarded through a router to the Internet. A device wanting to communicate over the Internet must have its own individual, public IP address, which will be allocated by an Internet service provider (ISP). An ISP can allocate either a dynamic IP address, which can change during a session, or a static address, which normally comes at an additional monthly fee.

Ports

A port number defines a particular service or application so that the receiving device (for example, a network speaker) will know how to process the incoming data. When a computer sends data tied to a specific application, it usually automatically adds the port number to an IP address.

Port numbers can range from 0 to 65535. Certain applications use port numbers that are pre-assigned to them by the Internet Assigned Numbers Authority (IANA). For example, a web service via HTTP is typically mapped to port 80 on a network audio device.

Setting IPv4 addresses

For a network audio device to work in an IP network, an IP address must be assigned to it. Setting an IPv4 address for an Axis network device can be done automatically using DHCP (Dynamic Host Configuration Protocol) which requires a DHCP server on the network. Alternatively, the address can be set manually. One way to do this is to use the product's web page to enter the static IP address, subnet mask, and the IP addresses of the default router, the DNS (Domain Name System) server and the NTP (Network Time Protocol) server. Another way to set the IP address is to use a management software tool such as AXIS Device Manager.

A DHCP server manages a pool of IP addresses, which it can assign dynamically to network devices, with this function often being performed by a broadband router. The router in turn is typically connected to the Internet and gets its public IP address from an Internet service provider. Using a dynamic IP address means that the IP address for a network device may change from day to day. With dynamic IP addresses, it is recommended that users register a domain name for the network audio product at a dynamic DNS server, which can always tie the domain name for the product to any IP address that is currently assigned to it. (A domain name can be registered using some of the popular dynamic DNS sites such as *www.dyndns.org*.)

Using DHCP to set an IPv4 address works as follows. When a network device comes online, it sends a query requesting configuration from a DHCP server. The DHCP server replies with the configuration requested by the network device. This normally includes the IP address, the subnet mask, and IP addresses for the router, DNS server and NTP server. The product first verifies that the offered IP address is not already in use on the local network, assigns the address to itself and can then update a dynamic DNS server with its current IP address so that users can access the device using a domain name.

With AXIS Device Manager, the software can automatically find and set IP addresses and show the connection status. The software can also be used to assign static and private IP addresses for Axis network devices. This is recommended when using audio management software to access network audio products. In a network audio system with potentially hundreds of devices, a software program, such as AXIS Device Manager, is necessary to effectively manage the system.

NAT (Network address translation)

When a network device with a private IP address wants to send information via the Internet, it must do so using a router that supports NAT. Using this technique, the router translates the private IP address into a public IP address, for public exposure on the Internet.

IP multicast

IP multicast allows a host to send audio streams to many other hosts, preferably in the same broadcast domain, without any unnecessary data-packet duplication. The sender can transmit to members in a group, defined by a Class D address within the range of 224.0.0.0 – 239.255.255.255. By allowing multiple users to access the same data stream, multicast conserves bandwidth compared to unicast, which requires one data stream per connected client.

Port forwarding

To access devices that are located on a private LAN via the Internet, the public IP address of the router should be used together with the corresponding port number for the network device on the private network.

Since a web service via HTTP is typically mapped to port 80, what happens when there are several network devices using port 80 for HTTP in a private network? Instead of changing the default HTTP port number for each network device, a router can be configured to associate a unique HTTP port number to a particular device's IP address and default HTTP port. This is called port forwarding.

Port forwarding works as follows. Incoming data packets reach the router via the router's public (external) IP address and a specific port number. The router is configured to forward any data arriving at a predefined port to a specific device on the private side of the router, by replacing the router's public address with the private address of the device. The reverse happens with outgoing data packets. The router replaces the private IP address of the device with the router's public IP address before the data is sent over the Internet. For the external client, it looks like it is communicating with the router when in fact the sent packets originate from the device on the private network.

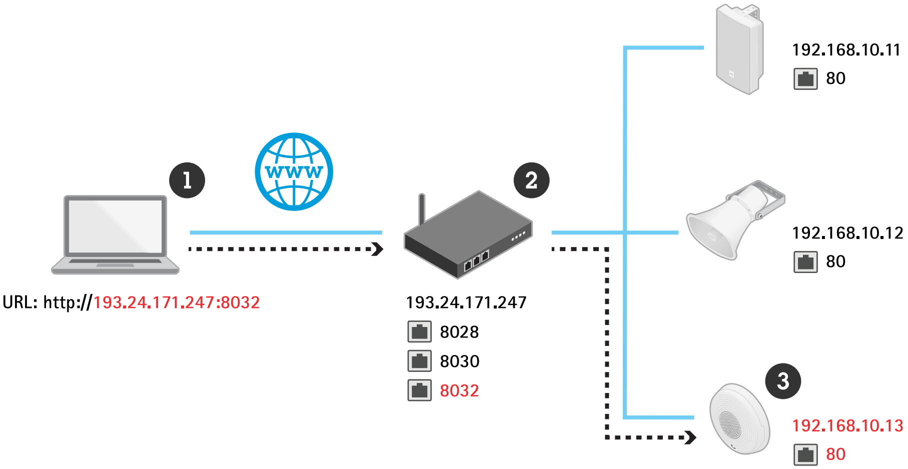


Figure 8.2a *Port mapping in the router. In this illustration, data (requests) from an external client (1) is forwarded by the router (2) with external IP address 193.24.171.247 on port 8032 to a network speaker (3) with a private IP address of 192.168.10.13 on port 80.*

Port forwarding is done by first configuring the router. Different routers have different ways of doing this, but the process usually involves bringing up the router's interface in a browser and entering the router's public (external) IP address and a unique port number that is then mapped to the internal (private) IP address of the specific network device and its port number. Specialized websites such as *www.portforward.com* offer step-by-step instructions for different brands.

To make port forwarding easier, Axis offers the NAT traversal feature in its network audio products. When enabled, NAT traversal will attempt to configure port mapping in a NAT router on the network using UPnP (Universal Plug and Play). On the network audio product's web page, users can manually enter the IP address of the NAT router. If a router is not manually specified, the network audio product will automatically search for NAT routers on the network and select the default router. In addition, NAT traversal will automatically select an HTTP port if none is manually entered.

8.2.3 IPv6 addresses

An IPv6 address is written in hexadecimal notation with colons subdividing the address into eight blocks of 16 bits each; for example, 2001:0da8:65b4:05d3:1315:7c1f:0461:7847.

The major advantages of IPv6, apart from the huge number of IP addresses it provides, include enabling a device to automatically configure its IP address using its MAC address. For communication over the Internet, the host requests and receives, from the router, the necessary prefix of the public address block, as well as any additional information. The prefix and host's suffix are then used, so DHCP for IP address allocation and manual setting of IP addresses is no longer required with IPv6. Port forwarding is also no longer needed. Other benefits of IPv6 include renumbering to simplify switching entire corporate networks between providers, faster routing, point-to-point encryption according to IPSec, and connectivity using the same address in changing networks (Mobile IPv6).

An IPv6 address is enclosed in square brackets in a URL and a specific port can be addressed in the following way: http://[2001:0da8:65b4:05d3:1315:7c1f:0461:7847]:8081/

Setting an IPv6 address for an Axis network device is as simple as checking a box to enable IPv6 in the device. The device then receives an IPv6 address according to the configuration in the network router.

8.2.4 Data transport protocols

The Transmission Control Protocol (TCP) and the User Datagram Protocol (UDP) are the IP-based protocols used for sending data. These transport protocols act as carriers for many other protocols, for example, HTTP (Hyper Text Transfer Protocol), as used to browse web pages, is carried by TCP.

TCP provides a reliable, connection-based transmission channel. It ensures that data sent from one point is received at the other. TCP's reliability through retransmission may introduce significant delays, but in general TCP is used when reliable communication is preferred over reduced latency.

UDP is a connection-less protocol and does not guarantee delivery of the transmitted data, thus leaving the whole control mechanism and error-checking to the application itself. UDP does not re-transmit lost data and, therefore, does not introduce any further delay.

Protocol	Transport protocol	Port	Common usage	Network audio usage
FTP (File Transfer Protocol)	TCP	21	Transfer of files	Transfer of audio to an FTP server or an application
SMTP (Send Mail Transfer Protocol)	TCP	25	Protocol for sending email	The audio device sends images or alarm notifications via built-in email client.
HTTP (Hyper Text Transfer Protocol)	TCP	80	Browsing the web: getting web pages from web servers	The audio device functions as a web server, making audio available for the user or application server.
HTTPS (HTTP over Transport Layer Security)	TCP	443	Secure access to web pages using encryption	Secure transmission of audio to/from network audio devices.
RTP (Real Time Protocol)	UDP/TCP	Not defined	RTP standardized packet format for delivering audio and video over the Internet, often used in streaming media systems or video conferencing	Transmission of audio, and synchronization of audio and video. RTP provides sequential numbering and timestamping of data packets, enabling correct reassembly. Unicast or multicast.
RTSP (Real Time Streaming Protocol)	TCP	554	Set up and control of multimedia sessions over RTP	

Table 8.2a *Common TCP/IP protocols and ports used for network audio.*

8.2.5 SIP

Session Initiation Protocol (SIP) is a text-based protocol, similar to HTTP and SMTP, for communication over IP networks. It is used to start, change, and end media stream sessions, which can include voice and video elements. SIP is the standard protocol used in Voice over IP (VoIP) applications and unified communication platforms, for video conferencing, call control, and instant messaging. SIP constitutes a way to connect, integrate, and control Axis network products.

SIP calls can be set up in many ways, but there are three main types:

> Peer-to-per calls (also called local calls)

> SIP server calls (also called private branch exchange [PBX] calls)

> SIP trunk calls

Peer-to-peer calls are calls between two devices (such as computers, network speakers, softphones, door stations, cameras, or IP desk phones) that belong to the same network. The call is made to the SIP address of the device.

To make SIP server calls, the devices must be connected to a SIP server that handles the call exchanges. A SIP server, or a PBX, is a hub that works like a traditional switchboard. It can be hosted on an intranet or by a third-party service provider. The SIP-enabled devices register with the SIP server and can contact each other through their SIP addresses. A PBX can show call status, allow call transfers, handle voicemail, and redirect calls among other things.

SIP addresses (also known as SIP uniform resource identifiers [URIs] or SIP numbers) are used to identify users within a network, just like phone numbers or email addresses. Like email addresses, SIP addresses are a type of URI that includes two user-specific parts, a user ID or extension, and a domain name or IP address. Together with a prefix and the @ symbol, they make up a unique address. In the case of a peer-to-peer call, the SIP address would include the IP address rather than the domain name.

With a service provider that offers SIP trunking, the traditional telephone network can be used to make calls and traditional phone numbers can be assigned to the SIP devices. This way calls can be made from a network speaker or a network door station to a cell phone or the other way around.

To make a SIP call a sequence of steps is performed to exchange information between the user agents initiating and receiving the call.

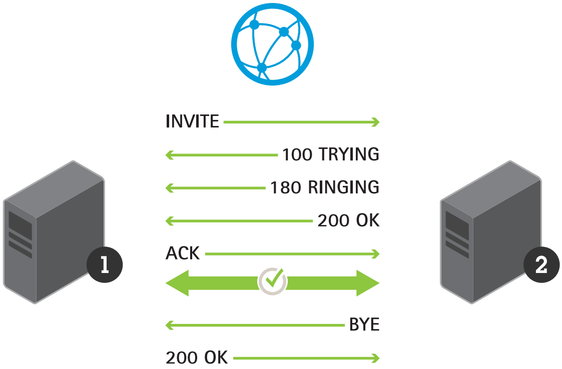


Figure 8.2b *Information is exchanged between the initiator user agent (1) and the recipient user agent (2) in a SIP call sequence.*

When initiating a call, the initiator sends a request or an INVITE to the recipient's SIP address. The INVITE contains a Session Description Protocol (SDP) body describing the media formats available and contact information for the initiator of the call.

Upon receiving the INVITE, the recipient immediately acknowledges this by answering with a 100 TRYING response.

The receiving UA then compares the offered media formats described in the SDP with its own. If a common format can be decided on, the UA alerts the recipient that there is an incoming call and sends a provisional response back to the initiating UA - 180 RINGING.

When the recipient decides to pick up the call, a 200 OK response is sent to the initiator to confirm that a connection has been established. This response contains a negotiated SDP indicating to the initiator which media formats should be used and to where the media streams should be sent.

The negotiated media streams are now set up using the Real-time Transport Protocol (RTP) with parameters based on the negotiated SDP and the media travels directly between the two parties. The initiator sends an acknowledgement (ACK) via SIP to acknowledge that it has set up the media streams as agreed. The SIP session is still active but it is no longer involved in the media transfer.

When one of the parties decides to end the call, it sends a new request, BYE. Upon receiving a BYE, the receiving party acknowledges this with a 200 OK and the RTP media streams are then stopped.

In a more complex SIP infrastructure setup, the initiation looks a little different, as the SIP session is set up step-by-step for each hop. However, once the SIP session is set up, traffic is not normally routed, but instead travels directly between the different parties.

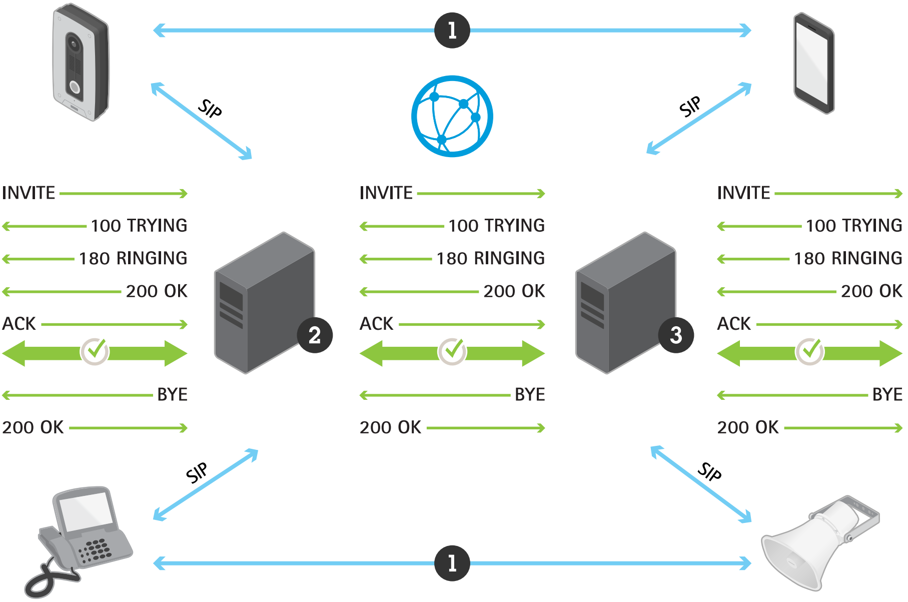


Figure 8.2c *Call setup in a complex SIP infrastructure. Media RTP streams (1) travel directly between the parties once the SIP session has been set up between initiator (2) and recipient (3).*

In a more complex network environment, it may be necessary to utilize Network Address Translation (NAT) traversal techniques. NAT is a way of translating IP addresses on a private local network to a public representation. This means that all units in a private sub-network share a common IP address prefix, for example, 192.168.1.XXX. This is the address they use when communicating with each other. When they communicate with another network, this address is translated to the router's public address, appended with a port mapping.

8.3 VLANs

When a network audio system is designed, there is often a desire to keep the network separate from other networks, both for security as well as performance reasons. At first glance, the obvious choice would be to build a separate network. While the design would be simplified, the cost of purchasing, installing and maintaining the network would often be higher than using a technology called virtual local area network (VLAN).

VLAN is a technology for virtually segmenting networks, a functionality that is supported by most network switches. This can be achieved by dividing network users into logical groups. Only users in a specific group can exchange data or access certain resources on the network. If a network audio system is segmented into a VLAN, only the servers located on that VLAN can access the network audio devices. VLANs provide a flexible and more cost-efficient solution than a separate network, and can be used to, for example, separate an audio network from an office network.

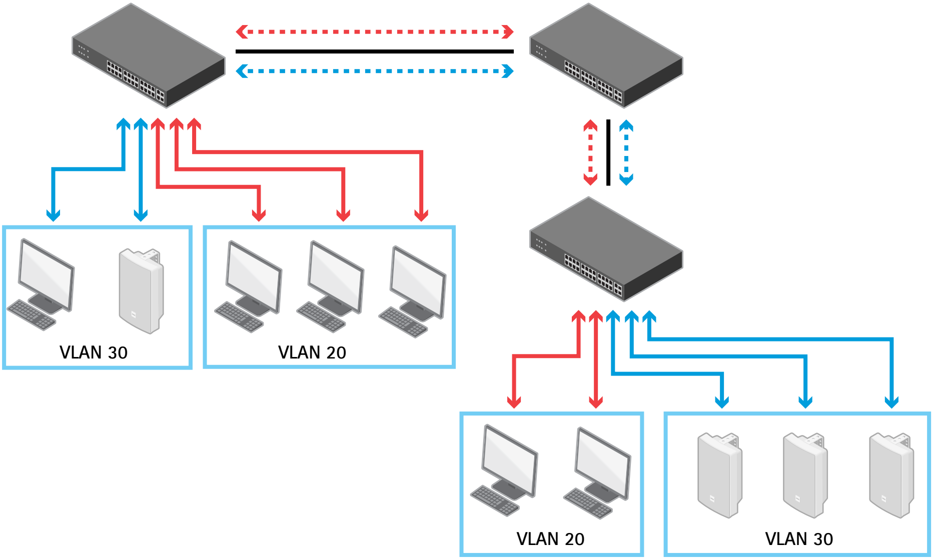


Figure 8.3a *In this illustration, VLANs are set up over several switches. The two different LANs are segmented into VLAN 20 and VLAN 30, and only members of the same VLAN can exchange data, either within the same network or over different networks.*

VLANs can communicate with each other through inter-VLAN routing, supported by layer 3 switches. VLANs logically segment the switch into different subnets, and when a router is connected to the switch, the router can forward the traffic between the VLANs configured on the switch.

The primary protocol used when configuring VLANs is IEEE 802.1Q, which tags each frame or packet with extra bytes to indicate which virtual network the packet belongs to.

8.4 Quality of Service

Since different applications—for example, telephone, email and audio—may be using the same IP network, there is a need to control how network resources are shared to fulfill the requirements of each service. One solution is to let network routers and switches operate differently for different kinds of services (voice, data, and audio) as traffic passes through the network. By using Quality of Service (QoS), different network applications can co-exist on the same network without consuming each other's bandwidth.

The term Quality of Service refers to a number of technologies, such as Differentiated Service Codepoint (DSCP), which can identify the type of data in a data packet and so divide the packets into traffic classes that can be prioritized for forwarding. One main benefit of a QoS-aware network include the ability to prioritize traffic to allow critical flows to be served before flows with lesser priority. Another is greater reliability in the network, by controlling the amount of bandwidth an application may use and thus controlling bandwidth competition between applications. A prerequisite for the use of QoS in an audio network is that all switches, routers and network audio products must support QoS.

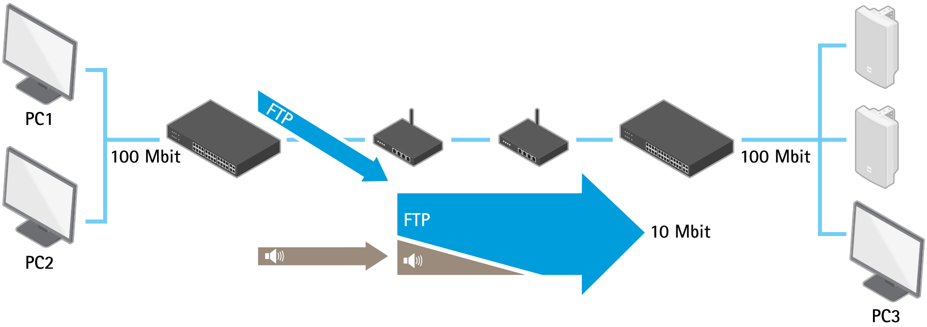


Figure 8.4a *Standard (non-QoS aware) network. In this example, FTP and audio streaming have 10 Mbit/s to share and the audio bandwidth cannot be guaranteed.*

In the illustration, PC1 is sending two audio streams to the two speakers, with each audio clip streaming at 2.5 Mbit/s. Suddenly, PC3 starts a file transfer from PC2. In this scenario, the File Transfer Protocol (FTP) will try to use the full 10 Mbit/s capacity between the routers, while the audio streams will try to maintain their total of 5 Mbit/s. The amount of bandwidth given to the audio system can no longer be guaranteed and the audio quality will probably be reduced. At worst, the FTP traffic will consume all the available bandwidth.

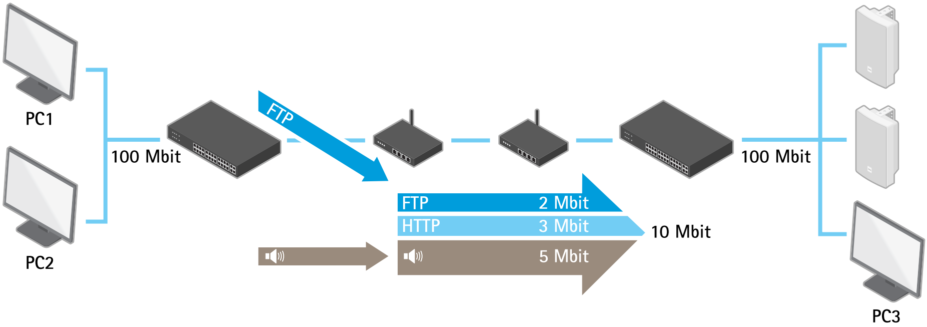


Figure 8.4b *QoS aware network. In this example, audio streaming has a guaranteed bandwidth of up to 5 Mbit/s.*

In this illustration, the first router has been configured to use up to 5 Mbit/s of the available 10 Mbit/s for streaming audio. FTP traffic can use 2 Mbit/s, and HTTP and all other traffic can use a maximum of 3 Mbit/s. Using this division, audio streams will always have the necessary bandwidth available. File transfers are considered less important and get less bandwidth, but there will still be bandwidth available for web browsing and other traffic. Note that these maximums only apply when there is congestion on the network. If there is unused bandwidth available, this can be used by any type of traffic.

9. System protection

Cyber threats are commonly associated with hackers and malware. But in reality, negative impact is often a result of unintentional misuse. To be protected against active attacks, a system needs to be both well configured and well maintained.

A recommended approach is to work according to a well-defined IT protection standard, such as ISO 27001 or NIST throughout the deployment of a system. While this may be overwhelming for smaller organizations, having even minimal policy and process documentation is far better than nothing. This could include simple things like defining roles and responsibilities, defining protection levels for the system and its components, defining system maintenance intervals, and keeping staff informed about common do's and don'ts based on best practices and common sense.

9.1 Network protection

A network needs protection form several types of malicious threats. All network packages sent on the network may be collected by other computers on the same network. If the payload in the packages is sent in clear text the data can be easily compromised, through what is called network sniffing. Another threat is network spoofing, which refers to when an attacking computer tries to impersonate a legitimate server, computer, or network device in order to get access to the network. Encrypted connections and CA-signed certificates provide protection.

9.1.1 IEEE 802.1X

Many Axis network products support IEEE 802.1X, which is a method used to protect a network against connections from unauthorized devices. IEEE 802.1X establishes a point-to-point connection or prevents access from the LAN port if authentication fails. IEEE 802.1X prevents what is called "port hijacking"; that is, when an unauthorized device gets access to a network through physical access to a network port/socket. IEEE 802.1X is useful in network audio applications, since network speakers are often located in public spaces where an openly accessible network socket can pose a security risk. In today's enterprise networks, IEEE 802.1X is becoming a basic requirement for anything that is connected to a network.

In a network audio system, IEEE 802.1X can work as follows: 1) A network speaker that is configured for IEEE 802.1X sends a request for network access to a switch; 2) the switch forwards the query to an authentication server; for instance, a RADIUS (remote authentication dial-in user service) server such as a Microsoft Internet Authentication Service server; 3) if authentication is successful, the server instructs the switch to open the port, to allow data from the network speaker to pass through the switch and onto the network.

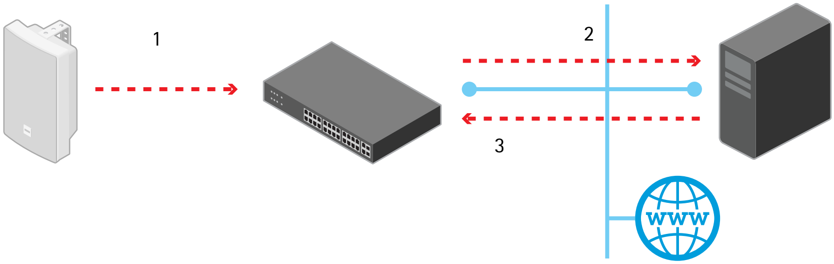


Figure 9.1a *IEEE 802.1X enables port-based security.*

9.1.2 HTTPS (HTTP over TLS)

HTTPS (Hyper Text Transfer Protocol Secure) is a secure communication method that sends HTTP inside a Transport Layer Security (TLS) connection. This means that the HTTP connection and the data itself are encrypted.

To enable an Axis network speaker to communicate over HTTPS, a digital certificate and an asymmetric key pair must be installed in the product. The key pair is generated by the Axis product. The certificate can either be generated and self-signed by the Axis product, or it can issued by a certificate authority. In HTTPS, the certificate is used for authentication and encryption, which means that the certificate allows a browser to verify the identity of the product, and it encrypts the communication using keys that are generated by public-key cryptography.

9.1.3 VPN (Virtual Private Network)

A VPN is a secure "tunnel" created between two communicating devices, enabling secure communication over the Internet. In such a setup, the original packet, including the data and its header, which may contain information such as the source and destination addresses, the type of information being sent, the packet number in the sequence of packets and the packet length, is encrypted. The encrypted packet is then encapsulated in another packet that shows only the IP addresses of the two communicating devices (i.e., routers). This set up protects the traffic and its contents from unauthorized access, and only devices with the correct "key" will be able to work inside the VPN. Network devices between the client and the server will not be able to access or view the data.

In SSL/TLS only the actual data of a packet is encrypted, whereas in VPN, the entire packet is encrypted and encapsulated to create a secure tunnel. Both technologies can be used in parallel, but this is not recommended, since each technology will add an overhead and reduce system performance.

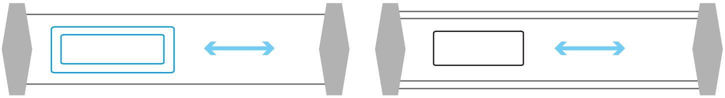


Figure 9.1b *The difference between SSL/TLS encryption (left) and a VPN tunnel (right). Double lines indicate security. With SSL/TLS, only the actual data is encrypted, while with VPN, the entire packet is encrypted.*

9.1.4 SRTP (Secure RTP)

SRTP (Secure realtime transport protocol or Secure RTP) is an extension to RTP (Realtime transport protocol) that incorporates enhanced security features. Like RTP, it is intended particularly for VoIP (Voice over IP) communications. SRTP uses encryption and authentication to minimize the risk of denial-of-service (DoS) attacks. SRTP can achieve high throughput in diverse communications environments that include both hard-wired and wireless devices. Provisions are included that allow for future improvements and extensions.

9.1.5 CA-signed certificates

A Certificate Authority (CA) is a service that issues (signs) server certificates to be installed in devices. The audio management system uses the certificate to validate the identity of the device. A publicly trusted CA, such as Comodo and Symantec, is typically used for public services, such as public web and email servers. Public trusted CA root certificates are preinstalled on clients, such as Windows, Linux, Mac, and mobile devices. A private CA is a trust point for private network services and issues certificates for private servers. The private CA root certificate must be installed in clients that need to validate the signed certificates in devices. To meet demands for end-to-end encryption, the audio management system also needs to have a server certificate so that audio clients can validate that they are accessing a legitimate management system. AXIS Device Manager has a built-in CA service that can cost-efficiently issue and deploy server certificates to the speakers.

9.1.6 Network isolation

Network isolation, or network segmentation, is a way to separate critical network resources from each other in order to reduce the risk of each of them having a negative impact on each other. This is an especially relevant tactic if different resources don't need to interact with each other – or should not. Network segmentation can be virtual (VLAN) and require an infrastructure of managed switches, or the networks can be separated with different cabling and network gear.

9.2 Device protection

Each speaker and its resources must be duly protected. Protection may in this case refer to both physical protection, such as placing speakers out of reach, and protection of the speaker's software and firmware. Axis hardening guide provides more information about device protection.

9.2.1 User account management

Device passwords tend to spread within the organization. For example, during device maintenance someone requests the password in order to adjust something. A couple of days or weeks later, someone new has the same request. Within short, many new (or temporary) users know the password to all your devices, and you have lost control over who can access them. The strength of the password makes no difference in this scenario. Managing devices should have multiple accounts (role-based) and temporary accounts should be created for occasional maintenance/troubleshooting.

For account accesses, use a principle of least privileged accounts. This means that user access privileges are limited to only the resources needed in order to perform their specific work tasks. AXIS Device Manager helps you easily and efficiently manage multiple accounts and passwords for Axis devices belonging to different privilege levels.

9.2.2 IP address filtering

Axis network audio products provide IP address filtering, which allows or denies access to defined IP addresses. A typical configuration is to configure the network speakers to allow only the IP address of the server hosting the audio management software to access the product.

IP filtering acts like a local firewall in the speaker. The only computer or server that should be accessing speakers during normal operations is the audio management server. The speakers can be configured with an IP filter to only respond to whitelisted IP addresses, typically the audio management server and administration clients. IP filtering helps mitigate risks if the speaker password is compromised, from unpatched speakers and for brute-force attacks.

9.2.3 Keeping software and firmware up to date

In most cases, using the latest firmware versions will ensure you utilize security patches for all newly discovered vulnerabilities. Leaving the system unpatched for a longer period of time will increase the risk of an adversary exploiting the vulnerabilities and possibly compromising the system, application, or devices. Running devices with up-to-date firmware versions mitigate common risks for devices, as the latest firmware versions will include patches for known vulnerabilities that attackers may try to exploit. To protect from the specific threats of firmware tampering and supply-chain tampering, the signed firmware and secure boot features are widely available in Axis devices.

9.3 Physical protection

While a device can never be 100 % physically protected, there are various physical aspects to consider. By using a speaker design that blends into the environment, placing the speaker high up on a wall or in a ceiling, and always running the speaker cable directly through the wall or ceiling behind the speaker, you may come a long way regarding protection from both spontaneous vandalism and more planned attacks. Important network gear (routers, switches, etc.) should be placed in locked areas.

10.Designing a network audio system

One of the main benefits of a network audio system is flexibility and scalability: the freedom to mix and match the most appropriate components and the power to optimize or expand the system to any size. Still, there are many considerations to make when you plan a network audio system.

10.1 Know the purpose of your system

Before you start designing your system, you need to be clear about its purpose. Only then can you dimension the network properly and enable the right integration possibilities. Do you need audio for security, integrated with a surveillance system? Do you need audio for safety, or audio for improving operational efficiency?

When evaluating an audio system, also look at other systems, such as video, access control, intercom, and intrusion detection, to determine if there is a way to construct an integrated system that covers all physical security needs. Axis network speakers can integrate with several video management software applications.

10.2 Plan the network

For a solid foundation, a cabled network infrastructure is recommended in network audio. If you need to use WLAN (wireless local area networks), make sure your infrastructure is equipped to handle RToWLAN (real-time traffic over WLAN).

Most networks today carry a range of network traffic types, and it is common that more and more IP-based devices are added to the network over time. Without appropriate action, there are risks of delay, jitter, and packet loss.

The bandwidth requirements for audio applications depend on the transmitted audio quality – a higher quality generally requires more bandwidth. For example, an uncompressed stereo audio signal of 48 kHz sample rate and 16 bits per sample needs approximately 1.5 Mbit/s in bandwidth.

Different codecs can be used to compress the transmitted signal and reduce the bandwidth usage. This will lower the audio quality to a certain degree depending on the codec and how hard the codec compresses the signal. The bandwidth consumption will be determined by the bandwidth needed for the audio signal multiplied by the number of streams you need. With unicast, every receiver must be addressed with a single stream. To avoid network capacity issues the support of multicast is recommended, since this reduces the bandwidth requirements significantly.

Another measure is to segment the network into virtual LANs (VLANs). The sending and receiving network audio devices (such as the audio management system and the speakers) need to be on the same VLAN. VLAN is a service separation sharing the same IP subnet, so it is independent of where the devices are physically located.

To ensure the right traffic prioritization, Quality of Service (QoS) should be implemented, both for resource reservation (integrated services) in terms of network resources allotted according to an application QoS request, and for prioritization (differentiated services) in terms of classified and allotted network resources according to bandwidth management policy criteria.

10.3 Plan and design the site

Considerations for this stage are outlined in the following subsections. Use AXIS Site Designer to help you. On axis.com you can also find a quickguide for speaker coverage calculations which can assist you in determining how many speakers you need. Furthermore, Axis provides coverage shapes for Microsoft® Visio® and GLL files for EASE Evac design tool, which can help you plan your audio solution.

10.3.1 Make a site survey

To successfully plan an audio installation, you must first be clear about the purpose of the system. If you need to make announcements in a classroom, one speaker may be enough. If instead you want to combine making announcements with playing background music in a retail store, you need several speakers for a good listening experience, even if both rooms have the same size.

When you have defined your purpose, you need to make a walkthrough of the site to make some measurements. If available, you should also use CAD drawings of the premises in a system design software.

Measure the room dimensions. Use a laser distance meter. The ceiling height and the possible mounting heights are the most important measures. Higher mounting means better coverage.

Note the reflectance. Rooms with reflective surfaces may have reverb, while larger spaces may encounter delay. This may have impact on how many speakers you need.

Measure the background noise. Use a professional dB meter. Important voice messages that everyone need to hear should be up to 12 dB above the ambient noise. If the background noise level is high, you must make sure to choose a speaker with a high enough SPL.

Check the mounting possibilities. What mounting spots are available, regarding physical room characteristics as well as connectivity? Are there any limitations to how and where the speakers can be installed? Remember that it is the mounting height of the speaker, not the ceiling height, that will determine the spread of sound.

10.3.2 Find and compare speakers

The following subsections provide some guidelines to help you determine what type of speaker you need and which functionalities it should have. You can then use Product selector on axis.com to find and compare Axis products that are suitable for your installation. AXIS Site Designer is also a helpful online tool in this stage. It lets you plan and design an audio installation, by providing guidance on which speakers to use, how many speakers are needed, their optimal placement, and so on, with regard to the conditions at the site.

10.3.2.1 Speaker types

To determine which types of speakers are suitable and how many speakers are needed, the site environment and the purpose of the speaker installation must first be considered. Considerations include the following:

- **Indoor or outdoor.** If you are going to use speakers outdoors, it is important that you choose an outdoor-ready speaker. The speaker must also be approved for use in the temperature range that the particular site exhibits.
- **Area of coverage.** How far and wide the sound spreads depends on the type and model of the speaker. Horn speakers typically have a higher sound pressure and may have a long but rather narrow area of coverage, while ceiling and cabinet speakers may be designed to have lower sound pressure with wider coverage. Check the speaker specification for its SPL value and coverage information.
- **Overt or highly discreet installation.** Speakers come with very different form factors, and the shape is mainly related to the functionality. But more often than not, it is possible to find a speaker that matches your operational needs while also complying with your requirement of either a non-discreet or a discreet installation.

10.3.2.2 Speaker functionalities

Network speakers from Axis have additional functionalities apart from providing an audio stream. Some examples are:

> **Integrated microphone.** Makes it possible to run self-tests of the speaker functionality and use audio analytics that listen in on events.

> **Audio analytics for detection.** When integrated microphones detect sounds above a certain level, they can trigger the playing of a voice message or trigger a camera to start recording.

> **Input/output (I/O) connectors.** Connecting external input devices to a speaker (such as a door contact, infrared motion detector, radar device, glass-break sensor, or shock sensor) enables the speaker to react to an external event by, for example, sounding an audio message. Outputs enable the speaker or a remote operator to control external devices, for example, alarm devices, door locks, or lights.

10.3.3 Decide how even sound you need

Axis recommends different solutions depending on the customer requirements. Depending on how even the audio level needs to be you can choose from a basic or a premium setup, or anything in between.

The basic setup will give you the right number of speakers needed to cover a certain area. Going below this number of speakers is never recommended, since the evenness of audio levels would vary too much within the area.

The premium setup will give you twice the number of speakers compared to a basic solution. A more even audio level will be maintained throughout the area.

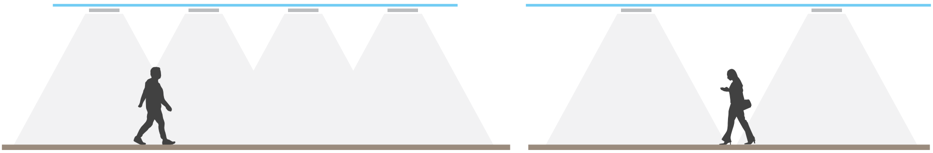


Figure 10.3a *Left: in a premium solution, sound is even throughout the area. Right: in a basic solution, sound is less even, and spots with lower sound volume are allowed.*

For announcements, a basic solution will be enough in most situations. If the ambient audio level is very high (like a noisy manufacturing site), a premium solution is recommended. For background music in retail (like a grocery store), a basic solution will be enough. A retail business choosing a premium solution could be a high-end fashion store where the customer experience is critical. Some installations might also require a combination of basic and premium. This could be a larger project with many different types of areas, such as a school campus, shopping mall, or manufacturing site.

10.3.4 Determine the speaker placement pattern

In an outdoor solution the spread of audio is better than in an indoor solution, due to less reflections, and you can often use a lower number of speakers than in an indoor solution.

When placing speakers indoors, the general rule is to, if possible, point the sound along the room. That is, if you have a rectangular room, try to place the speakers on the short walls pointing out along the longer walls. This will let the sound spread as far as possible before being reflected on the walls. However, it is not recommended to place a speaker in a corner, since that would unevenly amplify the bass sound.

Cluster placement. If you prioritize simple and low-cost installation, you can install the speakers in clusters. This will minimize cabling but might not be the best way to get a good spread of the sound.

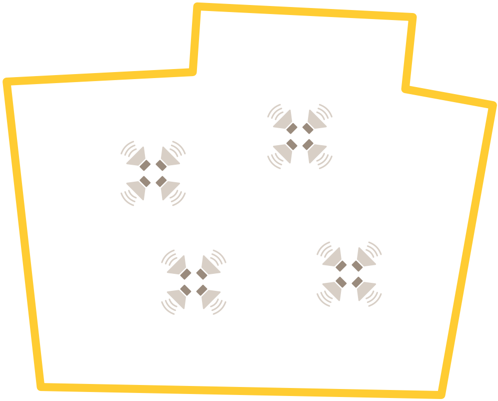


Figure 10.3b *Speakers can be placed in clusters, here as seen from above.*

Wall placement. If the room dimensions allow, and you do not mind the extra cabling, a wall placement solution will probably spread the sound better. With the same number of speakers as in the cluster placement example above, the installation might look like the below figure. If the room is large, however, the reach of the speakers might be too short.

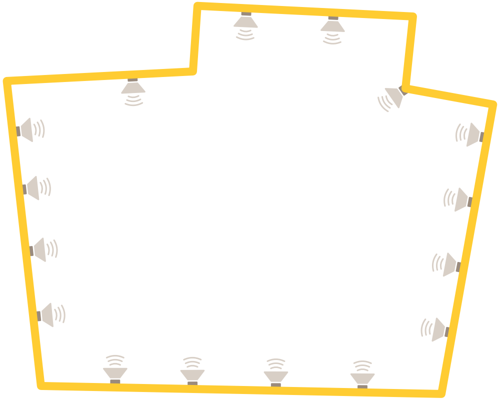


Figure 10.3c *Wall placement of speakers, as seen from above.*

Ceiling placement. If the room has a drop ceiling, or if it is possible to install built-in ceiling speakers, a ceiling placement can be a discreet solution. However, this placement is very sensitive to the ceiling height. The lower the ceiling, the more speakers you need in order to cover a certain area.

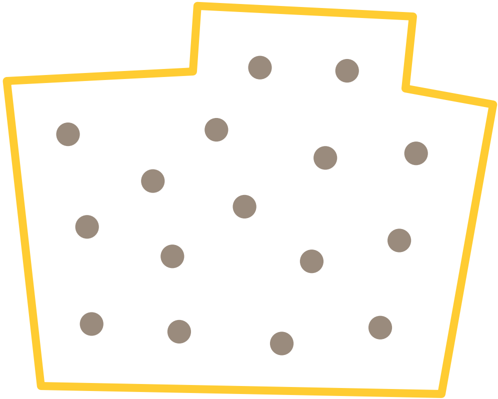


Figure 10.3d *Ceiling placement of speakers, as seen from above.*

10.3.5 Choose speaker mounts

Many network speakers come with integrated mounts that facilitate surface installation on a wall or in a ceiling. Other types of mounts may be available as separate accessories.

Ceiling mounts. All Axis network speakers can be mounted in the ceiling, which means that the sound will go from ceiling to floor. The preferred way to get the most even coverage of audio levels is a ceiling-mounted solution. This means that if you play a message it can be clearly heard at any position. However, sometimes a ceiling mounted solution is not possible due to a very solid material of the ceiling, the height of the ceiling, or obstacles between the ceiling and the floor.

Speakers can be mounted on ceilings by use of:

- Surface mount: mounted directly on the surface of a ceiling and therefore completely visible.
- Recessed mount: mounted inside the ceiling with only parts of the speaker visible. This mount is also known as a flush mount or drop-ceiling mount.
- Pendant mount: enabling the speaker to be hung from a ceiling.

Wall mounts. Wall mounts are a common and straightforward choice for many speaker installations, both indoor and outdoor. With wall-mounted speakers the audio is directed from the speaker straight out into the room. A wall-mounted solution may require fewer speakers to spread the sound than a ceiling-mounted solution, but it may require more cabling. When considering a wall-mounted indoor solution we need to account for the distance the speaker can reach into the room. For a large room, the audio levels in the center of the room might be too low if messages are to be clearly heard at any position.

Pole mounts. Some types of speakers can be mounted on poles by use of a steel strap mount. This is often used outdoors.

10.3.6 Determine mounting height and number of speakers

The coverage of an area depends on the number of speakers and their mounting height. A quick guide for speaker coverage calculations is available on axis.com. To calculate coverage of an entire room or building, use AXIS Site Designer or other available tools.

10.3.7 Choose audio management software

Audio management software from Axis is used to manage the audio content as well as the audio devices. It features functionalities such as zone and content management, scheduling, prioritization of audio sources, and system health control.

For small and midsized systems with basic use cases, AXIS Audio Manager Edge provides all the features and functionality needed for efficient and intuitive management. It allows, for example, zoning and scheduling of audio content to different areas from a single user interface. Larger and more complex systems are better managed through AXIS Audio Manager Pro, which can handle long-term scheduling and advanced priority settings.

Device management is an important part of an audio management system. It helps you manage installation, security, and operational tasks of your devices, for example, device configuration and firmware upgrade.

11. Online tools

Axis provides tools to facilitate audio installations at *axis.com/tools*

Find and compare products:

> **Product Selector** helps you find and compare Axis products.

> **AXIS Site Designer** helps you plan and design an audio installation (as well as a video installation), including which speakers to use and how many speakers are needed.

Plan and design sites:

> As a first step, we recommend the document **Quickguide for speaker coverage calculation**. It provides rules of thumb to help you estimate the number of speakers needed on a site.

> As the second step, use **AXIS Site Designer** which helps you plan and design an installation, including which speakers to use and how many speakers are needed.

> As a third step, you can use **Axis coverage shapes for Microsoft® Visio®**. Using a standard version of Microsoft® Visio® you can add speakers from the Axis coverage shapes stencils into your floor plan to visualize speaker coverage (based on sound pressure level and coverage angle) and facilitate speaker placement planning.

> If you need even more advanced design help, Axis provides **GLL files for EASE® Evac design tool**. These are speaker input files which can be imported into the commercial design tool EASE® Evac to optimize speaker placement for a carefully designed sound. Similarly, you can use **Axis plugin for Autodesk® Revit®** to place Axis products in Autodesk® Revit® building plans.

Install and manage systems:

> **AXIS Device Manager.** Helps you manage all major installation, security, and operational tasks of your devices, for example, device configuration, firmware upgrade, restore settings, and cybersecurity controls.

T10144682/EN/M1.56/

About Axis Communications

Axis enables a smarter and safer world by creating network solutions that provide insights for improving security and new ways of doing business. As the industry leader in network video, Axis offers products and services for video surveillance and analytics, access control, intercom and audio systems.

Axis has more than 3,800 dedicated employees in over 50 countries and collaborates with partners worldwide to deliver customer solutions. Axis was founded in 1984 and has its headquarters in Lund, Sweden.

For more information about Axis, please visit our website axis.com.