White paper

# Introduction to audio

## Acoustics, speakers and audio terminology

OCTOBER 2017

# Table of contents

# 1. Introduction

The audio quality that we can experience in a certain room is affected by a number of things, for example, the signal processing done on the audio, the quality of the speaker and its components, and the placement of the speaker. The properties of the room itself, such as reflection, absorption and diffusion, are also central. If you have ever been to a concert hall, you might have noticed that the ceiling and the walls had been adapted to optimize the audio experience.

This document provides an overview of basic audio terminology and of the properties that affect the audio quality in a room. It also presents a background on different speaker types and their optimal placement for an audio installation.

# 2. Audio frequency

## 2.1 Audible frequencies

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 range recorded voice.

Click the icons below to hear a pure, sinusoidal, tone at different frequencies. Use your headphones or external computer speakers if possible.

*100 Hz* *200 Hz* *1,000 Hz* *5,000 Hz* *10,000 Hz*

*NOTE: if reading a printed version of this white paper, please visit www.axis.com/products/audio for a online copy of this white paper. We recommend opening the white paper PDF in Internet Explorer or Adobe for the sound files to work correctly.*

## 2.2 Sampling frequency

The sampling frequency is the number of audio "snapshots" taken per second of the analog input audio, in order to digitally reconstruct it. In audio files and CDs, 44.1 kHz is a commonly used sampling frequency, thus using 44,100 samples per second. The sampling frequency must be at least twice as high as the highest input audio frequency that should be reconstructed.

Listen to the difference between a human voice recorded at different sampling frequencies, 8000 Hz and 44,100 Hz. Since we can only record frequencies lower than (or equal to) half the sampling frequency, the 8000 Hz sampling can only reconstruct voice frequencies up to 4000 Hz.

*Test voice 8,000 Hz* *Test voice 44,000 Hz*

## 2.3 Frequency and wavelength

There is a simple, inverse, relation between frequency (f, in Hz) and wavelength (λ, Greek letter lambda, in m):

$$\lambda = v/f$$

The wavelength is equal to the speed of sound (v=340 m/s in air) divided by the frequency. For quick conversion between wavelength and frequency, there are also online tools that can be used.

To provide some examples of audio wavelengths: a frequency of 20 Hz corresponds to a wavelength of about 17 m (56 feet), while a higher frequency of 20 kHz corresponds to a shorter wavelength of about 1.7 cm (0.7 inches). Obviously, there is a wide spread of the wavelengths of audio that we can perceive.

# 3. Acoustics and room dimensions

## 3.1 Echoes

In a room that is completely empty, there will be reverb and/or delay in the sound. This is, of course, 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.

Listen to the three recordings below to hear the difference between a dry audio, one with reverb, and one with delay.

*Dry sound* *Reverb* *Delay*

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. 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.

In small rooms at low frequencies, the room can be said to dominate the sound, whereas at higher frequencies, the speaker dominates the sound. For small rooms, the room transition frequency is often around 300 Hz. This is the frequency where the audio can be said to transcend from behaving like a wave to behaving like a ray.

## 3.3 Professional solutions for neutral 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 1. Curtains can be used to improve room acoustics.*

# 4. Measures of sound

This section deals with human perception of sound, different measures of sound, and how these relate to each other.

## 4.1 Human sound perception and phon

Even though the human 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.

Figure 2 below shows 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.

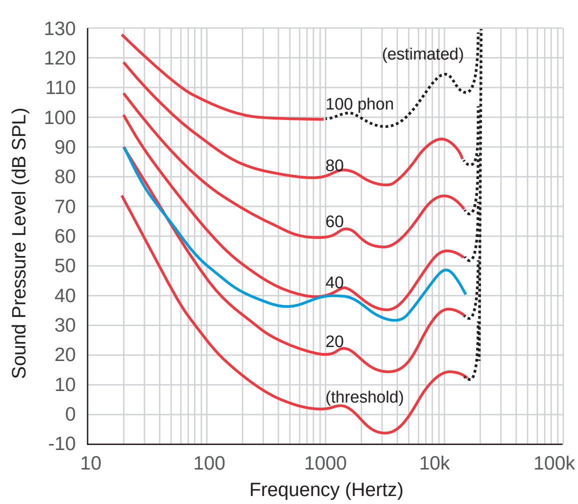


*Figure 2. Sound pressure levels needed at different frequencies in order to make a sound perceived as equally loud over all frequencies. The curves are originally from the ISO standard ISO 226:2003.*

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 - 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.

Listen to the 15 seconds long frequency sweep of a sinusoidal pure tone from 20 Hz to 20 kHz below. Note how certain parts are perceived as louder than others although all frequencies have the same power. Use headphones or external speakers to be able to hear as many frequencies as possible.

*Sweep 20 Hz - 20 kHz*

## 4.2 Watts

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.

The human ear does not perceive a 10 W sound to be twice as loud as a 5 W sound. In fact, the sound power has to be 10 times higher (50 W) for the ear to perceive it to be twice as loud. This is where the decibel comes in.

## 4.3 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 3 shows familiar sound sources and their power levels in dB.

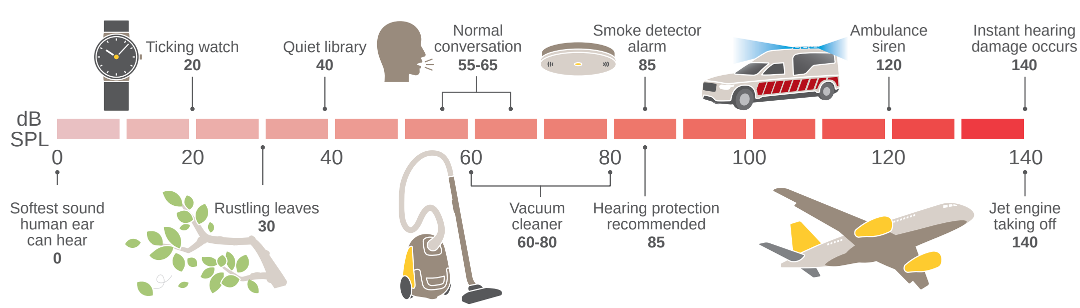


*Figure 3: Sound levels in dB SPL from familiar audio sources.*

A sound pressure level given in the weighted dBA scale has been compensated for the human ear's frequency-dependent perception of sound, as discussed in section 4.1. 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.

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.

### 4.4 Sound pressure level

Sound pressure level (SPL) is the RMS 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 the decreases by 6 dB with each doubling of the distance from the source, as illustrated in Figure 4. However, for more detailed information about the sound levels of a certain speaker, we need to look at its polar response as exemplified in section 6.1.

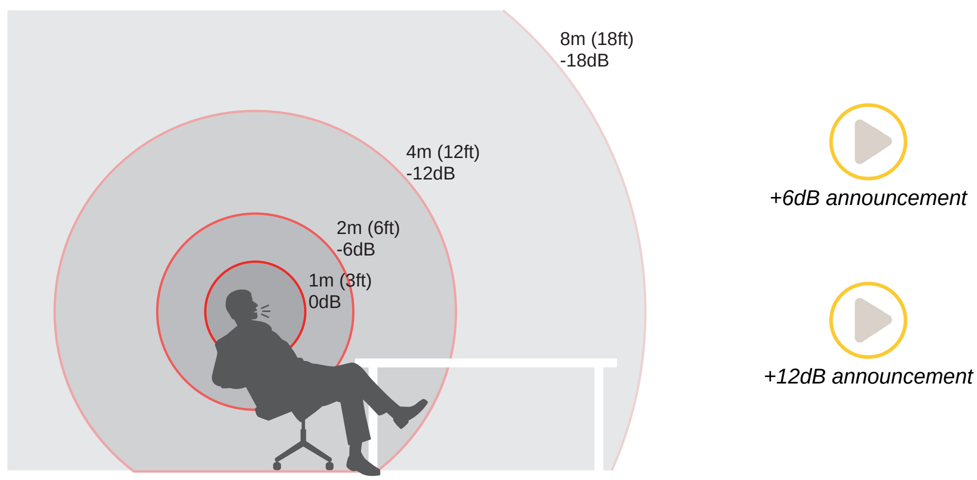


*Figure 4. The sound pressure level from an audio source decreases by 6 dB with each doubling of the distance from the source.*

## 5. Dynamic range, compression and loudness

Listen to the recording below. It is also visualized in Figure 6, where loudness is on the y-axis and time is on the x-axis. The recording has a large dynamic range, meaning that there are large differences between the quietest and the loudest part.


*Figure 5. Visualization of a large dynamic range recording, without compression.*

Listen below to the same recording after dynamic range compression was applied.

*Compression of dynamics*

The quietest parts become louder, while the loud parts either stay the same or become less loud. The differences between peaks and dips are smaller, which makes us perceive this recording as louder. As can be seen in Figure 6, the dynamic range is decreased.

*Figure 6. Visualization of the same recording as below, now after compression.*

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.

# 6. Speakers

An audio speaker can have different physical shapes depending on its purpose. The component that distributes the audio, the speaker element, is usually cone-shaped but can have other form factors if it should reconstruct high frequencies. 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.

## 6.1 Polar response

The polar diagram in Figure 8 shows how different frequencies spread out differently from a generic example speaker, placed in the center of the diagram. It shows that lower frequencies have a wide spread (even behind the speaker, at 180 degrees) while higher frequencies are more directional.

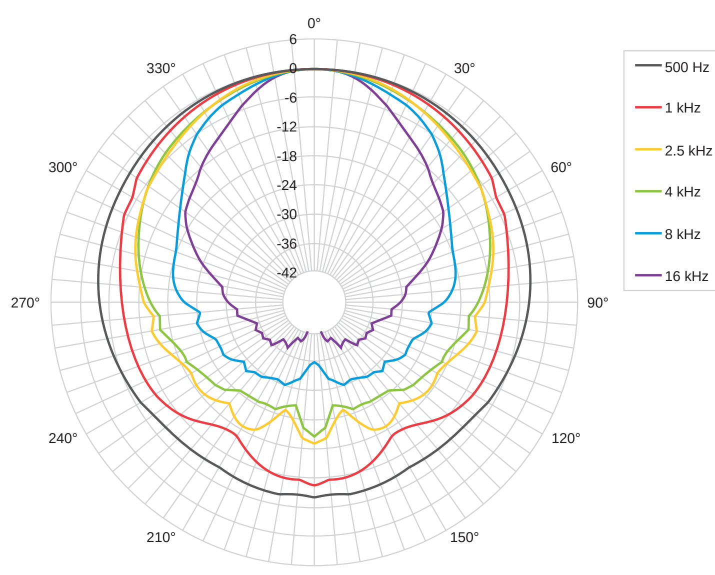


*Figure 7. A polar diagram showing the spread of different frequencies from a speaker. All sound pressure levels were normalized to 0 dB in the forward direction (at 0 degrees).*

## 6.2 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 10-inch speaker is not necessarily better than an 8-inch 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.

## 6.3 Speaker types

### 6.3.1 The hi-fi speaker

In hi-fi equipment, so-called '2-way' or '3-way' speakers are common. These speakers use several different speaker elements, in order to accurately reproduce as many frequencies between 20 Hz and 20 kHz as possible. One element might be responsible for reproducing sound up to 500 Hz, a second one for frequencies from 500 Hz to 9 kHz, and a third for frequencies above 9 kHz. These border frequencies are called 'crossover frequencies'. A hi-fi speaker is designed to reproduce audio very accurately at high loudness.

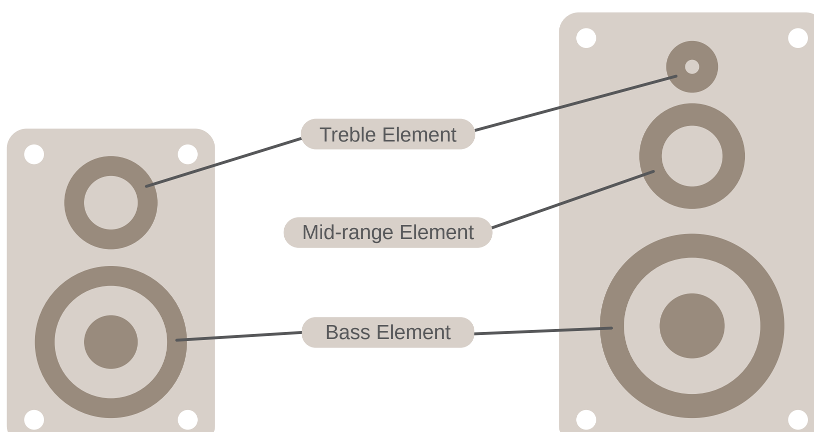


*Figure 8. Hi-fi speakers.*

### 6.3.2 The horn speaker

*Figure 9. AXIS C3003-E Network Horn Speaker*

The horn speaker has a completely different usage than a HiFi speaker, and should not cover a large frequency range. Its purpose is instead to maximize the loudness of those frequencies to which the human ear is the most sensitive, so that the speaker can convey a message (a human voice or a siren, for example) as clearly as possible. The horn directs all sound in one direction, which further enhances its sound pressure. AXIS C3003-E Network Horn Speaker is an example of this type of speaker.

### 6.3.3 The background music speaker

This speaker type has been designed to play background music with relatively low loudness. It can be a 2-way cabinet version, or a speaker designed for integration in drop ceilings. Both AXIS C1004-E Network Cabinet Speaker and AXIS C2005 Network Ceiling Speaker are good examples, using Power over Ethernet for their relatively low power consumption.

*Figure 10. AXIS C1004-E Network Cabinet Speaker and, to the right, AXIS C2005 Network Ceiling Speaker.*

## 6.4 Placement of speakers

There are many possible ways to place the speakers. The general rule is to, if possible, always 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.

### 6.4.1 The 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.

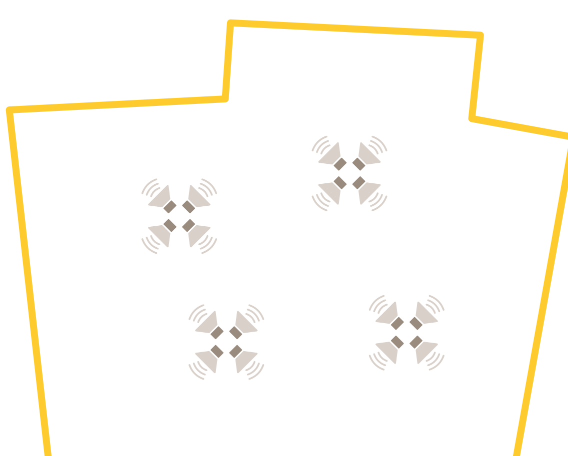

*Figure 11. Cluster placement of speakers.*

#### 6.4.2 The 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.

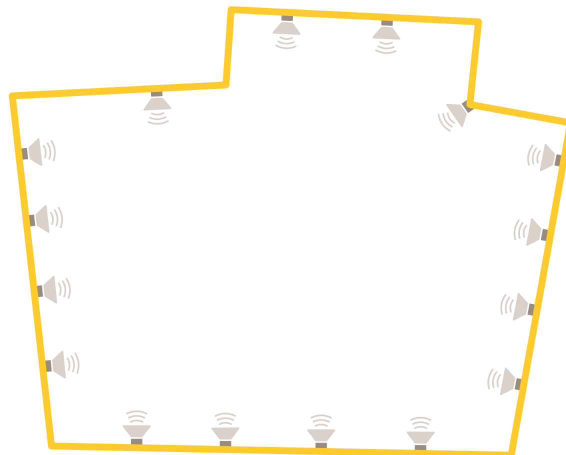

*Figure 13. Wall placement of speakers.*

#### 6.4.3 The 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.

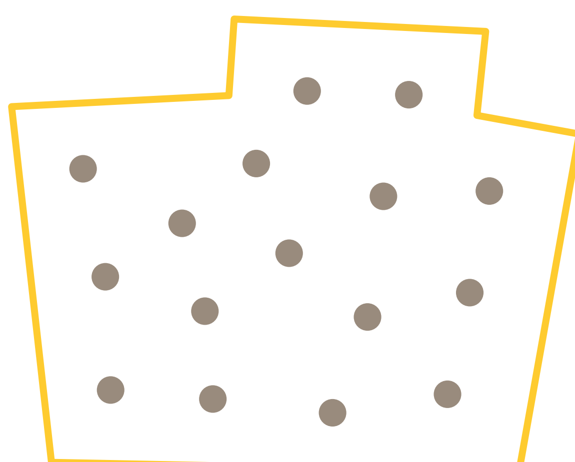

*Figure 14. Ceiling placement of speakers.*

### 6.5 AXIS Site Designer

AXIS Site Designer (**https://sitedesigner.axis.com**) is a helpful online tool for planning and designing an audio installation (as well as a video installation), including which speakers to use, how many speakers are needed, their optimal placement, and so on, with regard to the conditions at the site.

71069/EN/R1/1709

# About Axis Communications

Axis offers intelligent security solutions that enable a smarter, safer world. As the market leader in network video, Axis is driving the industry by continually launching innovative network products based on an open platform - delivering high value to customers through a global partner network. Axis has long-term relationships with partners and provides them with knowledge and ground-breaking network products in existing and new markets.

Axis has more than 2,700 dedicated employees in more than 50 countries around the world, supported by a global network of over 90,000 partners. Founded in 1984, Axis is a Sweden-based company listed on NASDAQ Stockholm under the ticker AXIS.

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

©2017 Axis Communications AB. AXIS COMMUNICATIONS, AXIS, ETRAX, ARTPEC and VAPIX are registered trademarks or trademark applications of Axis AB in various jurisdictions. All other company names and products are trademarks or registered trademarks of their respective companies. We reserve the right to introduce modifications without notice.

AXIS® COMMUNICATIONS