Engineering Psychology

Prof. Naveen Kashyap

Department of Humanities and Social Sciences

Indian Institute of Technology, Guwahati

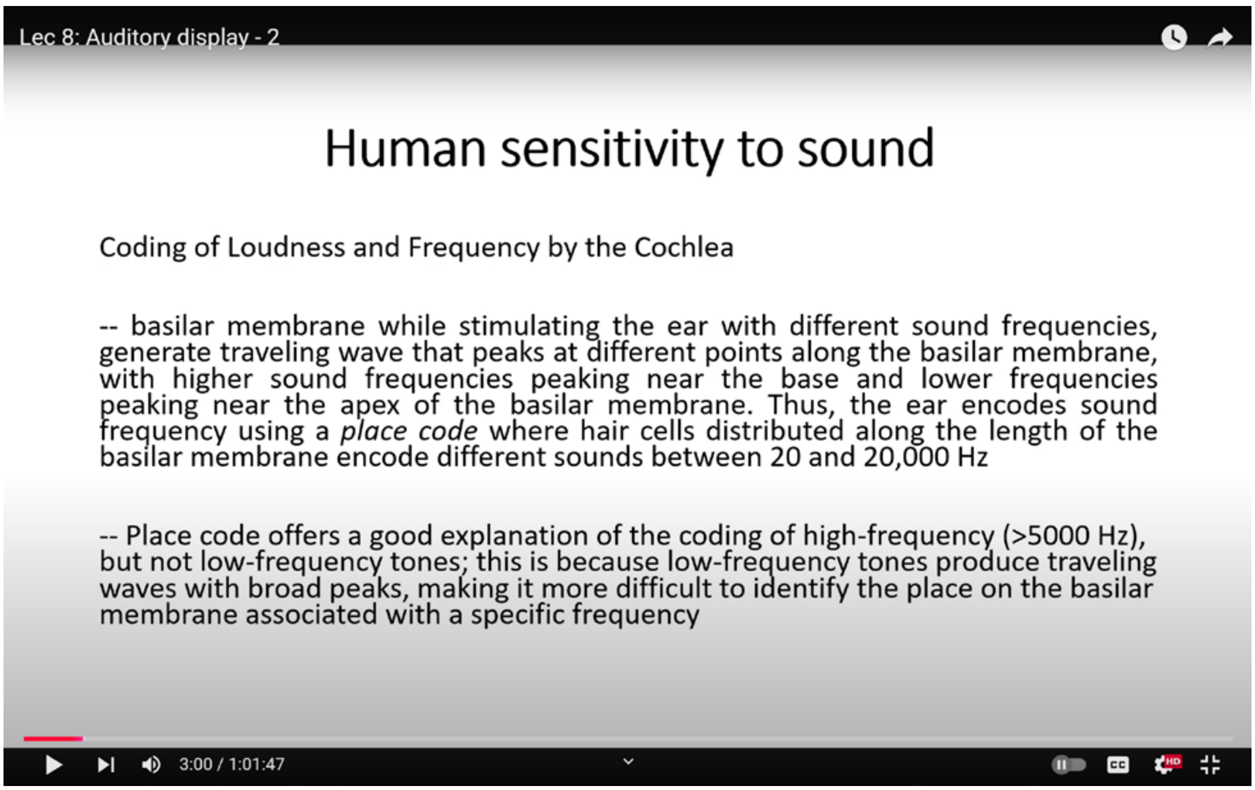
Week-04

Lecture-08

Auditory display - 2

Namaskar, friends. Welcome back to the second part of the lecture on audition. This series of lectures is designed for the course on engineering psychology, and over the past few sessions, we have been exploring how engineering psychology operates. In the previous lecture, we discussed the sensory capabilities of the auditory system.

(Refer Slide Time: 03:00)



Continuing from where we left off, today I will explain some features of the auditory sensory system. As a quick review, in the last lecture, we explored what sound is and how it is produced.

We examined how sound is measured and how complex sounds can be broken down into simpler components using the Fourier theorem. Additionally, we delved into the physiology of the human ear, specifically the outer, middle, and inner ear. I explained how sound is registered and interpreted by the human ear, and we also discussed the vestibular system, which is associated with the ear and helps maintain balance and posture.

The vestibular system plays a crucial role in helping humans maintain an upright position and track head and torso movements. We looked at how the vestibular system captures acceleration and body movements, as well as how it maintains balance and orientation. Finally, we covered motion sickness and discussed the two main theories explaining its cause.

In today’s lecture, we will explore some properties and characteristics of sound, focusing on how to harness these properties to design effective warning and alert systems for human operators. Let's begin.

To recap from the last class, we mentioned that sound can be explained in terms of two main attributes: frequency and amplitude. Frequency corresponds to pitch, while amplitude corresponds to the intensity or loudness of sound. So, how does the human ear code pitch and loudness?

As I mentioned earlier, sound perception occurs in the inner ear through a structure called the cochlea. The cochlea is responsible for perceiving and interpreting sound. Its structure resembles a snail, and it is through this cochlea that sound is registered and processed.

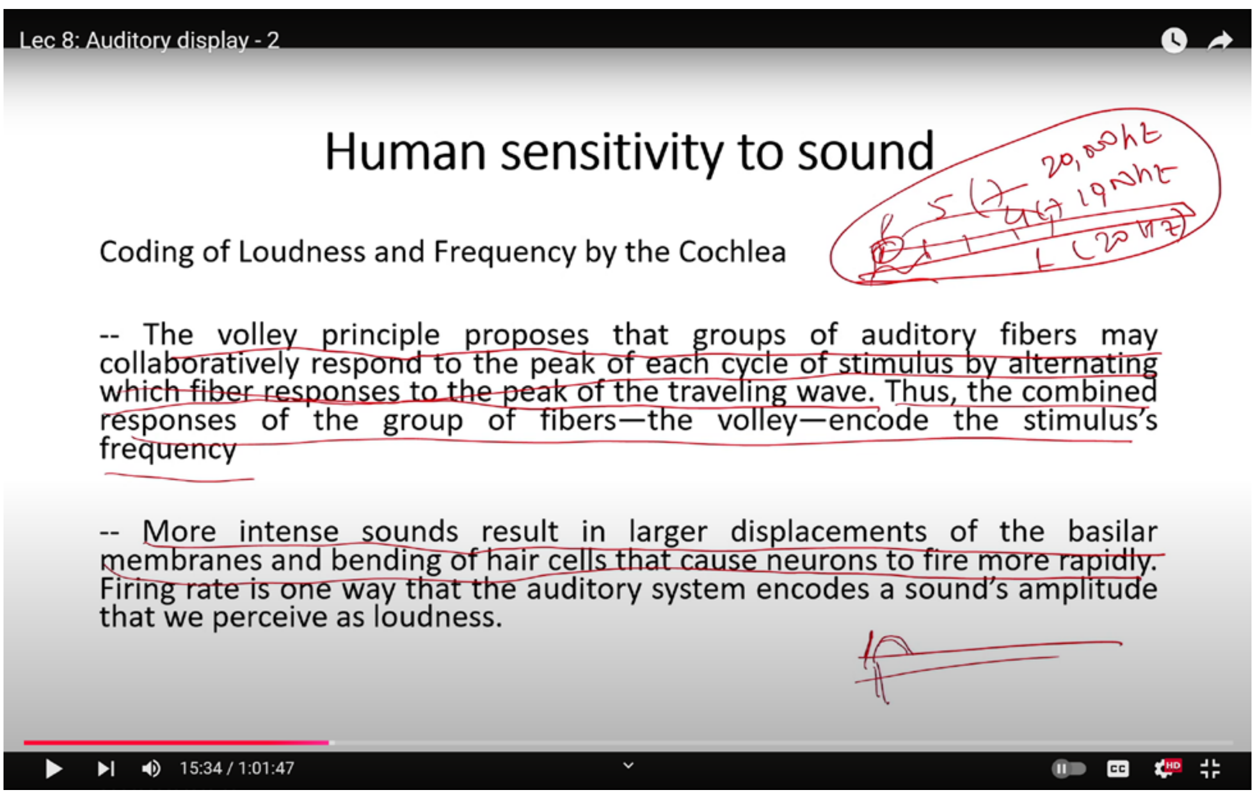
If you recall from the previous lecture, the cochlea consists of three layers. The innermost layer is the basilar membrane. Sound waves travel through the auditory canal, striking the eardrum, which then vibrates. This vibration is transmitted to the inner ear via the oval window, which causes a traveling wave to form in the fluid-filled cochlea. This fluid movement is centered around the basilar membrane.

The basilar membrane is where the interpretation of sound frequencies and loudness takes place. If you were to straighten out the human cochlea, it would be approximately 1.5 feet long, and the basilar membrane runs along its length. The basilar membrane contains hair cells that move in response to the fluid motion within the inner ear, which is caused by the vibration of the eardrum and oval window. The movement of these hair cells is what allows the cochlea to detect sound frequencies.

As different sound frequencies stimulate the eardrum, they generate a traveling wave along the basilar membrane. These waves peak at different points along the membrane depending on the frequency. Higher sound frequencies tend to peak near the base of the cochlea (closer to the oval window), while lower frequencies peak near the apex.

This arrangement of frequency peaks along the basilar membrane is known as the tonotopic arrangement. For example, sound frequencies of around 20,000 Hz peak at the base of the basilar membrane, while frequencies as low as 20 Hz peak at the apex. This organization enables the cochlea to encode sound frequencies using a "place code," where hair cells positioned at different points along the basilar membrane are sensitive to different frequencies.

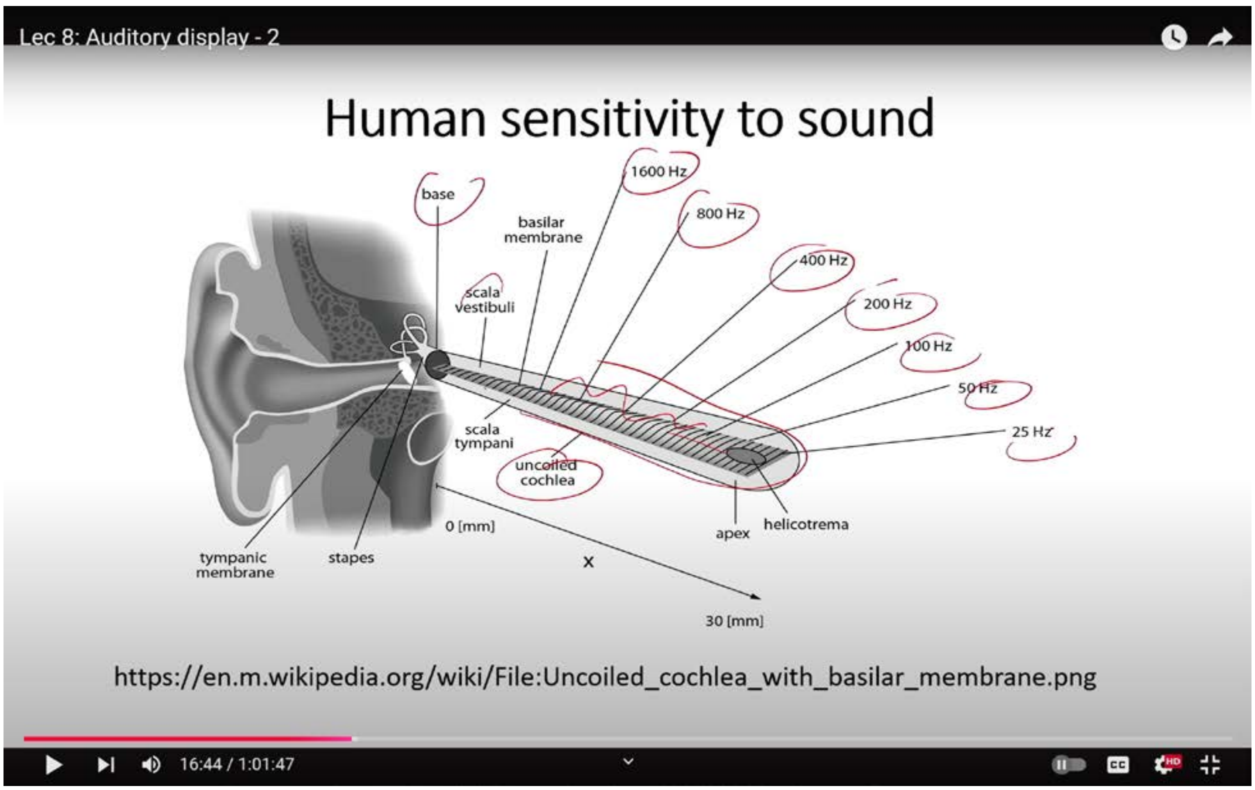
(Refer Slide Time: 15:34)



In this way, the cochlea encodes sounds ranging from 20 to 20,000 Hz. Hair cells at the base of the basilar membrane are responsible for encoding high-frequency sounds, while hair cells towards the apex detect lower frequencies. The specific location of a hair cell along the membrane determines which sound frequency it will encode, this is the concept of place coding.

The place code theory works exceptionally well for high-frequency sounds, particularly those above 5,000 Hz. However, for lower-frequency sounds, the place code does not provide a sufficient explanation. To account for low-frequency sound encoding, another theory known as the volley theory has been proposed. This theory suggests a different mechanism for encoding low-frequency tones, complementing the place code theory's explanation for higher frequencies.

(Refer Slide Time: 16:44)



The simple reason high-frequency tones override lower-frequency tones is that low-frequency tones produce broader amplitude peaks, which makes them harder to be accurately coded by the hair cells. Due to this, a different mechanism is needed to encode low-frequency tones. Let’s first understand what really happens. Low-frequency tones generate traveling waves with wider peaks.

For example, a 20,000 Hz tone will produce sharp, narrow peaks, while a 20 Hz tone will produce much broader peaks. This difference means that if both frequencies are processed simultaneously or sequentially, the lower-frequency tone, with its broader peak, will be less distinct and harder to detect. This makes it difficult to identify the precise location on the basilar membrane responsible for encoding a specific frequency. So, how are lower-frequency sound waves perceived?

They are perceived through the collective firing of multiple hair cells. The volley principle suggests that groups of auditory fibers may work together, with different fibers responding to the peak of each traveling wave in an alternating manner. This group response, or "volley," encodes the stimulus frequency. This theory indicates that one group of cells does not code for just one unique frequency. Instead, multiple groups of hair cells work together to encode different frequencies.

For example, let’s say there are five groups of hair cells encoding a 20,000 Hz tone. If only four groups are excited during the next peak of the traveling wave, it might code for a 10,000 Hz tone. If fewer groups are excited, it might code for a frequency as low as 20 Hz. The number of fiber groups excited determines the frequency of the sound wave traveling through the liquid in the cochlea. This is the explanation provided by the volley principle.

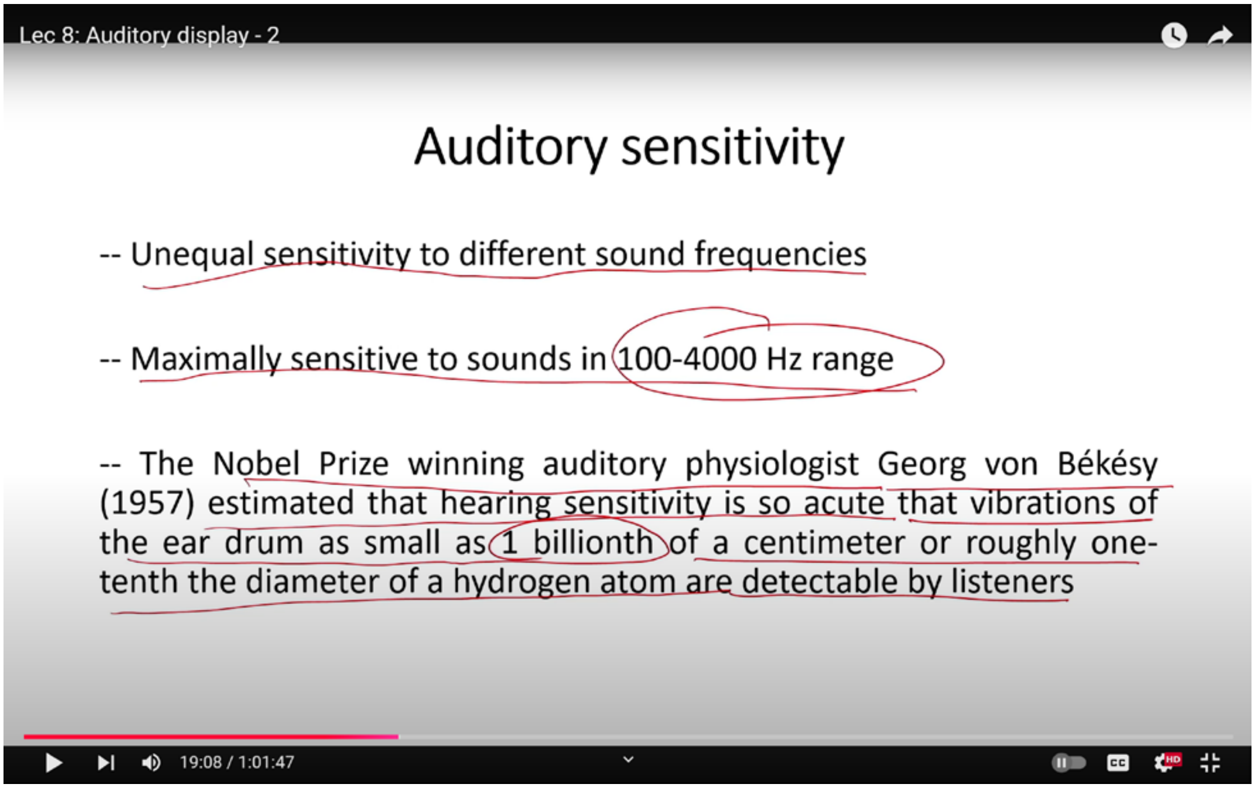
Additionally, more intense sounds cause greater displacement of the basilar membrane and more bending of the hair cells, which leads to neurons firing with greater precision. Therefore, frequencies are coded by the location or group of fibers that respond to the peaks of the traveling waves.

But how is sound intensity or loudness encoded? Loudness is encoded by the extent of hair cell bending and basilar membrane displacement. The more displacement that occurs, the stronger the traveling wave and the more forceful the liquid movement. This, in turn, results in greater bending of the hair cells. For example, if a hair cell bends significantly, the attached neurons will fire more frequently. The more bending that occurs, the more neurons are excited, leading to the perception of louder sounds. So, intense sounds cause greater displacement of the basilar membrane, leading to increased hair cell bending and more rapid firing of neurons.

This firing rate is one way the auditory system encodes the amplitude of sound, which we perceive as loudness. While hair cells encode frequency by their location or grouping, the force with which the hair cells are bent encodes loudness.

Here is a diagram that illustrates this process. As I explained, the cochlea is uncoiled, and the basilar membrane has hair cells specialized for encoding different frequencies.

(Refer Slide Time: 19:08)



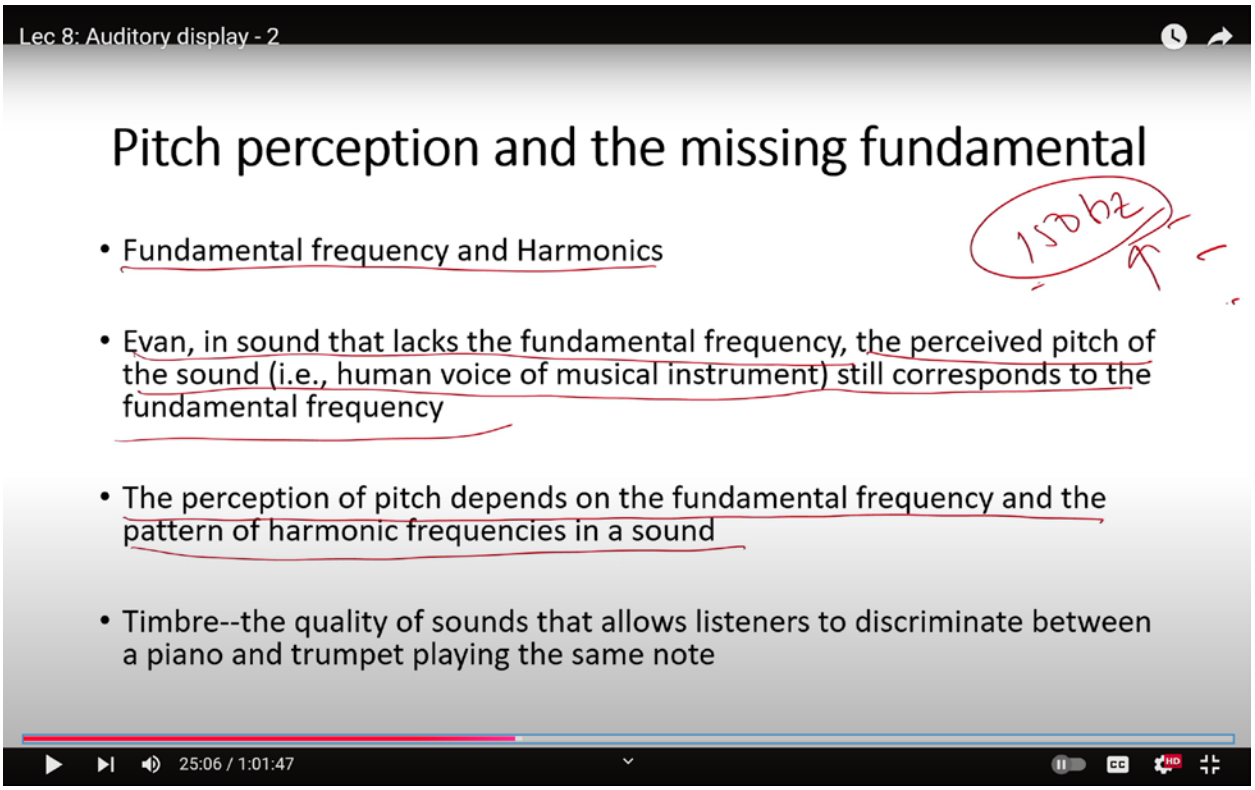
Now, let's talk about auditory sensitivity, which is directly linked to how effectively a person can perceive warning signals. Auditory sensitivity varies across different sound frequencies. For some frequencies, the auditory system is highly sensitive, while for others, sensitivity is lower. This unequal sensitivity has been experimentally studied using the theory of magnitude estimation by S.S. Stevens.

In these experiments, a standard sound frequency is produced, and subjects are then asked to rate a sound of a different frequency in comparison to the standard sound. It has been found that sensitivity varies more for lower-pitched sounds compared to higher-pitched sounds. The human ear has maximum sensitivity within the 100 Hz to 4,000 Hz range.

To further illustrate how sensitive the auditory system is, Nobel Prize-winning auditory physiologist George Wexley estimated that hearing sensitivity is so acute that vibrations as small as one-billionth of a centimeter (roughly one-tenth the diameter of a hydrogen atom) can be detected by the human ear. However, this sensitivity varies across frequencies.

This variability in sensitivity can be explained by Weber's law, which I have discussed in another course on human behavior. The law states that the amount of change needed to detect a difference in a stimulus depends on the magnitude of the original stimulus. For example, increasing the price of a 5-rupee item by 1 rupee feels like a significant change, whereas increasing the price of a 1-lakh rupee item by 1 rupee does not feel as impactful. A similar principle applies to auditory sensitivity, where the perception of sound changes depends on the magnitude of the original sound.

(Refer Slide Time: 25:06)



Now, how do people perceive pitch? It's important to note that humans don't perceive each individual frequency; instead, they perceive harmonics. Harmonics are multiples of a fundamental frequency. For example, if the fundamental frequency is 200 Hz, the harmonics would be 400 Hz, 600 Hz, 800 Hz, and so on, by multiplying the base frequency by integers.

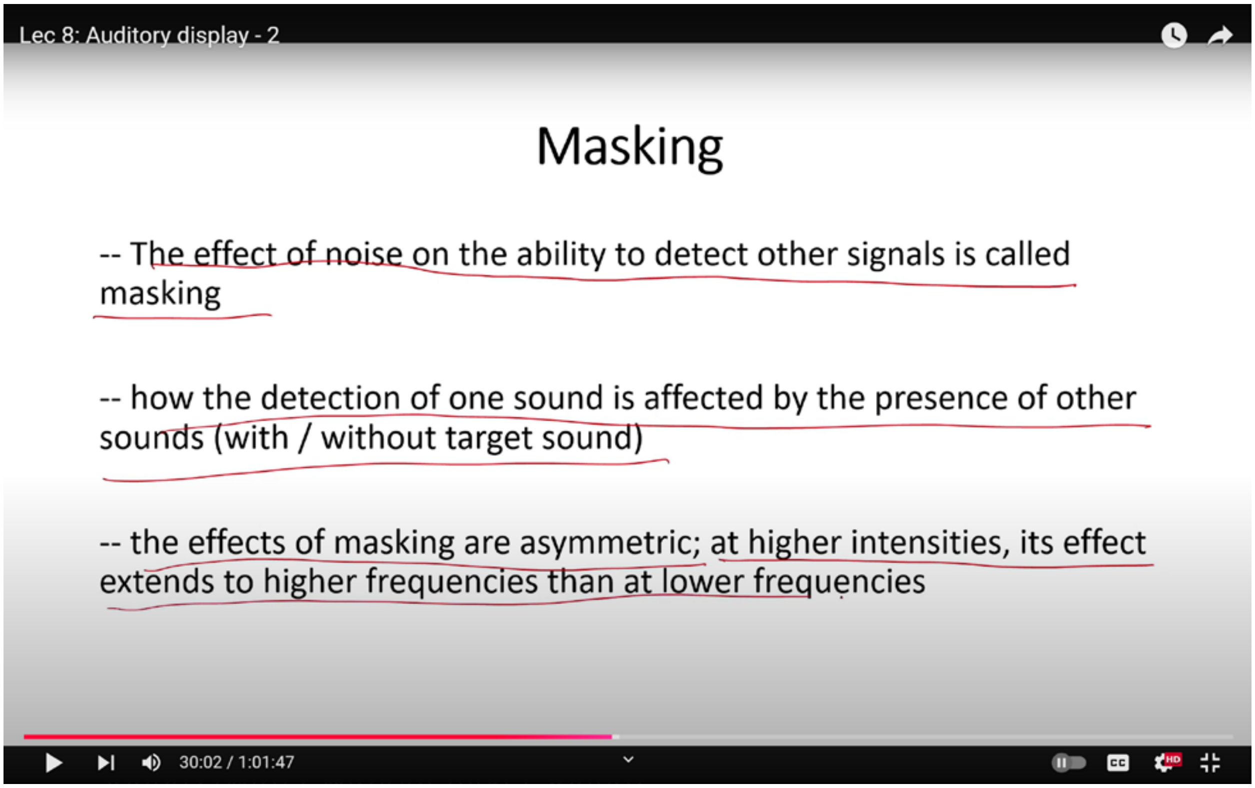
This concept helps improve the perception of pitch and sensitivity of the human ear. Musical instruments, too, have fundamental frequencies, and the multiplication of these frequencies by integers results in harmonics.

Even when a sound lacks its fundamental frequency, humans can still perceive the corresponding pitch. This is known as the phenomenon of "missing fundamental." For example, if a cell phone speaker cannot produce the 150 Hz fundamental frequency of a male voice, listeners are still able to distinguish male from female voices based on the harmonic frequencies.

Pitch perception is determined by the fundamental frequency and the harmonic frequencies of a sound. For instance, if 200 Hz is the fundamental frequency, multiplying it by 2, 4, 6, and 8 creates one complex sound, while multiplying it by 3, 6, and 9 creates another. These combinations of odd and even multiples result in different harmonic structures and, therefore, different complex sounds.

Timbre, the quality of sound, allows us to differentiate between instruments like a piano and a trumpet playing the same note. Timbre also depends on the fundamental frequency and its harmonics.

(Refer Slide Time: 30:02)

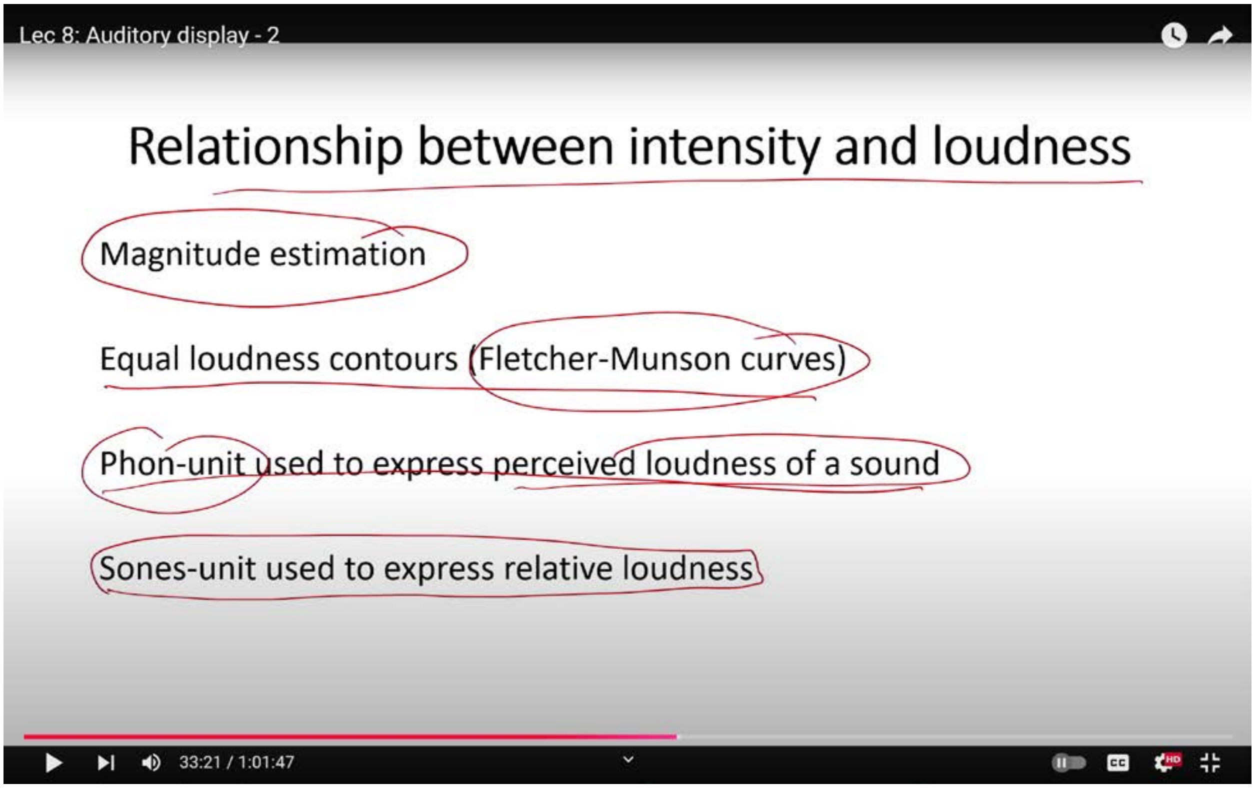


Had this not been the case, people would not be able to differentiate between the minor or major sounds produced by two different instruments. Even though two instruments might play the same note, like a C or D minor, listeners can still distinguish between them due to the perception of the fundamental frequency. Timbre, by the way, refers to the quality of sound, which helps us understand how the basic characteristics of sound are mapped and recorded by the auditory system.

Now, let's explore the concept of masking and noise.

Masking occurs when two sounds are presented, and one sound interferes with the perception of the other. The sound that interferes is called noise, while the sound of interest is referred to as the stimulus or sound of interest. Masking refers to the effect that noise has on the ability to detect other signals. It's important to note that noise is not always detrimental. For example, if the noise has the same frequency as the sound of interest, it can create greater interference. However, noises that are either higher or lower in frequency than the sound of interest tend to produce less interference.

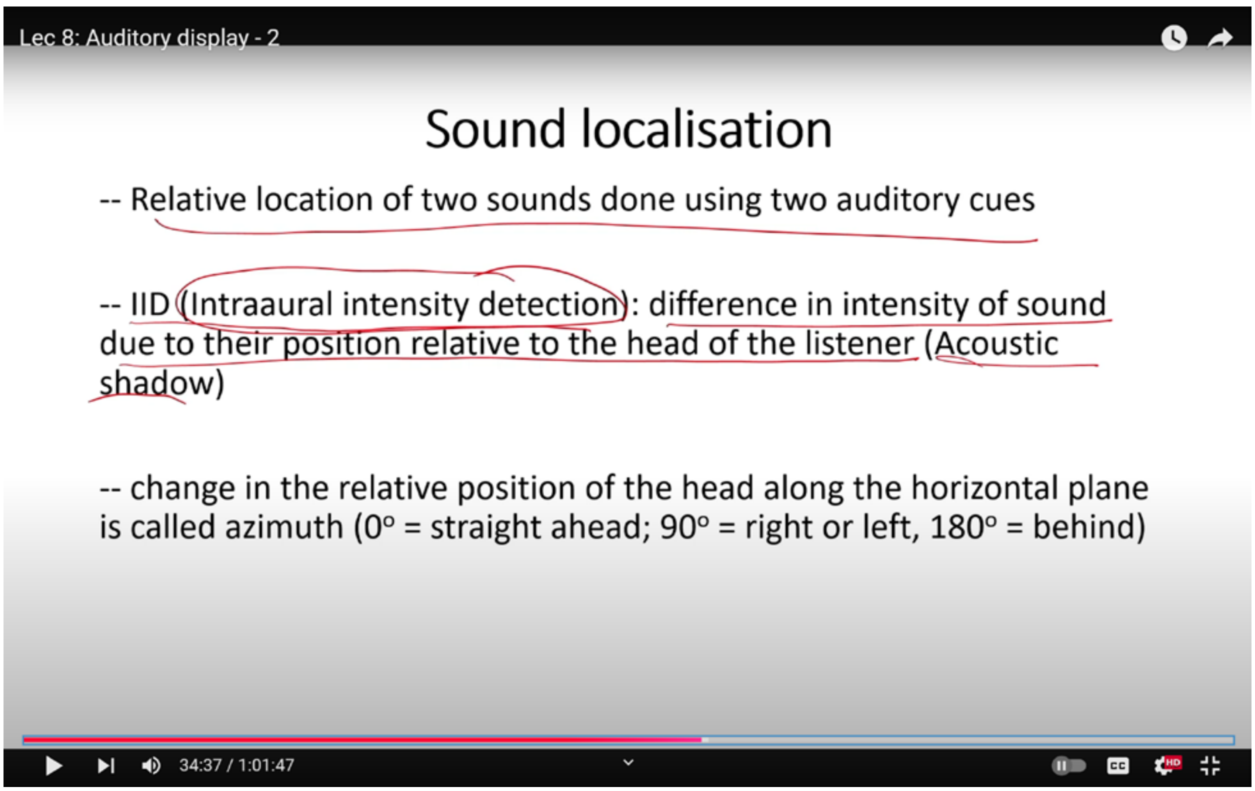
(Refer Slide Time: 33:21)



In a noisy environment, such as a factory, people can still communicate even when surrounded by machine noise because the background noise has a different intensity and frequency. In contrast, at a party where everyone is talking at a similar frequency, it becomes harder to hear a conversation clearly. This difference in how noise affects sound detection can be studied experimentally by presenting sounds of interest at different frequencies alongside noise and comparing the results.

The effects of masking are asymmetric. At higher intensities, masking has a more pronounced effect on higher-frequency sounds, while at lower frequencies, the effect is less significant. This asymmetry is important for understanding how different sound frequencies are affected by noise.

(Refer Slide Time: 34:37)

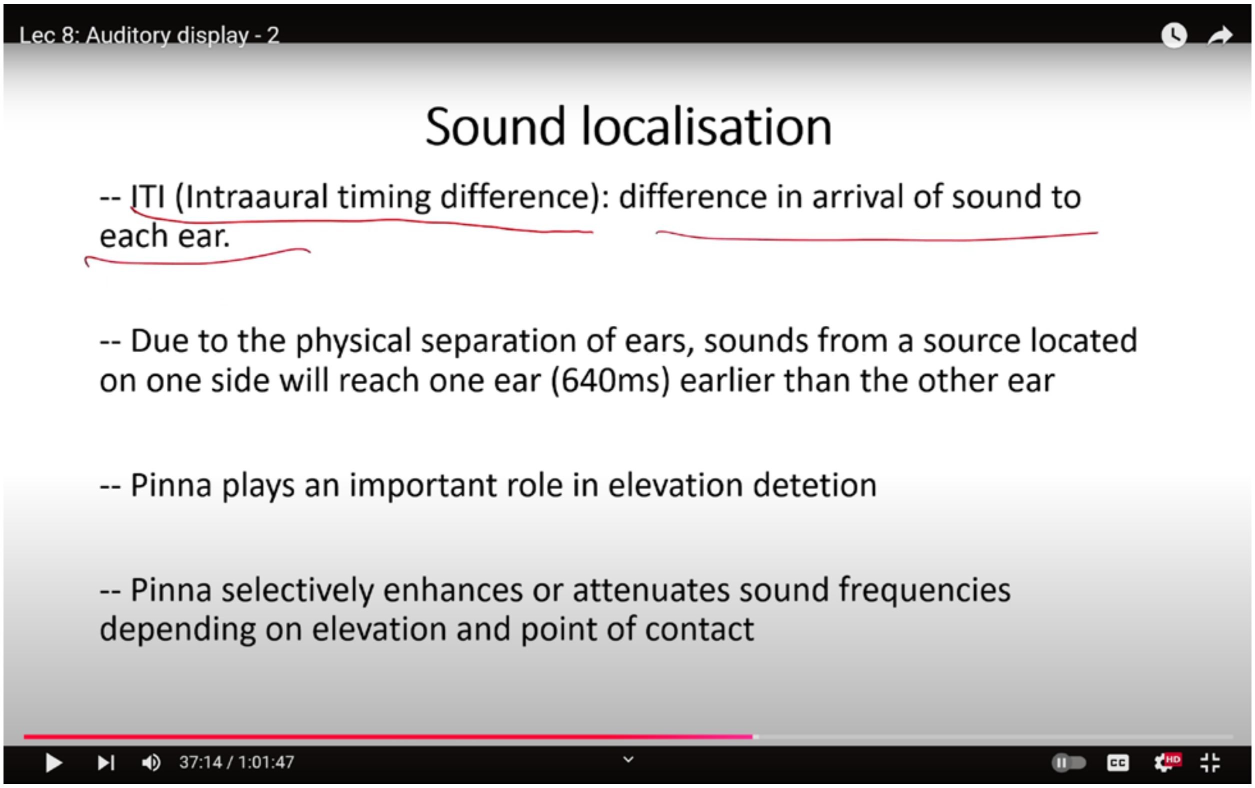


Now, let's revisit the relationship between intensity and loudness, which was discussed earlier. Intensity and loudness can be measured through magnitude estimation, where a standard sound is presented, followed by sounds of varying frequencies and loudness. The subjects are then asked to compare the difference between the new sound and the standard. For instance, if a 20 Hz or 200 Hz sound is presented as the standard, the next sound is produced and compared. The subject must determine how intense or different the sound is from the standard.

Several sounds are presented in ascending or descending order, and the point at which 50% of subjects perceive the target sound as being equivalent to the standard sound is identified. This difference is referred to as the just noticeable difference (JND), which quantifies the smallest perceptual difference between two sounds. This concept can also be studied using equal loudness contours and the Fletcher-Munson curves, which are covered in detail in the prescribed book.

The perceived loudness of sound is expressed in phons, while the relative difference between sounds is measured in sones. Phons represent perceived loudness, while sones express the relative loudness difference between two sounds.

(Refer Slide Time: 37:14)



Now, how does the human ear localize sound? This process involves two primary mechanisms: interaural intensity differences and interaural timing differences. These cues help the auditory system determine the relative location of sounds.

1. Interaural intensity difference (IID): This refers to the difference in sound intensity between the two ears. The position of the head creates an acoustic shadow, which affects the intensity of sound reaching each ear. For instance, if a sound originates from the left side, the left ear will receive a more intense sound than the right ear, which receives a slightly attenuated sound due to the acoustic shadow. This intensity difference helps localize the sound's source.

2. Azimuth: The azimuth is the angle of sound relative to the listener’s head on the horizontal plane. A sound coming from directly in front has a 0-degree azimuth, while sounds from the left or right have a +90-degree or -90-degree azimuth, respectively. A sound coming from behind has a 180-degree azimuth. The ear uses the difference in intensity between the front and back directions to determine the sound’s origin.

3. Interaural timing difference (ITD): This refers to the difference in the time it takes for sound to reach each ear. If the sound comes directly in front (0-degree azimuth), both ears receive the sound simultaneously. However, if the sound originates from the left or right, one ear will perceive it slightly earlier than the other. While this time difference is minimal, the brain is sensitive enough to detect it, allowing for the localization of the sound's source.

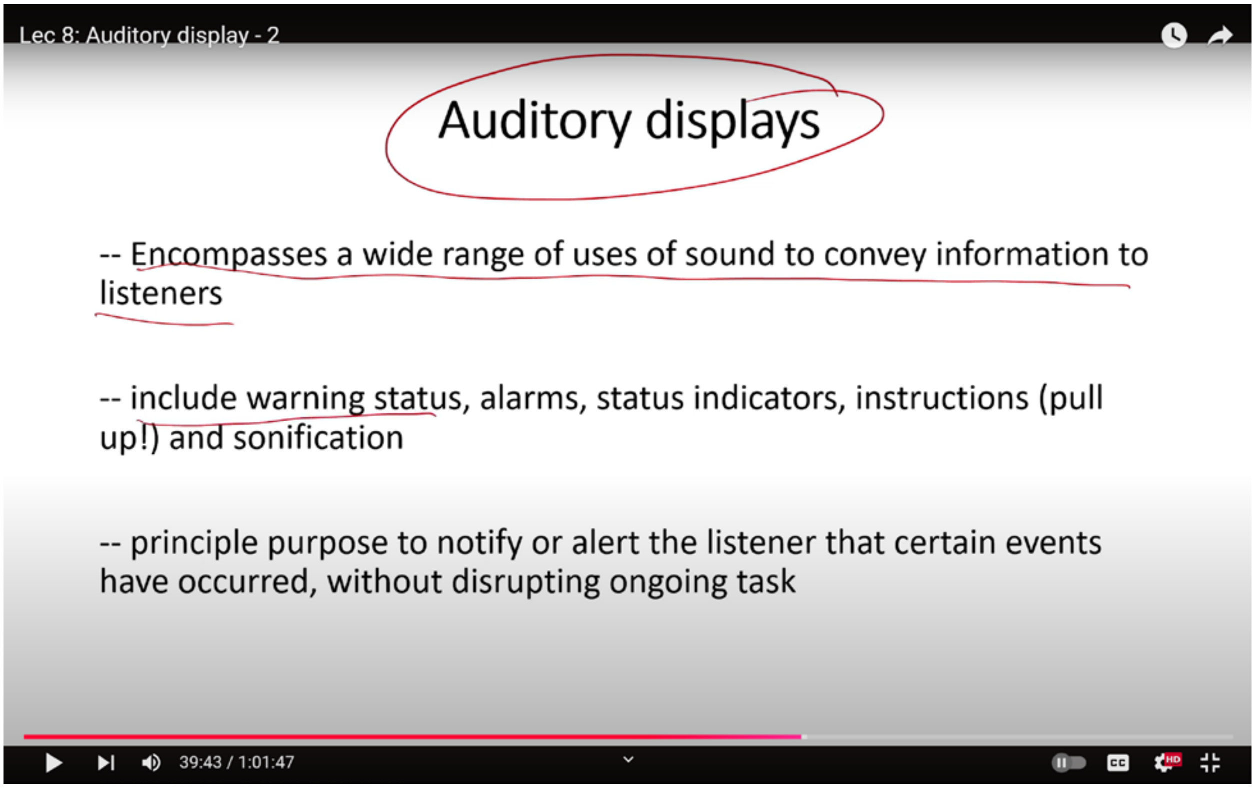
Due to the physical separation of the ears, sound from one side will reach one ear 640 microseconds earlier than the other. The pinna also plays a significant role in detecting sound elevation (whether a sound is above or below you). Its shape enhances certain frequencies depending on the direction the sound comes from. This helps in determining the elevation of the sound source.

This concludes the section on how sounds are perceived and understood by the auditory system and the physiological structures involved in encoding sound. Now, let's apply this knowledge to the design of auditory displays.

Auditory displays use sound to convey information to listeners in a variety of ways, including warning signals (such as seatbelt reminders), alarms (like the morning clock alarm), and status indicators (e.g., the ticking sounds from machines that indicate operational status). These displays are designed to take advantage of the auditory system's capabilities and limitations in conveying critical information to users.

Most of you have probably heard the sound produced when a hard disk malfunctions. It intensifies, resembling the noise of something being rubbed against another surface. This sound acts as a status indicator, warning that the hard disk is not functioning correctly. Similarly, consider the dinging sound that occurs when cooking finishes in a microwave oven, it's another type of status indicator.

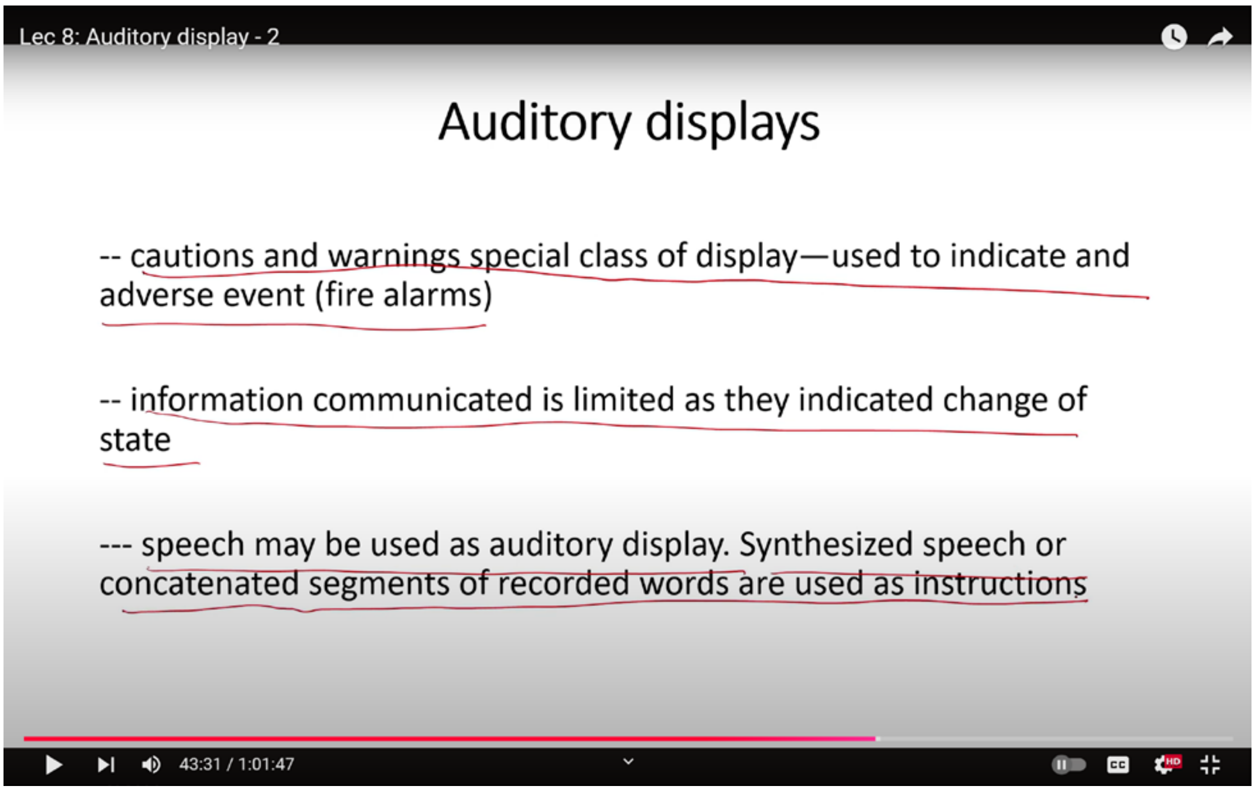
(Refer Slide Time: 39:43)



Instructions are also conveyed through auditory displays. For example, pilots receive auditory instructions before landing or when a procedural error is made. If you’ve ever sat near the front of an airplane and listened closely just before landing, you might have heard the flight control system issuing commands such as "40 meters, 30 meters, retract, retract, retract," which is a warning that the landing gear should be retracted for a smoother landing. Likewise, auditory signals like "pull-up" are provided by the aircraft’s onboard computers when the system detects that the plane is too close to the ground or in a potentially dangerous position, signaling the need for corrective action to avoid a crash.

Another crucial component in auditory displays is "sonification," which we will explore in detail. The primary purpose of these warning systems, or auditory displays, is to alert and notify the listener of specific events without disrupting ongoing tasks. For example, a pilot might be listening to the ACAS (Airborne Collision Avoidance System) or TCAS (Traffic Collision Avoidance System) broadcast while monitoring numerous visual dials and communicating with the flight engineer. The auditory display conveys only the most critical information, such as when the plane is at risk of stalling due to a steep nose-up position. The auditory system will sound an alarm warning the pilot to lower the nose or take other necessary corrective actions.

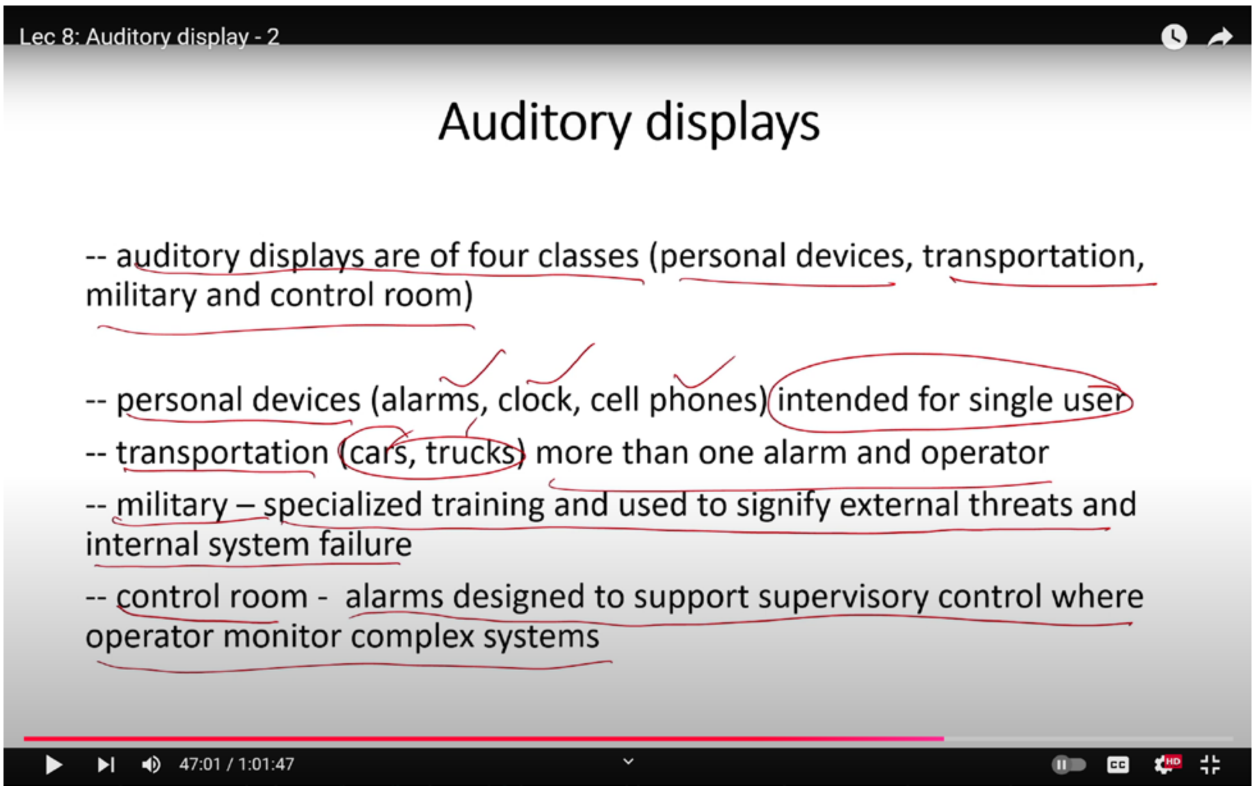
(Refer Slide Time: 43:31)



Cautions and warnings are special categories of auditory displays used to indicate adverse events. For instance, in the case of a critical issue, a specific warning, such as a fire alarm, will be issued. These alerts inform the operator that an urgent situation has arisen and requires immediate attention. While auditory displays are limited in the amount of information they can communicate, they are highly effective in signaling a change of state that demands attention.

Speech can also serve as an auditory display, using synthesized speech or concatenated segments of recorded words for instructions. At train stations and airports, you often hear a combination of sounds with mechanical fillers, like an automated announcement stating, "Train number XYZ is arriving at platform number N." In these systems, mechanical sounds are used between the pre-recorded sections to convey vital information such as train numbers, platform numbers, and arrival times.

(Refer Slide Time: 47:01)



Auditory displays can be classified into four main categories: personal devices, transportation, military, and control rooms.

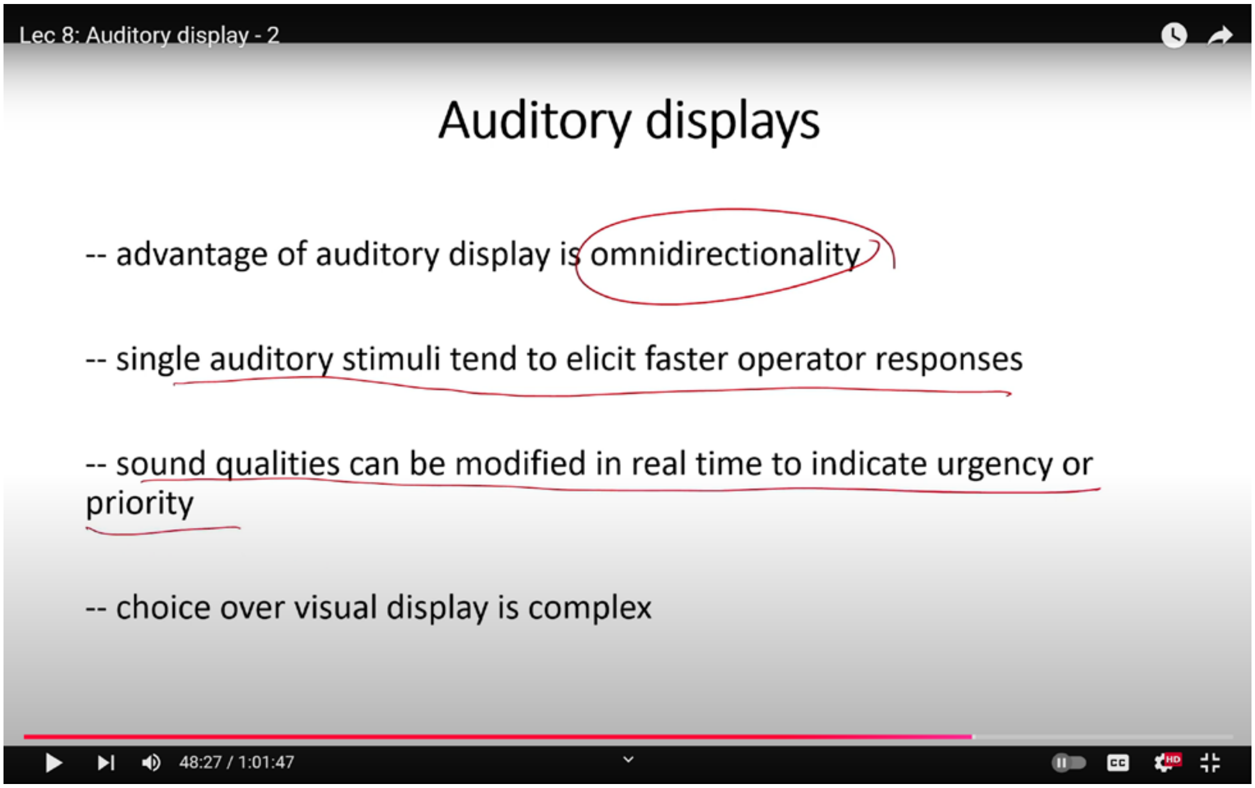
1. Personal Devices: Auditory displays in personal devices include alarms, clock alerts, and cell phone notifications. These are typically designed for individual users.

2. Transportation: In transportation, auditory displays cater to multiple operators, such as in cars or trucks where different drivers may receive warnings about seatbelts, low battery, or engine issues. In public transportation systems, auditory signals (e.g., "next station on the left" in metros) are meant for all passengers.

3. Military: In the military, auditory displays are used to signal both external threats and internal system failures. Specific codes and warning signals are employed, such as "I can see my 20, he is at your 6," indicating the relative position of an assailant to the counteracting force. These codes help convey the threat's location to the personnel involved.

4. Control Rooms: In control rooms, auditory displays assist operators who monitor complex systems. Since it is impossible to watch all dials simultaneously, auditory alerts help direct the operator’s attention to critical areas that need monitoring, such as sudden changes in pressure or temperature.

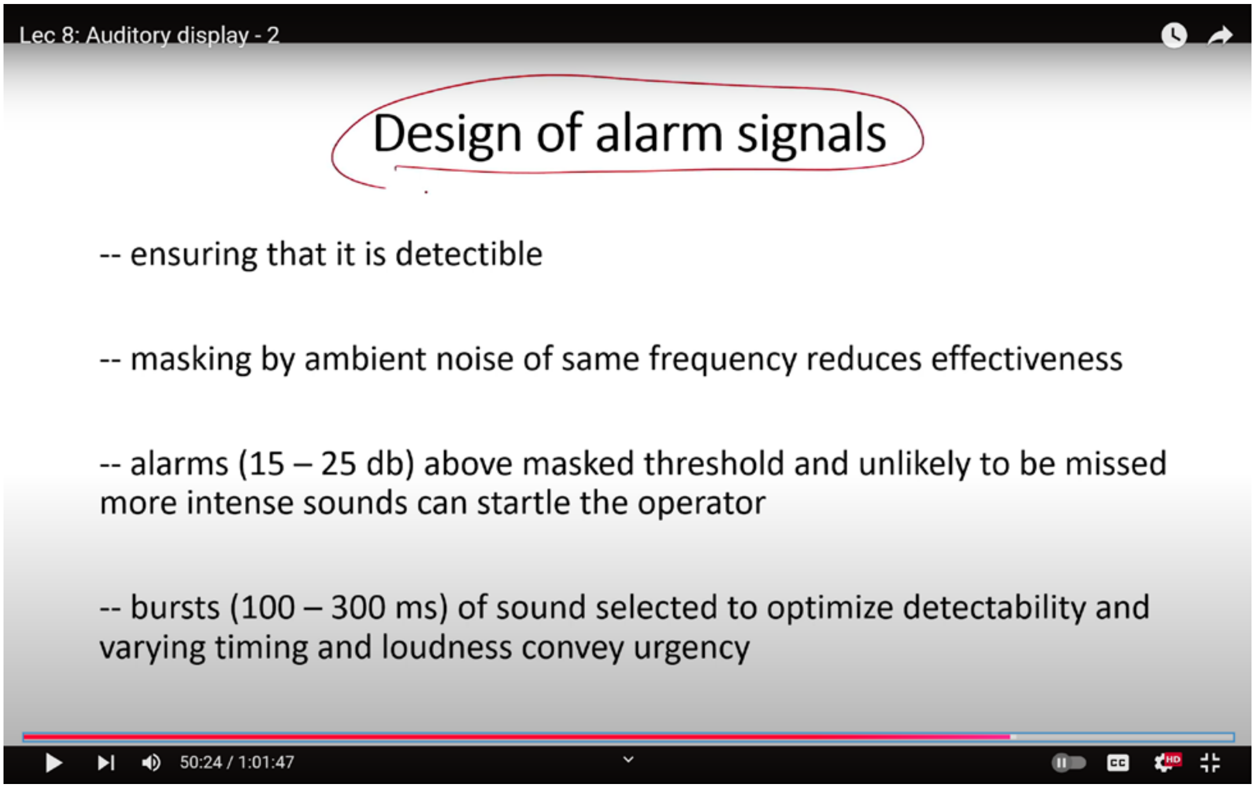
(Refer Slide Time: 48:27)



The key advantage of auditory displays is their omni-directionality. Sound can be heard from any direction, making it easier for people to receive auditory cues. Additionally, auditory stimuli generally elicit faster responses compared to visual stimuli. For example, in situations requiring quick reflexes, auditory signals like beeping can be more effective in conveying crucial information than visual stimuli.

Auditory displays also have some drawbacks. They are transitory, meaning the listener has limited time to process the information as the sound occurs in real-time. Additionally, processing auditory signals places a high demand on memory, as the listener must retain and interpret the information quickly.

(Refer Slide Time: 50:24)



Designing an Effective Alarm System:

1. Alarms should be detectable by ensuring the pitch and frequency values are within a range that humans can perceive.

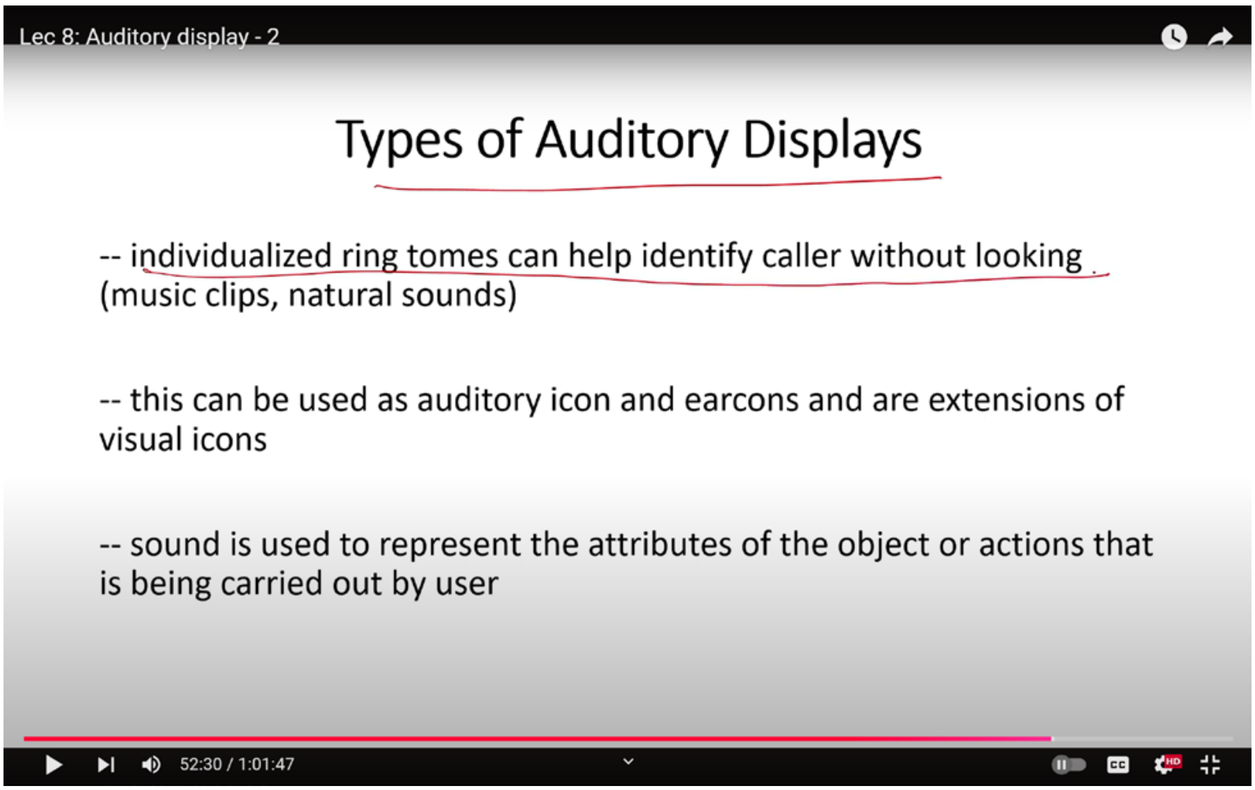
2. Avoid masking by ambient noise of similar frequencies, as this reduces the alarm's effectiveness.

3. Alarms should be 15 to 25 decibels above the ambient noise level to ensure they are not missed, but overly intense sounds can startle the operator, leading to errors.

4. High-frequency sounds may cause startle reactions, so caution is needed in alarm design.

5. Bursts of sound lasting 100 to 300 milliseconds are optimal for detectability, and varying the timing and loudness of these bursts can indicate the urgency of the situation. For instance, in a car's crash detection system, the closer the vehicle gets to an object, the more intense the alarm becomes, signaling the need for urgent action.

(Refer Slide Time: 52:30)

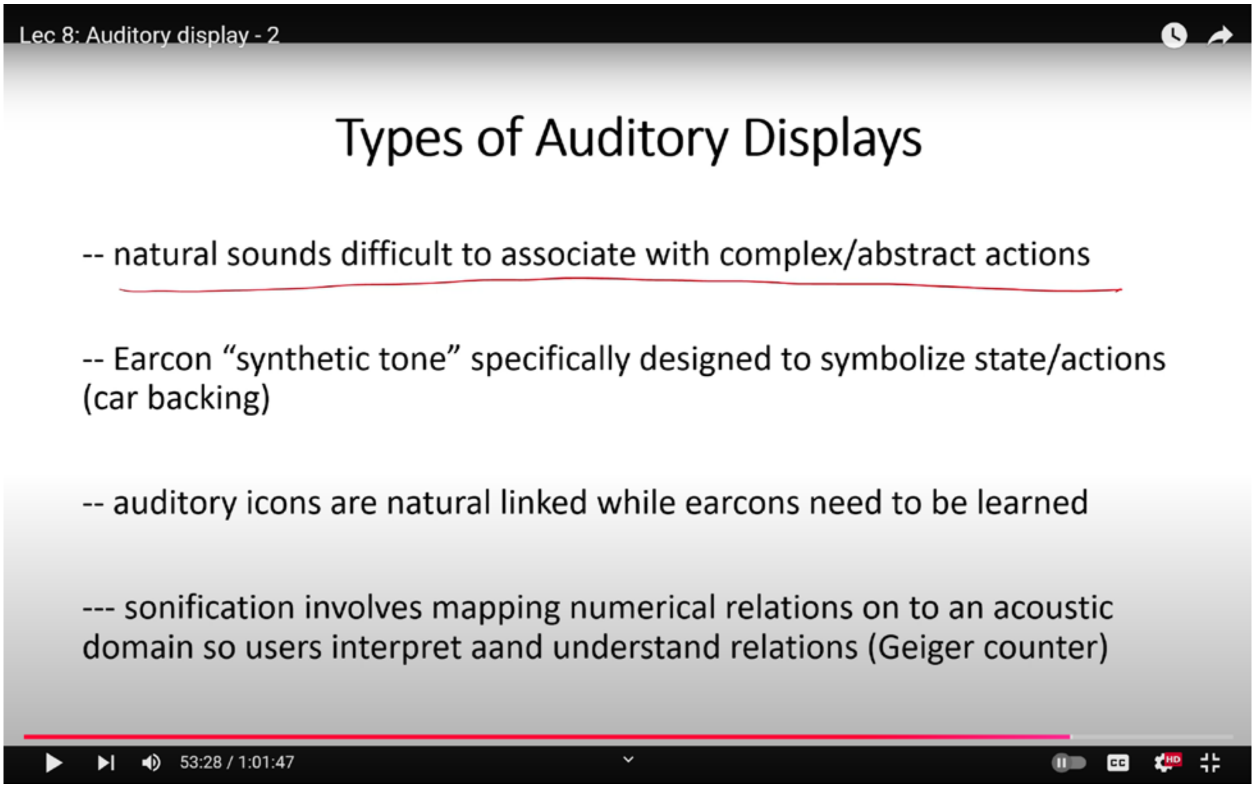


Types of Auditory Displays:

Individualized ringtones can help identify a caller without needing to look at the phone. These sounds can include music clips or natural sounds. People often associate specific sounds with particular individuals, such as the sound of a cat, a dog, or a bottle opening. These kinds of personalized sounds are set as ringtones to signify who is calling. This represents one form of auditory display.

Auditory displays can take the form of auditory icons and earcons, which are extensions of visual icons. Auditory icons are an extension of visual icons in that they personalize certain types of sounds for certain people. These sounds are used to represent attributes of objects or actions performed by the user. User actions can also be conveyed through specific sounds.

(Refer Slide Time: 53:28)



Natural sounds, however, are difficult to associate with complex or abstract actions. For instance, abstract or complex actions cannot be easily associated with natural sounds. Take the example of a car backing up, what sound should be used? The "tick, tick, tick" sound you hear when a car is backing up was artificially created and linked to this action, and now it is universally understood. This type of artificially created sound, designed to represent an action, is called an earcon. An earcon is a synthetic tone designed to symbolize a state or action, such as a car reversing.

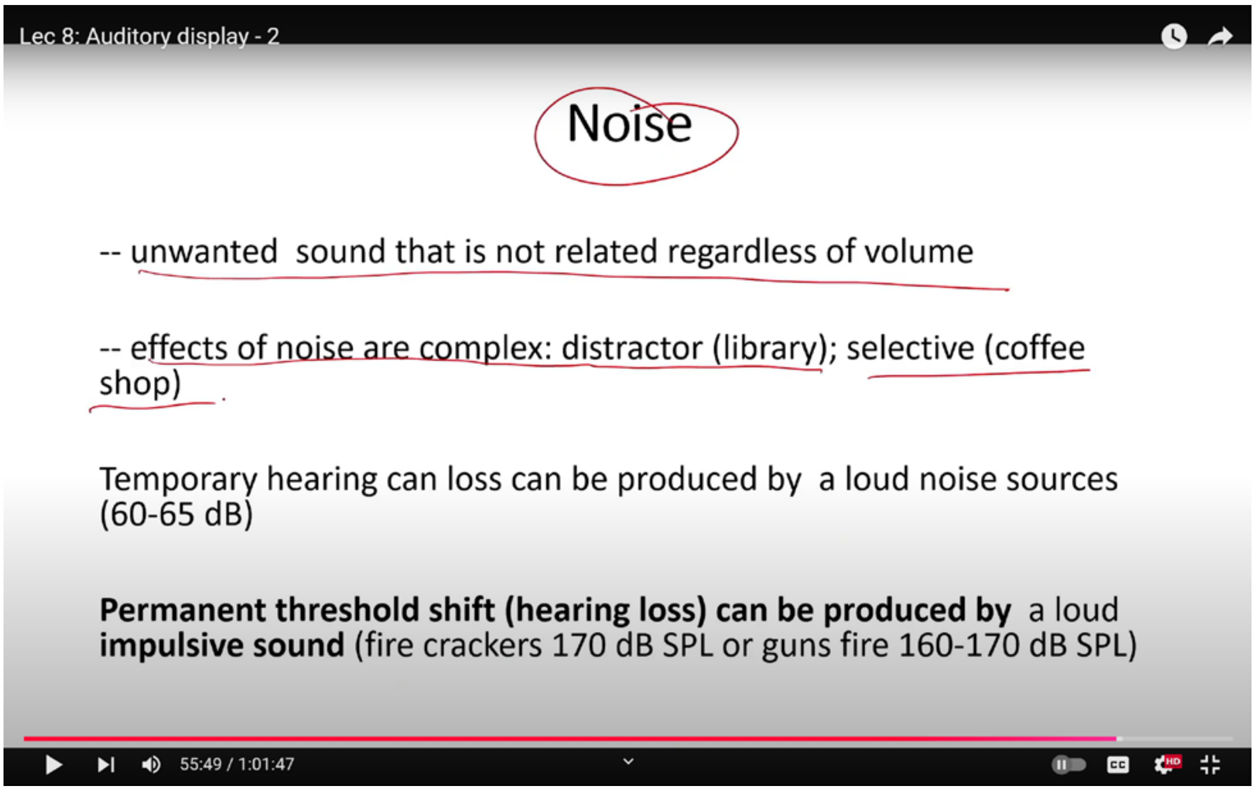
Auditory Icons and Earcons:

Auditory icons naturally link to their sounds, making them easier to understand. For instance, if a person is known to cry often, the sound of a baby crying could be set as their ringtone. If someone is loud-mouthed, a corresponding sound could be used to represent them. Earcons, on the other hand, are linked to actions that do not have inherent sounds. As in the earlier example of a car backing up, a sound was created and linked to this action, and this is known as an earcon.

Sonification:

Sonification involves mapping numerical relationships to an acoustic domain, allowing users to interpret and understand the relationships through sound. To put it simply, the more intense the sound, the more intense the situation is perceived to be. For example, a Geiger counter increases its beeping frequency in response to higher radiation levels in a room. This kind of mapping, where the intensity of the beep increases with the intensity of the situation, is the essence of sonification.

(Refer Slide Time: 55:49)

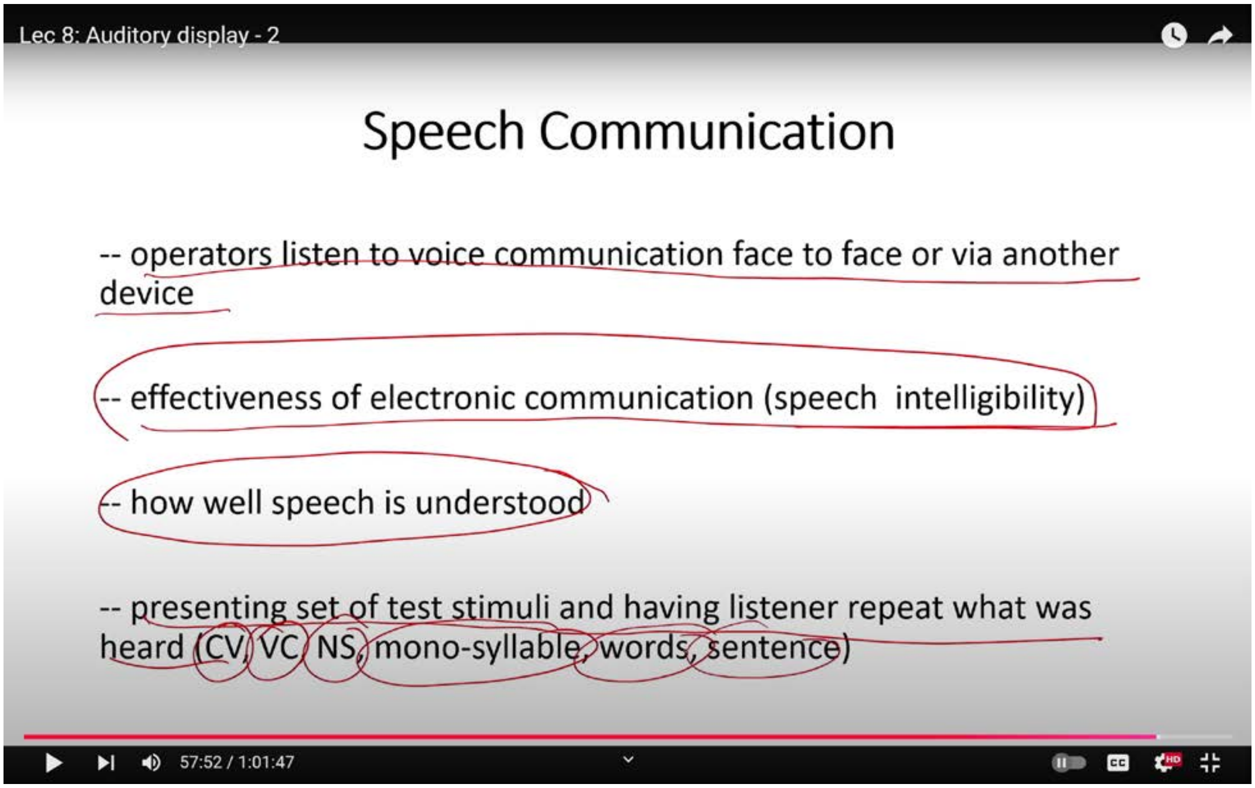


Noise:

Noise is defined as an unwanted sound that is unrelated to the target sound, and this is true regardless of the volume of the noise. The effects of noise can be complex, sometimes it is beneficial, and other times it is disruptive. For instance, noise in a library might act as a distractor, while in a coffee shop, a "ding" sound signaling that the coffee is ready serves as a demonstration of selective attention.

Noise can cause temporary hearing loss and is often produced by loud sources. Temporary hearing loss can result from exposure to sounds between 60 and 65 decibels, while permanent threshold shifts (or permanent hearing loss) can be caused by loud, impulsive sounds such as a firecracker at 160 to 170 decibels. Temporary hearing loss can recover over time, but permanent hearing loss cannot. For example, a gun fired near your ear or the loud barking of a dog could lead to permanent hearing impairment.

(Refer Slide Time: 57:52)

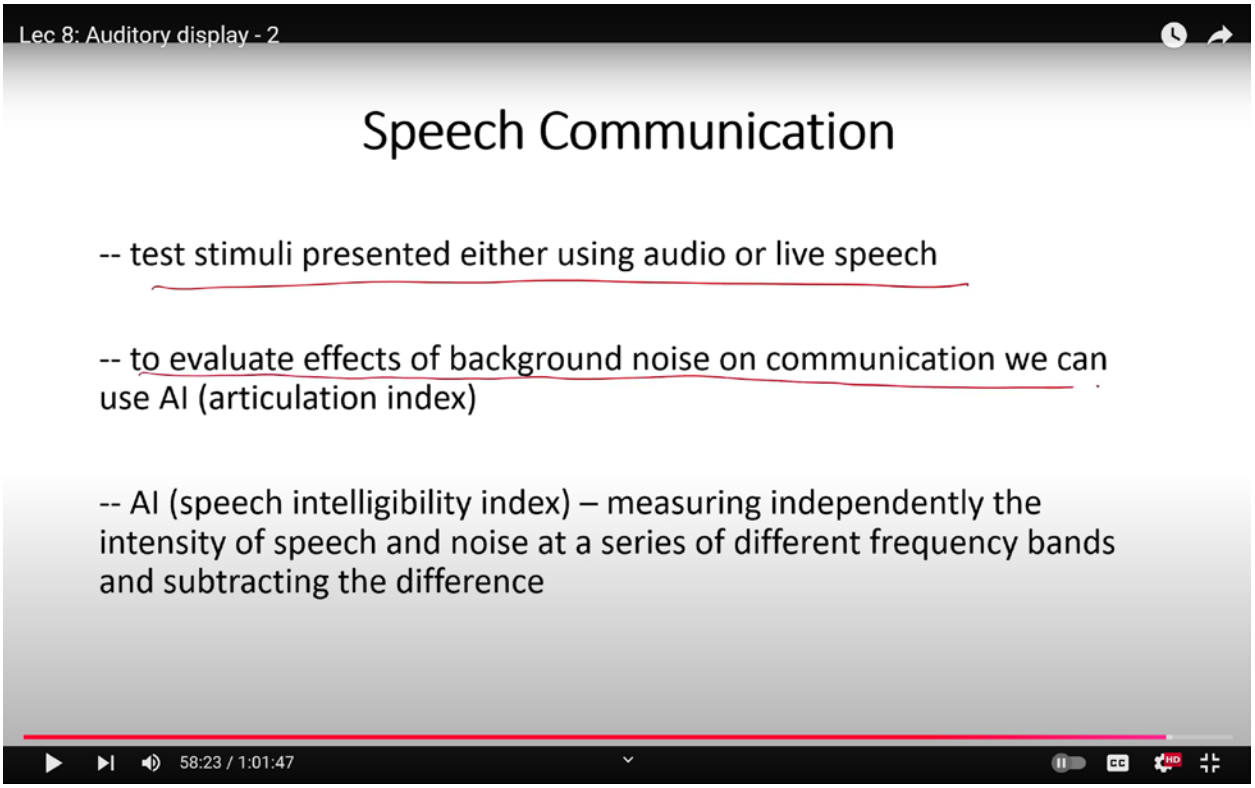


Speech as Warning Signals:

Speech can also be used as a warning signal. Operators often rely on face-to-face communication or communication via devices to perform tasks, and warnings can be conveyed through this form of speech communication. The effectiveness of this electronic communication depends on speech intelligibility. If the speech is not properly decoded, it can hinder communication. Speech intelligibility refers to how well a spoken message is understood.

Testing Speech Intelligibility:

(Refer Slide Time: 58:23)



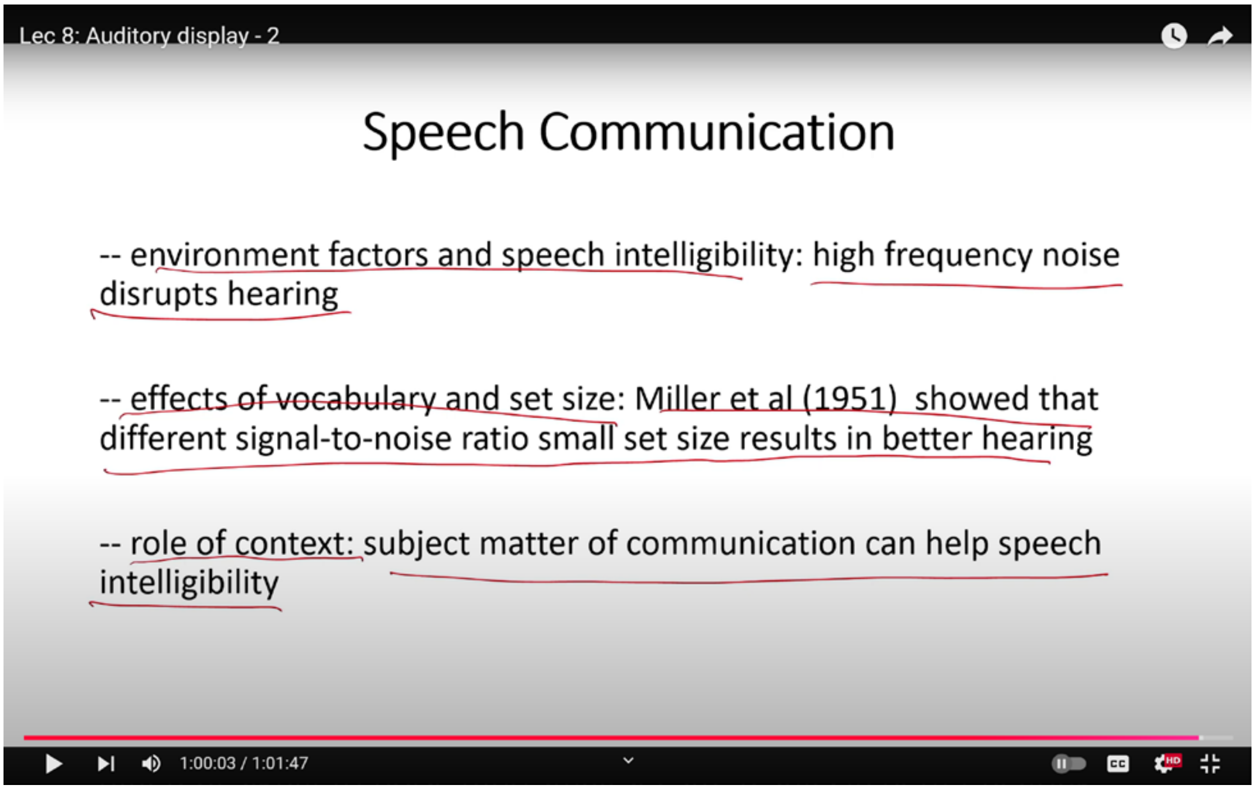
To test speech intelligibility, listeners are asked to repeat what they have heard. The test stimuli can include different types of speech, such as consonant-vowel combinations, vowel-consonant combinations, or even complex words and sentences. The listener's ability to accurately repeat what was heard indicates the level of speech intelligibility.

It could involve a monosyllabic word, a single letter (A-word), a complex word, or a sentence, and individuals are asked to repeat them. The participants hear these sounds either face-to-face or through an electronic device, and then they must reproduce what they heard. The more accurately they hear and reproduce what was said, the more effective the speech communication becomes. This information can then be applied to design better auditory displays.

Test stimuli are presented using either audio or live speech. To assess the effects of background noise on communication, we use the articulation index. If background noise affects speech intelligibility, this can be measured using the articulation index. What is it? The articulation index, also known as the speech intelligibility index, measures the intensity of both speech and noise at a range of frequency bands. Then, the intensity of the noise is subtracted from the intensity of the target speech. The result is called the articulation index.

This difference represents how well the speech is articulated. Certain environmental factors also influence speech intelligibility. For example, high-frequency noise can disrupt hearing. Someone shouting or creating a high-frequency sound can interfere with hearing.

(Refer Slide Time: 1:00:03)



Additionally, the effects of vocabulary and set size are important. Miller conducted an experiment that demonstrated how different signal-to-noise ratios and smaller sentence sizes improve hearing. Using smaller sentences or smaller sets of words leads to better understanding than using longer sentences for communication.

There is also the role of context in speech communication. The subject matter of the communication can significantly aid in speech intelligibility. For instance, if the audience knows the subject being discussed, say, a psychology class, then even if they do not fully hear or understand certain sounds being made, they can infer what was probably said. This is known as the role of context. The context, or the background against which the speech is being delivered, can help fill in the gaps in understanding missed words or phrases.

Reading the surrounding sentences, two lines before and two lines after, along with considering the context in which the speech is delivered, can help decipher and understand the unclear word or phrase.

In summary, this lecture explored various features of sound, the concept of auditory displays, and how to improve speech communication through better design. That's all for today’s lecture. I will now sign off from the MOOC studio. Thank you and Namaskar. Amen.