The Auditory Scene

- The loudness of a sound is determined by intensity and frequency
- Timbre is influenced by the number and intensities of harmonics and their attack and decay
- Gestalt-like laws influence how sounds are grouped into auditory objects
- The ability to locate sound in space is based on differences in the sound at the two ears: intensity and phase

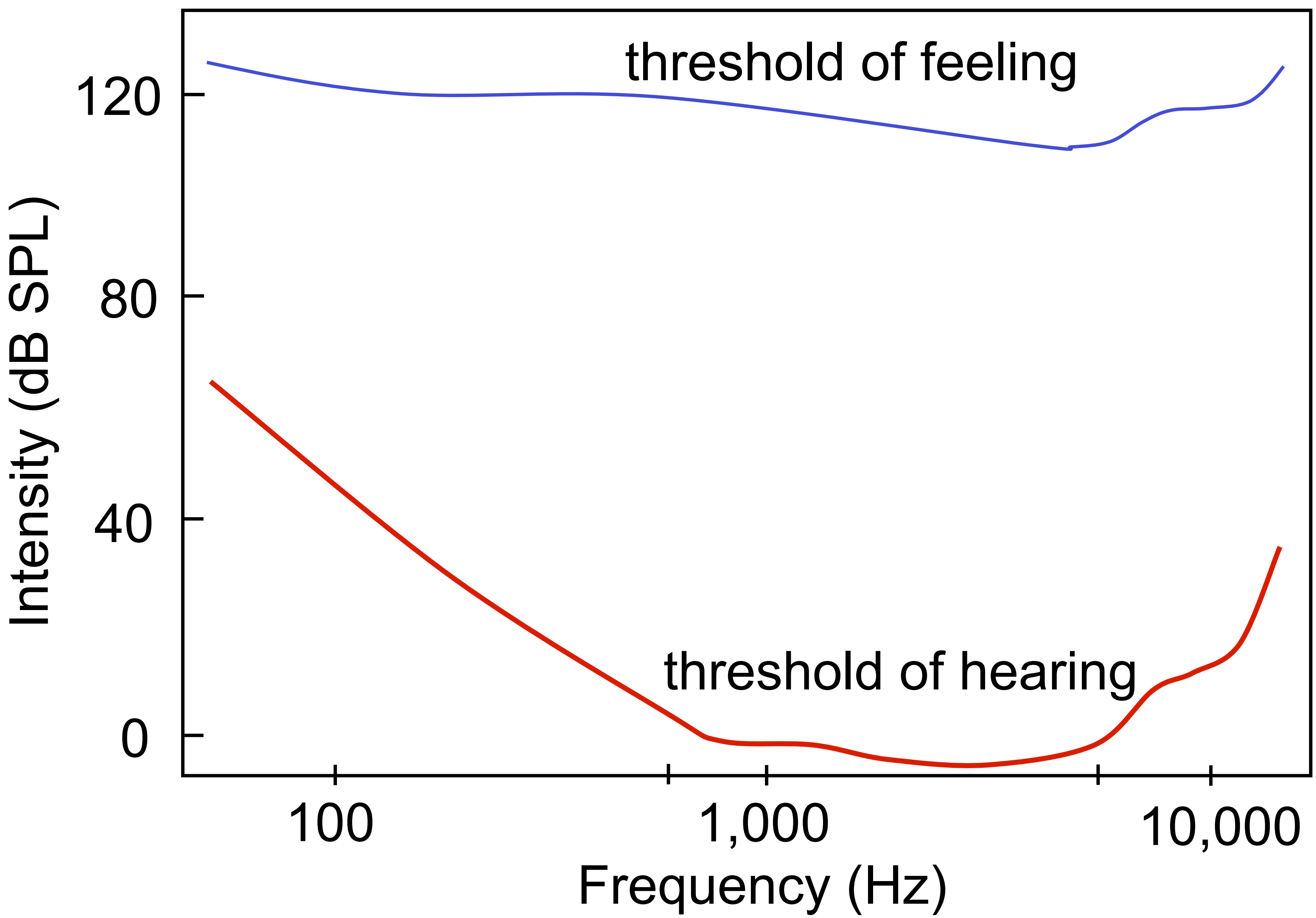
Loudness

While we refer to our perception of intensity as loudness, it is influenced by both intensity and frequency.

For example, the intensity at which a sound is just audible, the threshold, is different for different frequencies.

A sound must be 1,000 time more intense to be at threshold at 40 Hz than at 1,000 Hz.

Equal Loudness Curves



Hearing for Different Animals

For humans, the audible frequency range is 20 to 20,000 Hz (20 Hz to 20 kHz).

For elephants, it is about 10 Hz to 10 kHz.

Dogs about 40 Hz to 50 kHz.

Dolphins about 100 Hz to over 100 kHz.

In all of these cases, the threshold function has the same basic shape as humans (see Figure 12.3).

Neural Code for Loudness

1. As loudness is increased, auditory neurons fire faster.

However, they saturate (reach their maximum) over only a 40 dB range. Also, remember that firing rate can signal frequency via volleying and phase locking.

2. As loudness and intensity increase, the traveling wave covers more of the basilar membrane and more neurons fire. So, the number of units may also code loudness.

3. Efferent (descending neural projections) activity may play a role in the neural code for loudness.

Timbre

Timbre is that quality which makes two sounds with the same fundamental sound different.

1. The intensities of the harmonics. The relative intensities of the harmonics differs across instruments. For a guitar, the second harmonic may be most intense while for a bassoon, the first is more intense and for a saxophone, the fundamental, first and second harmonics are roughly equal. A flute has very few harmonics while piano notes have a rich harmonic structure.

Timbre - 2

2. Attack and decay. This refers to the time for the note (fundamental and harmonics) to reach maximum intensity and then fall back to silence. For example, plucking a violin string produces a very rapid attack while bowing the string produces a slower attack.

If a clarinet and flute play the same note, they sound different. If we remove the attack and decay and just play the steady part to listeners, it is hard to tell which is the flute and which is the clarinet. This indicates that attack and decay are the most important aspects of timbre for identifying musical instruments.

Auditory Scene Perception

In listening, we are often faced with a situation where there are multiple sound qualities present. These qualities change over time. Our perceptual system must group those together that belong to the same source.

For example, in an orchestra, one can listen to the lead violin. In a crowded room, we can follow one conversation. This implies a perceptual grouping process is active during listening.

Like the Gestalt laws in visual perception, there is a set of principles for grouping sound elements together in listening. We can also locate and separate sounds because we have two ears (sound localization).

Auditory Localization

Listeners can locate sounds in space. They are better for sounds in front of them than for sounds behind them.

Much of this ability is based on differences in the sound that arrives at the two ears. These are the binaural cues.

1. Interaural time differences. The difference in time between when sound reaches the left and right ears.

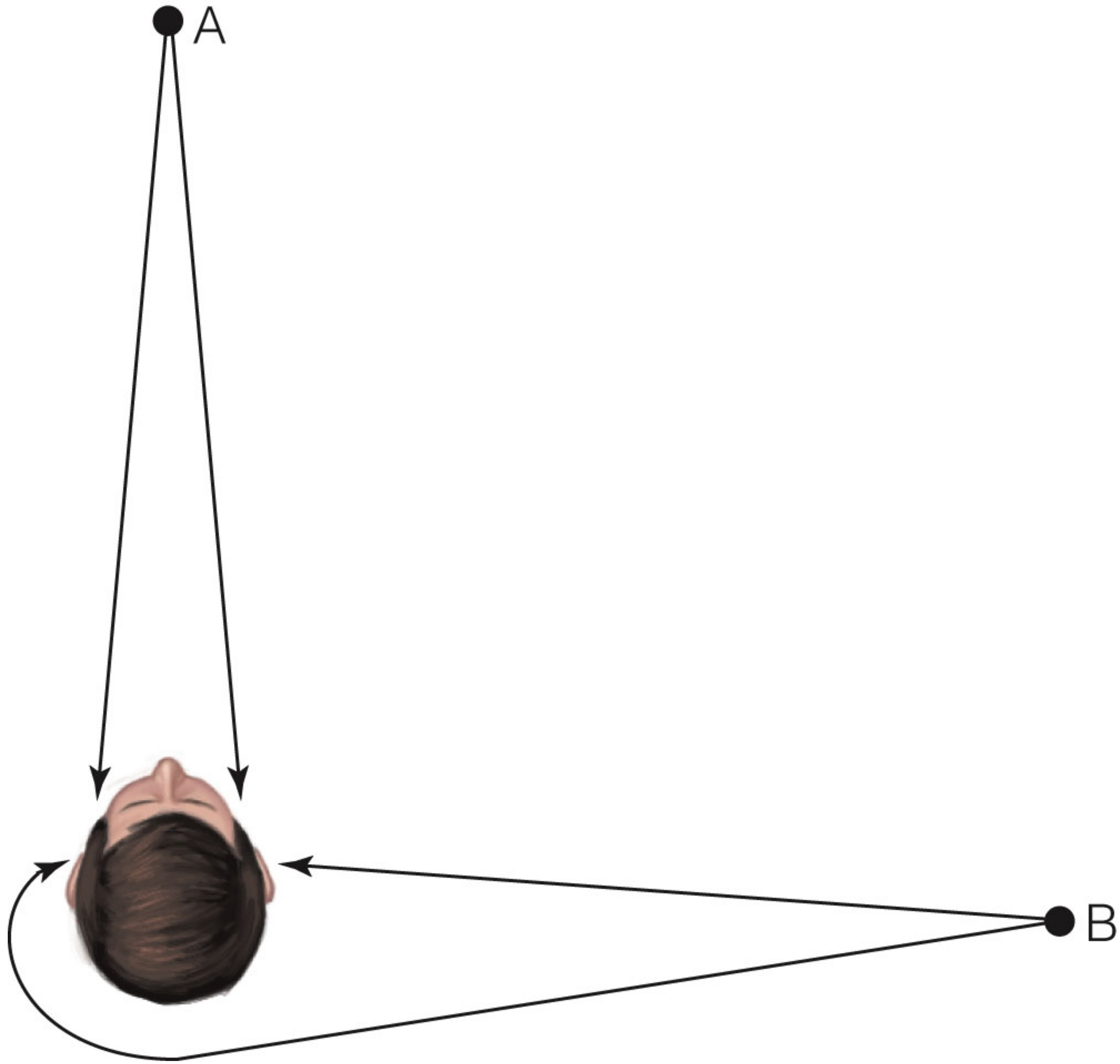
2. Interaural intensity differences. The difference in intensity between the sound at the left and right ears.

Interaural Time

If the sound is directly in front, overhead or behind (on the midline), it arrives at the two ears at the same time.

If it is off to one side, it arrives at the closer ear first. The difference in time is one cue to the location of the sound in space.

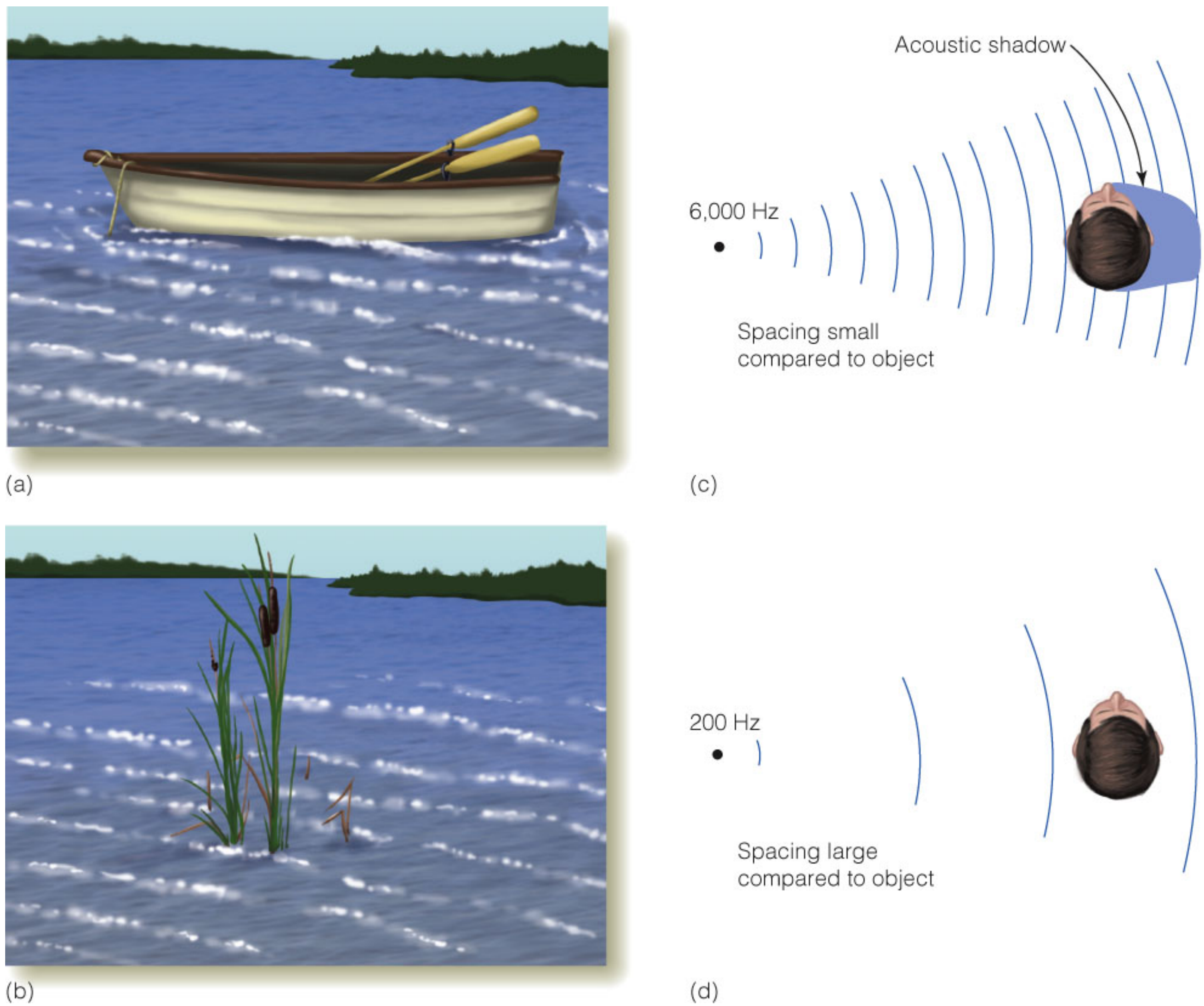
For a sound to one side, this time difference is about 600 microseconds for the average size head. Differences as small as 10 microseconds can be reliably detected by human observers.


A
B

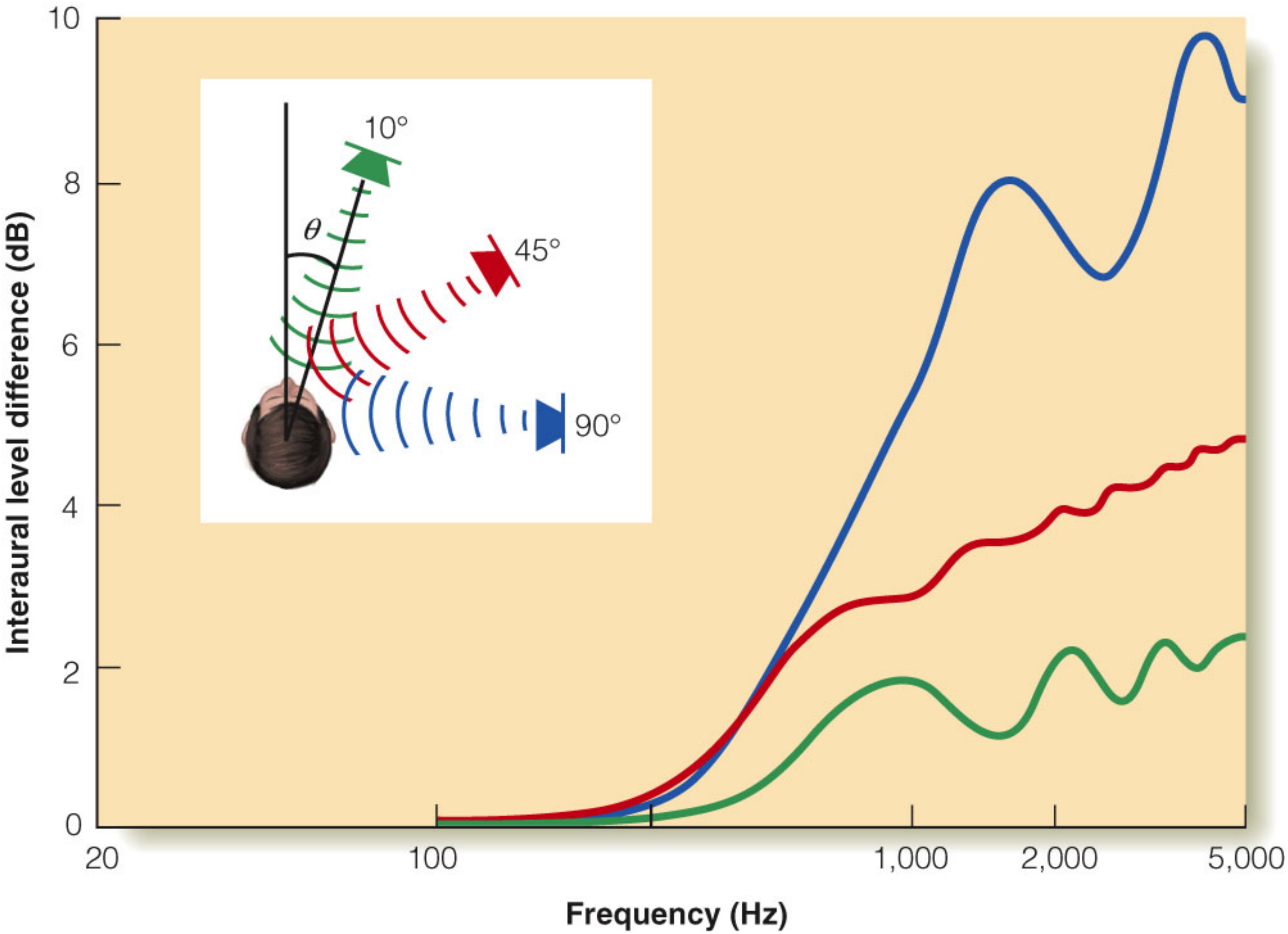
Interaural Intensity

For a sound on the midline, the path of the sound to the two ears is the same and the intensity at the two ears is the same.

For a sound off to one side, the head casts an “acoustic shadow”. That is, the head partially obstructs the sound and the intensity at the far ear is lower. The size of this intensity difference varies with frequency. The intensity difference is greater at high frequencies (about 20 dB at 10 kHz) and smaller at low frequencies (about 2 dB at 250 Hz).



Acoustic shadow
6,000 Hz
Spacing small
compared to object
200 Hz
Spacing large
compared to object
(a)
(b)
(c)
(d)



10
8
6
4
2
0
Interaural level difference (dB)
20
100
1,000
2,000
5,000
Frequency (Hz)
10°
θ
45°
90°

Binaural Cue Summary

Generally, differences in interaural time and intensity co-occur. These two binaural cues are effective in indicating position in the horizontal plane (azimuth).

They do not help with regard to the dimension of height (elevation) or distance.

Elevation and the Pinnae

The complex shape of the pinnae provides a monaural cue to location. Inserting ear plugs with different shapes that make the overall shape of the pinnae smoother reduces the ability to locate sounds in space.

The pinnae seems to alter the intensity of different frequencies and create a distinctive pattern of echoes. The precise changes depend upon the elevation and azimuth of the sound source.

When combined with the binaural cues, this gives the listener the ability to localize the direction (azimuth and elevation) of a sound source.

Cues to Distance of a Sound

Listeners can judge the distance of a sound. We know relatively little about this ability. There are at least four sources of information available to the listener:

1. Pressure - Sound pressure decreases with distance.

2. Frequency - High frequencies are absorbed by the atmosphere more than low frequencies.

3. Movement Parallax - Nearby sounds shift their apparent location faster when moving.

4. Reverberation - The ratio between direct sound and indirect sound decreases with distance. (reverberation increases)

Physiology and Localization

Before considering the response properties of single cells, the overall role of the auditory cortex should be described.

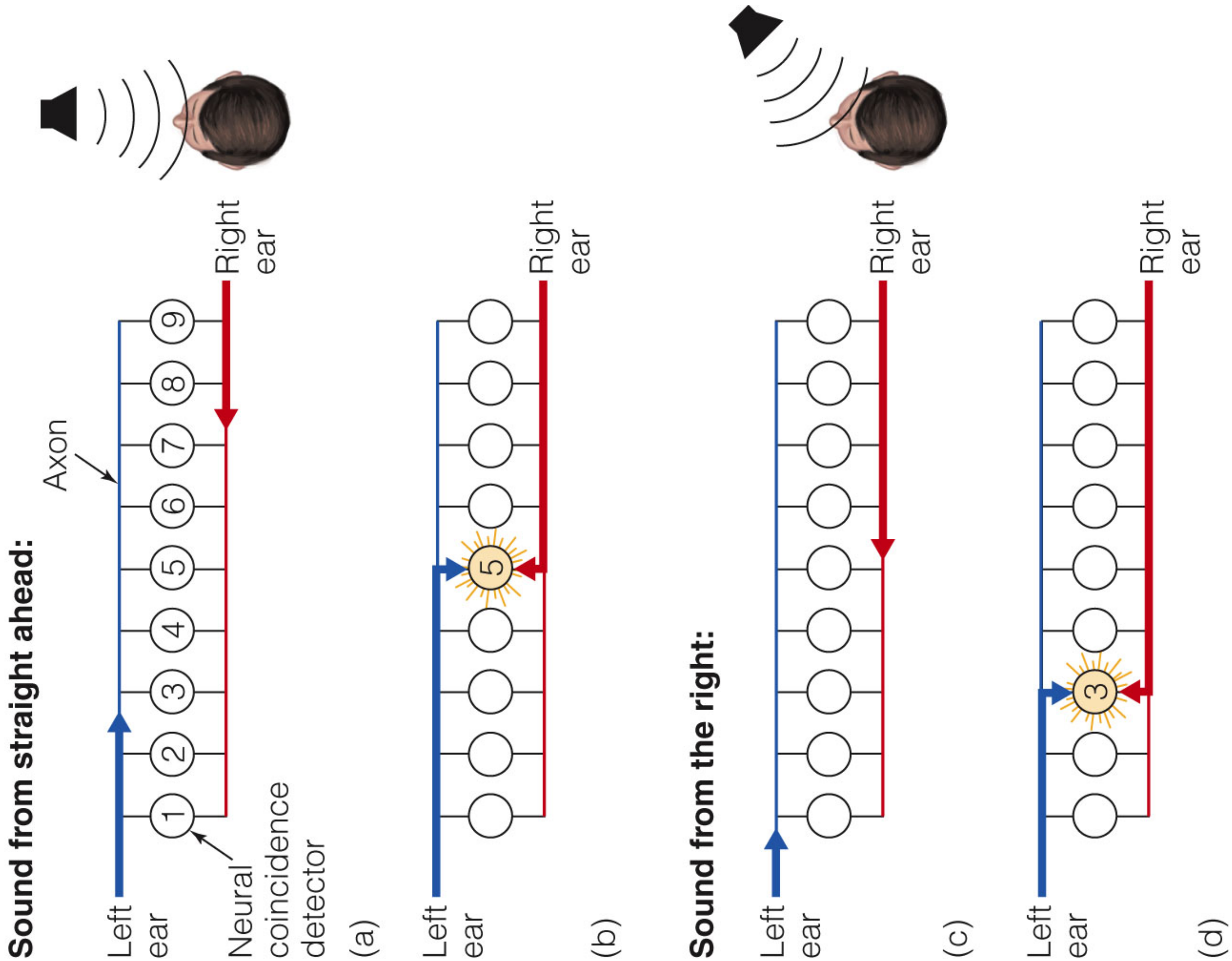
1. Removal of all or part of A1 degrades auditory localization. This implies that part of the coding by cells in A1 is for localization of sound.

2. The auditory cortex (A1) has a tonotopic map. If part of the cortex is lesioned, then the cat's ability to localize sound is impaired for the frequencies coded by that particular part of A1. Localization of other frequencies was unaffected.

Interaural Time Cells

In monkey A1, some cells respond to specific differences in when the sound arrives at the two ears.

Cells that respond to interaural time differences are also found in the ascending auditory system, starting at the superior olivary nucleus. These are the first cells to receive input from both ears.


Sound from straight ahead:
Axon
Left ear
Right ear
Neural coincidence detector
1
2
3
4
5
6
7
8
9
(a)
Left ear
Right ear
5
(b)
Sound from the right:
Left ear
Right ear
(c)
Left ear
Right ear
3
(d)

Azimuth and Motion Cells

In monkey A1, cells are found that:

1. Respond to sound from a specific point in space left to right (specific azimuth). The tuning of these cells is fairly broad (45 to 50 degrees). This indicates that the ability to locate a sound left to right (azimuth) is probably based on the pattern of response across a set of these cells (distributed coding).

2. Respond to sound whose location changes in a specific direction. That is, these cells respond to the motion of the sound source in a particular direction.

Panoramic Neurons

Some cells respond to sounds no matter what azimuth they originate from. However, the timing of the firing differs for different azimuth. For example, a cell may fire early and frequently to a sound directly behind and a bit later (relative to when the sound starts) and less frequently to a sound at the front.

The pattern of response across these types of cells would indicate the azimuth.

Echo Location in Humans

Humans can judge distance to an object via sound. While this ability may be more acute (practiced) in blind individuals, it can be demonstrated in blindfolded individuals.

Blindfolded listeners are positioned at different distances from an obstacle and asked to walk toward it until they perceive it (or collide). With practice, they learn to use the sound of footsteps to stop before collision.

Blind listeners have also been shown to be able to judge the distance between objects from the echo of sound.

Auditory Grouping (Gestalt Laws in Hearing)

1. Location
2. Similarity of Timbre
3. Similarity of Pitch
4. Temporal Proximity
5. Temporal Synchrony in Onset and Offset.
6. Good Continuation
7. Experience (Familiarity)

Location

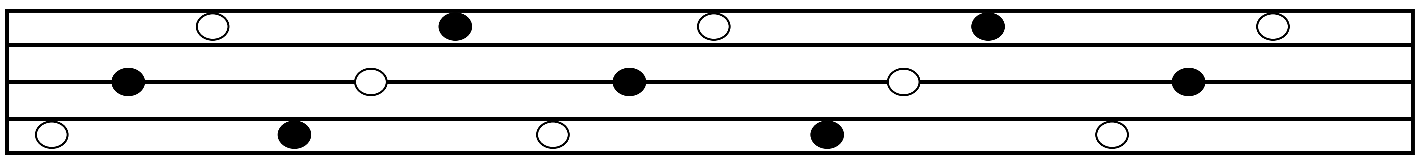
If sounds elements come from the same location in space, they will tend to be grouped together. If the sound source changes slowly, the elements will be grouped together (as if the sound source were moving).

This grouping process is based on our perceptual ability to localize sounds in the 3-dimensional world we live in.

We will return to this after discussing the other principles for perceptual grouping.

Similarity of Timbre

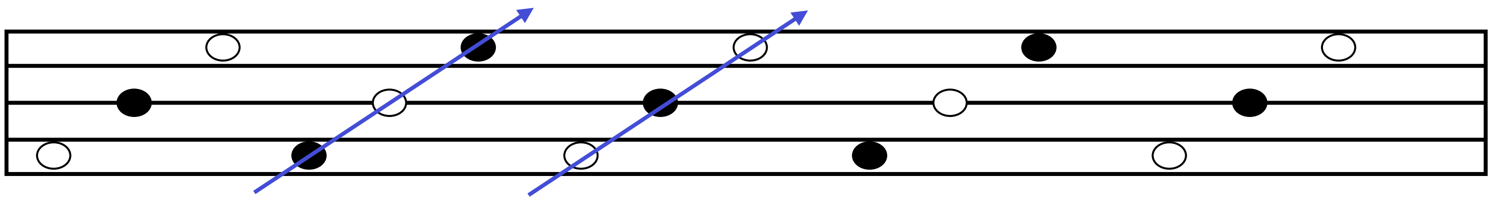
Sounds that have the same timbre often have the same source.



In this example, the black and white notes stand for notes played with different timbre (as if by different instruments such as a clarinet and a trumpet). They are played in the sequence shown. What listeners report depends upon the timing (played slowly or rapidly).

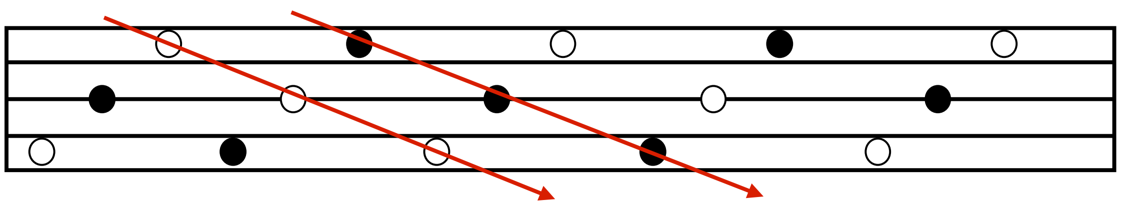
Similarity of Timbre - 2

(a)



If played slowly, listeners report an ascending sequence by different (alternating) instruments.

(b)



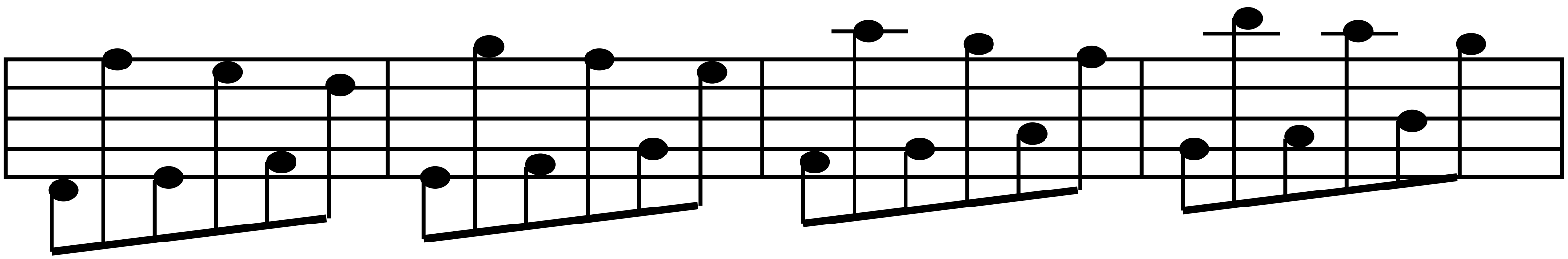
If played rapidly, listeners report two separate descending sequences, played by different instruments. This is an example of auditory stream segregation where grouping has produced two separate sequences.

Similarity of Pitch

In grouping by similarity of pitch, sounds that have a similar pitch are grouped together. In general, sounds from the same source would have the same or a similar pitch over time.

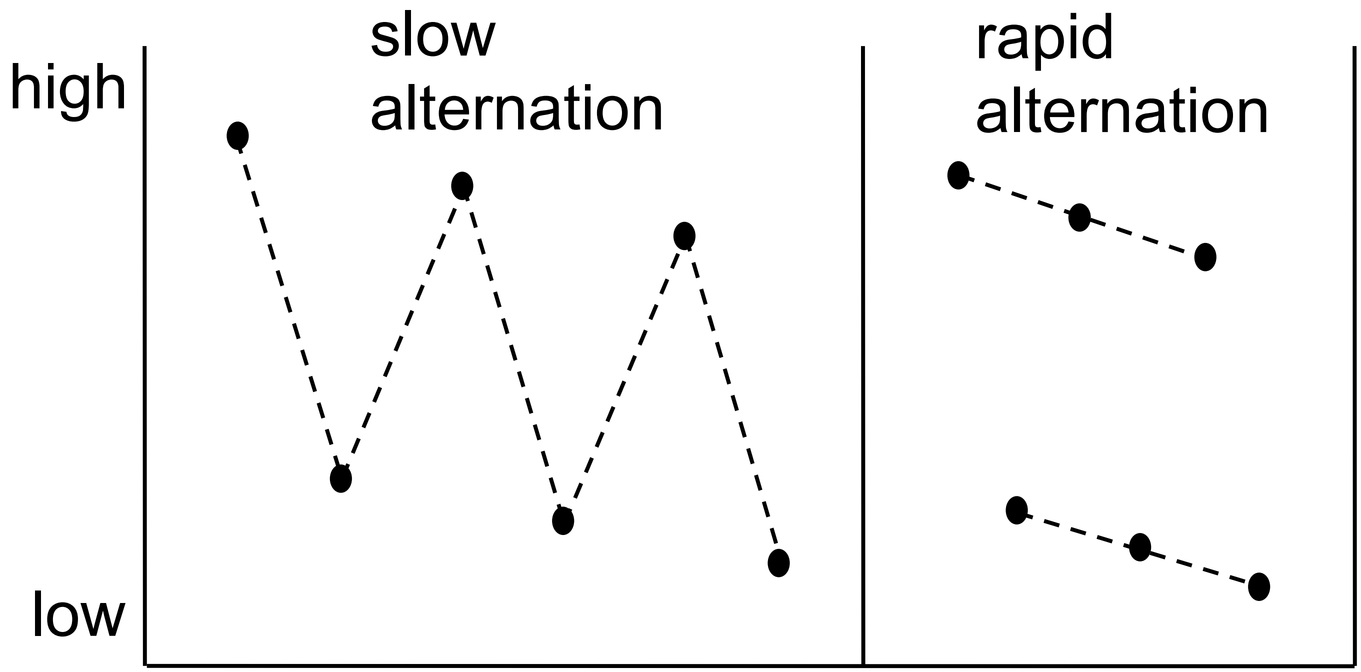
In music, this principle has been known for hundreds of years. If notes on a single instrument are rapidly alternated between high and low, listeners will perceive two separate melodies - one based on the high pitch notes and one based on the low pitched notes. This is called *implied polyphony* or a *compound melodic line*.

Similarity of Pitch - 2



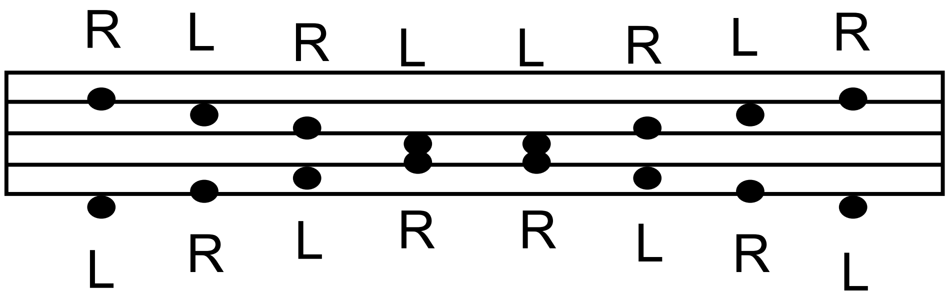
This four measure sequence, by J. S. Bach, sounds like two separate melodic lines when played rapidly. The high notes group together into a single sequence and the low notes group together into a separate sequence.

Similarity of Pitch - 3

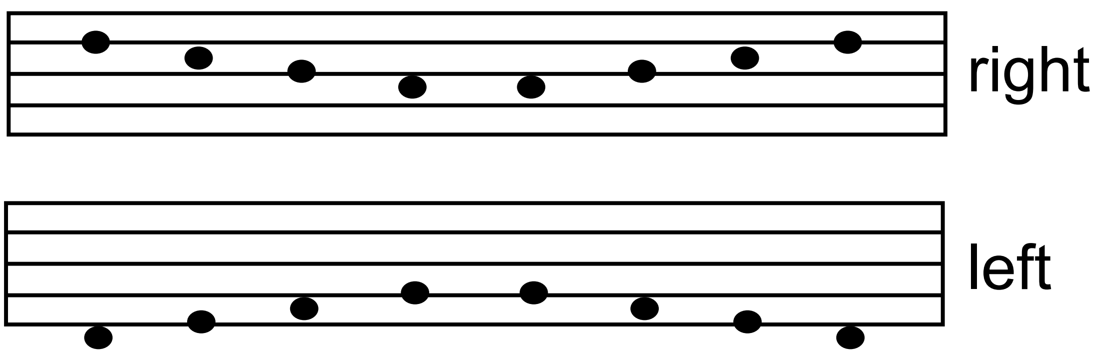


If alternating high and low tones are played slowly, listeners report hearing a alternating sequence (left). If the sequence is played rapidly, listeners report hearing two separate streams (descending, on right).

Similarity of Pitch - 4



Take the two sequences of notes here and play the notes alternately to the left and right ears.



Listeners report one smooth sequence in the right ear and another in the left.

Here, grouping by pitch has created the illusion of two smooth sequences, one in each ear.

Temporal Proximity

For sound elements to be grouped together, they must be close to one another in time. In the examples of grouping by pitch and timbre, stream segregation occurred when notes similar in pitch or timbre were close to one another in time.

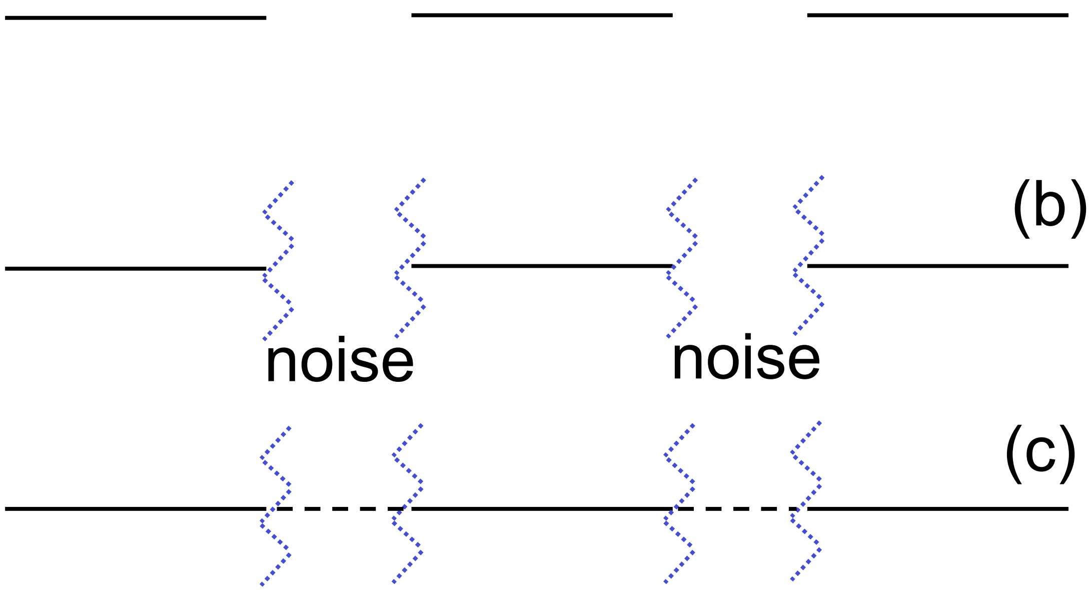
If well spaced, the elements will form separate groups, even if similar in pitch or timbre.

Temporal Synchrony in Onset and Offset

In general, sound elements from a single source start and stop at the same time.

Sound elements that start and stop at different times generally represent different sources.

Good Continuation



(a) Sequence of three tones is heard as three tones

(b) Fill silent gaps between tones with noise

(c) Listeners report a single, continuous tone that continues through the noise

Experience

The notes from “Three blind mice” and “Mary Has a Little Lamb” are interleaved. When played to listeners, it sounds like a meaningless sequence of notes. When told the names of the melodies, they were able to follow one melody at a time.

This ability to separate one sequence out of the “jumble” is based on experience or familiarity.

Grouping Summary

In the laboratory, we can manipulate the sound to test each of these principles separately.

In nature, these sources of information work together to guide auditory perceptual grouping.

The Ecological Approach

In our description thus far, we have emphasized an analytical approach to listening. We have emphasized pitch and timbre as qualities that a listener might focus on. This is quite appropriate for music.

In everyday listening, however, our impression is that we hear events. We hear air conditioners, squeaky chairs, footsteps. Listeners can be very good at this, such as distinguishing between male and female based on footsteps.

The Ecological Approach - 2

The key to the ecological approach is to understand the relationship between the event that produces the sound and the qualities present in the sound.

For example, when played a recording of a file drawer closing, it is often confused with the sound from a bowling alley. Listeners describe a rolling sound followed by an impact.

The next step is to determine the relation between the basic qualities of the sound and the percept. This is the focus of most current work on auditory perception.

Auditory Scene Perception - Summary

Perhaps the most complex sound that we readily perceive is speech. Here, like with everyday sounds, the percept is strongly related to the source of the sound.

Like a melody, the sounds of speech change rapidly over time. From this, we recover the words, the emotional content of the voice and the identity of the speaker.

In addition, the basic properties of perceptual grouping and localization in space are critical to our ability to understand spoken language, particularly when a conversation occurs against a background of other sound.

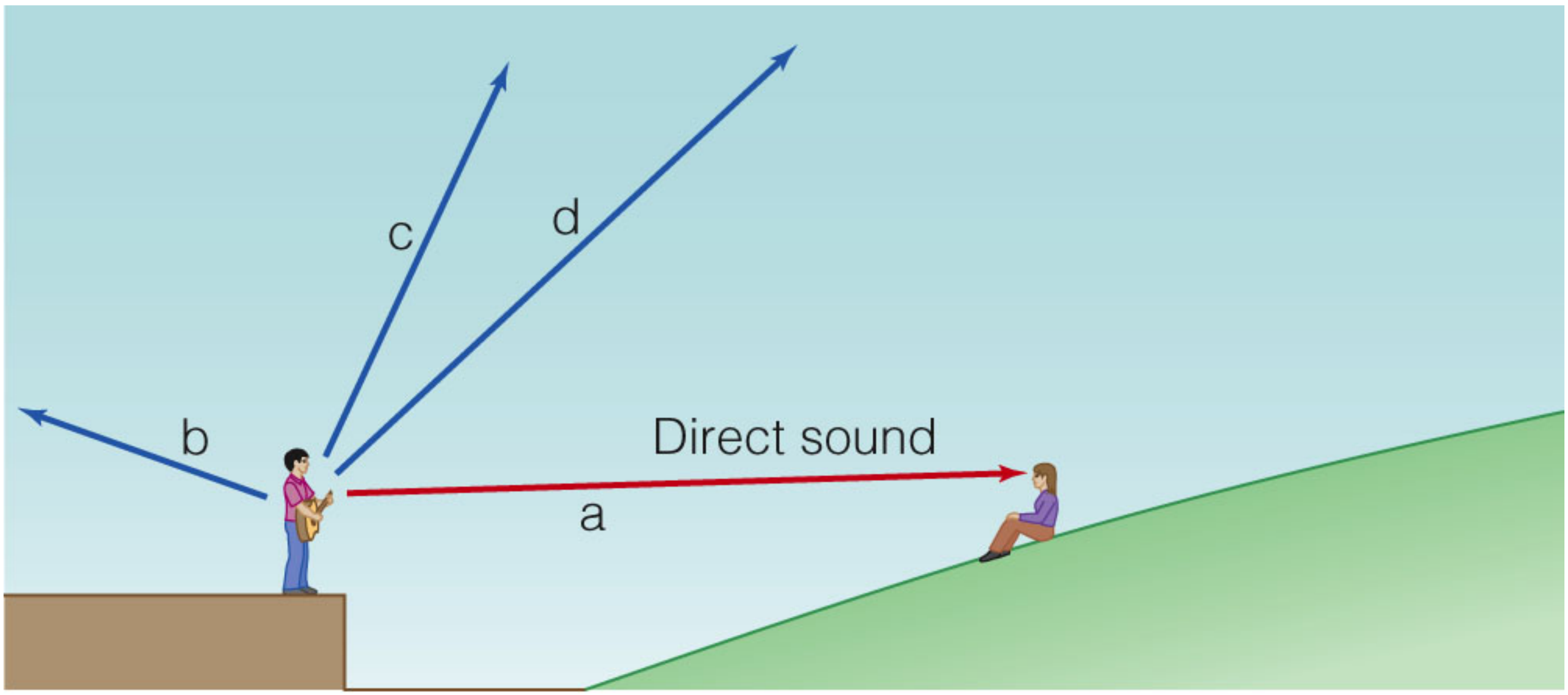
Room Acoustics

Finally, lets consider the influence of room acoustics on our experience of sound.

Here we want to distinguish between the direct path for sound to reach the observer and indirect paths where sound bounces off objects and walls before reaching the observer.

Direct Sound

In an open field, sitting next to someone play an instrument, almost all of the sound that reaches our ear is direct. There is little to reflect the sound.

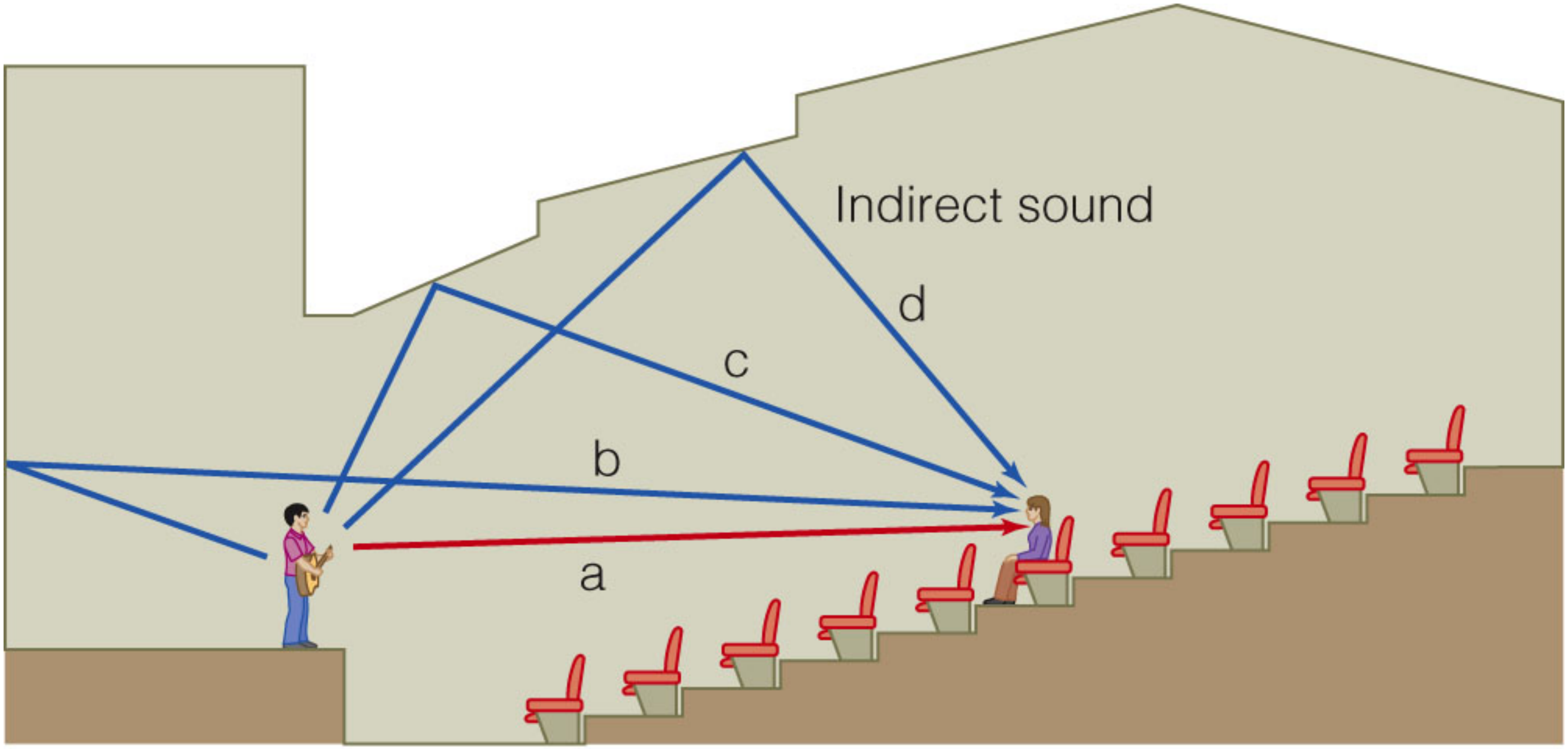


(a)

Indirect Sound

In a room, the sound reaching us is a combination of the direct and indirect, reflected off of the walls, ceiling, floor, etc. Surfaces can absorb or reflect sound. In addition, the intensity of high frequency sounds and low frequency sounds are affected differently by different surfaces.

The perceived quality of the sound is influenced by this combination and the time difference between when direct and indirect path sound arrives - the reverberation time.



(b)

Architectural Acoustics

Architectural acoustics is the art of designing a room to have certain reverberation, time between direct sound and the first reflection (intimacy), ratio of low to middle frequency intensity of reflected sound (bass ratio) and relative intensity of reflected sound (spaciousness factor).