More about Hearing

Josh McDermott
9.35
April 29, 2004

PITCH

Pitch is one of the fundamental dimensions of our auditory experience.

Low frequencies sound low; high frequencies sound high.

Image removed due to copyright considerations.

Please see: Sekuler, Robert and Blake, Randolph. Perception (Fourth edition). New York: McGraw-Hill, 2001.

Figure 10.11, p. 349.

The interesting thing is that complex tones also have pitch, even though they consist of many frequencies.

In fact, all instrument sounds are complex tones.

Image removed due to copyright considerations.

Please see: Sekuler, Robert and Blake, Randolph. Perception (Fourth edition). New York: McGraw-Hill, 2001.

Figure 10.12, p. 351.

Instrument sounds, and many other natural sounds, are harmonic.

This means the frequencies they contain are integer multiples of some low frequency called the fundamental frequency of the tone.

(Image removed due to copyright considerations.)

Where does the pitch of a complex tone come from?

It seems to be determined by the fundamental frequency.

When asked to choose a pure tone (sine wave) that has the same pitch as a complex tone, people choose the pure tone with the same frequency as the complex tone's fundamental.

Image removed due to copyright considerations.

Please see: Sekuler, Robert and Blake, Randolph. Perception (Fourth edition). New York: McGraw-Hill, 2001.

Figure 10.12, p. 351.

The guitar and sax have very different amplitude spectra, but have the same pitch when played such that they have the same fundamental frequency.

Image removed due to copyright considerations.

Please see: Sekuler, Robert and Blake, Randolph. Perception (Fourth edition). New York: McGraw-Hill, 2001.

Figure 10.13, p. 352.

Here's the amazing thing - you can remove the fundamental frequency in an experiment, and the resulting sound has the same pitch as the original!

Images removed due to copyright considerations.

Please see: Sekuler, Robert and Blake, Randolph. Perception (Fourth edition). New York: McGraw-Hill, 2001.

Figure 10.12, p. 351.

This is called the missing fundamental illusion. SHOW DEMO.

The fundamental seems to determine pitch even when it's not present.

In the missing fundamental illusion, the brain seems to infer what the fundamental must be from the frequencies of the harmonics.

The fundamental is the frequency the harmonics are all multiples of.

Why does the brain do such a complicated computation? Why bother to extract pitch?

Animals (birds, cats, and monkeys) all perceive the missing fundamental illusion, so it's not a music-specific adaptation.

Pitch is probably important because it reveals physical properties of the object generating the sound.

- -size
- -gender of speaker
- -used for intonation in speech

TIMBRE

Timbre is the other dimension, besides loudness and pitch, along which our perception of sounds can vary.

It is not just one dimension, and we don't know how many dimensions there are.

One factor that determines timbre is the relative amplitude of overtones.

Images removed due to copyright considerations.

Please see: Sekuler, Robert and Blake, Randolph. Perception (Fourth edition). New York: McGraw-Hill, 2001.

Figure 10.13, p. 352.

However, the high and low tones from a musical instrument generally have different amplitude spectra, but they still sound like the same instrument.

E.g. a piano - low notes have more energy in the high overtones, whereas high notes have more energy in the fundamental.

If you take one note played on an instrument and simply scale all the frequencies up or down , it doesn't sound like the same instrument.

Bassoon Demo.

This effect may be learned from exposure to instruments, but we don't know for sure.

Another stimulus property that influences timbre is the envelope of a sound - the way the intensity changes over time.

This is different for something plucked vs. struck, blown etc.

The importance of the envelope can be demonstrated by playing things backwards, as the amplitude spectrum is unchanged.

Piano Demo.

Small co-modulations of all the harmonics in a harmonic stack makes it sound like a voice.

(Image removed due to copyright considerations.)

So, individual sounds are typically described in terms of their loudness, pitch, and timbre.

All of these are perceptual variables that tell us things about the physical events in the world that produced the sounds, which is the point of hearing them in the first place.

Timbre is the richest, least defined, and least understood of these dimensions.

But what happens when more than one sound occurs at once?

AUDITORY SCENE ANALYSIS

The Problem:
Many acoustic events happen in the world at once, but the ear only receives one sound wave.

This sound wave is the sum of those that would have been produced by the individual events.

Somehow, the brain must distinguish multiple events from the single signal it receives.

We are remarkably good at doing this. How do we do it?

The problem the auditory system solves is akin to asking someone to solve this equation:

$$1532 = x + y$$

Obviously there are many combinations of x and y that sum to 1532.

But for every sound wave that the ear receives, there are many possible sets of sound sources that could have generated it.

Example: two people talking. How do you know which frequencies belong to which voice?

The brain has to make its best guess as to how many sources there are, and what they are, based on what it knows about the world.

Sounds in the world are often harmonic. The brain seems to assume that frequencies that are harmonically related belong to the same acoustic event.

When one frequency is not harmonically related to a bunch of others, it segregates perceptually.

(Image removed due to copyright considerations.)

Sudden changes in intensity are cues to new events.

(Image removed due to copyright considerations.)

Sudden onsets of new frequencies are interpreted as new, separate events. Frequencies that were present before are heard as continuing. (Bregman's "old-plus-new" heuristic).

(Image removed due to copyright considerations.)

This "old+new" heuristic can even override the tendency to hear harmonically related frequencies as part of the same thing.

(Image removed due to copyright considerations.)

Sound localization cues also contribute to the grouping and segmentation of sounds.

A pure tone in one ear masked by noise in the same ear can paradoxically be "unmasked" by the addition of noise to the other ear. Only works if noise in two ears is the same.

(Image removed due to copyright considerations.)

This binaural masking-level difference can be large (thresholds can differ by as much as 20 dB), and is attributed to segmentation processes related to the perceived location of the two sounds.

Suppose the tone and noise were present in both ears. Could anything be done to reduce the masking in this case?

(Image removed due to copyright considerations.)

In vision, occlusion often renders parts of objects invisible. The shape of the occluded object can nonetheless often be inferred, and is represented and perceived via *amodal completion*.

(Image removed due to copyright considerations.)

An analogous process of filling in occurs in audition…

Amodal Completion of Sounds

When noise occurs during a gap in a sound, we hear the sound continue “behind” the noise. The brain is inferring that the sound’s absence is best explained by masking, and so it is filled in.

(Image removed due to copyright considerations.)

This only works if the noise contains frequencies close to those of the sound, so it is probably related to the old + new heuristic.

In vision, amodally completed shapes can aid recognition.

The same is true in hearing. Speech Demo.

(Image removed due to copyright considerations.)

Auditory amodal completion also occurs in monkeys.

(Image removed due to copyright considerations.)

(Image removed due to copyright considerations.)

So far we've been talking about how the brain decides to assign sound energy received at the ear to different sources in the world.

Several heuristics (harmonicity, common onset, temporal continuity, location) are used to decide how to segment the sound energy received at a given point in time.

The brain also has mechanisms that group sound energy across time. This is called streaming.

Alternating high and low tones can be perceived as part of a single process (a "galloping" rhythm is heard), or as two separate streams.

If the difference in frequency is too great or the time between tones too small, two separate streams are heard. (Demo 3).

(Image removed due to copyright considerations.)

Not surprisingly, sounds similar in timbre tend to be heard as part of the same stream.

(Image removed due to copyright considerations.)

When a sequence of sounds is parsed into two streams, the sounds are represented in a different way; this prevents you from judging the time relations between tones of the two streams.

(Image removed due to copyright considerations.)

So some of the factors known to influence streaming are pitch differences, speed, and timbre differences.

Streaming is poorly understood but probably a key part of speech recognition, among other things.

We have no idea how different streams might be represented in the brain.