CHAPTER 22

AUDITORY SCENE ANALYSIS

22.1. AUDITORY STREAM SEGREGATION

In audition, scene analysis results in the perception of sounds from different sources, their significance, and their relationships.

22.1.1. The "cocktail party effect". In the presence of multiple "competing"sounds that originate simultaneously from different sources, we are able to isolate one stream of sound (from one source. e.g., one speaker) and follow it.

22.1.2. Segregation of sounds by source. Streaming, or stream segregation, is the process by which a single complex waveform is separated into different "streams", or sequences, that belong together. Gestalt theory can be applied to hearing as well as vision.

a. Law of similarity: Sounds that are similar belong together.

Example: Man's voice vs woman's.

b. Law of proximity: Sounds that are close to one another in <u>frequency</u> or <u>time</u> seem to belong together.

Example: low pitched melody line vs. high pitched melody line.

Example: Words that follow one another rapidly in time form a sentence.

c. Law of common fate: Sound components that change together are perceived as belonging together.

Example: The collection of harmonics that make up the sound of one person's voice are perceived as coming from one source.

d. Good continuation, or filling in: Under the right conditions, gaps in a stream of sounds are not perceived. Good continuation can also arise through patterns of expectation.

Example: We learn to anticipate what is most likely to come next in music or speech and can perceive patterns even when notes or words are obscured by noise.

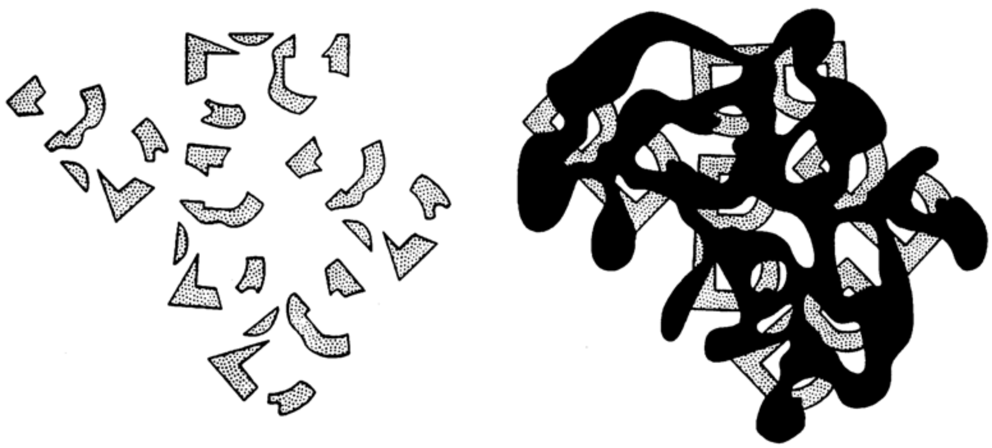

Figure 22-1. There is a visual illusion in which object segments by themselves are meaningless (left), but take on meaningful shapes when covered with another pattern (right).

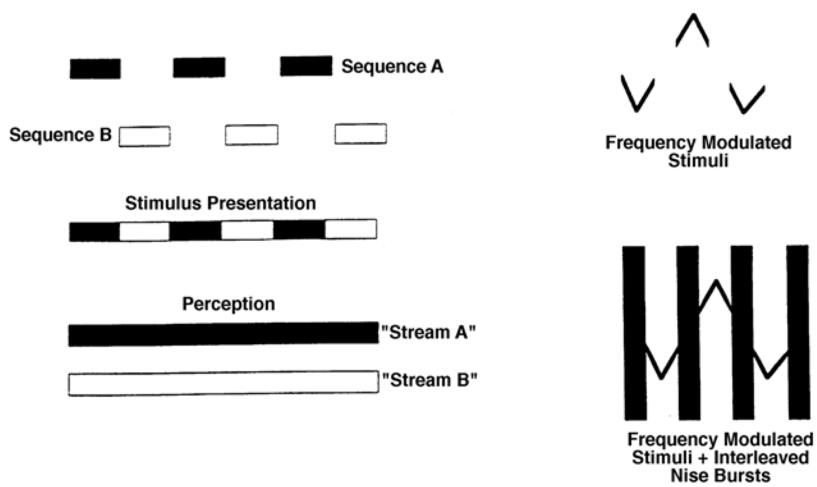

Figure 22-2. The same phenomenon can happen with sounds. For example, interleaved short tone bursts are perceived as two continuous tones. Frequency sweeps are perceived as separate when presented alone, but as continuous when interleaved with noise bursts.

22.2. AMPLITUDE MODULATION IN THE AUDITORY SYSTEM.

The human auditory system is capable of resolving amplitude modulation (change in the intensity of a sound) up to a rate of more than 600 cycles per second, an order of magnitude higher than the visual system.

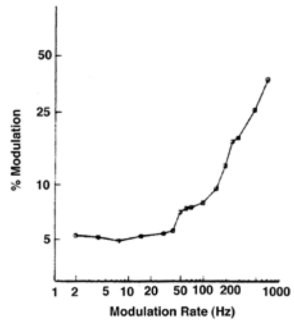

Figure 22-3. Amount of change in amplitude (% modulation) of a broad band noise needed in order for a listener to detect the fact that it is amplitude modulated. Note that the Y-axis on this graph is "backwards" - the lower the number (% modulation), the higher it is on the axis. This is done in order to clearly convey the idea that at higher modulation rates, greater contrast is needed in order for the modulation to be detectable.

The higher the modulation rate, the greater the amount of change in amplitude needed in order for a listener to detect the modulation.

22.3. AUDITORY MASKING

Any time two stimuli occur near one another in space and time, they can interact. When they are close enough in time, the perception of the two stimuli fuses into one, as in flicker fusion. One form of interaction that has been studied extensively in psychophysical experiments is called *masking*.

In *simultaneous masking,* the presence of one stimulus interferes with the perception of another one that occurs at the same time.

In *forward masking*, a stimulus presented first affects the perception of a subsequent stimulus by raising the threshold for perceiving it (i.e., decreasing the sensitivity to the second stimulus).

In *backward masking*, a stimulus presented second affects the perception of the previous stimulus by raising the threshold for perceiving it (i.e., decreasing the sensitivity to the first stimulus).

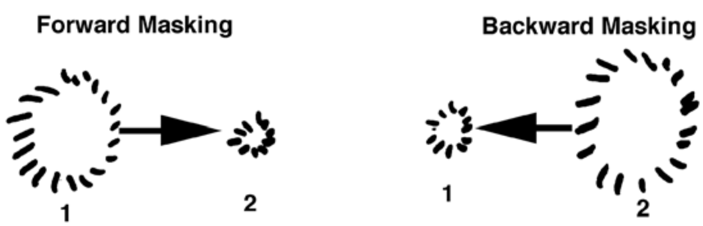

Figure 22-4. Forward masking (left) and backward masking (right). In forward masking, stimulus 1 interferes with the perception of stimulus 2. In backward masking, stimulus 2 interferes with the perception of stimulus 1.

Masking experiments are useful because they allow the experimenter to measure the rate at which various classes of information are extracted from a stimulus, to measure processing time in different populations of subjects, and to determine the extent to which different classes of information engage common neural processes.

22.3.1. Simultaneous masking.

When two sounds are present at the same time (simultaneously), the one that is higher in frequency is more likely to be masked by the one that is lower in frequency than vice versa.

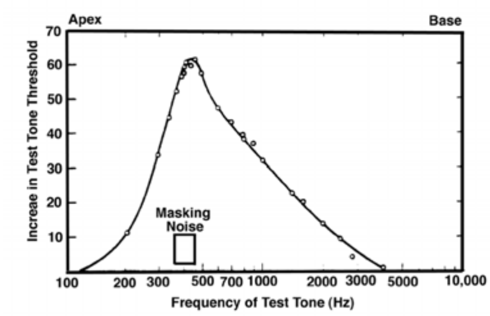

Figure 22-5. Masking of a pure tone of different frequencies by a narrowband noise centered around 400 Hz. The curve plots the increase in threshold for hearing the tone as a function of the tone's frequency.

22.3.2. Forward and backward auditory masking.

Forward masking occurs when one sound interferes with the perception of a subsequent sound.

Backward masking occurs when one sound interferes with the perception of a prior sound.

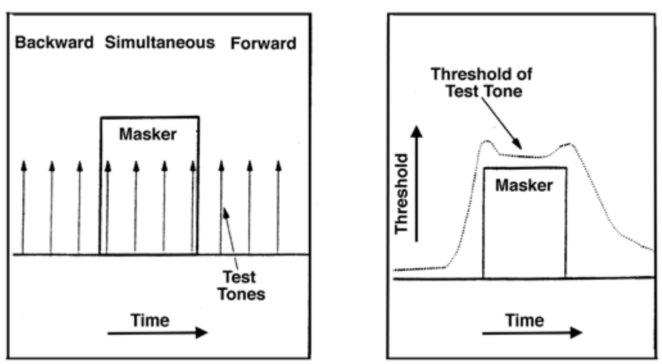

Figure 22-6. The threshold for hearing a sound depends on the temporal relation of that sound (probe signal) to the masker. The diagram on the left indicates the times at which the probe signal is presented (arrows) relative to the masker. In the diagram on the right, the dotted line indicates the threshold for hearing the probe tone. Note that the threshold is elevated when the probe tone occurs just before or just after the onset or offset of the masker.

22.3.3. The precedence effect.

Sounds that arrive within about 70 ms of one another are localized to the site from which the *first* sound originated.

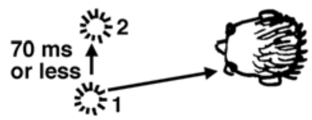

Figure 22-7. The precedence effect. One sound is perceived, and it is localized to the position of the first source.