

Auditory Scene Analysis

Authored by
mohammad looti

June 6, 2026

RECOMMENDED CITATION

mohammad looti (2026). *Auditory Scene Analysis*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=38125>

In psychophysics, auditory scene analysis (ASA) is a proposed model for the basis of auditory perception. This is understood as the process by which the human auditory system organizes sound into perceptually meaningful elements. The term was coined by psychologist Albert Bregman. The related concept in machine perception is computational auditory scene analysis (CASA), which is closely related to source separation and blind signal separation.

The three key aspects of Bregman's ASA model are: segmentation, integration, and segregation.

Background

Sound reaches the ear and the eardrum vibrates as a whole. This signal has to be analyzed (in some way). The model proposes that sounds will either be heard as "integrated" (heard as a whole - much like harmony in music), or "segregated" into individual components (which leads to counterpoint). For example, a bell can be heard as a 'single' sound (integrated), or some people are able to hear the individual components - they are able to segregate the sound. This can be done with chords where it can be heard as a 'color', or as the individual notes.

In many circumstances the segregated elements can be linked together in time, producing an auditory stream. This ability of auditory streaming can be demonstrated by the so-called cocktail party effect. Up to a point, with a number of voices speaking at the same time or with background sounds, one is able to follow a particular voice even though other voices and background sounds are present. In this example, the ear is segregating this voice from other sounds (which are integrated), and the mind "streams" these segregated sounds into an auditory stream. This is a skill which is highly developed by musicians, notably conductors who are able to listen to one, two, three or more instruments at the same time (segregating them), and following each as an independent line through auditory streaming. Organists also develop this skill having to stream up to five or more voices at a time.

Natural sounds, such as the human voice, musical instruments, or cars passing in the street, are made up of many frequencies, which contribute to the perceived quality (or timbre) of the sounds. When two or more natural sounds occur at once, all the components of the simultaneously active sounds are received at the same time, or overlapped in time, by the ears of listeners. This presents their auditory systems with a problem: Which parts of the sound should be grouped together and treated as parts of the same source or object? Grouping them incorrectly can cause the listener to hear non-existent sounds built from the wrong combinations of the original components.

Grouping and Streams

A number of grouping principles appear to underlie ASA, many of which are related to principles of perceptual organization discovered by the school of Gestalt psychology. These can be broadly

categorized into sequential grouping cues (those that operate across time - segregated) and simultaneous grouping cues (those that operate across frequency - integrated). In addition, schemas (learned patterns) play an important role.

Errors in simultaneous grouping can lead to the blending of sounds that should be heard as separate, the blended sounds having different perceived qualities (such as pitch or timbre) than any of the actually received sounds.

Errors in sequential grouping can lead, for example, to hearing a word created out of syllables originating from two different voices. The job of ASA is to group incoming sensory information to form an accurate mental representation of the individual sounds.

When sounds are grouped by the auditory system into a perceived sequence, distinct from other co-occurring sequences, each of these perceived sequences is called an "auditory stream". Normally, a stream corresponds to a distinct environmental sound pattern that persists over time, such as a person talking, a piano playing, or a dog barking, but perceptual errors and illusions are possible under unusual circumstances. One example of this is the laboratory phenomenon of streaming, also called "stream segregation". If two sounds, A and B, are rapidly alternated in time, after a few seconds the perception may seem to "split" so that the listener hears two rather than one stream of sound, each stream corresponding to the repetitions of one of the two sounds, for example, A-A-A-A-, etc. accompanied by B-B-B-B-, etc. The tendency towards segregation into separate streams is favored by differences in the acoustical properties of sounds A and B. Among the differences that favor segregation are those of frequency (for pure tones), fundamental frequency (for rich tones), frequency composition, spatial position, and speed of the sequence (faster sequences segregate more readily). An interactive web page illustrating this streaming and the importance of frequency separation and speed can be found here.

Experimental Basis

Many experiments have studied the segregation of more complex patterns of sound, such as a sequence of high notes of different pitches, interleaved with low ones. In such sequences, the segregation of co-occurring sounds into distinct streams has a profound effect on the way they are heard. Perception of a melody is formed more easily if all its notes fall in the same auditory stream. We tend to hear the rhythms among notes that are in the same stream, excluding those that are in other streams. Judgments of timing are more precise between notes in the same stream than between notes in separate streams. Even perceived spatial location and perceived loudness can be affected by sequential grouping.

While the initial research on this topic was done on human adults, recent studies have shown that some ASA capabilities are present in newborn infants, showing that they are built-in, rather than

learned through experience. Other research has shown that non-human animals also display ASA. Currently, scientists are studying the activity of neurons in the auditory regions of the cerebral cortex to discover the mechanisms underlying ASA.

ARABPSYCHOLOGY.COM