Auditory Scene Analysis (Auditory Pattern and Object Perception)

Tai-Shih Chi (冀泰石)
Department of Communication Engineering
National Chiao Tung University
Apr. 24, 2008
Introduction to Auditory Scene Analysis

- Sound source: a physical entity which gives rise to acoustic pressure waves.
- A stream: the percept of a group of simultaneous and/or successive sound elements as a coherent whole, appearing to emanate from a single source.
- Auditory Scene Analysis: the process of assigning (grouping) components emanating from different sources to different streams, a.k.a. “perceptual grouping”, “parsing”, “streaming”.
- Simultaneous / sequential grouping

FIGURE 8.2  The waveforms of four individual sounds (top four traces) and the waveform resulting from adding those sounds together.
Information used to separate auditory objects

- Fundamental frequency ($F_0$): people have no problems in distinguishing two musical instruments or two vowel sounds with different $F_0$, even though the harmonics may be interleaved or may even coincide. Ref: Broadbent and Ladefoged (1957); Scheffers (1983); Assmann and Summerfield (1987); McKeown and Petterson (1995).

- Harmonicity or spectral regularity (Roberts and Brunstrom (1998, 2001)? If a group of components form a regular spectral pattern, they tend to be heard as fused, while if a single component does not “fit” the pattern, it is heard to “pop out”. For example, a single mistuned harmonic in a harmonic complex tone or the middle component of the sequence [650, 850, 1050, 1250, 1450] shifted to 1008 Hz.
Information used to separate auditory objects (cont.)

- Onset disparities: it has been shown that a tone that stops or starts at a different time from a vowel is less likely to be heard as part of that vowel than if it is simultaneous with it (Darwin and Sutherland, 1984). For example, increasing the level of a single harmonic can produce a significant change in the quality of a vowel (what changed?). However, if that incremented harmonic starts before the vowel, the quality change is markedly reduced.

FIGURE 8.3 Schematic illustration of the stimuli used by Rasch (1978). Both the signal and the masker were periodic complex tones, with the signal having the higher fundamental frequency. When the signal and masker were gated on and off synchronously (panel a), the threshold for the signal was relatively high. When the signal started slightly before the masker (panel b), the threshold was markedly reduced. When the signal was turned off as soon as the masker was turned on (panel c), the signal was perceived as continuing through the masker, and the threshold was the same as when the signal did continue through the masker.
Correlated changes in amplitude or frequency (AM/FM):

- it seems quite clear that modulation of the components of one sound in amplitude or in frequency can aid the perceptual segregation of that sound from an unchanged background (McAdams, 1982, 1989; Summerfield and Culling, 1992; Moore and Bacon, 1993).

- Sounds with coherent amplitude changes tend to fuse perceptually, whereas sounds with incoherent changes tend to segregate (Bregman et al., 1985; Hall and Grose, 1990; Moore and Shailer, 1991; Moore et al., 1993).

- There is at present no clear evidence that the coherence of frequency modulation influences perceptual grouping when both sounds are modulated.

Sound location (see notes on “Space Perception”).
General principles of perceptual organization

- **Similarity**: elements will be grouped if they are similar. In hearing, similarity usually implies closeness of *timbre, pitch, loudness*, or subjective *location*.
  - For pure tones, frequency is the most important factor governing similarity.
  - For complex tones, differences in timbre produced by spectral differences seem to be the most important factor.

- **Continuation**: a smooth change in frequency, intensity, location or spectrum indicates a change within a single source, whereas a sudden change indicates that a new source has been activated.

- **Common Fate**: different frequency components arising from a single sound source usually vary in a highly coherent way. They tend to start and finish together (onset/offset), change in intensity (AM) and frequency (FM) together.
  - Principle of common fate: if two or more components in a complex sound undergo the same kinds of changes at the same time, then they are grouped and perceived as part of the same source.
General principles of perceptual organization (cont.)

- **Exclusive allocation (belongingness):** a single component in a sound can only be assigned to one source at a time.

- **Closure:** the masked sound tends to be perceived as continuous when the masking sound is present and no direct sensory evidence to indicate the masked sound has been interrupted. ⇒ “continuity phenomenon”.

![Diagram](image)

**Figure 1.15**: Tonal glides of the type used by Dannenbring (1976). Left: the stimulus with gaps. Right: the stimulus when the gaps are filled with noise.
Conclusions

- The auditory system is able to parse the acoustic input, so that the components deriving from each source are grouped together and form part of a single perceptual stream.

- The identification of a particular sound source depends on the recognition of its timbre (distribution of energy over frequency). The time structure of sounds can also have an influence on their timbre; onset transients and the temporal envelope are particularly important.

- Cues used for streaming: F₀, onset disparities, AM, FM, and sound location. They generally provide an excellent basis for the parsing of the acoustic input when used together.

- General principles to govern the perceptual organization of the auditory world: (see previous two slides)
Principle of Exclusive Allocation

Rubin’s vase-faces

“Perception is not determined simply by the stimulus patterns; rather it is a dynamic searching for the best interpretation of the available data.”

--- *Eye and Brain*, Richard Gregory, 1966, Oxford University Press
References
