

---

---

# Sequential Organization from an Ecological Perspective

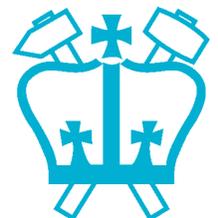
Dan Ellis

Laboratory for Recognition and Organization of Speech and Audio  
Dept. Electrical Eng., Columbia Univ., NY USA

dpwe@ee.columbia.edu

<http://labrosa.ee.columbia.edu/>

1. What is Sequential Organization / Streaming?
2. Why does Streaming Exist?
3. What are the Computational Implications?

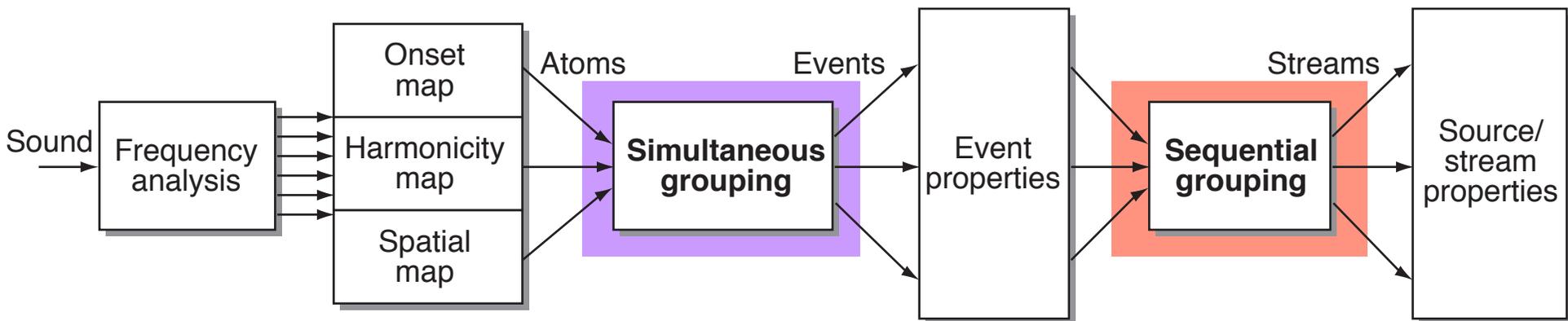


# I. Auditory Scene Analysis

Bregman'90

Darwin & Carlyon'95

- How do people analyze sound mixtures?
  - break mixture into **elements** (time-freq atoms)
  - elements are **grouped** in to sources using **cues**
  - sources have aggregate **attributes**
- **Grouping rules**
  - **cues**: common onset/modulation, harmonicity, ...

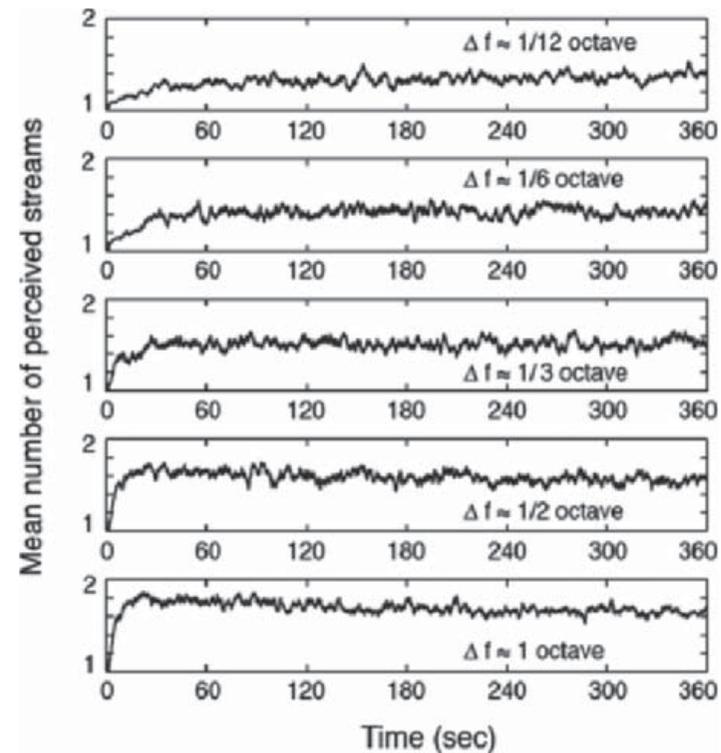
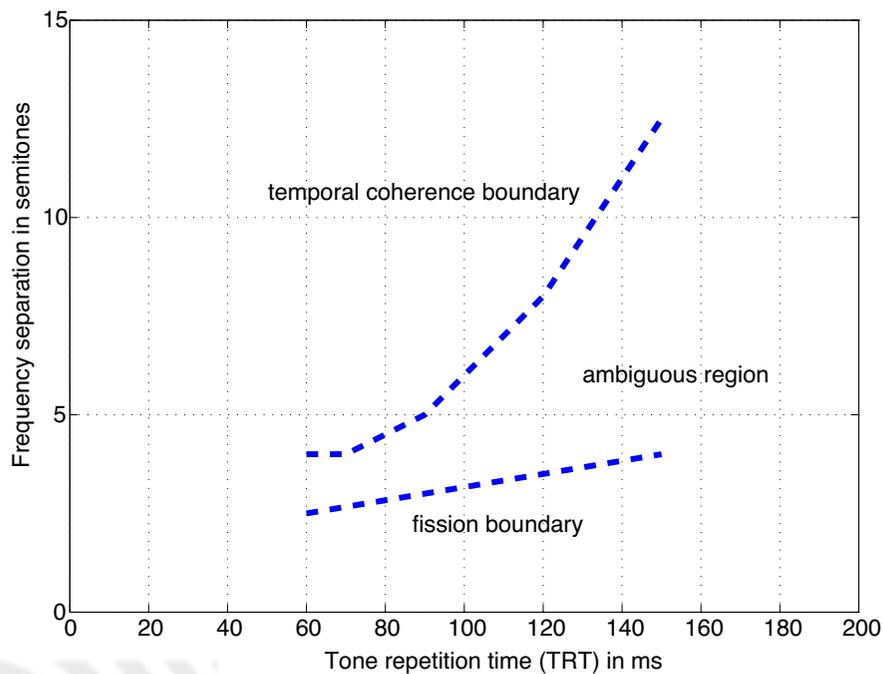
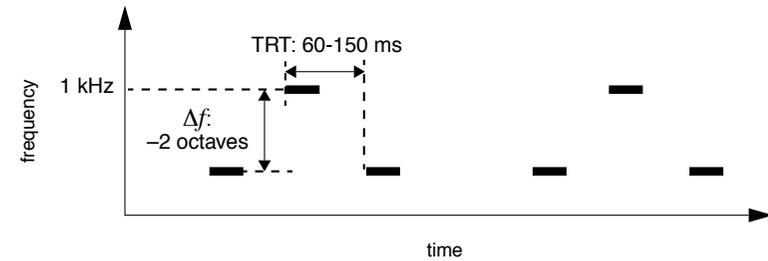


(after Darwin 1996)

# Auditory Streaming

Miller & Heise '50  
Bregman & Campbell '71  
van Noorden '75

- “Discovered” by musicians, beloved by psychologists...
- Ambiguity, buildup



Kashino et al. '07



# Relevance

- Does two-tone streaming tell us about the real world?

54 Chapter 2

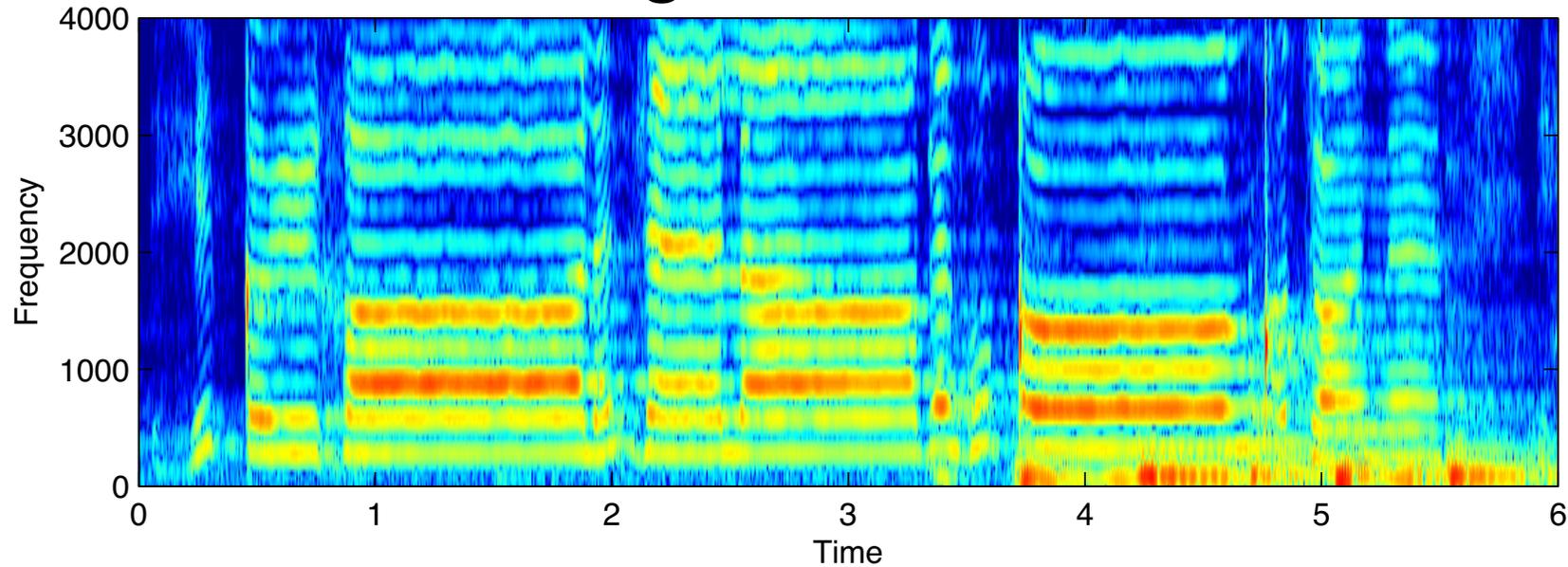
The method of repeating cycles has, however, one drawback: despite the continuous wiping out of briefly persisting sensory memories, some knowledge about the sequence gradually accumulates and it becomes predictable. Therefore, in all experiments in which repeating loops of sound are used, the observed effects that are observed may be influenced by this predictability as well as by the factors that are under study (this was pointed out by Jones, Kidd, and Wetzel).<sup>23</sup>

As a consequence, while the repeating loop method has important advantages, any results obtained by this method should be verified using less predictable patterns.

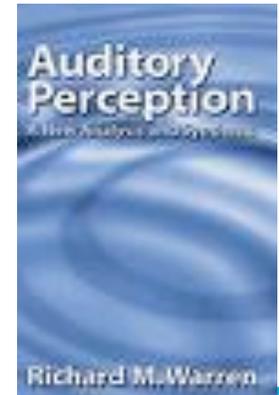
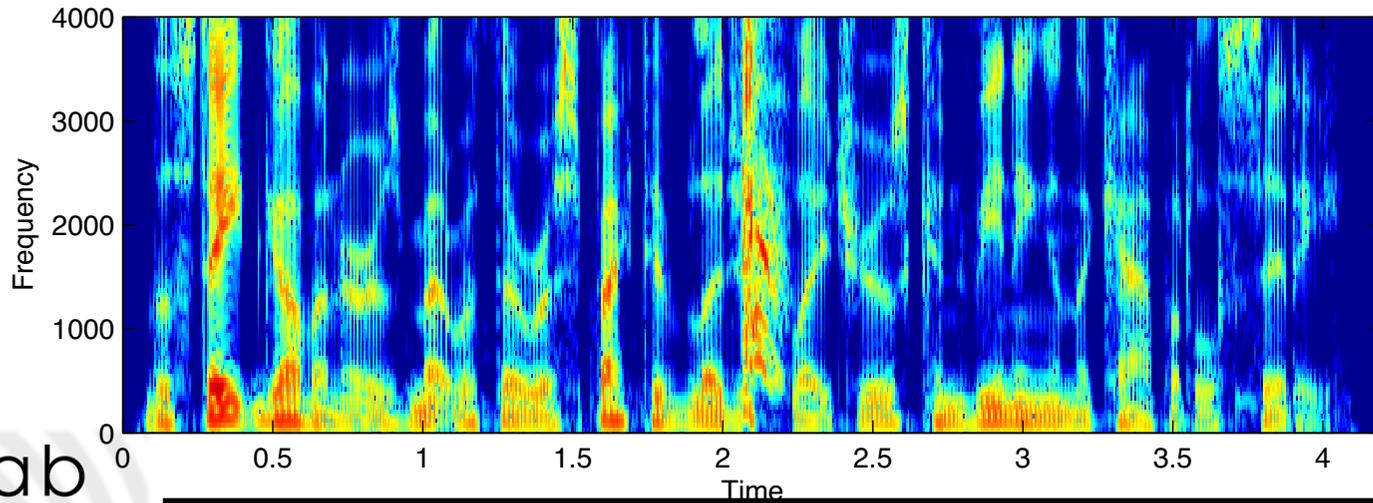
*Bregman '90*

# Ecological Streaming

- Streaming in the real world



*Miriam Makeba*



# Speech Streaming

Brungart et al. '02

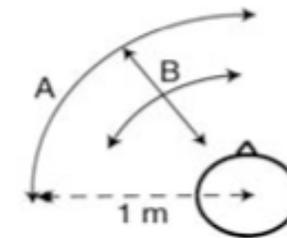
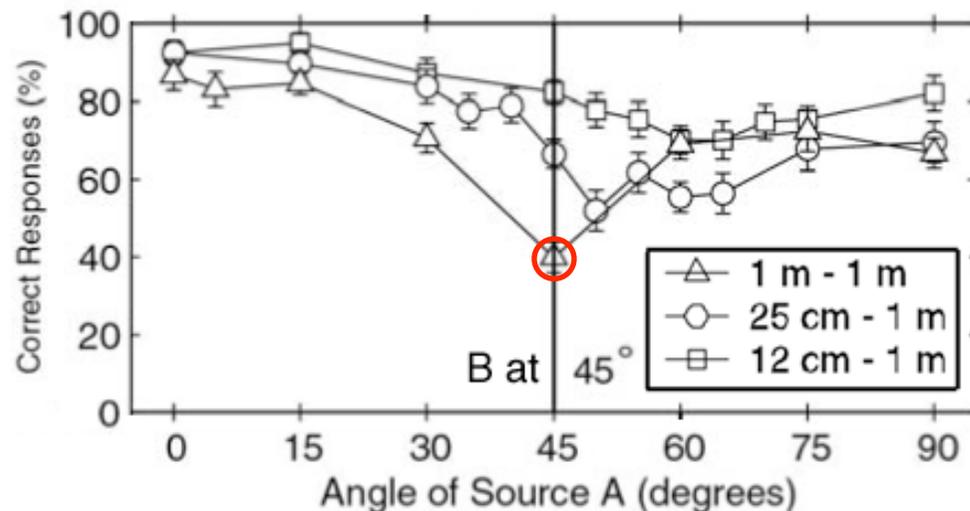
- **Task: Coordinate Response Measure**

- “Ready Baron go to green eight now”
- 256 variants, 16 speakers
- correct = color and number for “Baron”



crm-11737+16515.wav

- **Accuracy as a function of spatial separation:**



- A, B same speaker

- Range effect

## 2. Why Does Streaming Exist?

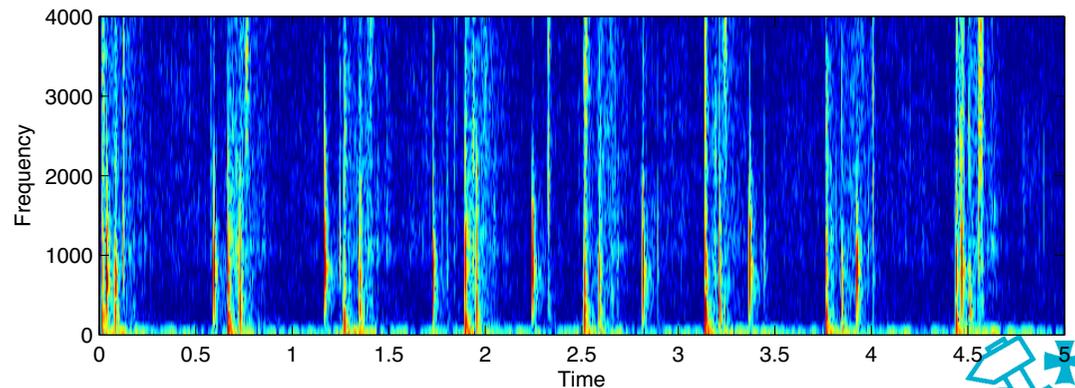
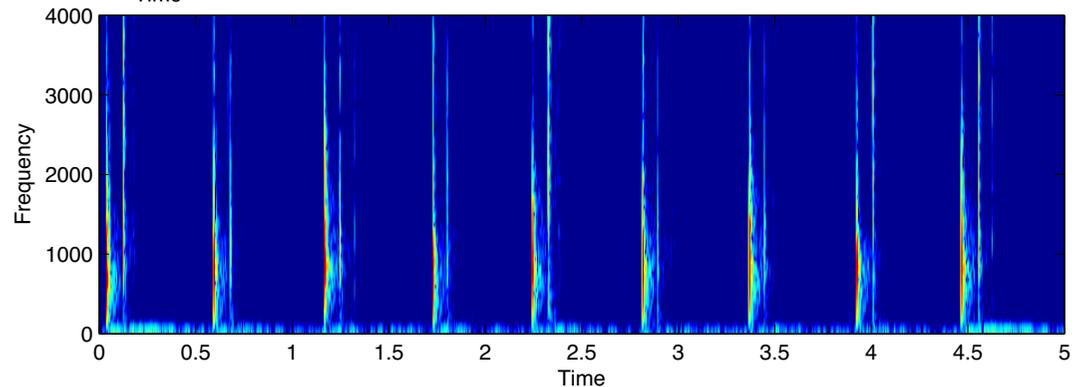
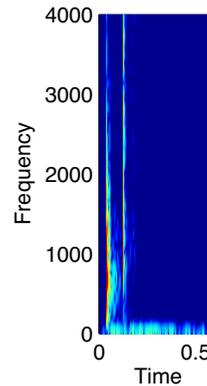
- The effect of streaming
  - “fission” of auditory percept into separate streams
  - **interferes** with judgments **between** streams
- But for perception, **context** is critical
  - to understand a sound event’s **meaning**, you need to know what comes before and after
  - .. and not be **confused** by random co-occurrences
- Need to **integrate** disparate evidence
  - streams as the common **hook** for glimpses
- So what **cues** can lead to streaming?
  - should be **anything that can distinguish sources**



# Time, Scale, & Context

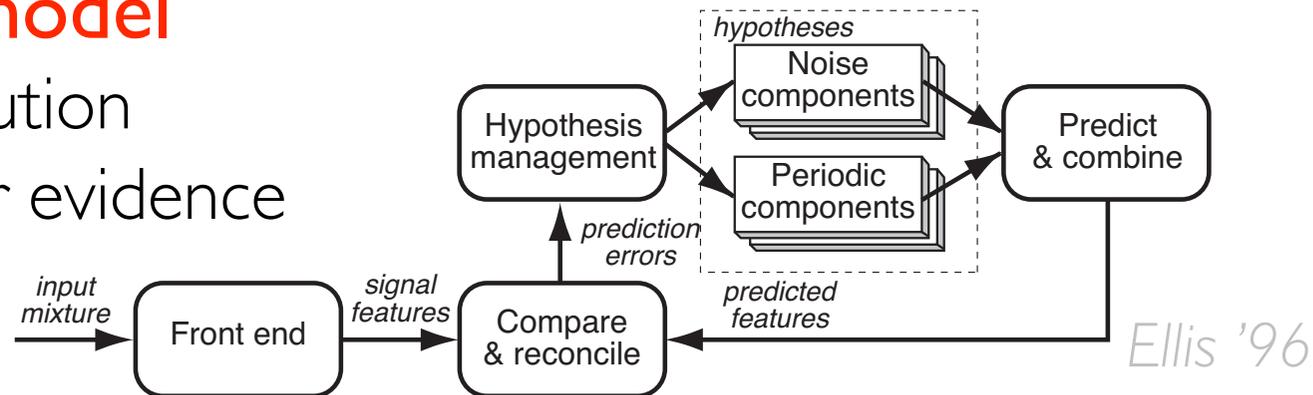
- Footsteps

- fall in the “ambiguous region” ?

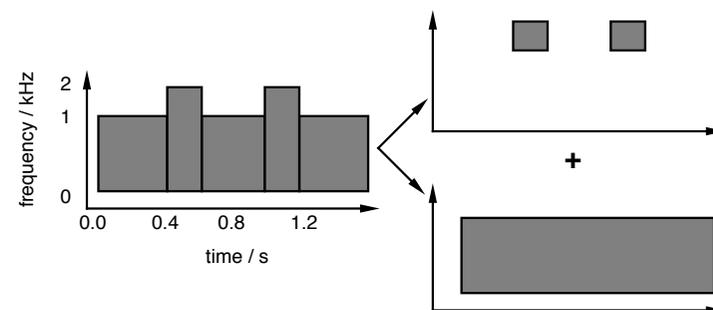


# 3. Computational Implications

- “World model”
  - state evolution
  - “hook” for evidence



- Just maximizing  $P(\text{observation} \mid \text{explanation})$ 
  - unifies with simultaneous organization
- Bottom up vs. top down
  - events depend on streams
  - same old problems of forming parts & organizing them



# Summary

- Objects may sound **intermittently**
    - makes a **stream** of relevant sound events / glimpses
  - **Meaning** relies on the full 'history' of sound events from a particular source
    - **between-event relations** are useful
  - Streaming is **critical**
    - forming streams **is** scene analysis
- ➔ If we are to correctly perceive something in the world, we must be able to **make a stream** out of it

