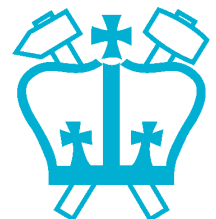

Integrating CASA with other approaches

Dan Ellis

Laboratory for Recognition and Organization of Speech and Audio
Dept. Electrical Eng., Columbia Univ., NY USA
dpwe@ee.columbia.edu

1. CASA cues
2. CASA in the Speech Fragment Decoder
3. Other approaches



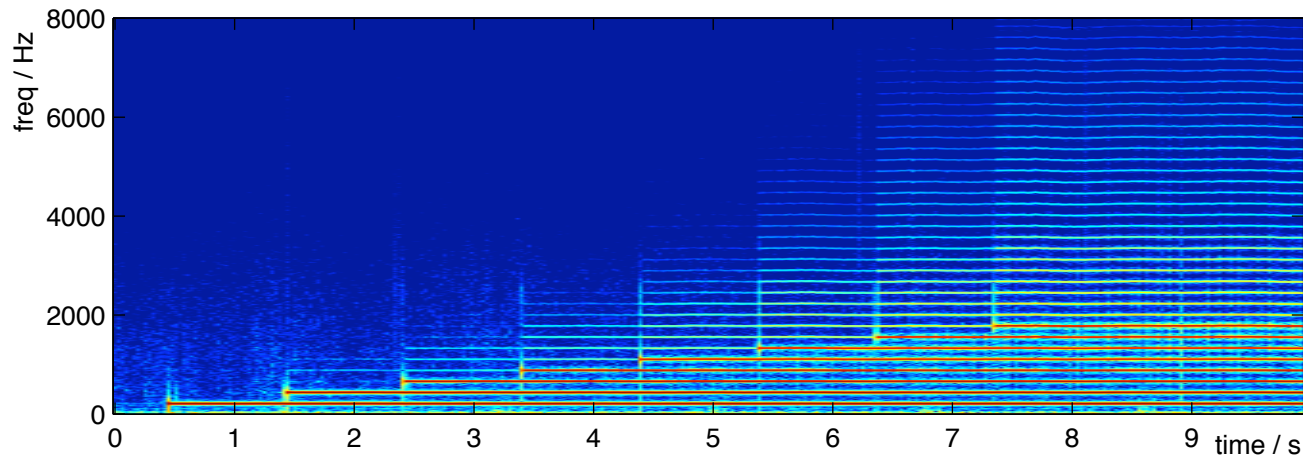
I. CASA Cues

- **Auditory Scene Analysis**
 - .. is based on 'grouping cues'
 - .. which reflect **real-world** sound characteristics
- **Computational ASA features**
 - common **onset**
 - **harmonicity** (common fundamental)
 - common modulation / **rate**
 - **spatial** cues (ITD/ILD i.e. perceived origin)
 - continuity
 - sequential similarity
 - learned patterns (schema)



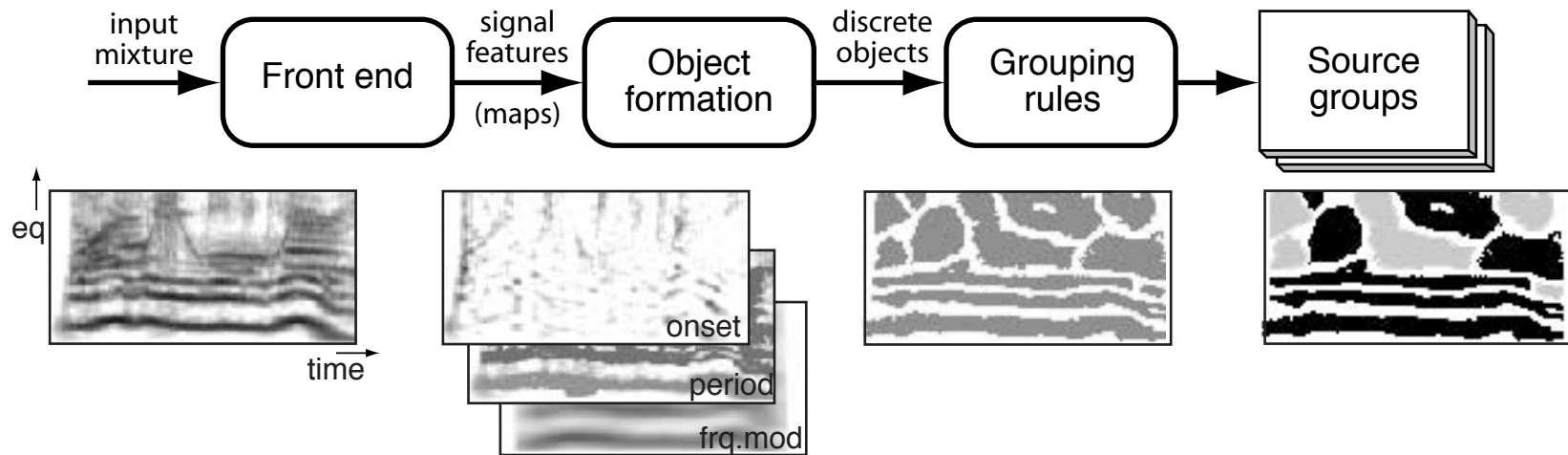
Grouping cues

- Onset & Harmonicity: Pierce example



- Multiple cues are not always consistent!

Pure CASA System (Brown'92)



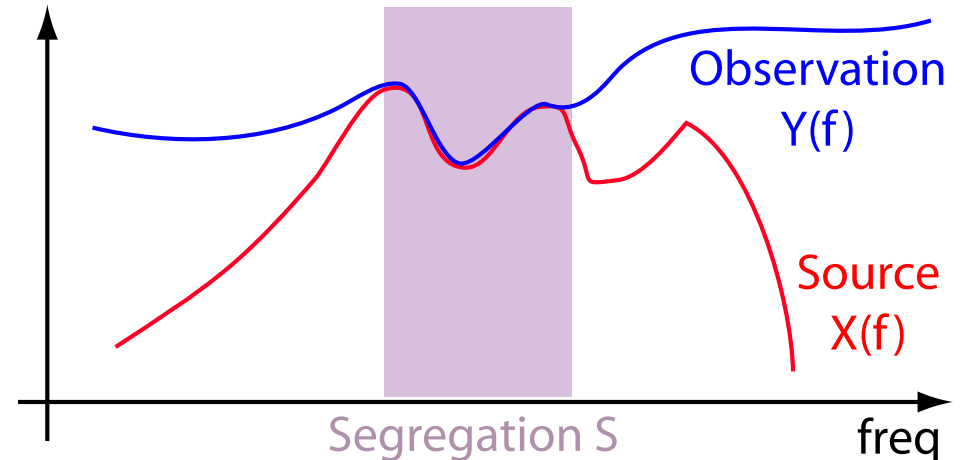
● Problems

- hand-built rules
- Harmonicity as principle cue:
how to integrate other factors?
- one continuously-voiced utterance

2. CASA within The Speech Fragment Decoder

Barker, Cooke, Ellis '04

- Match 'uncorrupt' spectrum to ASR models using **missing data**



- Joint search for **model M** and **segregation S** to maximize:

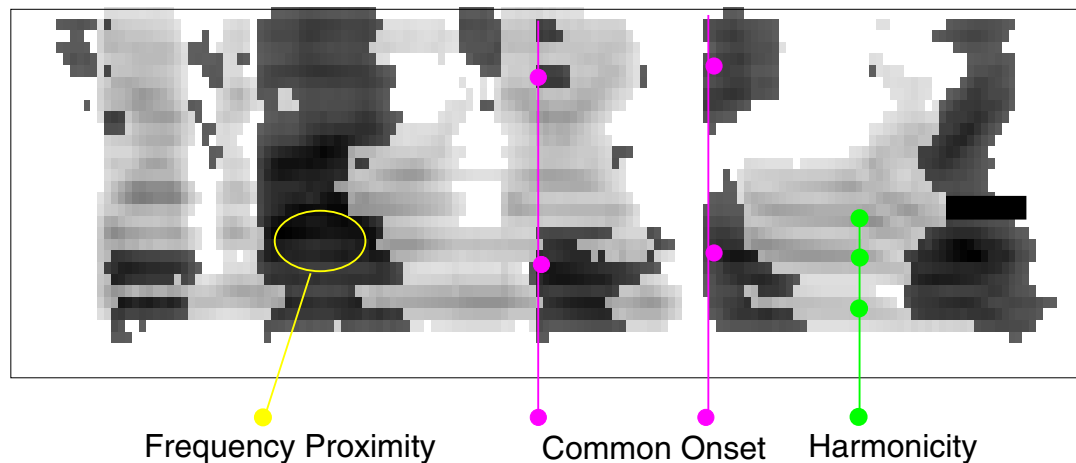
$$P(M, S|Y) = P(M) \int P(X|M) \cdot \frac{P(X|Y, S)}{P(X)} dX \cdot P(S|Y)$$

Isolated Source Model
Segregation Model

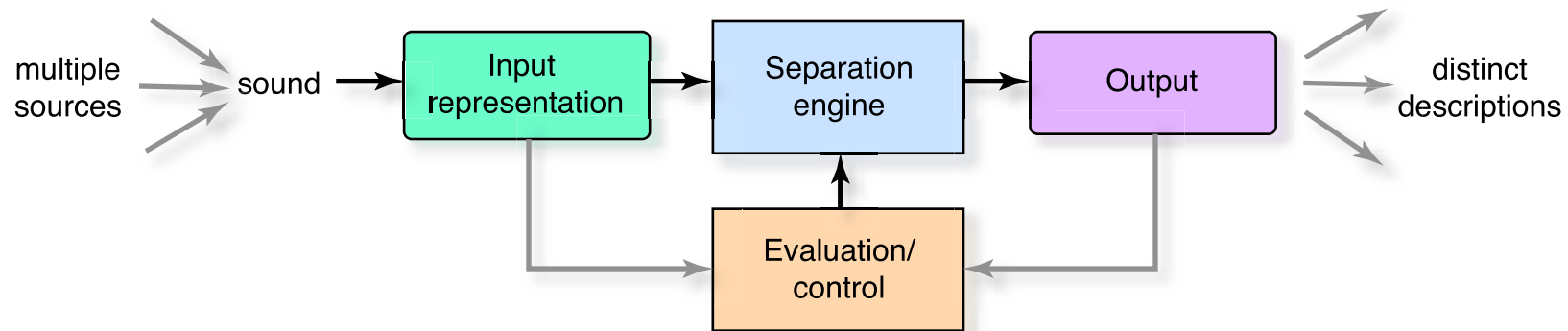
Using CASA cues

$$P(M, S|Y) = P(M) \int P(X|M) \cdot \frac{P(X|Y, S)}{P(X)} dX \cdot P(S|Y)$$

- **CASA helps search**
 - consider only segregations made from CASA chunks
- **CASA rates segregation**
 - construct $P(S|Y)$ to reward CASA qualities:



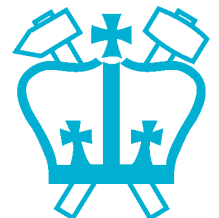
3. CASA + Other Approaches



- Put CASA in the **control** for any **engine**?
- CASA + ICA?
 - Te-Won's 5-year project
- CASA + Clustering (Francis Bach)
 - optimize 'affinity' based on auditory+... cues
- Nonnegative parts?
 - parts: local cues: local

Can Machine Learning Subsume CASA?

- ASA grouping cues **describe** real sounds
 - ..“anecdotally”
- **Machine Learning** is another way to find regularities in large datasets
 - can, e.g., Roweis **templates** subsume harmonicity, onset, etc.?
 - ... and handle **schema** at the same time?
 - “cut out the (grouping cue) middleman”
- **Trick is how to represent/generalize**
 - listeners can organize novel sounds



Conclusions

- **CASA cues:**
describe attributes of **real-world** sources
- **Integrating CASA:**
 - contribute to a 'reasonableness' metric (co.theory)
 - help constrain the search (algorithm)
- **Integration in general:**
 - how much 'different' information is available?
 - Session: Is there integration? Can there be?

