The Dangers (and Joys) of Parsimony in Query-by-Voice Applications

Colin Meek and William P. Birmingham

University of Michigan, Grove City College
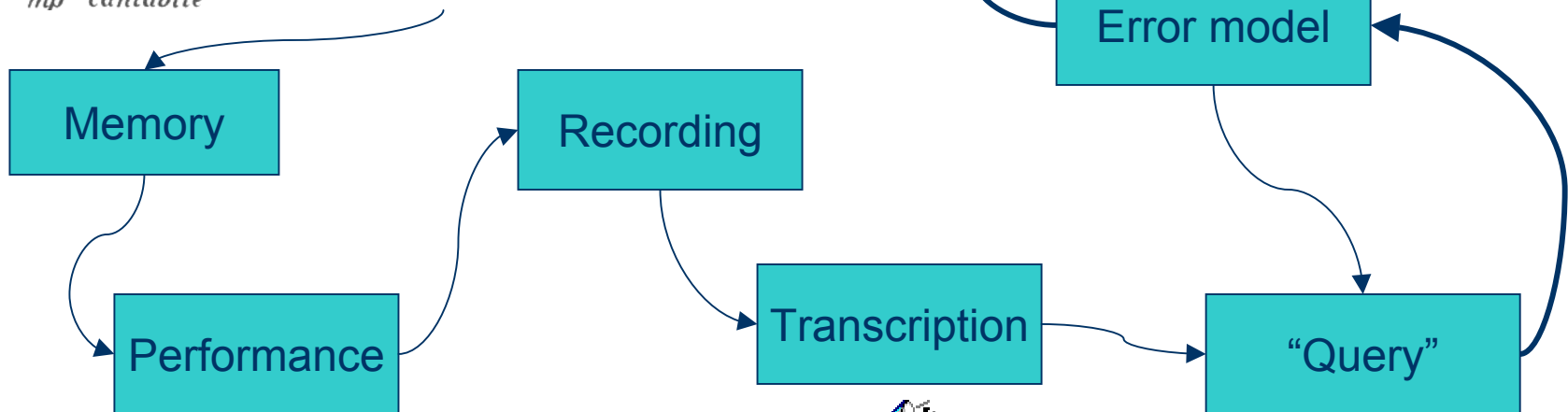
Musen Group

NSF IIS-0085945.

Modelling "Errors"

- Goal: What are they singing?
- Challenge: Several noisy steps between source and query…

Modelling "Errors"

- Goal: What are they singing?
- Challenge: Several noisy steps between source and query…

A sufficient (parsimonious) error model

- Given a query sequence $q_1\,q_2\,q_3\ldots\,q_m$ and a potential match $t_1\,t_2\,t_3\ldots\,t_n$:
  - Match strength = # of positions at which $q_i = t_i$
- Any transformation can be accounted for, but…
  - *Alignment*: What will they sing?
  - *Edits*: What happens if notes are added, or missing?
  - *Context*: What about key/register differences?
  - All errors are not created equal!
- How can we explicitly address these issues in an error model?
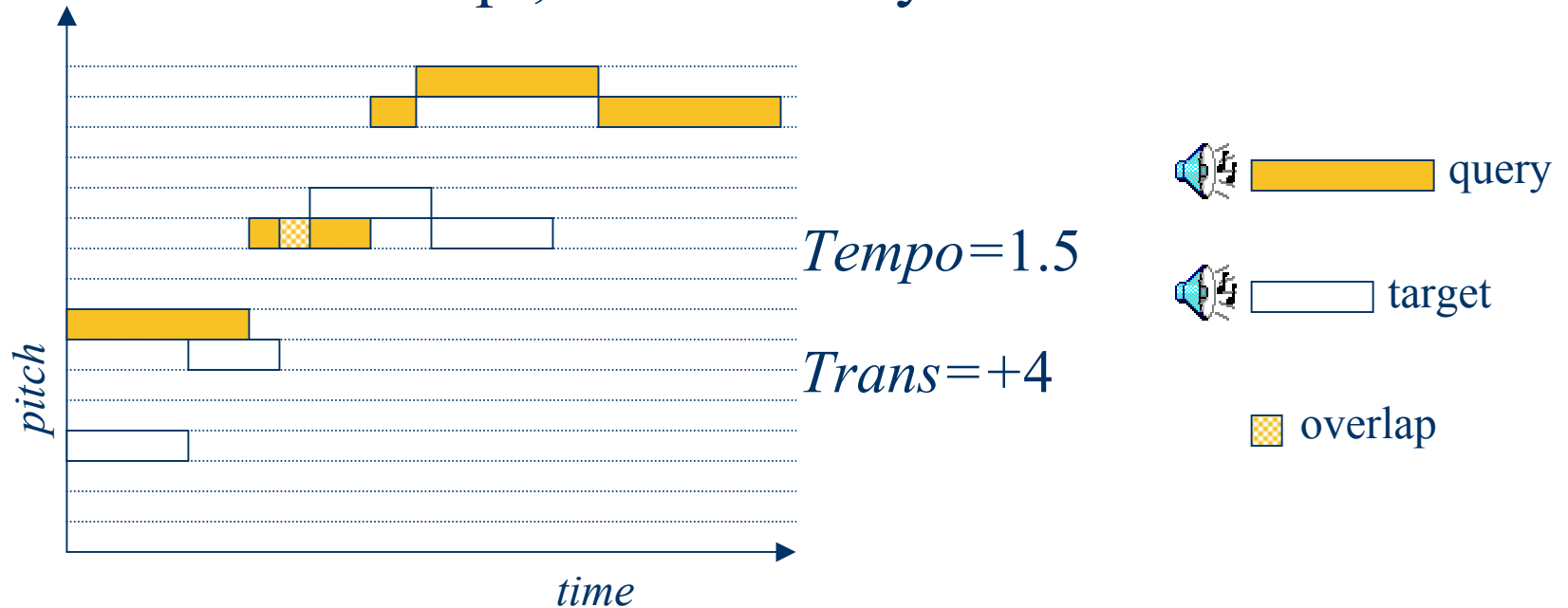
What will they sing?
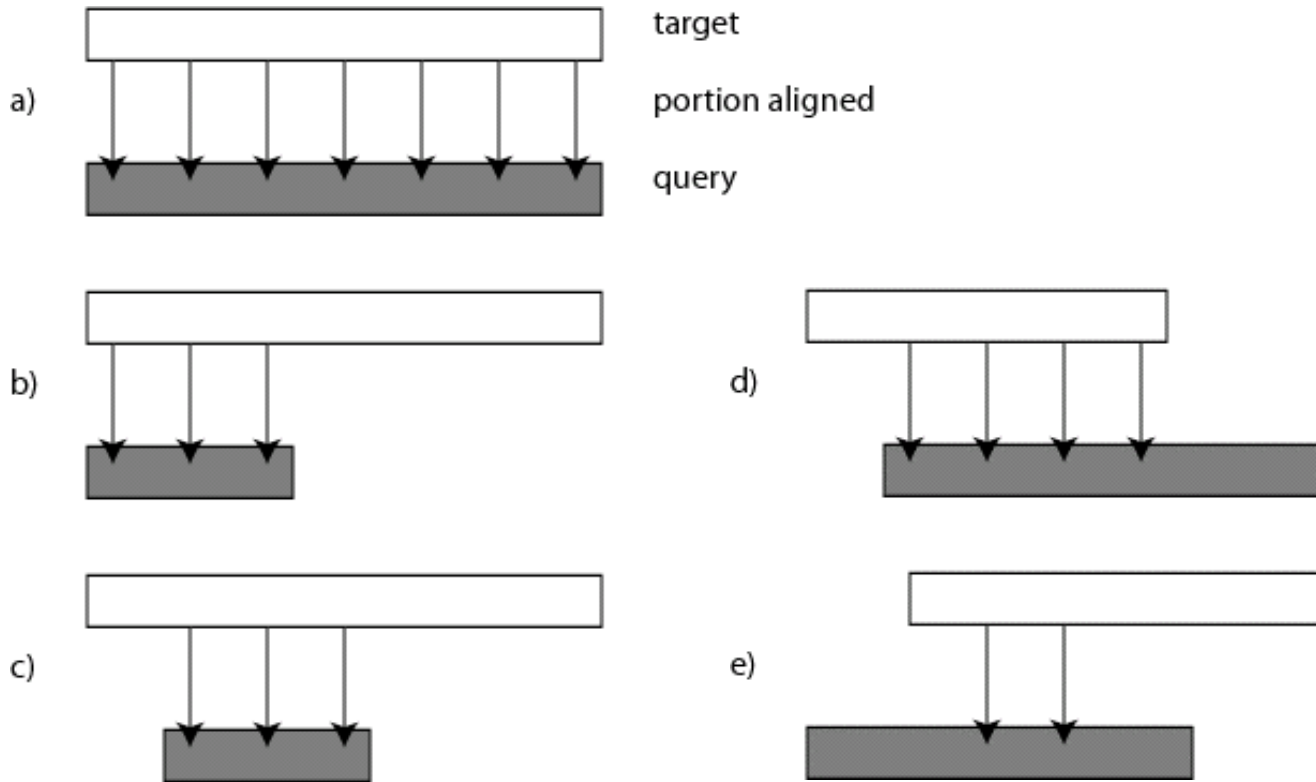
Edit errors

- Transcriber or singer may drop or add notes
- Represent using state transition model
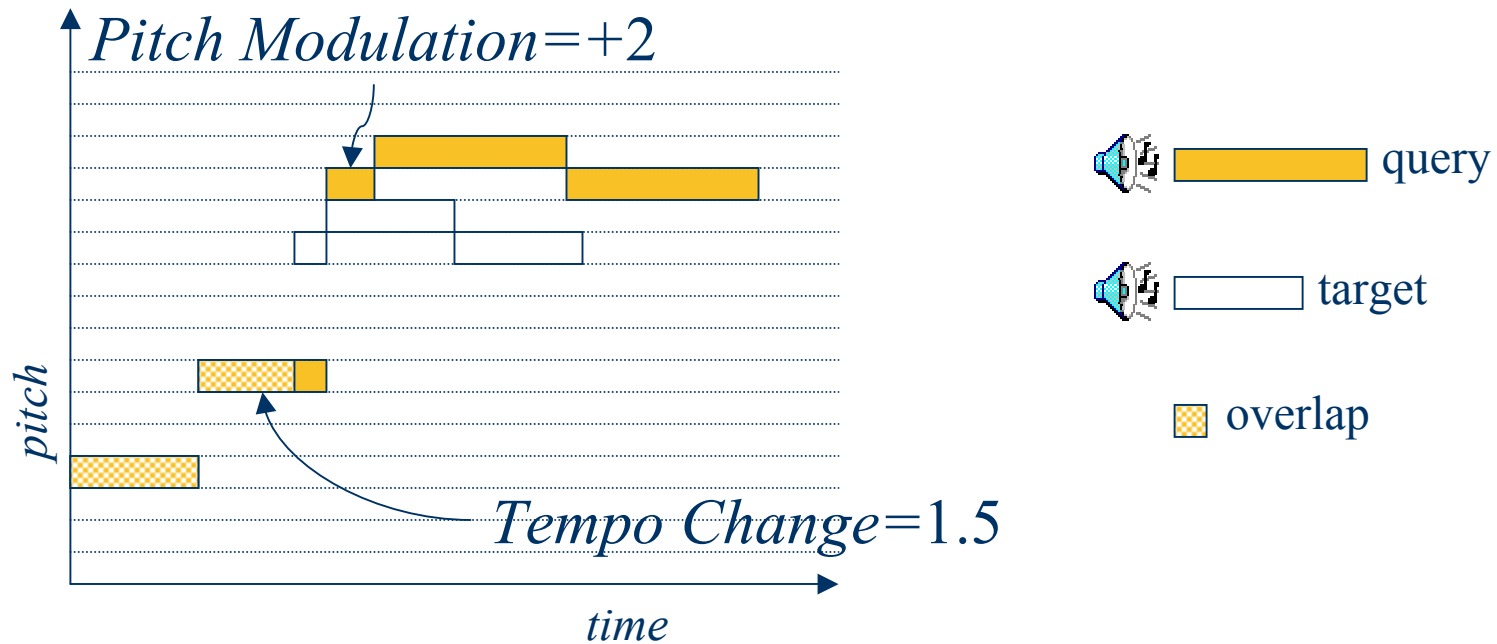- Why not conventional string edit?
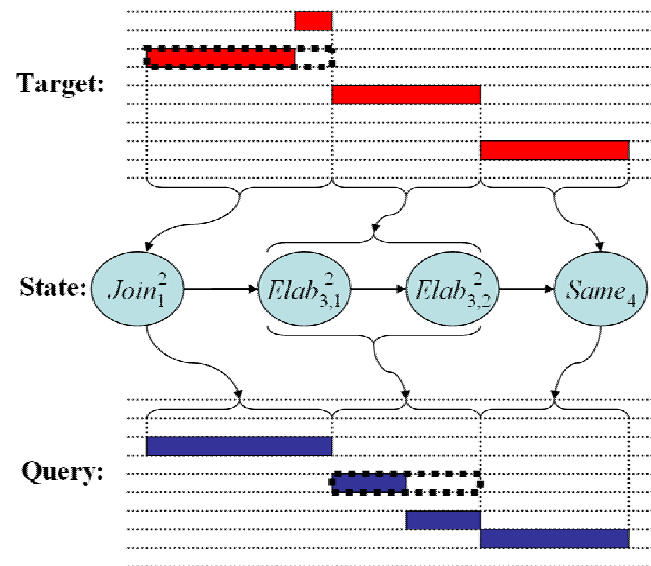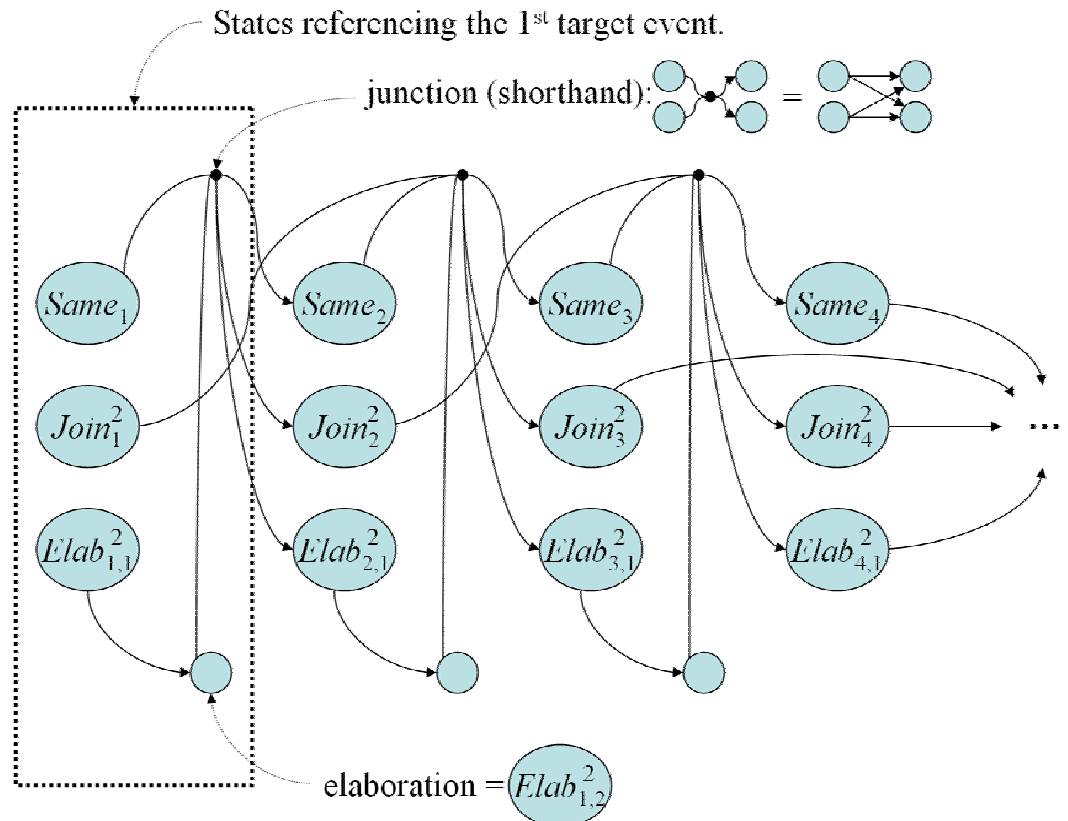
Context errors

- Different tempi, different keys

Tempo=1.5

Trans=+4

query

target

overlap

- Incorporate notion of "cluster" in state: <Edit, Transposition, Tempo>

Changing context errors

- Tempo and transposition can "wander"

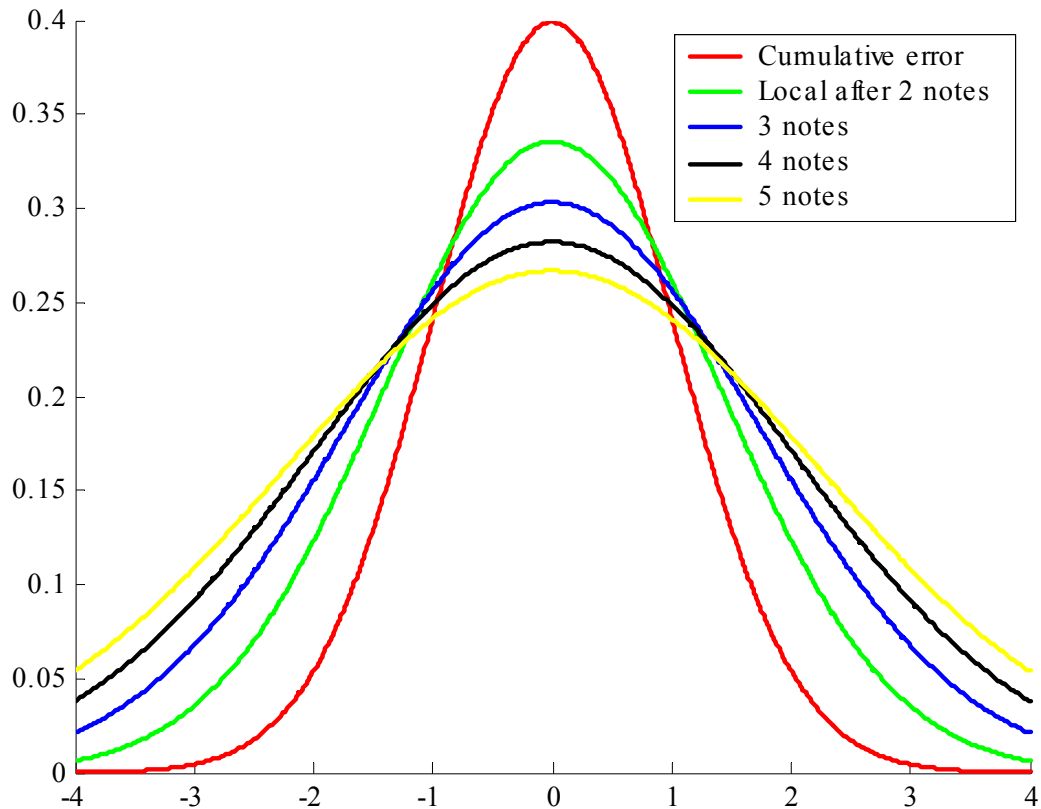# Local errors

- Out of tune, poor rhythm

query

target

overlap

$RError = 0.67$

$PError = -1$

pitch

time
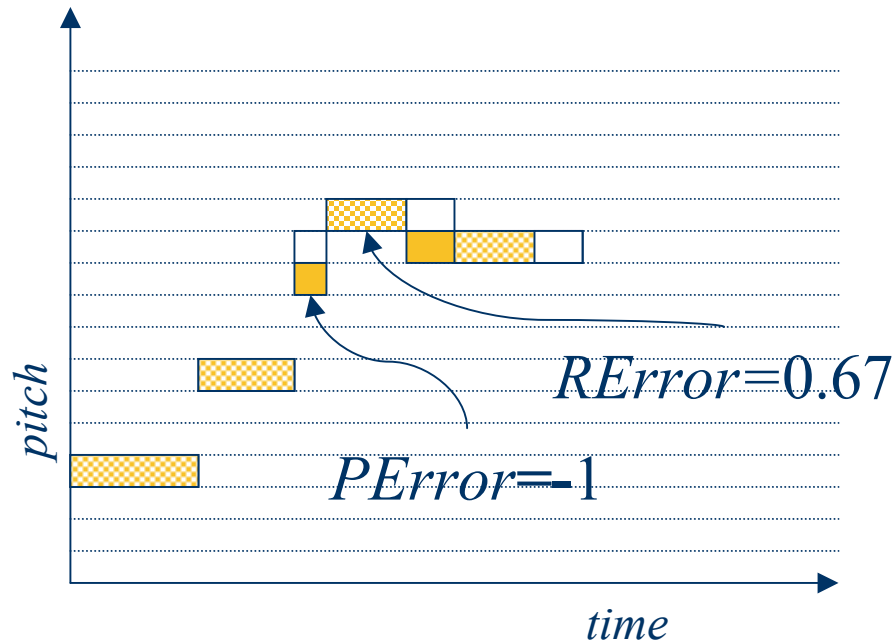
- Note competing explanations for "errors"!

Why not use a parsimonious model?
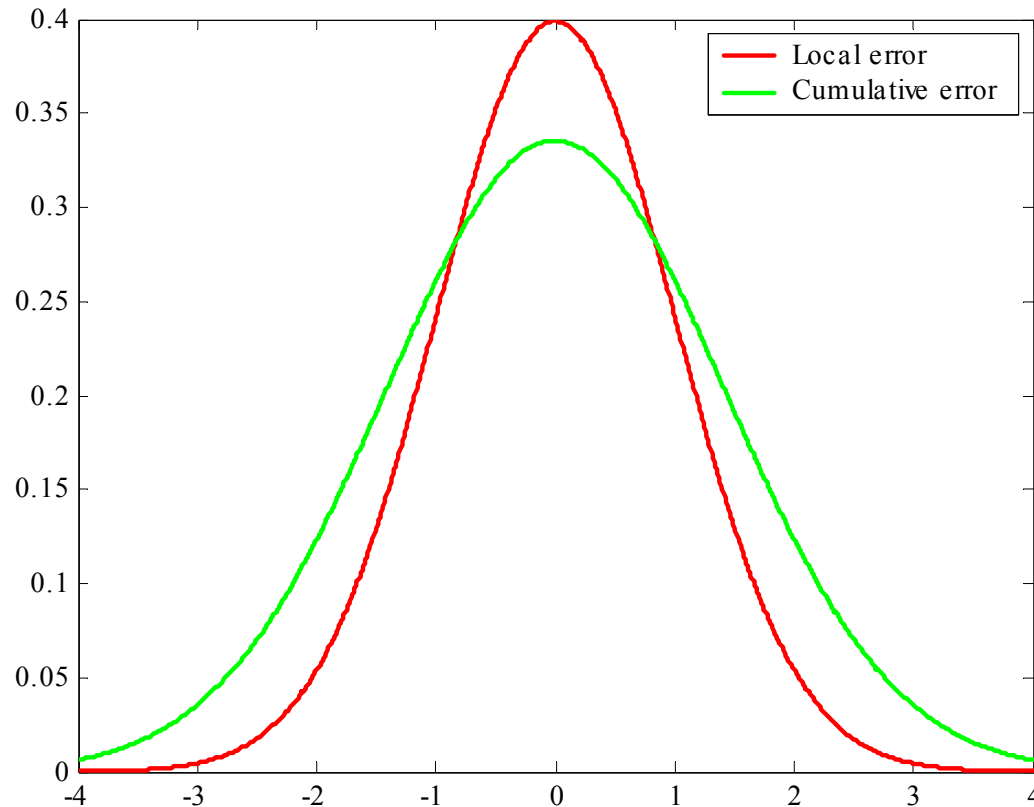
Paying interest…

Related work

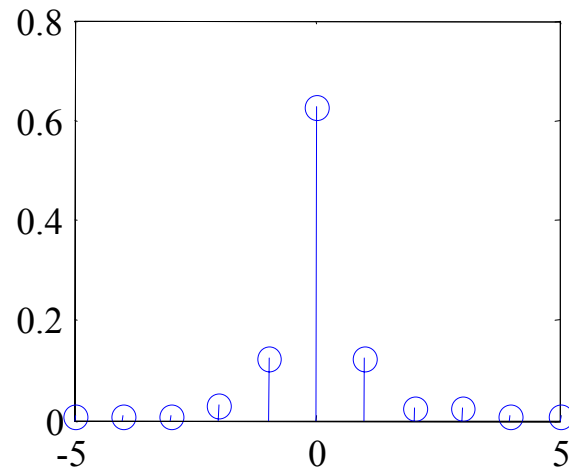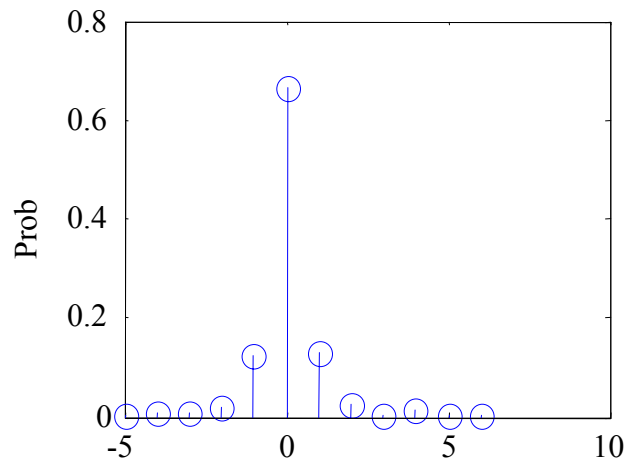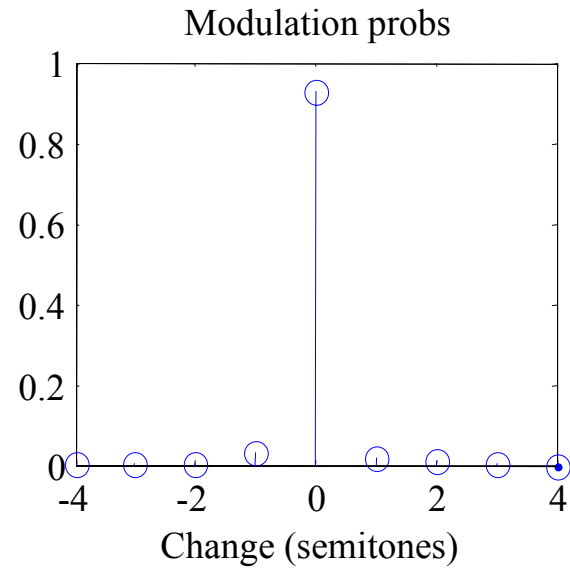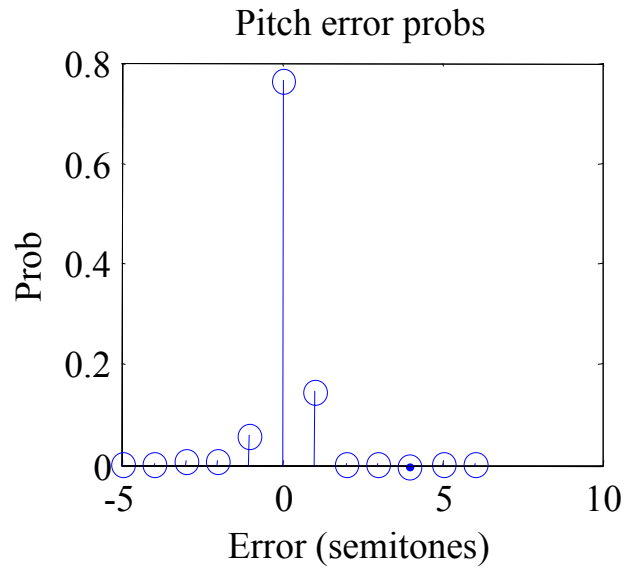- Kalman Filters: Support continuous changes in context

- HMM and Edit Algorithms: Support local errors XOR changes in context (representation dependent)

- Distributed HMM: Support discrete changes in context, branching

- Dynamic Bayesian Networks: Support continuous changes in context, branching

Model Features

Full model

Pitch error probs

Modulation probs
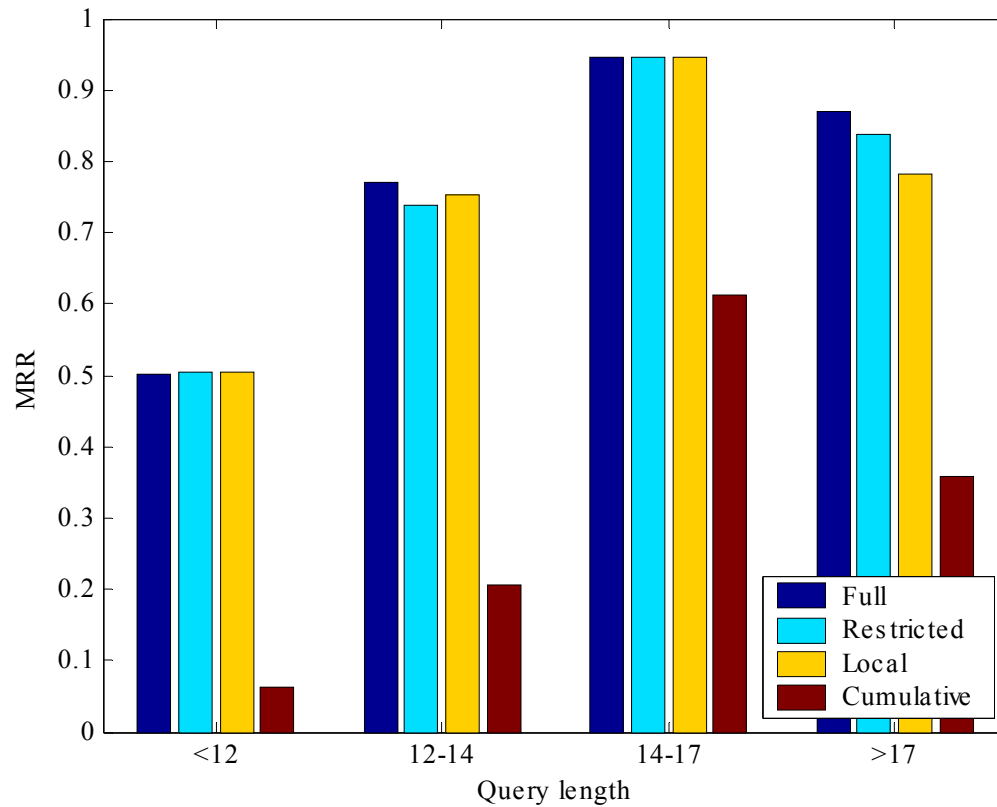
"Local" error only

"Cumulative" error only

Contributions

- Develop a unified and comprehensive model for QBV applications

- Demonstrate experimentally tradeoffs between model simplicity and retrieval performance

- Other dimensions along which QBV apps can vary:
    - Treatment of alignment
    - Continuous vs. discrete observation space
    - Symbolic (e.g., notes) or lower-level (e.g., pitch-track)