

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.



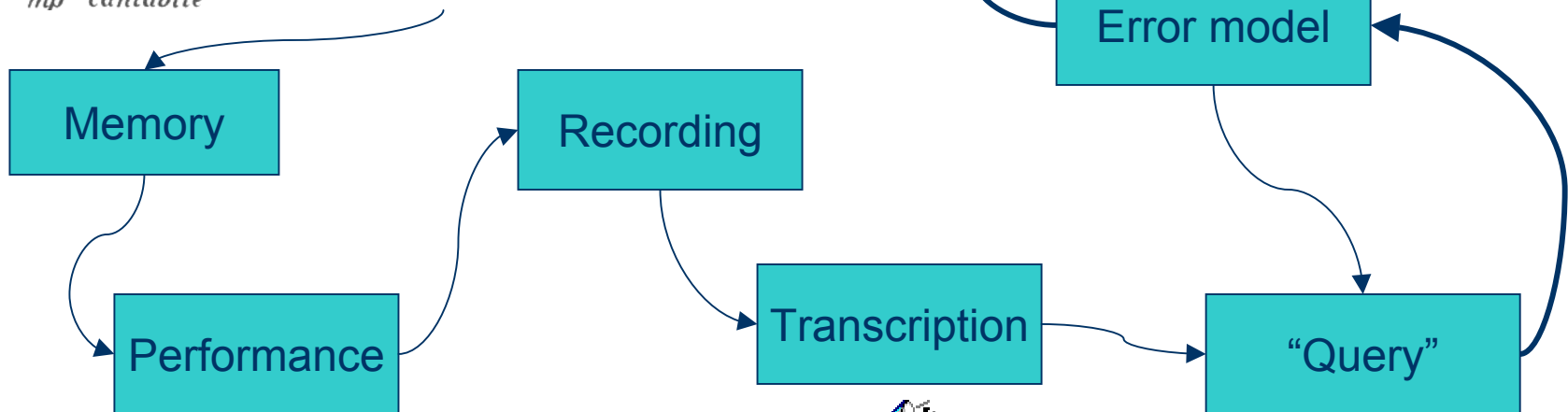
Modelling “Errors”

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

The image displays three staves of musical notation, likely for a piano or guitar, with various performance instructions and annotations. The first staff (measures 132-134) includes markings for *dim.* and *mp*, and a *pizz.* instruction with a triplet. The second staff (measures 135-137) features the tempo marking *Allegro ma non troppo*, dynamic marking *mf*, and the instruction *rit.* with a *senza sord.* marking. A red dashed box highlights a section of the second staff. The third staff (measures 142-144) includes the instruction *cresc.* and dynamic markings *mf* and *espr.*. Fingerings and articulation marks are present throughout the notation.

Modelling “Errors”

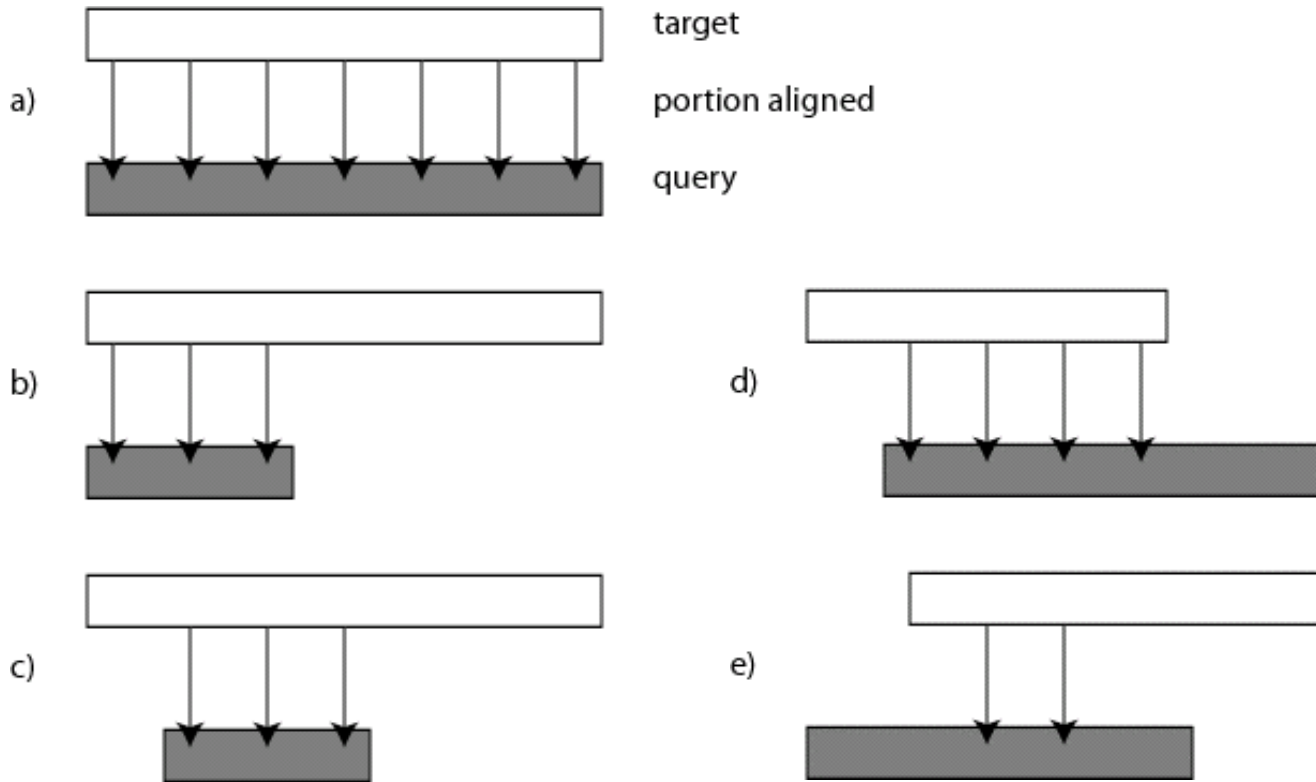
- Goal: What are they singing?
- Challenge: Several noisy steps between source and query...



A sufficient (parsimonious) error model

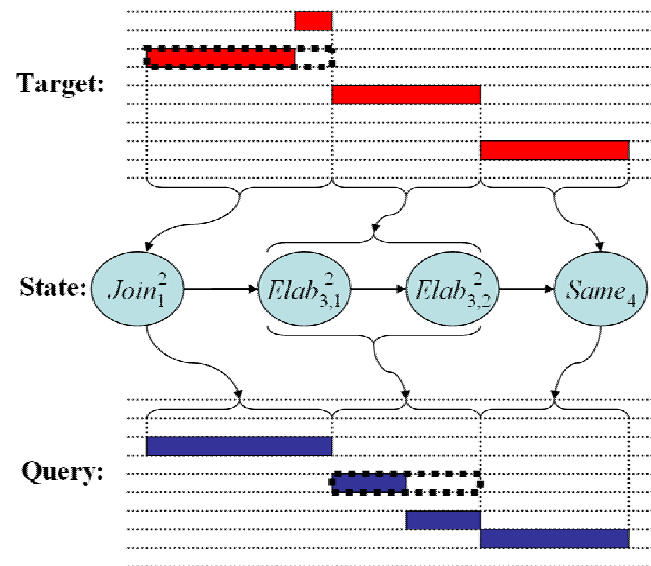
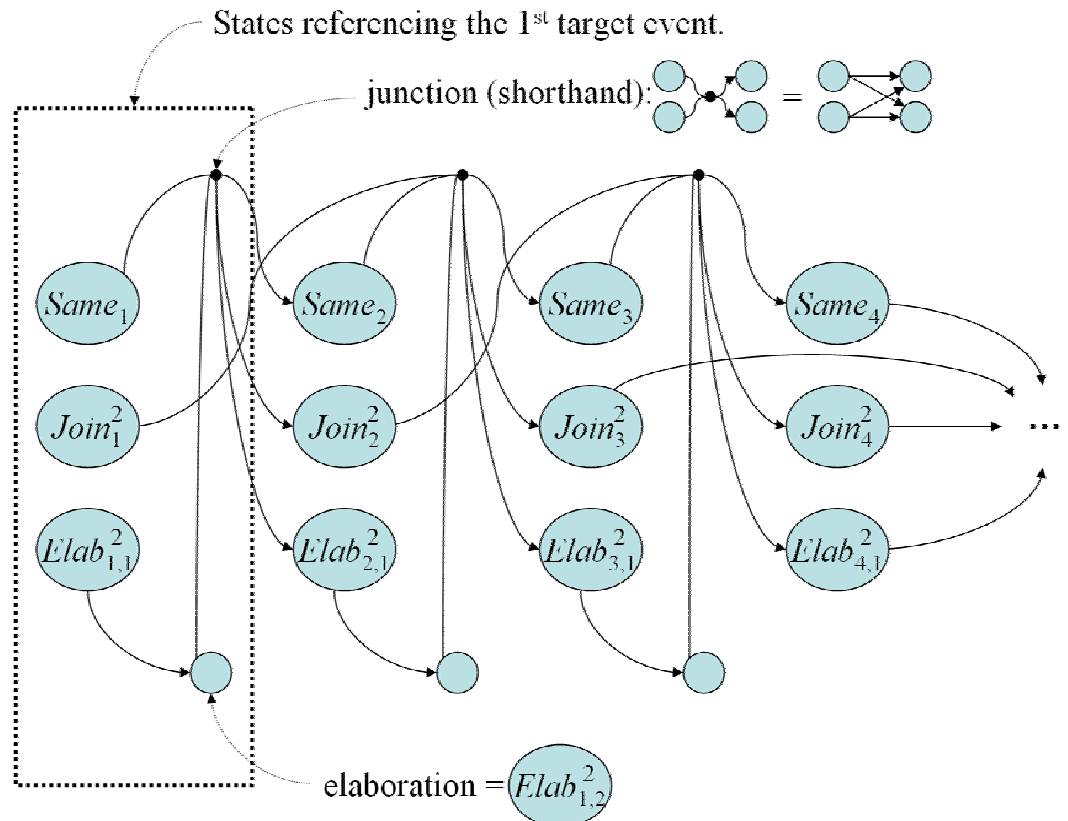
- Given a query sequence $q_1 q_2 q_3 \dots q_m$ and a potential match $t_1 t_2 t_3 \dots t_n$:
 - Match strength = # of positions at which $q_i = t_i$
- Any transformation can be accounted for, but...
 - *Alignment*: What will they sing?
 - *Edits*: What happens if notes are added, or missing?
 - *Context*: 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?



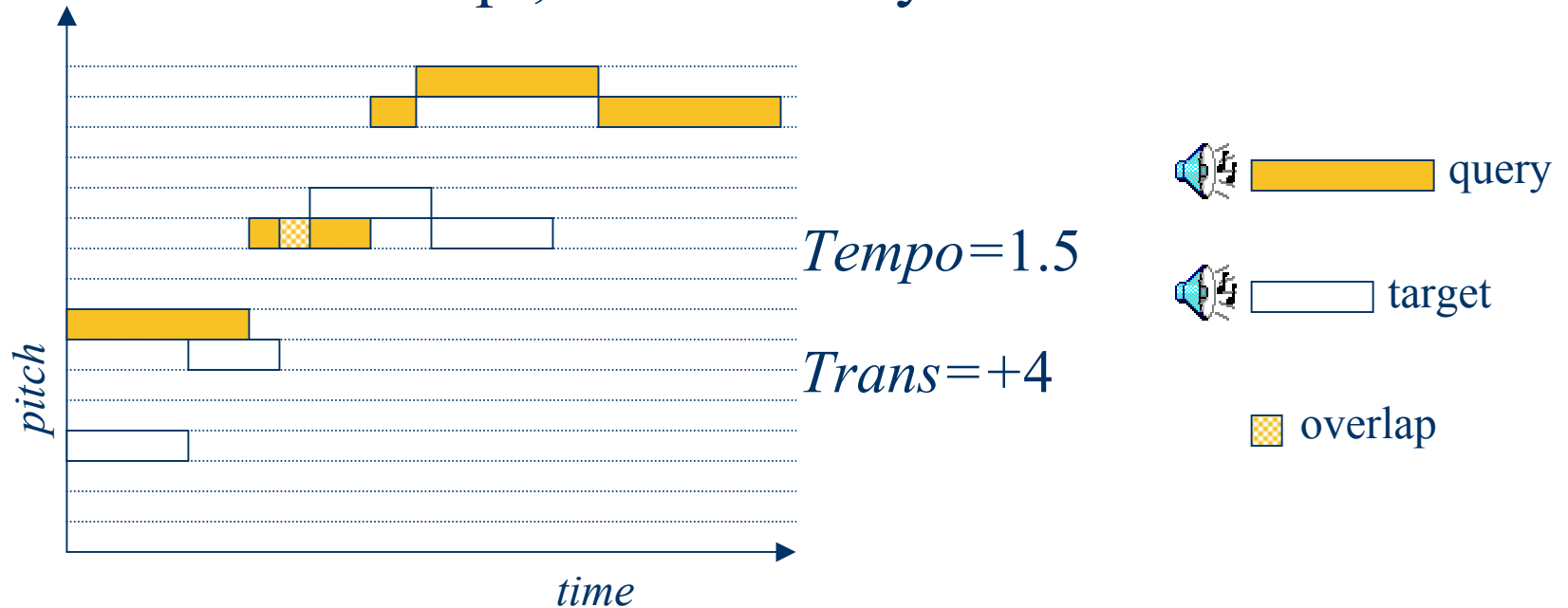
Edit errors

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



Context errors

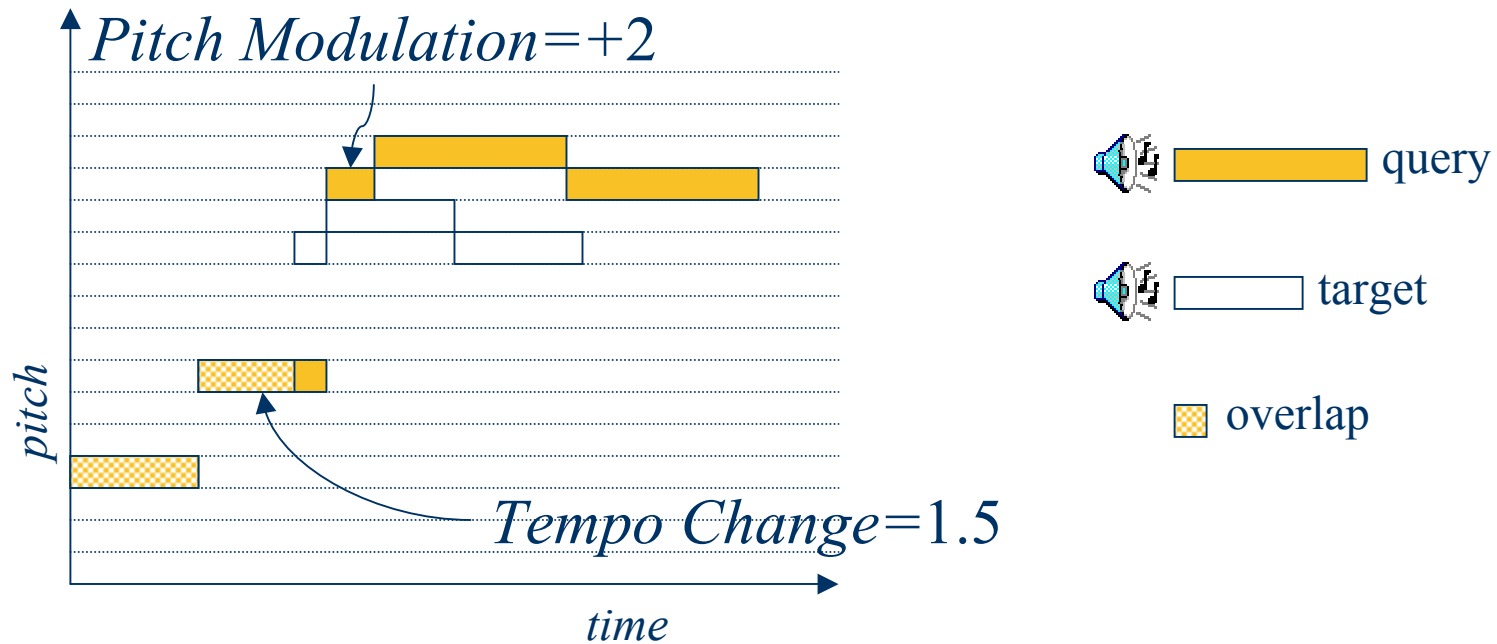
- Different tempi, different keys



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

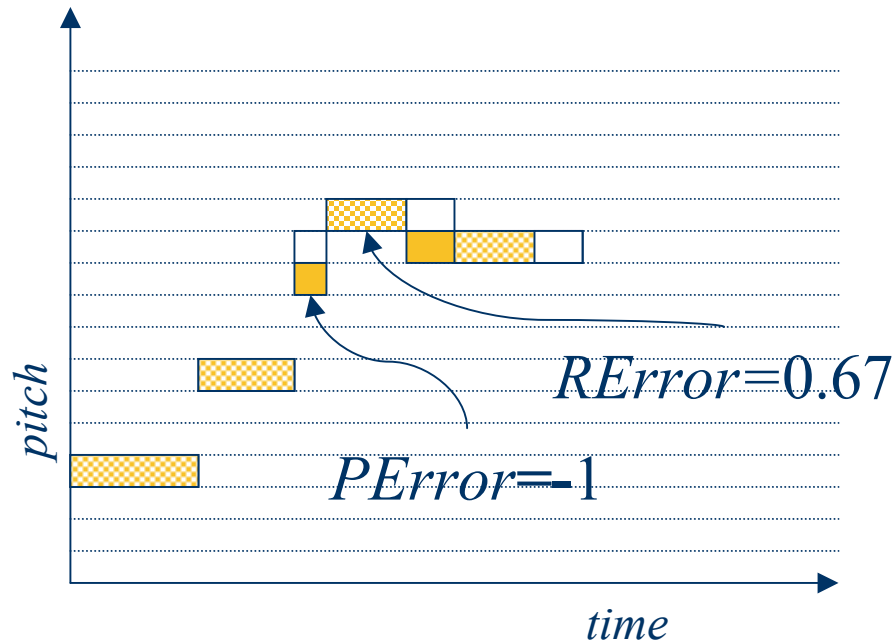
Changing context errors

- Tempo and transposition can “wander”



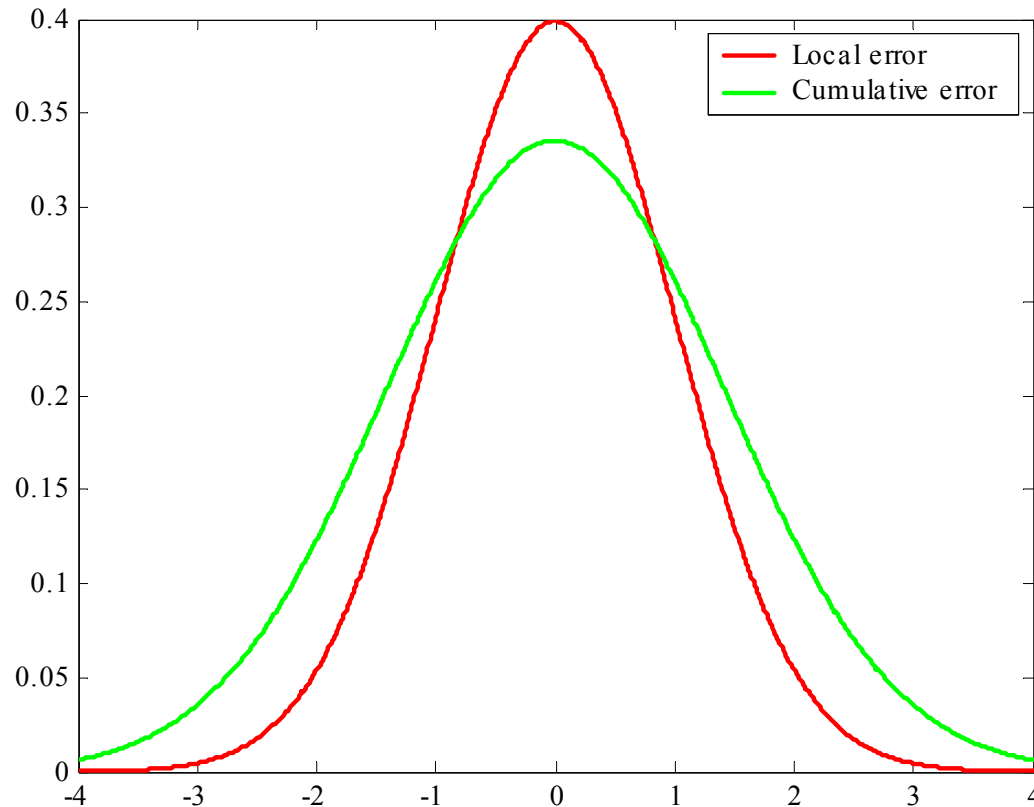
Local errors

- Out of tune, poor rhythm

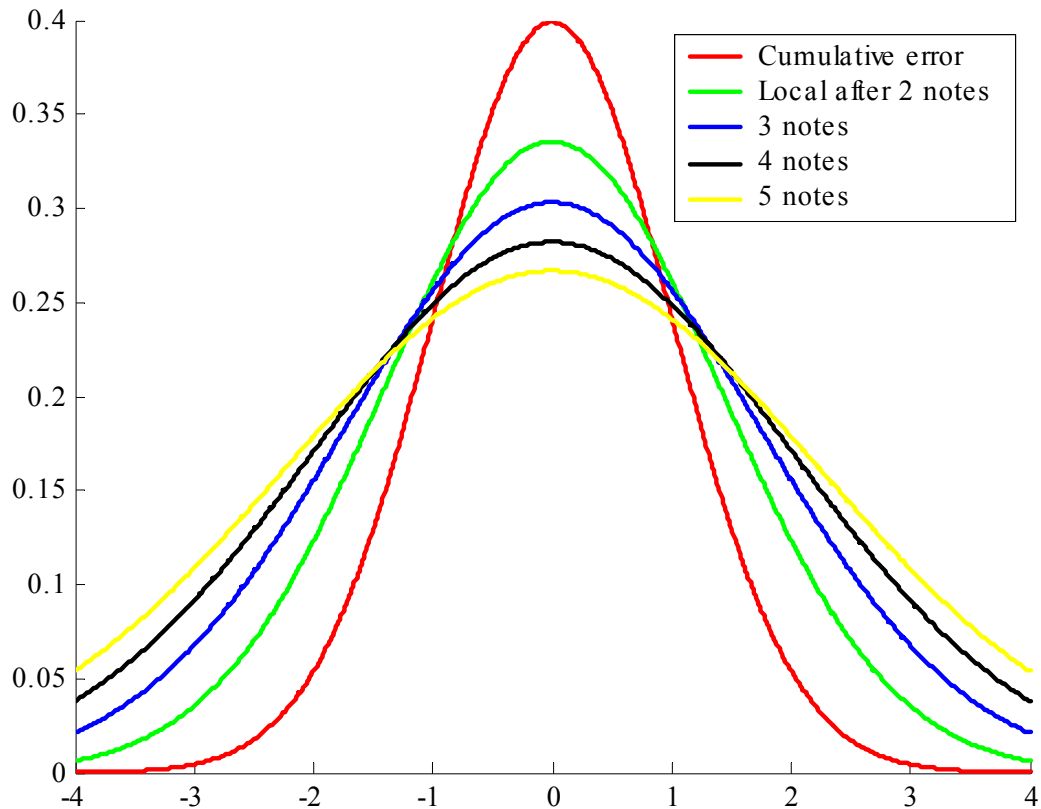


- Note competing explanations for “errors”!

Why not use a parsimonious model?



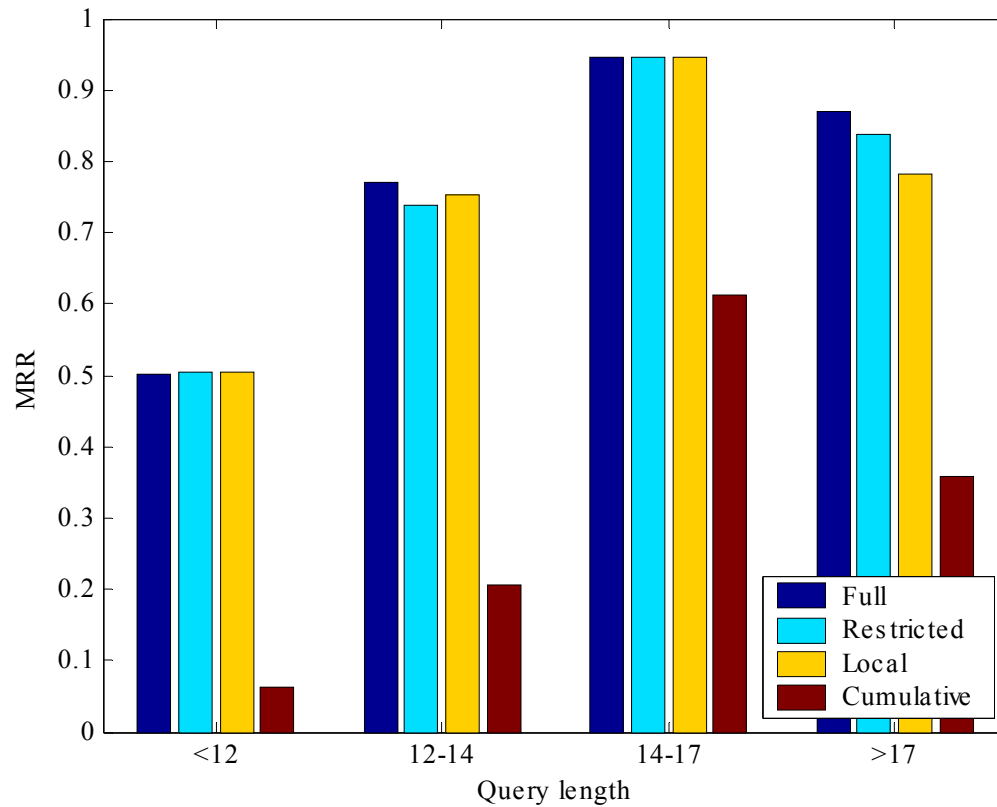
Paying interest...



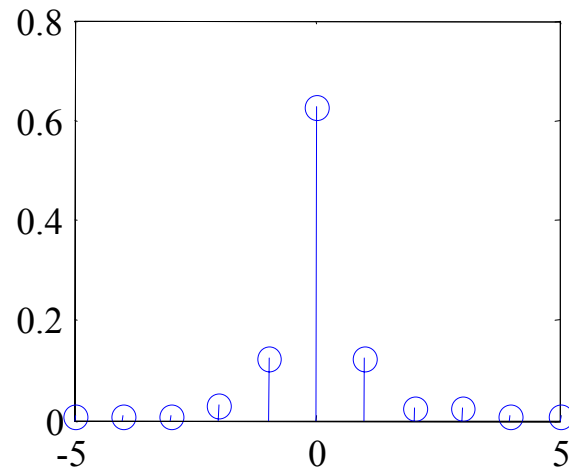
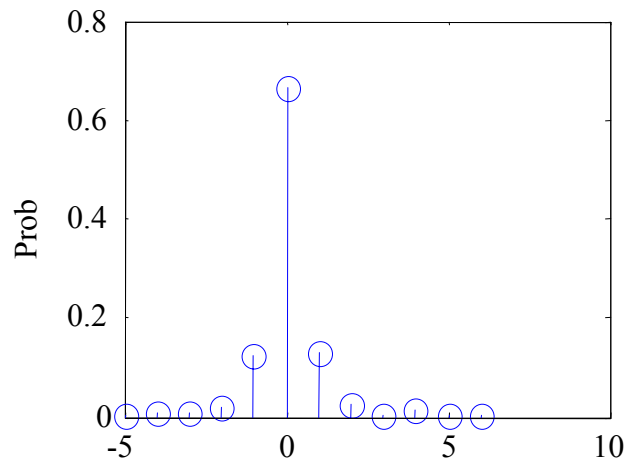
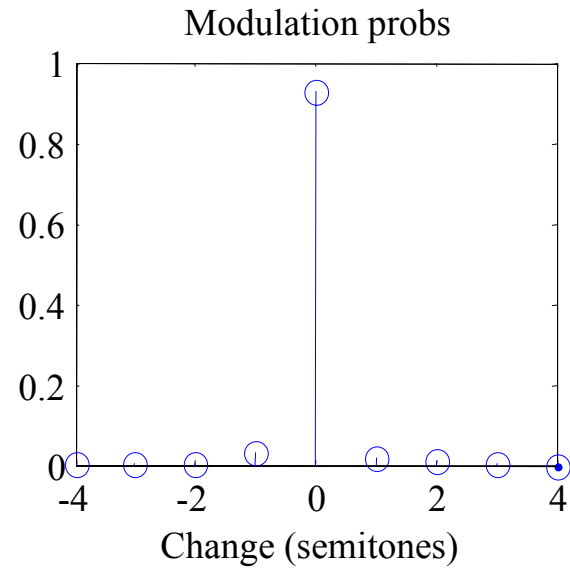
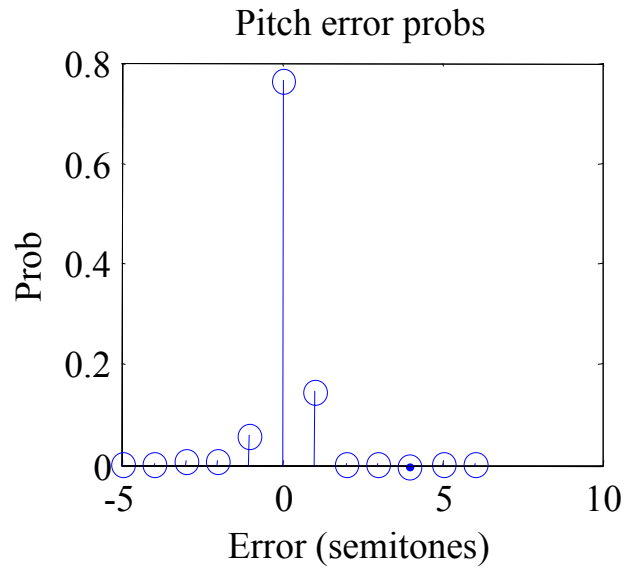
Related work

- 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



“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)