Blackboard Polyphonic Transcription

Rebecca Fiebrink

MUMT 611

17 February 2005

Presentation outline

- Overview of polyphonic transcription problems
- Define blackboard systems
- Present high-level details of one blackboard transcription system
- Discuss variations on this system
- Discuss persisting constraints and limitations of these systems

Limitations in Polyphonic Transcription, pre-1995

- Instruments
- Number of voices
- Relative range
- Absolute range
- Style
- Intervals

Blackboard System Basics

- "Bottom-up" systems:
  - Data-driven
- "Top-down" systems:
  - Prediction-driven
  - Many theories of psychology say the brain operates in a "top-down" way
- Blackboard systems
  - Combine bottom-up and top-down
  - Incorporate knowledge of the domain
  - **Blackboard systems in music seek to integrate musical, acoustic, physiological, and other knowledge, much like a human listener.**

Blackboard Components

- Blackboard
  - – Metaphor: Experts standing around a blackboard
  - – A shared knowledge space
- Knowledge sources
  - – The “experts”
  - – Precondition/action (“if/then”) pair
- Scheduler
  - – Governs when each expert accesses board and changes it

Background of Blackboards in music

- Blackboards used since mid-1970's; speech recognition, signal processing.
- Auditory scene analysis:
  - Ellis 1992+ (MIT)
  - Kashino et al. 1995 (Tokyo)
  - Klapuri 2001 (Tampere, Finland)
- Polyphonic transcription:
  - Martin 1996+ (MIT)
  - Godsmark and Brown 1999 (Univ. Sheffield)
  - Bello et al. 2000+ (London)

Martin’s system

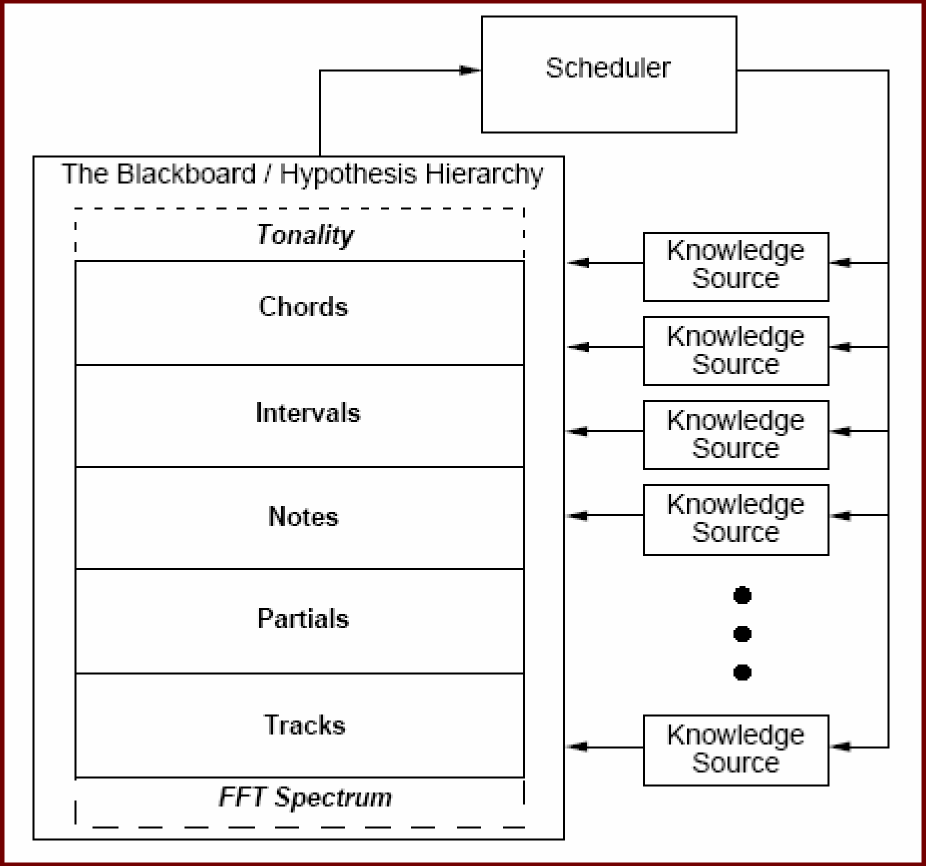
- Keith Martin: MIT Media Lab, 1996
- Goal: Transcription of synthesized piano performances of Bach chorales
- Incorporate knowledge about physics, musical practice (specific to four part tonal harmony), and “garbage collection”

System implementation

- Front end
  - Uses short-time running energy to detect note onsets
  - Uses STFT to get average spectrum for each chord
  - Outputs a set of time/frequency/magnitude “tracks”
- Blackboard system
  - Processes tracks into chords

The Blackboard Component

Abstraction hierarchy from tracks to chords:



From Martin 1996a

The Knowledge Base

- Each Knowledge Source (KS) is one “rule” about how to interact with the blackboard space
- The blackboard contains hypotheses (e.g. “we have the note C4 here,” “we have a C Major chord here”), rated according to their “strength”
- Hypotheses may have “Sources of Uncertainty” (SOUs) attached (e.g. “There is a track here that doesn’t belong to a partial”)
- KSs put hypotheses and SOUs on the blackboard, or take them off, according to their individual rules of action.
- Scheduler runs down list of KSs, one at a time

A Bottom-Up KS Example

- Track_NoExplanation
  - Precondition: There are active track hypotheses unattached to higher level hypotheses. (i.e. notes)
  - Action: Choose the unattached track with the lowest frequency. Divide by 1, 2, 3, 4, and 5, if this falls in the range of valid pitches, propose a note at this pitch. (i.e. create a note hypothesis)

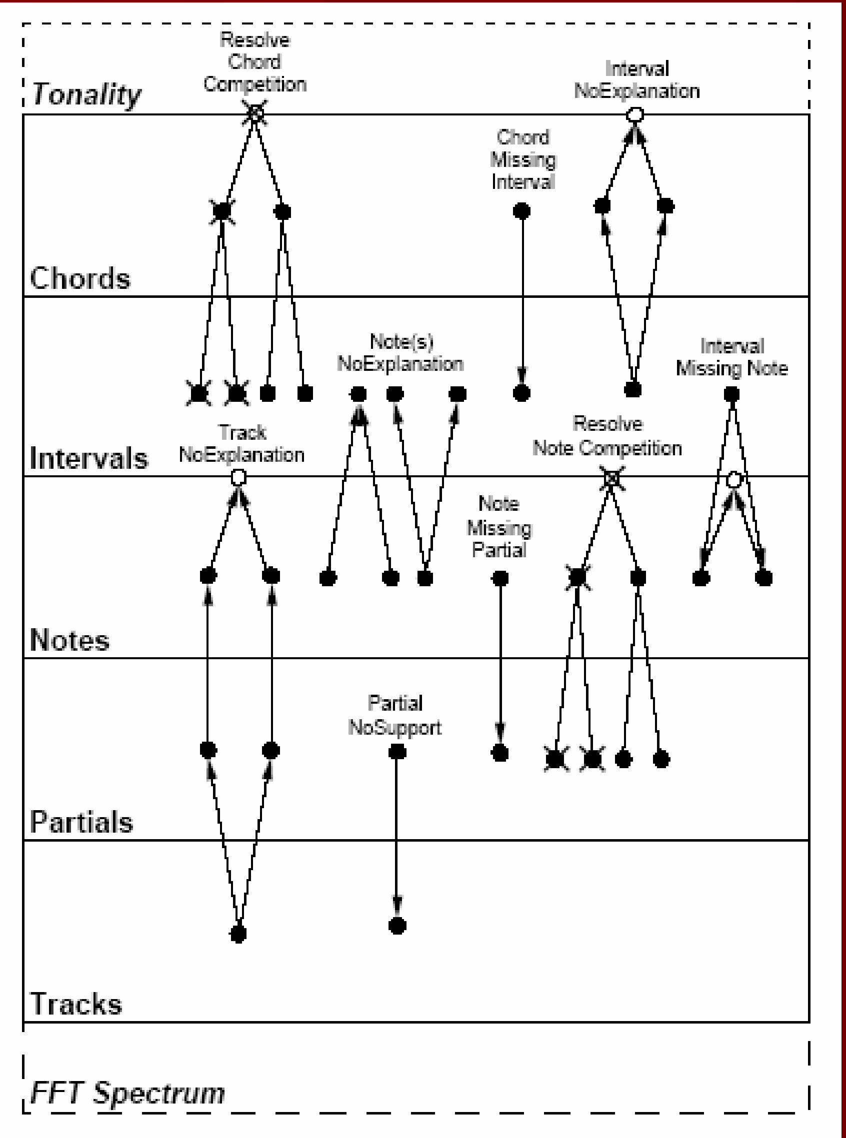
A Top-Down KS Example

- Note_MissingPartial
  - Precondition: There is a note hypothesis on the board with an empty partial slot. (Any of the first 10 partials)
  - Action: Create a partial hypothesis corresponding to the lowest missing partial, up to 2.5 kHz (top of front-end range)

A Garbage Collection KS Example

- Note_PoorSupport
  - Precondition: There is a note hypothesis on the board, with all partial slots filled, but which has a low rating. (i.e. track magnitudes of partials are very low)
  - Action: Remove the note hypothesis with all the supporting partial hypotheses

Knowledge Base/Blackboard interaction



From Martin 1996

Performance of Martin's System

- Many chords are transcribed exactly
- Upper members of octaves are missing
  – Partials between a note and its octave overlap
- Higher notes are missing
  – Upper partials for higher pitched notes on the piano are very weak
- Assumes all notes in chord occur simultaneously, and there is no pitch modulation
- Musical knowledge limited to a restricted genre
- A second system, using a correlation-based front end, still had several problems. (Martin 1996b)

Kashino et al.

- Developed 1995, separate from Martin's work
- "Music scene analysis" – rhythm, chords, source-separated notes, instrument identification
- Uses Bayesian probabilistic network
  – nodes for hypotheses "components," notes, and chords
  – information about a node propagates through the network (as opposed to tight, global knowledge control of Martin's blackboard system)

Knowledge Sources in Kashino

- Chord transition dictionary: based on statistical analysis of 206 traditional songs
- Chord-note relation database: probabilities of notes appearing in a given chord
- Chord naming rules: based on music theory
- Tone memory: frequency components appearing in a note for a given instrument, dynamic, range, and duration
- Timbre models: 11 dimensions of features, each instrument is separate in space
- Perceptual rules: “harmonicity” and “onset timing”

Performance of Kashino's System

- 1995 papers: between 30% and 100% correct, depending on test; demonstrates reduced note identification accuracy on octaves and fifths
- Inclusion of note and chord data sets does improve performance
- Correct instrument identification a problem
- 1999 paper focuses on improving instrument identification (using real instruments)

Bello, Monti, and Sandler

- 2000: Use a blackboard system containing a neural network trained to recognize whether a chord is present; if so, look for more than one note
- 2002: Use a blackboard system with a fuzzy inference system (FIS) to suggest new note candidates

Bello, Monti, and Sandler: Results

- 2000 system
  - – Octave errors still a big problem
  - – No quantitative analysis
- 2002 system
  - – About ½ notes aren’t transcribed at all
  - – 75% of transcribed notes are correct
  - – Octave errors still present

Shortcomings of Current Systems

- Octave errors
- Range errors
- Generalizability: instrument, style, range, ...
- Need note onset scheme
- Limited by spectral & temporal resolution of FFT method employed
- Errors when source instrument modulates in frequency
- Efficiency, complexity

Conclusion

- Blackboard algorithms are complex and powerful
- They are better than previous methods for transcription, but they don't work miracles.
- They could be applied to areas of MIR besides just transcription (e.g., OMR)
- Some researchers are moving towards the use of "model-based" architectures for transcription