Developing a symbolic music representation inspired by speech recognition

Johanna Devaney
Ohio State University

CIRMMT Workshop on symbolic music processing, semantic audio, and music information retrieval

Goal and Motivation

Prior Work

Predicting Phrase Labels

Next Steps

Conclusions

Goal

- Develop a symbolic representation that
  - provides an estimate of which notes are structurally significant
  - works for a range of musical textures
  - captures temporal relationships
  - facilitates the analysis of multiple levels of musical structure
  - is computationally tractable

Motivation

Speech recognition	↔	**Higher-level music analysis**
“Language model”	↔	Functional analysis
Phonemes	↔	Harmonic analysis
MFCCs and other representations	↔	*Representation*
Acoustic signal	↔	Musical surface

Prior Work: Representations

- Vertical slices of polyphonic music to form “chords”
  - per event (Quinn 2010)
  - per beat (Radicioni and Espositio, 2006)
- Working from chord labels (when available)
  - de Haas et. al (2011)

Prior Work: N-grams

- Large-scale analysis approaches are heavily influenced by text retrieval methods, namely N-grams (sequences of N contiguous symbols)
- N-grams work well for
  - melody retrieval in monophonic contexts (Pickens 2001)
  - chord retrieval in polyphonic contexts when the chords occur as distinct units (Scholz et al. 2009)
    - e.g., peachnote.com's N-gram viewer (Viro 2011)

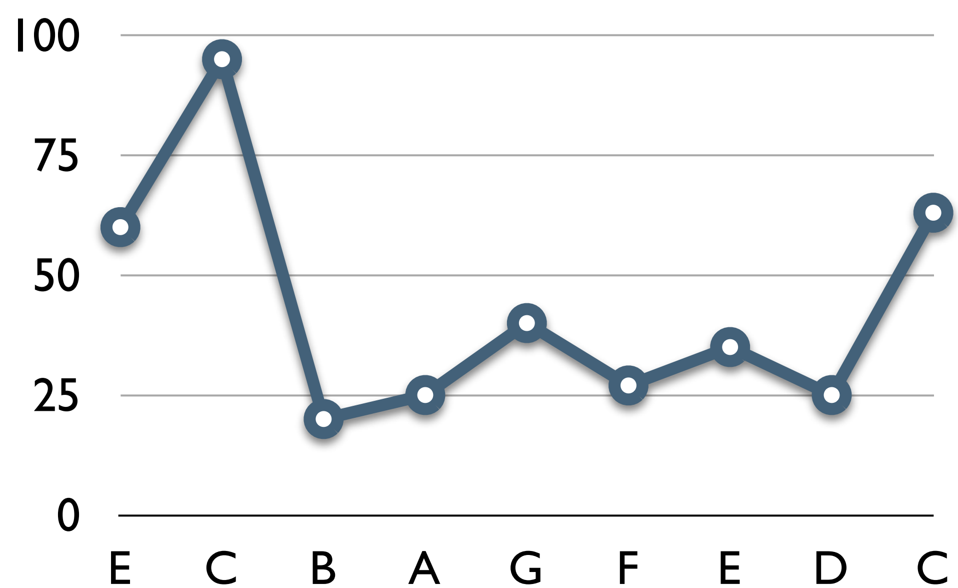
Prior Work: N-grams

- N-gram representations encounter problems with more complex textures
  - e.g., where the notes of chords are not played simultaneously

- N-grams cannot distinguish between what is structurally significant and what is not
  - e.g., between what is a chord-tone and what is a non-chord tone/ornamentation.

This needs to be encoded in the representation

What might this look like?



“Language Model” for Music

Functional harmonic analysis may be flawed but it can be useful

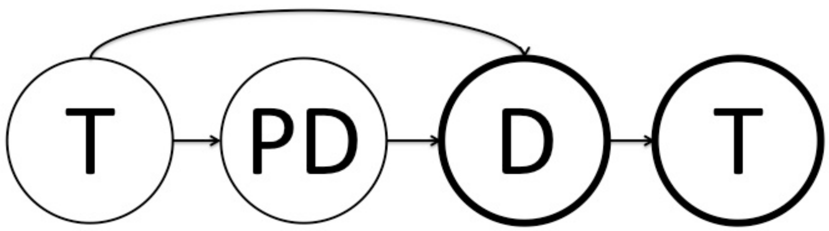
A model of phrase-level function and its relationship to roman numeral labels can be used as a “language model”

Musical Phrases

- Complete musical statements built from the ordered presentation of three harmonic functions
  - tonic, pre-dominant, and dominant functions
- End with a cadence
  - remaining on the dominant function for a half cadence
  - returning to the tonic function for an authentic or a deceptive cadence

Predicting Phrase Labels

- Hidden Markov model (HMM)
  - State space



- Encoded chord label and function for all examples from Laitz's textbook The Complete Musician (2011) into digital representation
  - used to train HMM transition probabilities and evaluate model

This part of the project was undertaken with Daniel Shananan

Predicting Phrase Labels

- Evaluation
  - 80/20 split of textbook data: 94.3% overall accuracy
    - tonic: 93%, pre-dominant: 93%, dominant: 89%
  - Workbook exercises: 87.5% overall accuracy
    - tonic: 96%, pre-dominant: 83%, *dominant: 52%***

      *** 36% of phrases ended with a dominant function in the workbook versus 17% in the training set*

This part of the project was undertaken with Daniel Shananan

Conditional Random Fields

- Like HMMs, CRFs are probabilistic, temporal models
- HMM observations are isolated, only containing information about its own label
  - i.e., uses a joint distribution over both label and observation sequences
- CRFs encode information about temporal context, including metrical position and features from previous time frames
  - i.e., uses conditional probability over label sequences given an observation sequence

Conditional Random Fields

- CRF Implementation
  - observations will be the output of a multi-layer perceptron run on the musical surface
  - states will have chord and function label
  - outputs possible state sequences
- Representation
  - the musical surface with likely structurally significant notes highlighted
  - additional processing will be needed to achieve this

Conclusions

- This talk has described an ongoing project for developing a multi-level model for western art music analysis
- The model is inspired by
  - music theory pedagogy
  - speech recognition
- The ultimate goal of this project is to develop a representation of the musical surface that can capture temporal relationships and works for a range of musical textures

Thank You!

References

de Haas, W.B., J.P. Magalhaes, R.C. Veltkamp, F. Wiering. 2011. HarmTrace: Improving Harmonic similarity estimation using functional harmony analysis. In *Proceedings of International Society of Music Information Retrieval* conference. 67–72

Laitz, S. G. 2011. *The Complete Musician*. Oxford: Oxford University Press, 3rd edition.

Pickens, J. 2001. A survey of feature selection techniques for music information retrieval. Center for Intelligent Information Retrieval, Department of Computer Science, University of Massachusetts, Technical Report.

Quinn, I. 2010. What's 'Key for Key'?: A Theoretically Naive Key–Finding Model for Bach Chorales. *Zeitschrift der Gesellschaft für Musiktheorie*, 7.

Radicioni, D.P. and Esposito, R. 2006. Learning tonal harmony from Bach chorales. In *Proceedings of the International Conference on Cognitive Modelling*.

Scholz, R., E. Vincent, and F. Bimbot. 2009. Robust modeling of musical chord sequences using probabilistic N-grams. In *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*. 53–6

Viro, V. 2011. Peachnote: Music score search and analysis platform. In *Proceedings of the International Society for Music Information Retrieval Conference*. 359–62.