Automatic analysis and comparison of musical performances

Johanna Devaney
Assistant Professor of Music Theory and Cognition
School of Music
The Ohio State University

Introduction
Motivations and challenges.
1

A brief history
Quantitative approaches to performance analysis.
2

Extracting Performance Data
MIDI-audio alignment for automatic analysis of recorded performances.
3

Experiments
Studies of intonation in the singing voice.
4

Developing a Representation of Symbolic Music
Comparing performances of different pieces.
5

Conclusions
6

Introduction

Why study musical performance?

- **Performances convey musicians' interpretations**
- **Performances are what listeners actually hear**
- **Studying performance can help us gain insight into**
  - how an individual's performance practice evolves as they gain more experience
  - how performance practices evolve over time
- **Observing how performance practices relate to musical materials can help us develop models of "expressive" performance**

Introduction

What do I mean by studying performance?

- **Using (live) recorded performances**
- **Measuring performance parameters**
  - timing
  - dynamics
  - **tuning**
  - timbre
- **Assessing relationship between performance of various parameters and musical materials**

Introduction
Motivations and challenges.

1

A brief history
Quantitative approaches to performance analysis.

2

Extracting Performance Data
MIDI-audio alignment for automatic analysis of recorded performances.

3

Experiments
Studies of intonation in the singing voice.

4

Developing a Representation of Symbolic Music
Comparing performances of different pieces.

5

Conclusions

6

Quantitative Performance Analysis

A brief history

Pioneers
Binet and Courtier
Sears
Miller

1895–1930

1920–40s

University of Iowa
Seashore and colleagues

Ethnomusicology
Charles Seeger

1960s

1980s and 90s

Piano
Gabrielsson
Todd
Clarke
Repp

Intonation
Fyk
Prame
Vurma

1990s and 2000s

Computational Models
Friberg
Mazola
Widmer

Introduction
Motivations and challenges.
1

A brief history
Quantitative approaches to performance analysis.
2

Extracting Performance Data
MIDI-audio alignment for automatic analysis of recorded performances.
3

Experiments
Studies of intonation in the singing voice.
4

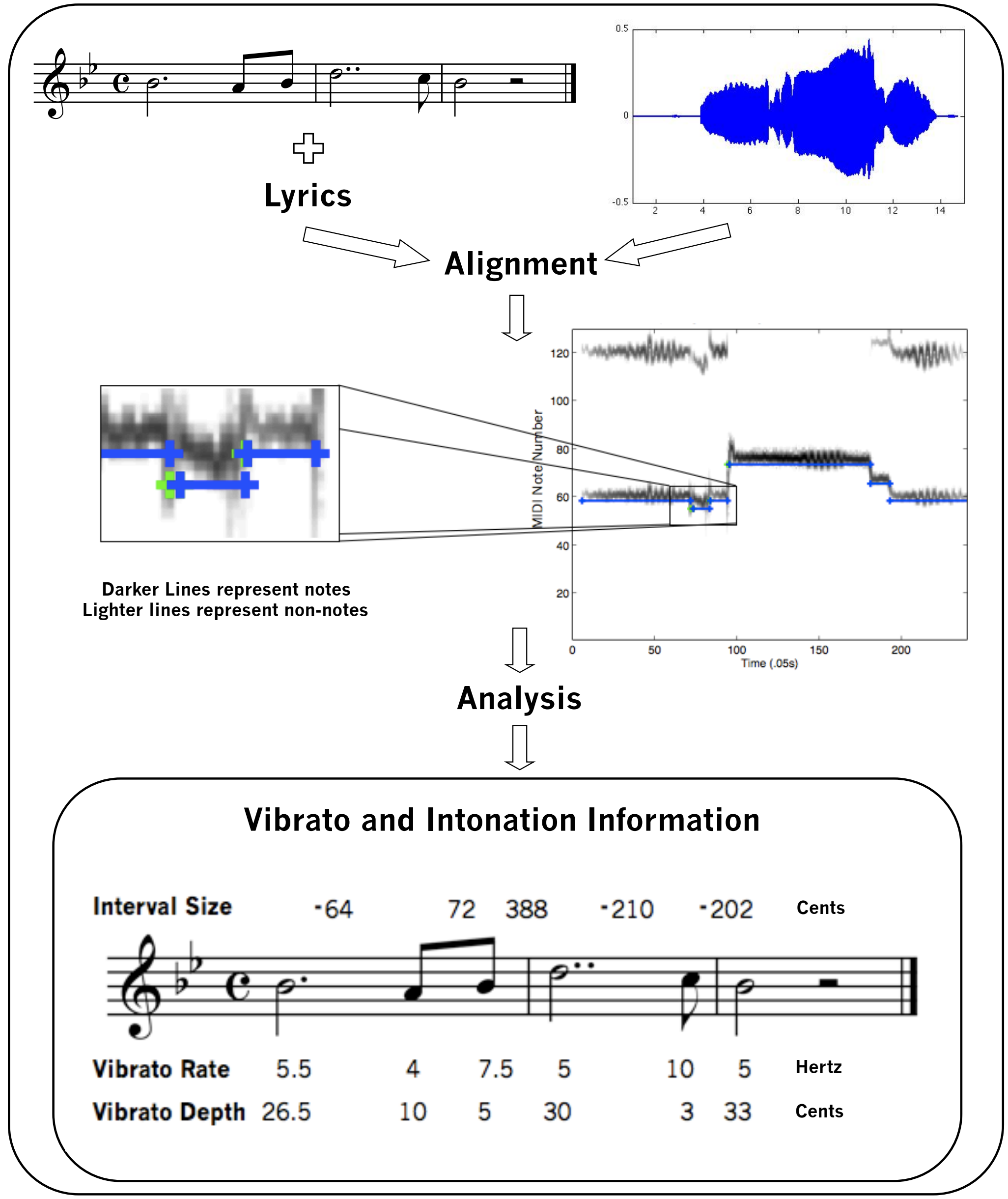
Developing a Representation of Symbolic Music
Comparing performances of different pieces.
5

Conclusions
6

AMPACT

Automatic Music Performance and Comparison Toolkit

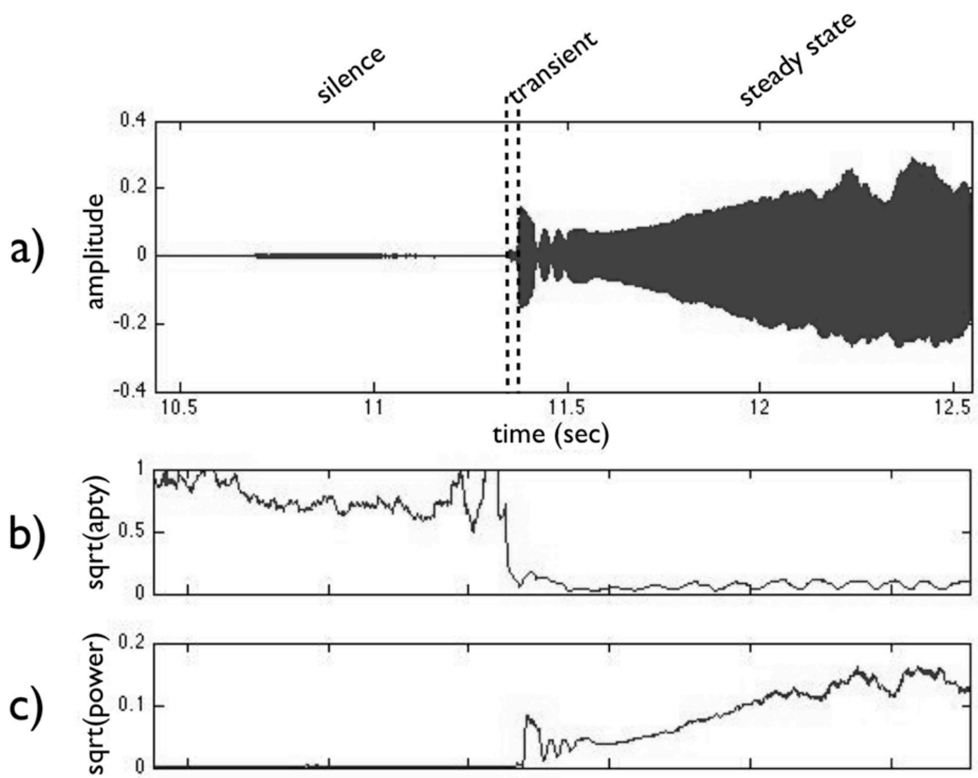
www.ampact.org



Monophonic audio

Identifying onsets and offsets

- **Multi-pass dynamic time warping (DTW)/hidden Markov model (HMM) algorithm**
- **HMM Observations: Periodicity, Power, and F_0**

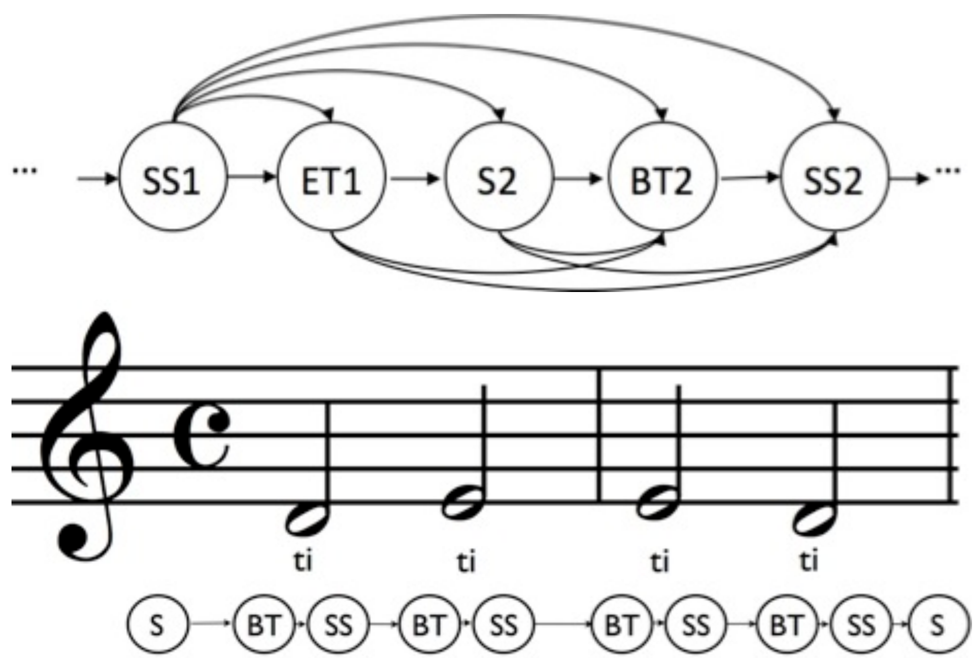


Devaney, Mandel, and Ellis (2009)

Monophonic audio

Identifying onsets and offsets

- **DTW used as prior to guide HMM**
- **HMM state path constrained by lyrics**

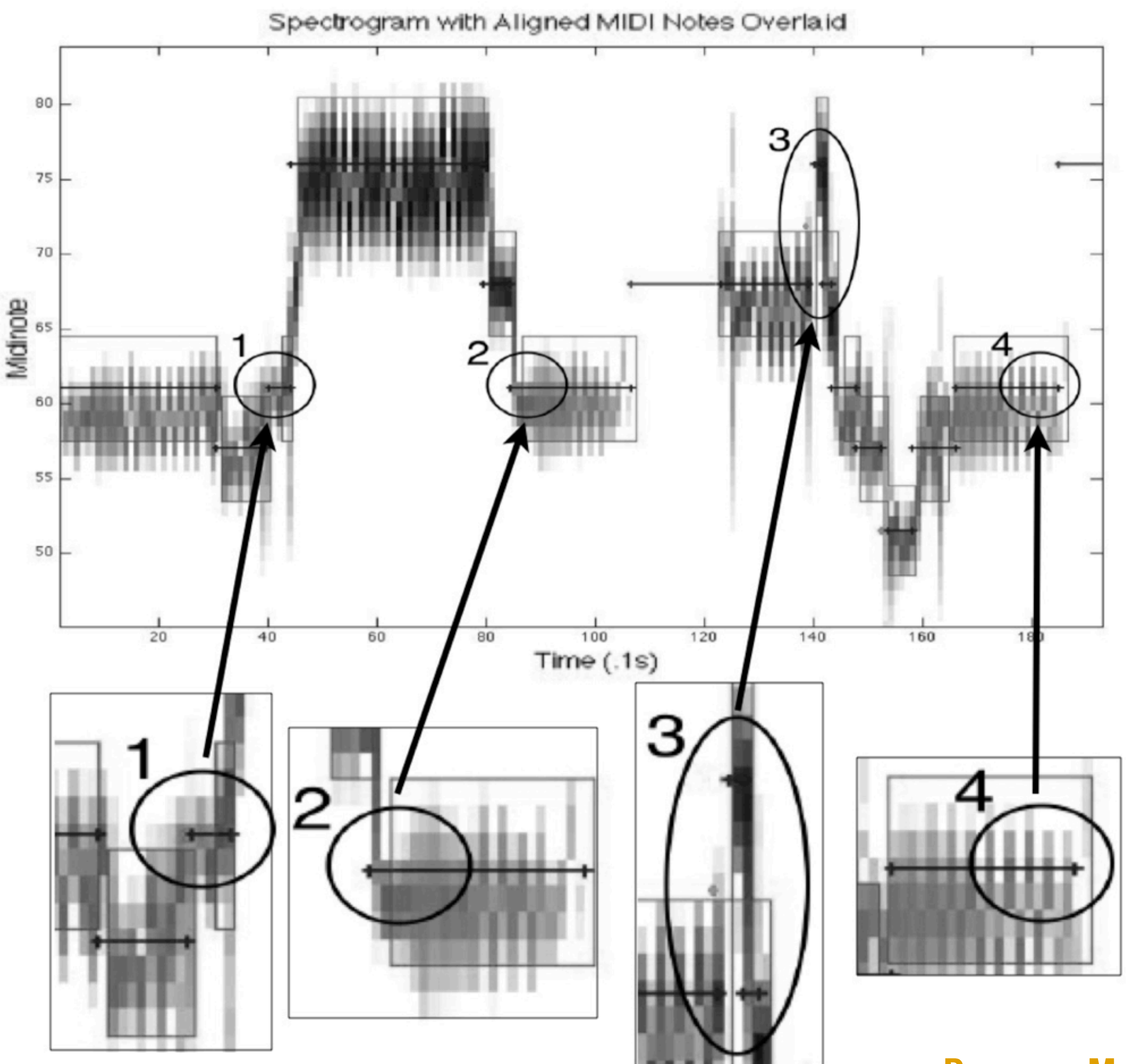


- **Improves median alignment error from 52 ms to 28 ms**

Devaney, Mandel, and Ellis (2009)

Monophonic audio

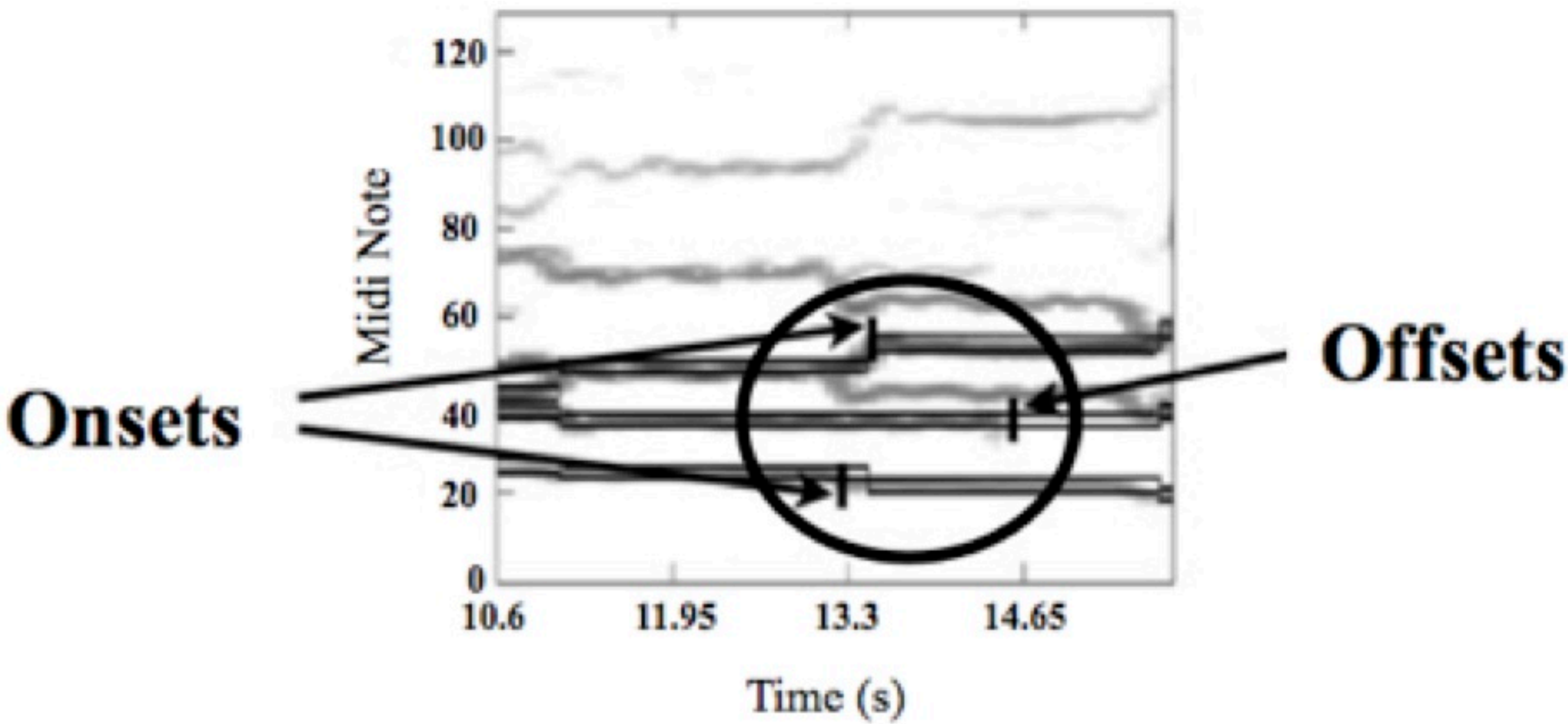
Identifying onsets and offsets



Devaney, Mandel, and Ellis (2009)

Polyphonic audio

Identifying asynchronies between voices



Devaney and Ellis (2009)

Polyphonic audio

Identifying asynchronies between voices

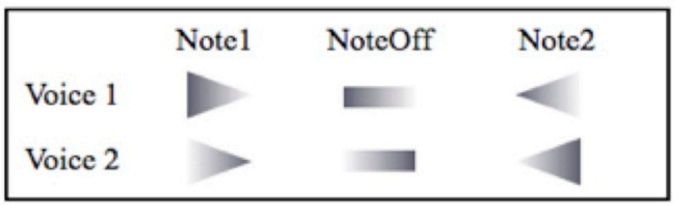
- **Multi-pass DTW/HMM algorithm**
- **DTW determines general note transitions**
  - providing a single offset/onset location for all of the musical lines
- **HMM finds the location of each line's onsets and offsets within a +/- 125 ms window around the DTW estimate**

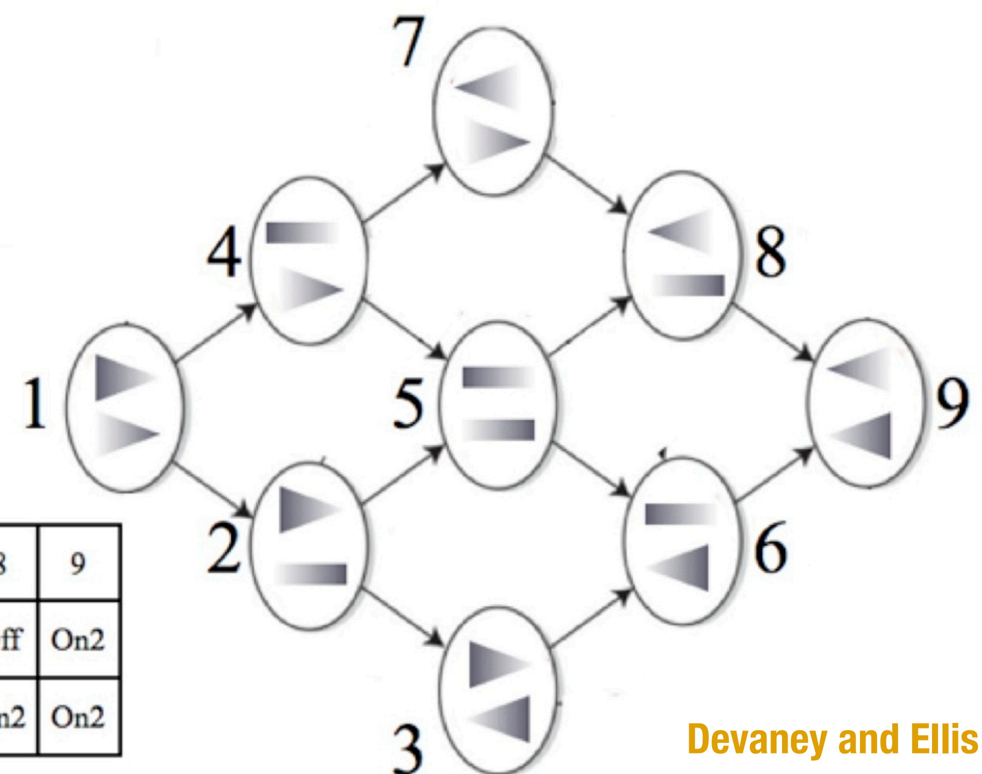
Devaney (2014)

Polyphonic audio

Identifying asynchronies between voices

- **HMM States: Note 1, Note Off, and Note 2 for each line**
  - number of states is 3*N* (where *N* is the number of lines)





State	1	2	3	4	5	6	7	8	9
Voice 1	On1	Off	On1	On2	Off	On1	On2	Off	On2
Voice 2	On1	On1	Off	On1	Off	On2	Off	On2	On2

Devaney and Ellis (2009)
Devaney (2014)

Polyphonic audio

Identifying asynchronies between voices

- **HMM Observations: power measurements from a constant-Q filter bank decomposition of the signal**
    - the power measurement is summed over a 3-semitone span around the fundamental of the ending and starting notes in each line in the DTW alignment
- **Improves median alignment for onsets from 118 ms to 77 ms for onsets and for offsets from 75 ms to 69 ms**

Devaney (2014)

Introduction 1
Motivations and challenges.

A brief history 2
Quantitative approaches to performance analysis.

Extracting Performance Data 3
MIDI-audio alignment for automatic analysis of recorded performances.

Experiments 4
Studies of intonation in the singing voice.

Developing a Representation of Symbolic Music 5
Comparing performances of different pieces.

Conclusions 6

Experiments with Performers

Overview

- Intonation in trained singers in the Western Art Music tradition
- Solo and small ensemble (2-4 voices)

Experiments with Performers

Why study the singing voice?

- In its most basic form singing is innate and universal
    - Training and enculturation refine specific practices of singing
- The voice is one of the most expressive instruments
- Singing research is complementary to speech research

Experiments with Performers

Research questions

- **Intonation data analyzed in regards to**
  - Tuning systems
  - Direction (ascending versus descending)
  - Musical context
  - *Effect of training*

Solo Singing

Overview

- **Schubert's "Ave Maria"**
  - 3x a cappella & 3x accompanied
- **12 solo singers**
  - 6 non-professional singers: undergraduate vocal majors
  - 6 professional singers: possess at least one graduate-level degree in voice performance
- **Melodic semitones and whole tones**

Devaney, Mandel, Ellis and Fujinaga (2011)
Devaney, Wild, and Fujinaga (2011)

Solo Singing

Significant trends

- **TUNING SYSTEMS**
  - No strict adherence, on average smaller than equal temperament (more so for semitones than whole tones)
- **DIRECTION**
  - Ascending semitones were 7–8 cents larger on average than descending semitones
- **MUSICAL CONTEXT**
  - Non-pros tended to compress leading tones
- **EFFECT OF TRAINING**
  - Pros were more consistent with one another
  - Pros’ semitones were 6 cents larger on average
  - Non-pros’ accompanied semitones were 3 cents larger than *a cappella* semitones

Three-Part Singing

Overview

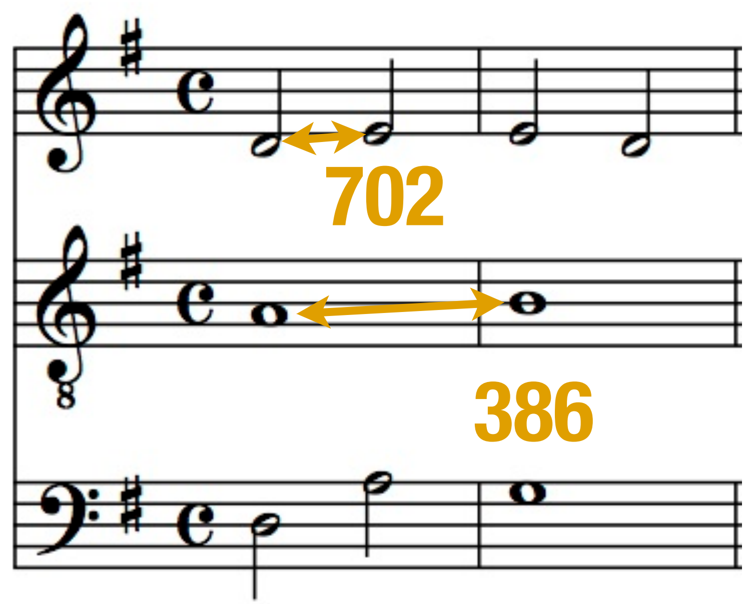
- Chord progression by Giambattista Benedetti
- 4 ensembles
- Melodic whole tones

Devaney, Mandel, and Fujinaga (2012)

Three-Part Singing

Significant trends

- **TUNING SYSTEMS:** No strict adherence, generally closer to equal temperament
- **DIRECTION:** no significant difference
- **MUSICAL CONTEXT:** melodic whole tones sung over a P5 were 15 cents larger on average than those sung over a M3



Four-Part Singing

Overview

- Praetorius' "Es ist ein Ros entsprungen"
- 3 ensembles
- Melodic semitone and whole tone intervals
- Vertical intervals in cadential contexts

Four-Part Singing

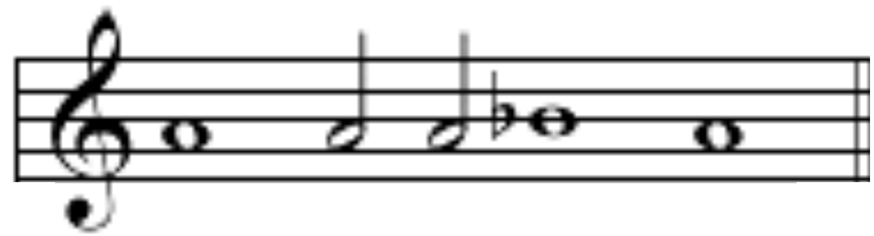
Significant trends

- **TUNING SYSTEMS**
  - No strict adherence, on average smaller than equal temperament (more so for semitones than whole tones)
- **DIRECTION**
  - Semitones – only one ensemble showed a significant difference (ascending 8 cents larger)
  - Whole tones – ascending 4 cents smaller
- **MUSICAL CONTEXT**
  - Melodic intervals – no effect of leading tone function
  - Vertical intervals in cadential contexts were significantly closer to Just Intonation than those in non-cadential contexts

Two-Part Singing

Overview

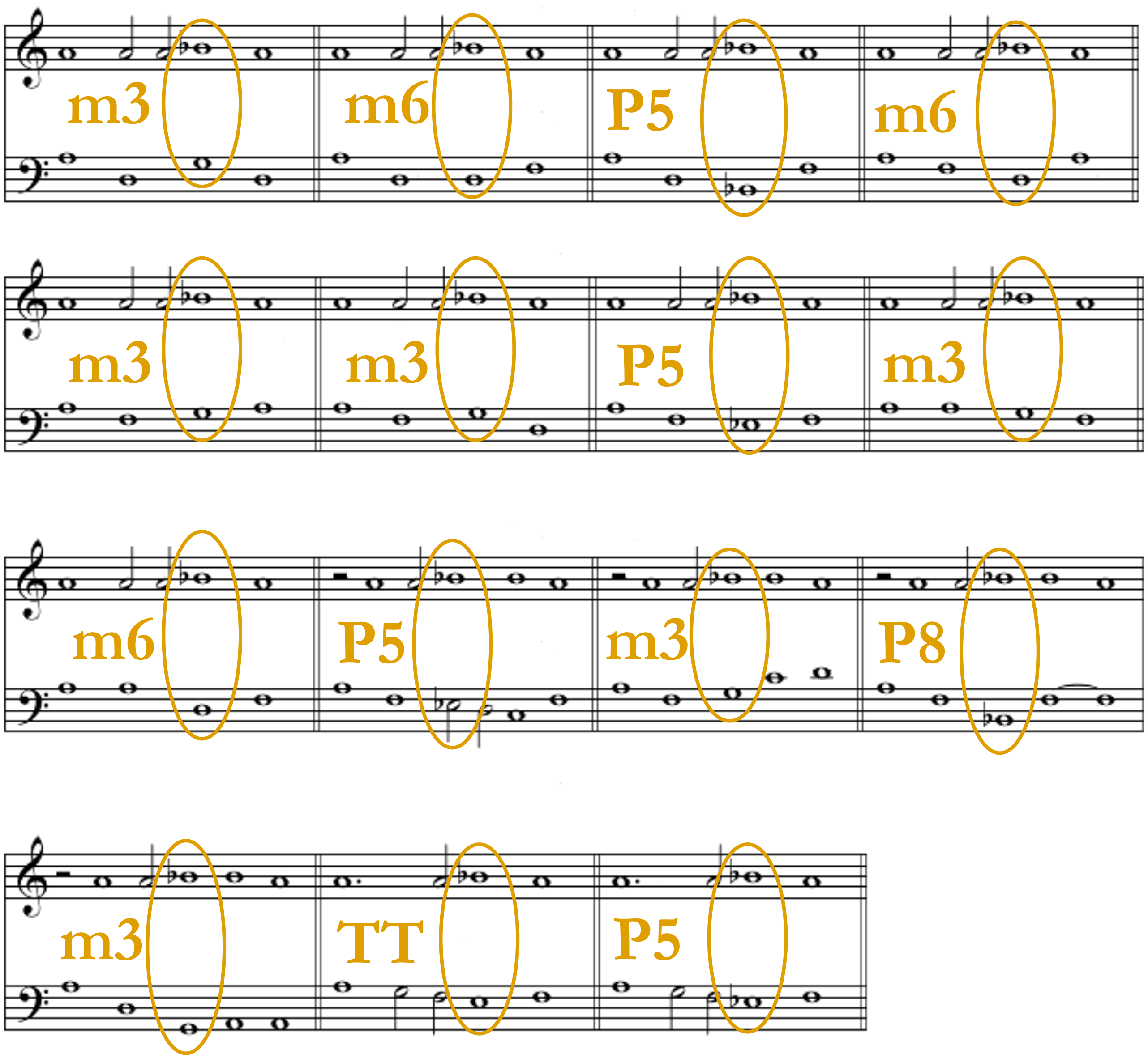
- **Semitone pattern sung against a recorded version of the lower-line that was tuned in three different systems at two pitch heights**



- **6 of 12 subjects *(analysis of remaining 6 subjects ongoing)***
    - 3 non-professionals: amateur singers
    - 3 professionals: possess at least one graduate-level degree in voice performance
- **Melodic semitones in vertical m3, TT, P5, m6, and P8 contexts**

Two-Part Singing

Exercises



Two-Part Singing

Significant trends

- **TUNINGS SYSTEM:** No strict adherence, on average smaller than equal temperament
- **DIRECTION:** Ascending semitones were 21 cents larger on average than descending semitones
- **EFFECT OF TRAINING:** Non-pros' semitones were 17 cents smaller on average than pros' semitones
- **DETUNING:** no significant effect
- **VERTICAL INTERVAL CONTEXT:** Semitones sung a perfect octave above the lower voice were 7 cents larger on average than those sung above other intervals
    - no significant differences for other intervals

Summary of Results

Solo vs. ensemble singing

- No overall adherence to a tuning system was observed
- A general trend of ascending semitones being larger than descending intervals was found in both solo and ensemble singing
- Results are variable for influence of specific vertical intervals on melodic intonation
    - 3-part experiment – melodic intervals sung over a P5 versus M3 showed a significant difference
    - 2-part experiment – melodic intervals only showed a significant difference when sung over a P8
    - Detuning of accompaniment did not influence melodic intonation in the short exercises studied

Next Steps

Where to go from here

- **Perform experiments on larger collections of recordings**
  - Develop more robust tools for automatic extraction of performance data from recordings
    - making the current tools more reliable and more accessible to other researchers (crowd-sourcing to improve algorithms)
  - Develop a representation of symbolic music for making automatic comparisons between different pieces

Introduction
Motivations and challenges.
1

A brief history
Quantitative approaches to performance analysis.
2

Extracting Performance Data
MIDI-audio alignment for automatic analysis of recorded performances.
3

Experiments
Studies of intonation in the singing voice.
4

Developing a Representation of Symbolic Music
Comparing performances of different pieces.
5

Conclusions
6

Representing Symbolic Music

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
- **This is useful for automatically determining similarities between different pieces**

Representing Symbolic Music

Inspiration from Speech Recognition

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

Representing Symbolic Music

N-grams

- **Large-scale music analysis approaches are heavily influenced by text retrieval methods, namely N-grams**
- **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)

Representing Symbolic Music

N-grams

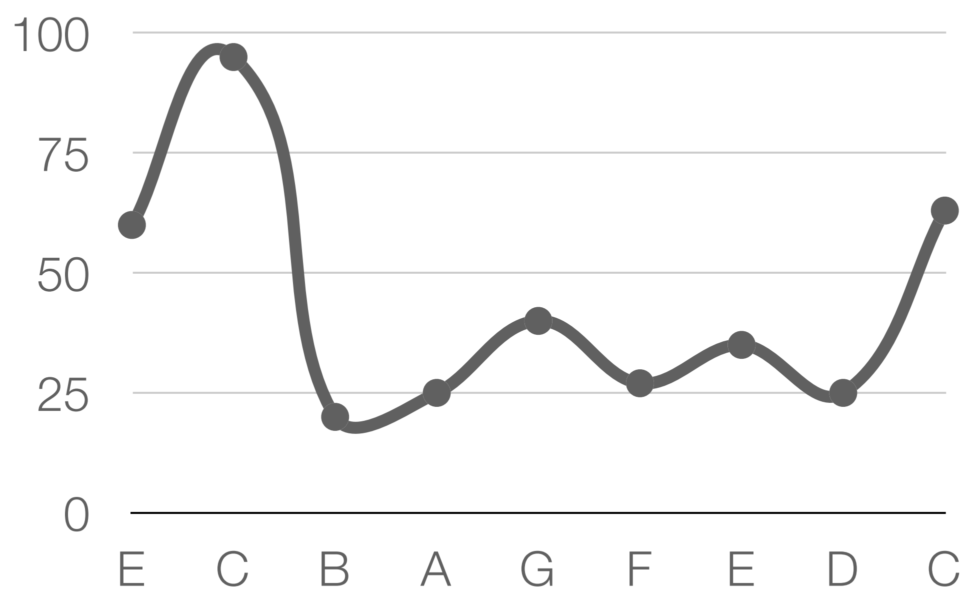
- **N-gram representations encounter problems with more complex textures**
  - e.g., where the notes of chords are not played simultaneously

- **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

Representing Symbolic Music

What might this look like?



“Language Model” for Music

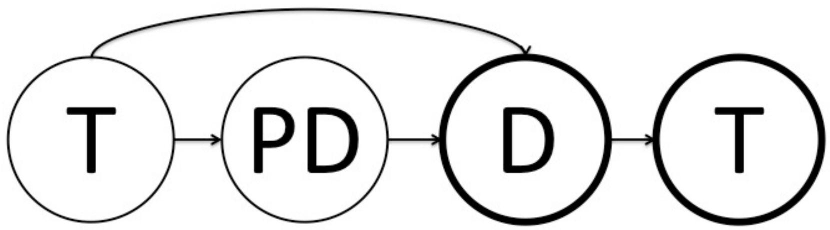
Using musical function

- **A model of phrase-level function and its relationship to roman numeral labels can be used as a “language model”**
- **Phrases are musical statements built from the ordered presentation of three harmonic functions**
  - tonic, pre-dominant, and dominant functions
- **Phrases end with a cadence**
  - remain on the dominant function for a half cadence
  - return to the tonic function for an authentic or a deceptive cadence

“Language Model” for Music

Pilot study

- **Hidden Markov model (HMM)**
  - State space



- **Used chord label and function for all examples from Laitz (2011) to train HMM transition probabilities and evaluate model**

- **Evaluation**
  - 80/20 split of textbook data: 94.3% overall accuracy
    - tonic: 93%, pre-dominant: 93%, dominant: 89%

“Language Model” for Music

Next steps

- Given the surface, jointly infer the chords and the functions
- Compute various features of the musical surface that capture pitch and metrical information
- Use the phrase model to constrain the space and sequence of possible chords

Introduction
Motivations and challenges. 1

A brief history
Quantitative approaches to performance analysis. 2

Extracting Performance Data
MIDI-audio alignment for automatic analysis of recorded performances. 3

Experiments
Studies of intonation in the singing voice. 4

Developing a Representation of Symbolic Music
Comparing performances of different pieces. 5

Conclusions 6

Summary

Where we have been

- **This talk has**
  - provided a brief overview of the history of quantitative performance analysis
  - discussed some of the challenges of automatically extracting performance data from recordings
  - summarized some of my recent work on vocal intonation practices in the western art music tradition
  - introduced an ongoing project on developing a representations of symbolic music that highlights structurally significant aspects of the surface

Acknowledgements

- School of Music and College of Arts and Sciences (OSU)
- Center for New Music and Audio Technologies (CNMAT)
- Distributed Digital Music Archives and Libraries (DDMAL)
- Centre for Research in Music Media and Technology (CIRMMT)
- Fonds de recherche sur la société et la culture (FQRSC)
- Social Sciences and Humanities Research Council of Canada (SSHRC)
- Advancing Interdisciplinary Research in Singing (AIRS)

Thank you!

References

Devaney, J. 2014. Estimating onset and offset asynchronies in polyphonic score-audio alignment. *Journal of New Music Research.* (In press).

Devaney J., M. Mandel, and D. Ellis. 2009. Improving MIDI-audio alignment with acoustic features. In *Proceedings of WASPAA*. 45–8.

Devaney, J., M. Mandel, D. Ellis, and I. Fujinaga. 2011. Automatically extracting performance data from recordings of trained singers. *Psychomusicology: Music, Mind and Brain* 21 (1–2).

Devaney, J., M. I. Mandel, and I. Fujinaga. 2012. Study of Intonation in Three-Part Singing using the Automatic Music Performance Analysis and Comparison Toolkit (AMPACT). *Proceedings of ISMIR*. 511–6.

Devaney, J., J. Wild, and I. Fujinaga. 2011 Intonation in solo vocal performance: A study of semitone and whole tone tuning in undergraduate and professional sopranos. In *Proceedings of the International Symposium on Performance Science*. 219–24.

Summary of Results

Comparison to earlier work

- **Schoen (1922) - solo**
  - sharper than equal temperament ✗
  - ascending intervals larger than descending intervals ✓
- **Prame (1997) - solo**
  - deviation from equal temperament ✓
- **Jers and Ternstrom (2006) - ensemble**
  - ascending intervals larger than descending intervals ✓
- **Vurma and Ross (2006) - solo**
  - ascending/descending semitones smaller than EQT ✓
- **Howard (2007a, 200b) - ensemble**
  - tendency towards Just Intonation ✗ ✓
- **Vurma (2010) - 2-part with synthesized lower voice**
  - singers' intonation did not change significantly when the synthesized voice was detuned ✓

Monophonic alignment

DTW prior

- **A rectangular window with half a Gaussian is placed on on each side over the DTW note position estimates**

	5% start	100% start	100% end	5% end
Silence (and Breath)	50% btwn N-1 On and N-1 Off	N-1 Off	N On	50% btwn N On and N Off
Opening Transient	N-1 Off	75% btwn N-1 Off and N On	25% btwn N On and N Off	N Off
Steady State	N-1 Off	N On	N Off	N+1 On
Closing Transient	N On	75% btwn N On and N Off	25% btwn N Off and N+1 On	N+1 On