Tuning In: Analyzing Recorded Vocal Performances

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

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.

Conclusions 5
Summary and future directions.

Introduction

Why study musical performance?

- **Performances convey musicians’ interpretations**
- **Performances are what listeners actually hear**
- **Studying performance can help us gain insight into**
  - the available range for extended techniques
  - 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**

Quantitative Performance Analysis

A brief history

Pioneers
Binet and Courtier
Sears
Miller

1895–1930 **1920–40s** **1960s** **1980s and 90s** **1990s and 2000s**

Quantitative Performance Analysis

A brief history

Pioneers
Binet and Courtier
Sears
Miller

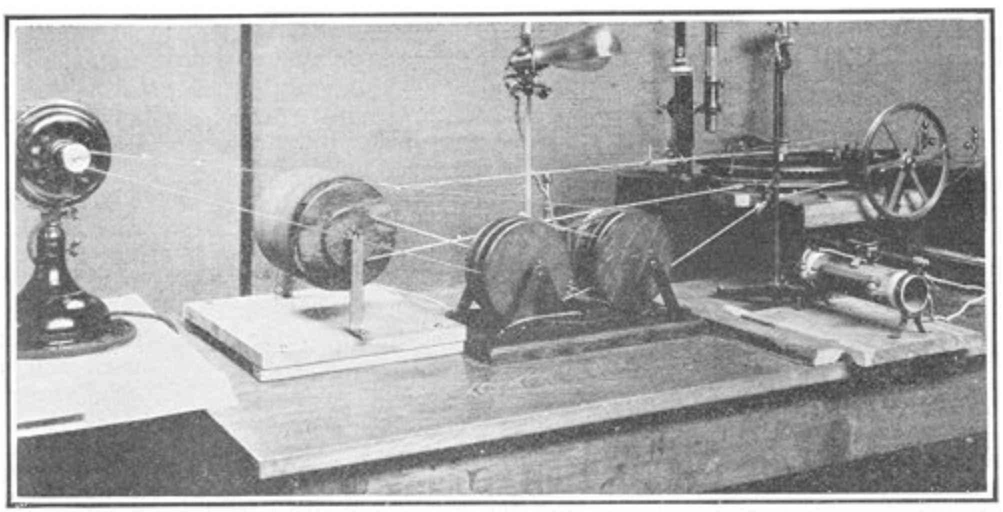
1895–1930 | **1920–40s** | **1960s** | **1980s and 90s** | **1990s and 2000s**

University of Iowa
Seashore and colleagues

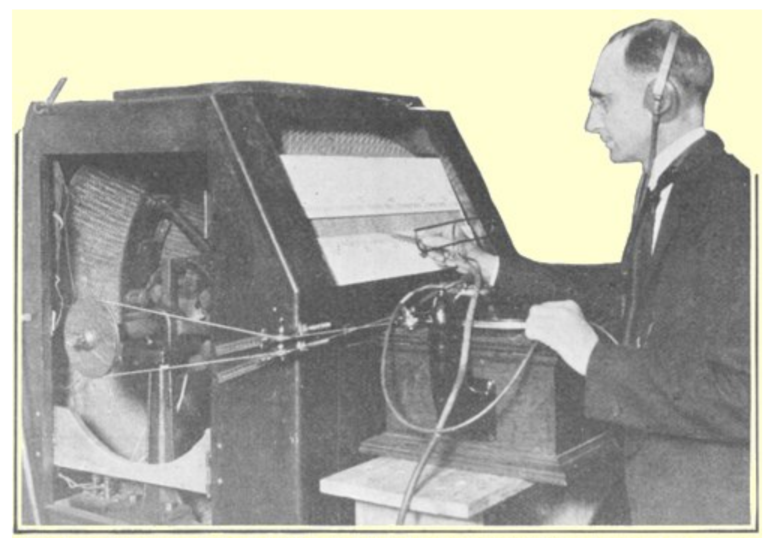
Quantitative Performance Analysis

University of Iowa

- **Carl Seashore (1938) and colleagues studied timing, dynamics, intonation, and vibrato in pianists, violinists, and singers**
  - Equipment: piano rolls, films of the movement of piano hammers during performance, phono-photographic apparati



Wave recorder for use with disk phonograph; the lever, acting like a pantograph, traces the waves on a revolving smoked drum

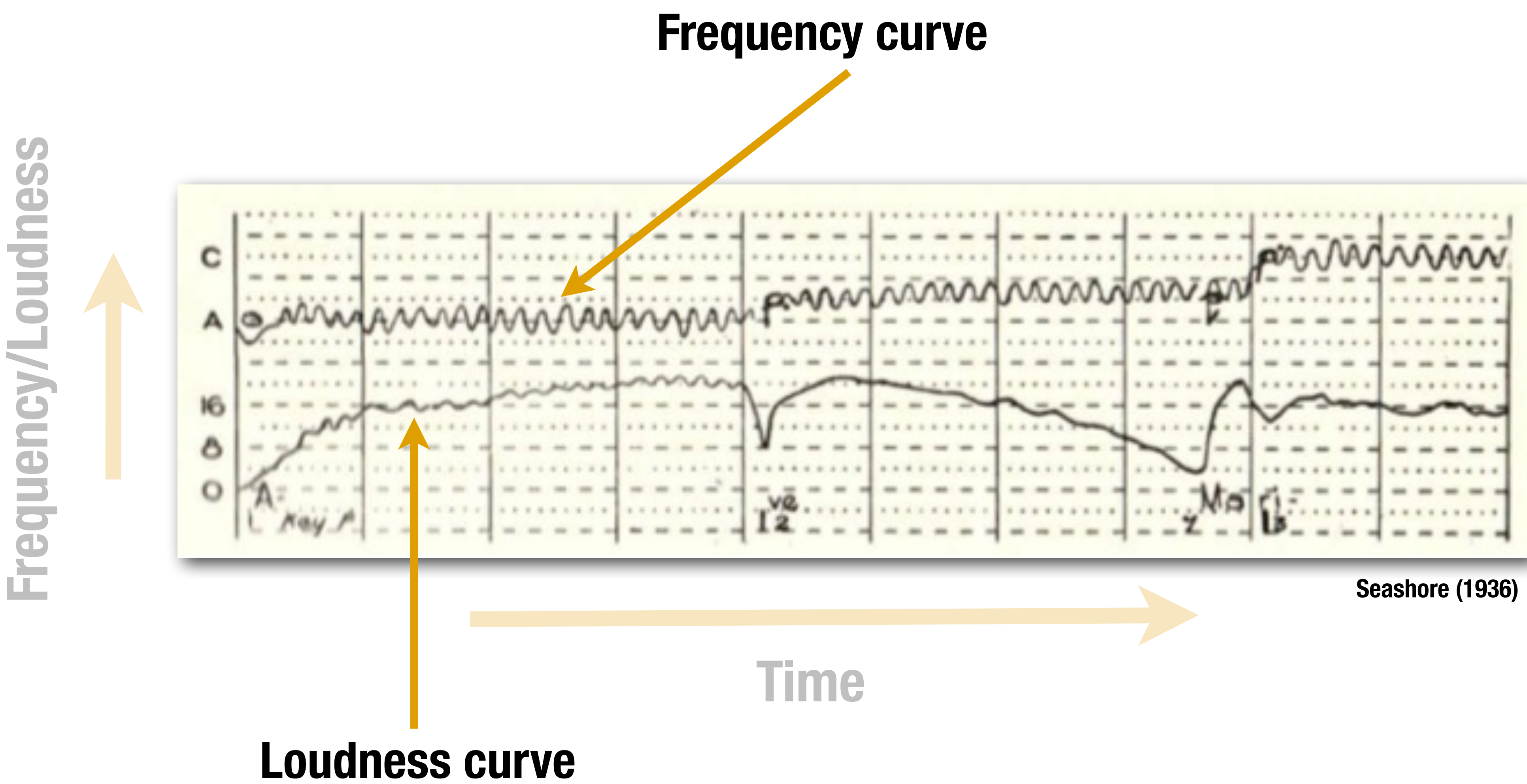


The tonoscope for analyzing the pitch of the tones on a disk phonograph record

Cary (1922)

Performance Scores

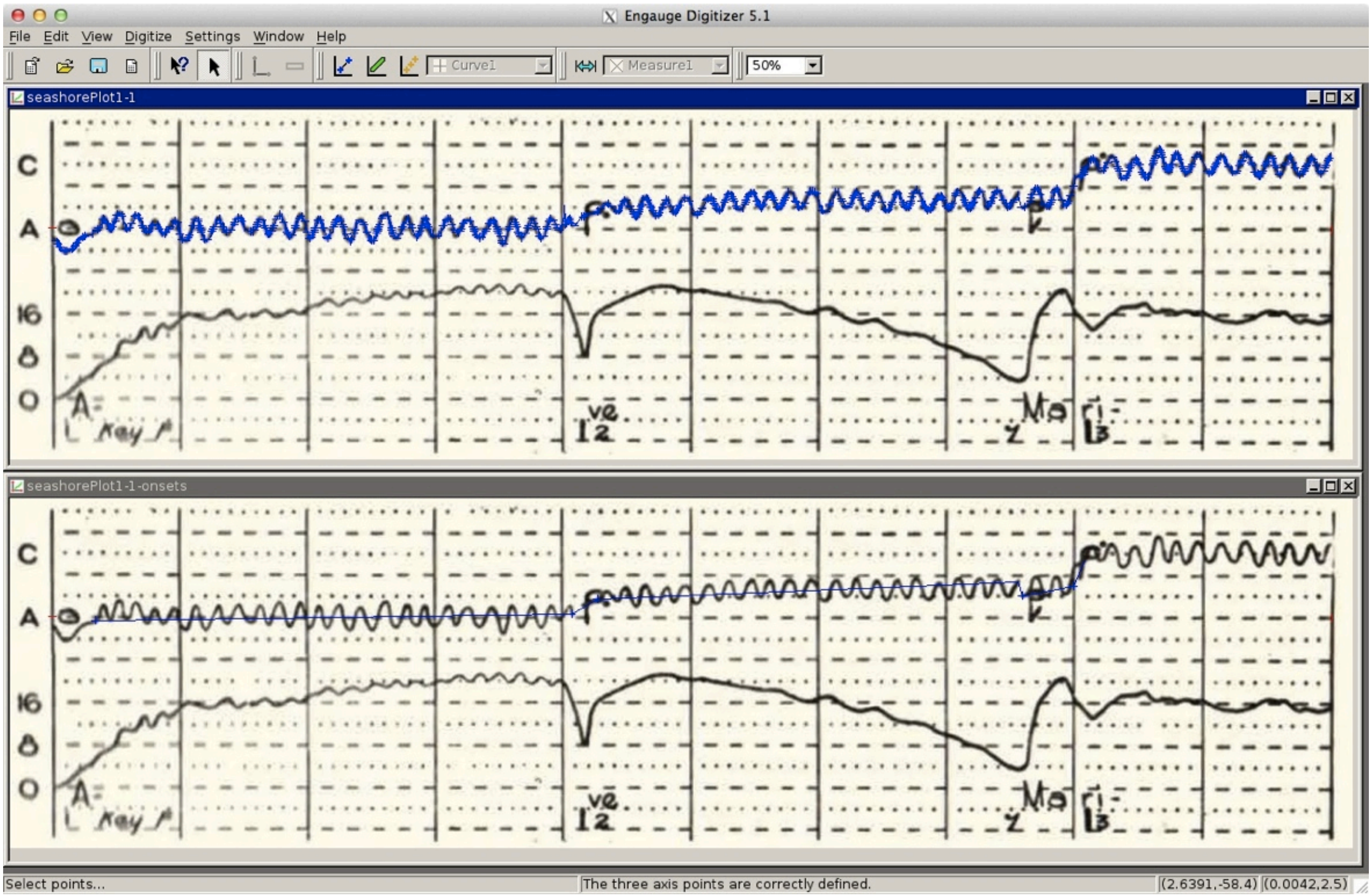
University of Iowa



Seashore (1936)

Performance Scores

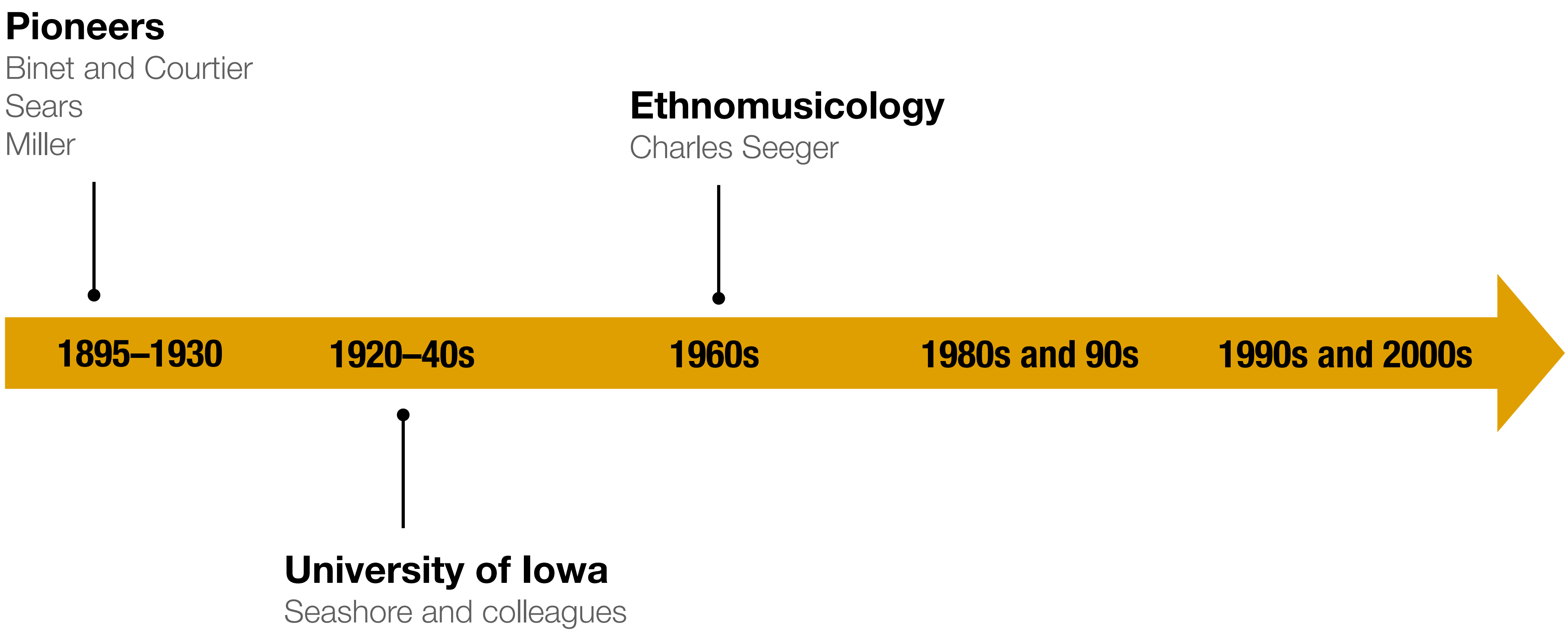
Digitizing the data



Devaney (Under Reivew)

Quantitative Performance Analysis

A brief history



Quantitative Performance Analysis

A brief history

1895–1930
Pioneers
Binet and Courtier
Sears
Miller

1920–40s
University of Iowa
Seashore and colleagues

1960s
Ethnomusicology
Charles Seeger

1980s and 90s
Piano
Bengtsson and Gabrielsson
Todd
Clarke
Repp

1990s and 2000s

Quantitative Performance Analysis

Popularity of the piano

- Large amount of solo repertoire
- Instrument's percussive nature
- Feasibility of using specially equipped pianos (e.g., MIDI)
  - cannot study existing recordings
  - new recordings are typically done in a lab environment

Bosendorfer SE piano at BRAMS, Montreal

Quantitative Performance Analysis

A brief history

1895–1930

Pioneers
Binet and Courtier
Sears
Miller

1920–40s

University of Iowa
Seashore and colleagues

1960s

Ethnomusicology
Charles Seeger

1980s and 90s

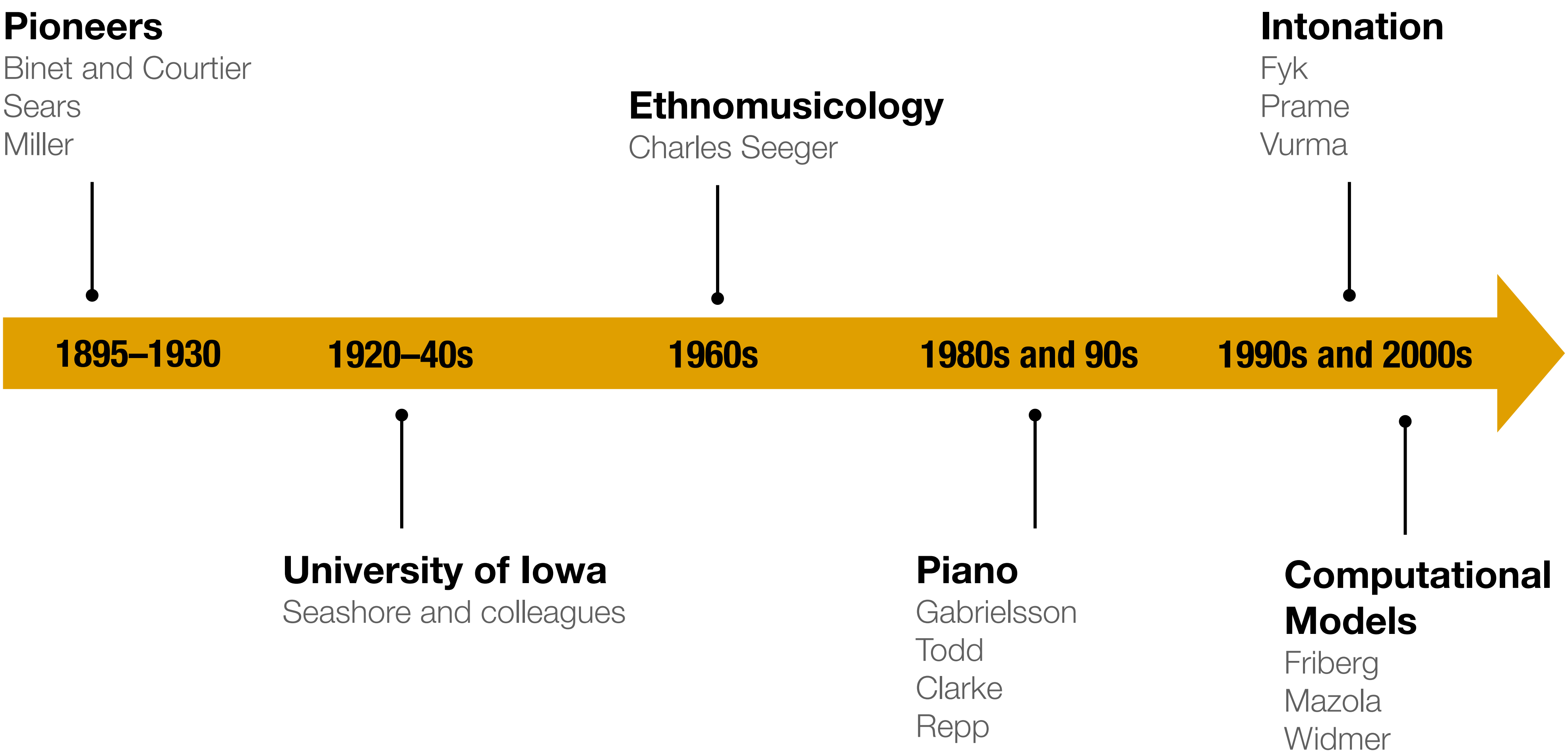
Piano
Gabrielsson
Todd
Clarke
Repp

1990s and 2000s

Intonation
Fyk
Prame
Vurma

Quantitative Performance Analysis

A brief history



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

Conclusions
Summary and future directions.

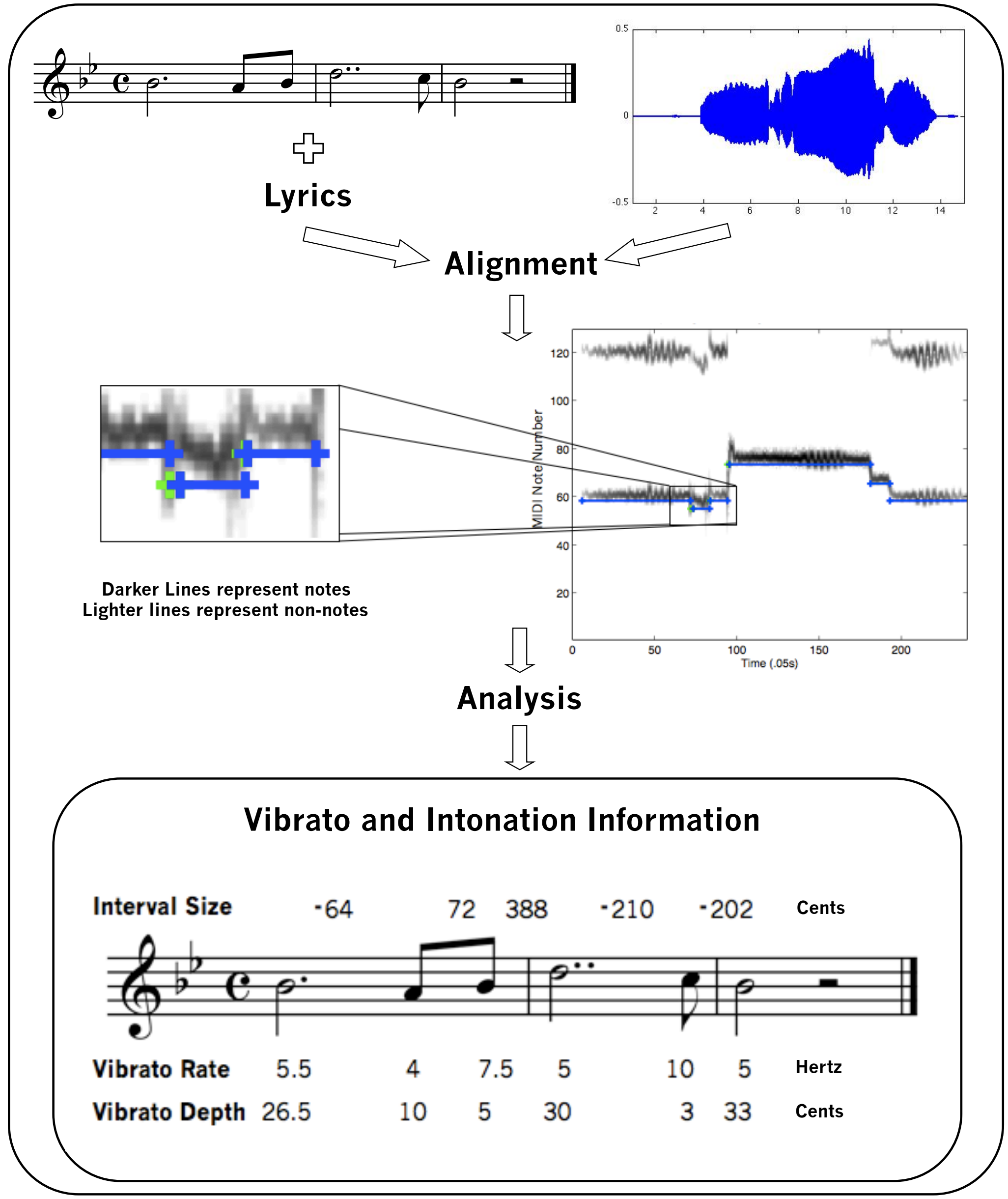
5

AMPACT

Automatic Music Performance Analysis and Comparison Toolkit

AMPACT

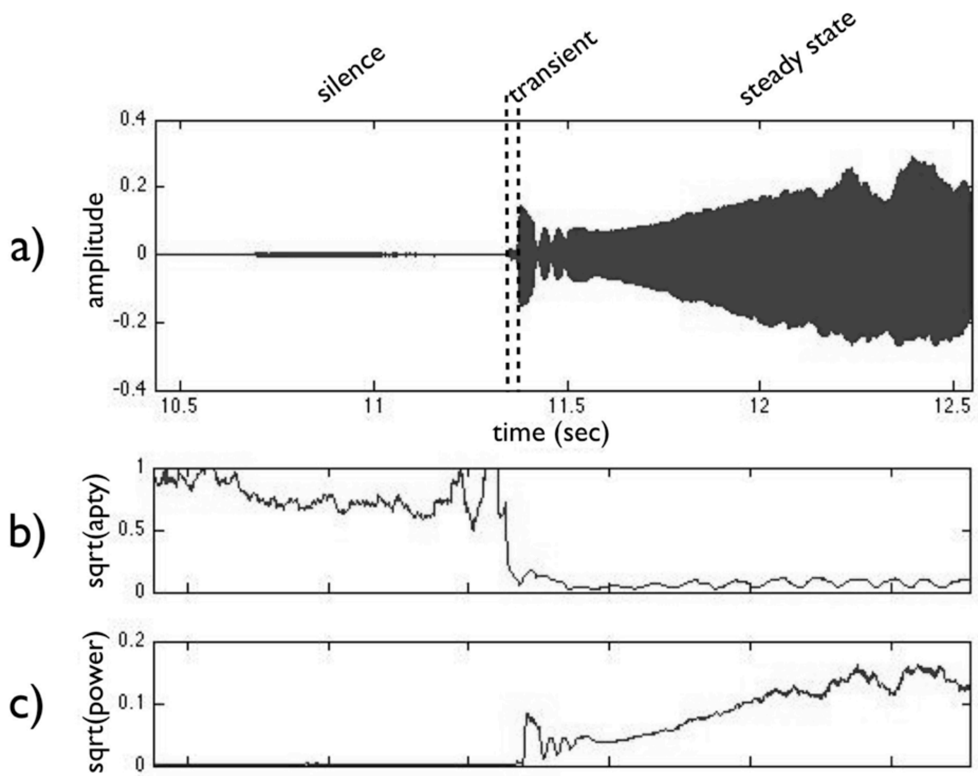
www.ampact.org



Monophonic audio

Identifying onsets and offsets

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



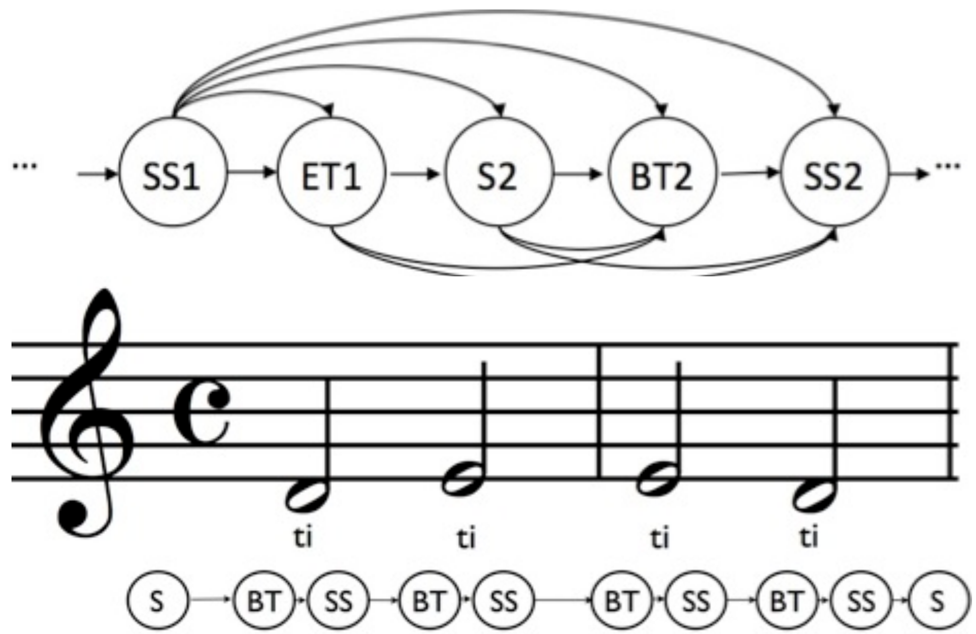
Devaney, Mandel, and Ellis (2009)

Monophonic audio

Identifying onsets and offsets

- **Dynamic time warping alignment (based on Orio and Schwarz, 2001) used as prior to guide a hidden Markov model**

- **HMM state path constrained by lyrics**

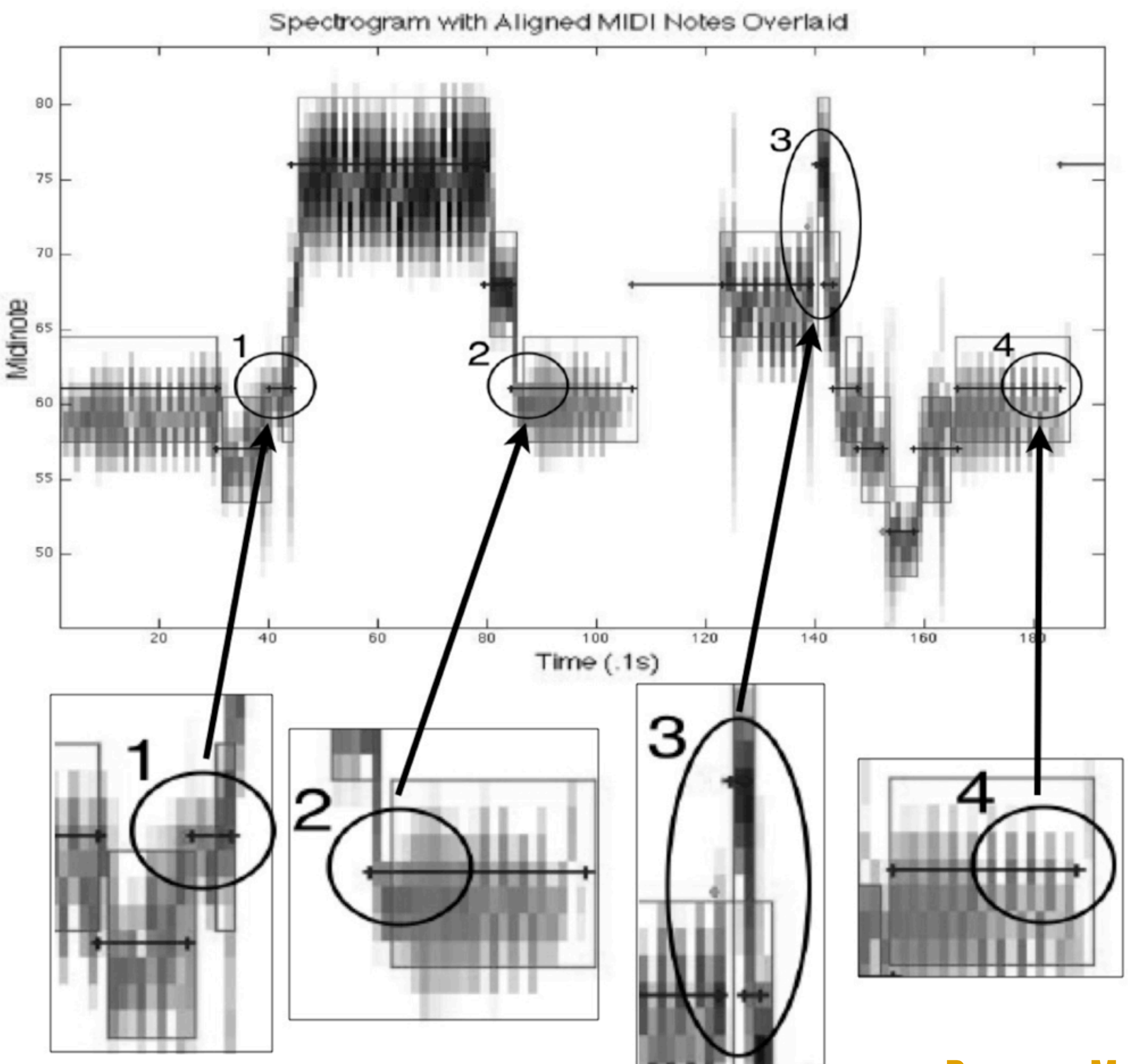


- **Improves median onset estimation error from 52 ms to 28 ms**

Devaney, Mandel, and Ellis (2009)

Monophonic audio

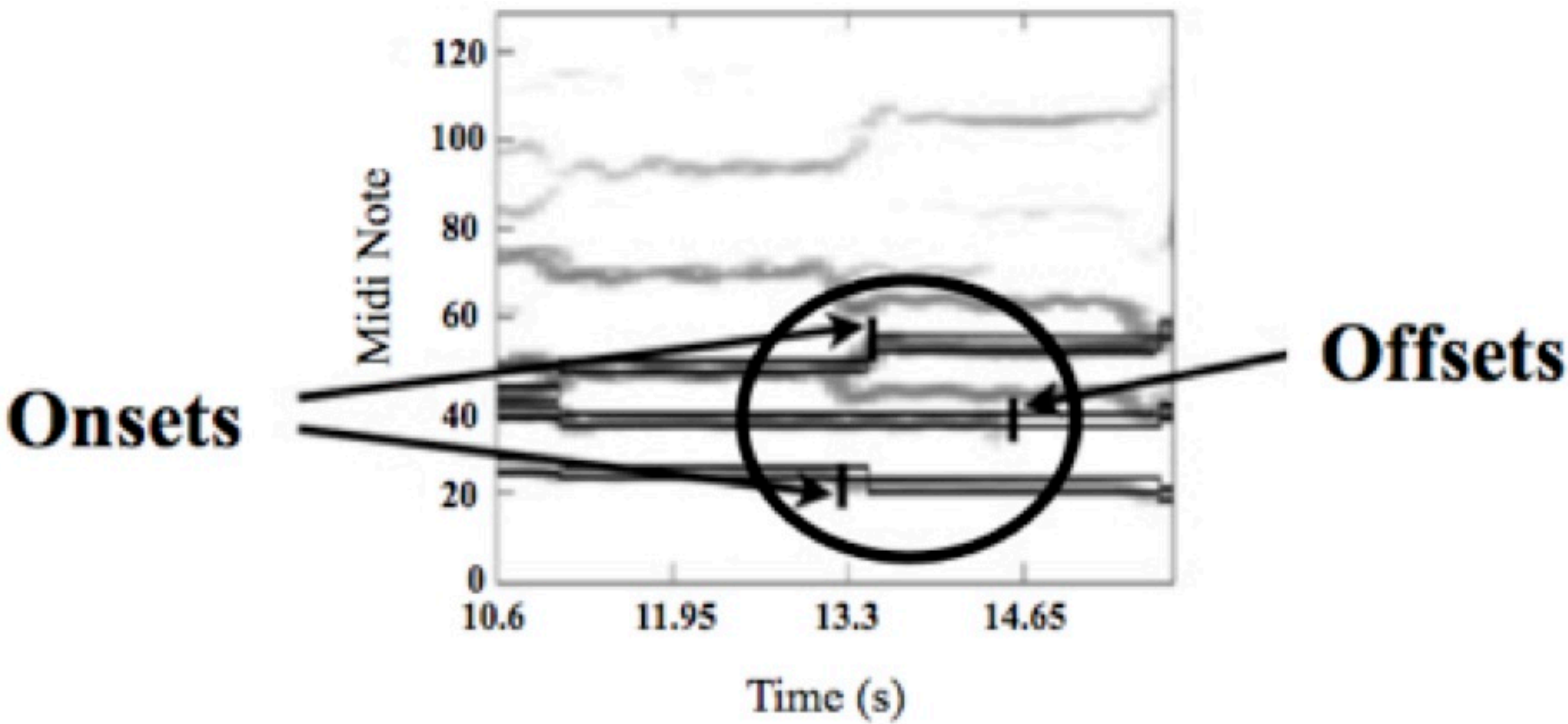
Identifying onsets and offsets



Devaney, Mandel, and Ellis (2009)

Polyphonic audio

Identifying asynchronies between voices



Devaney and Ellis (2009)

Polyphonic audio

Identifying asynchronies between voices

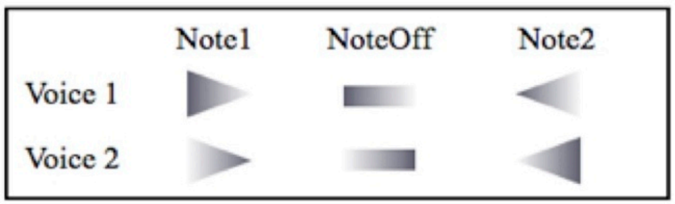
- **Also a multi-pass DTW/HMM algorithm**
- **DTW determines general note transitions**
  - providing a single offset/onset location for all of the musical lines
- **HMM finds the location of each line's onsets and offsets within a +/- 125 ms window around the DTW estimate**

Devaney (2014)

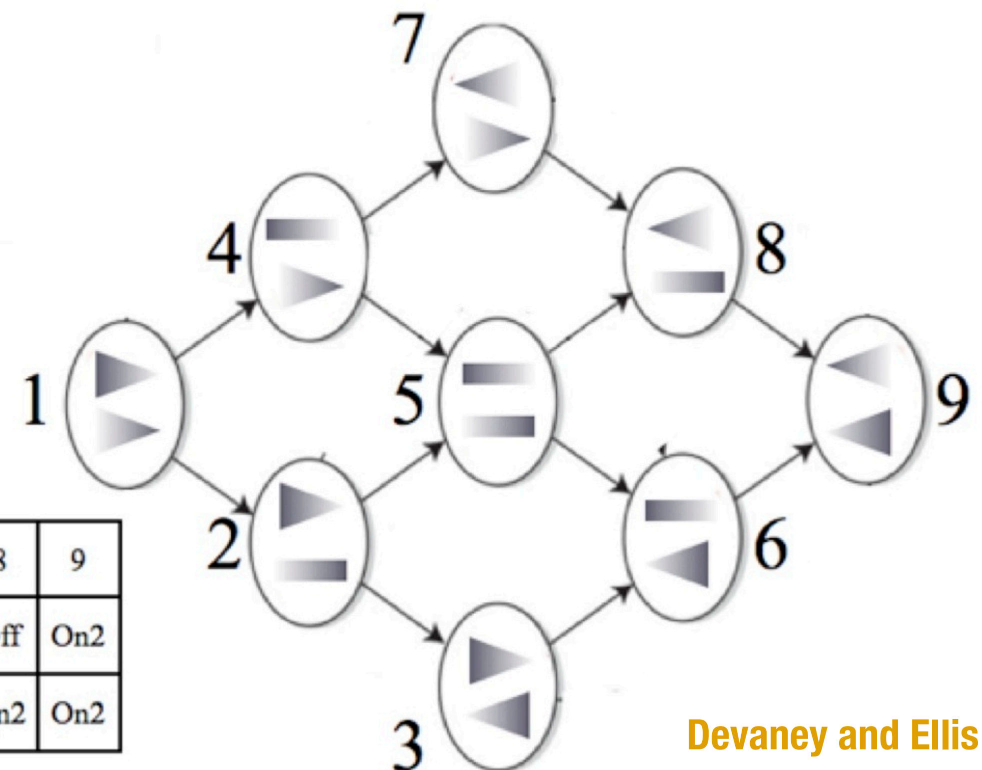
Polyphonic audio

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

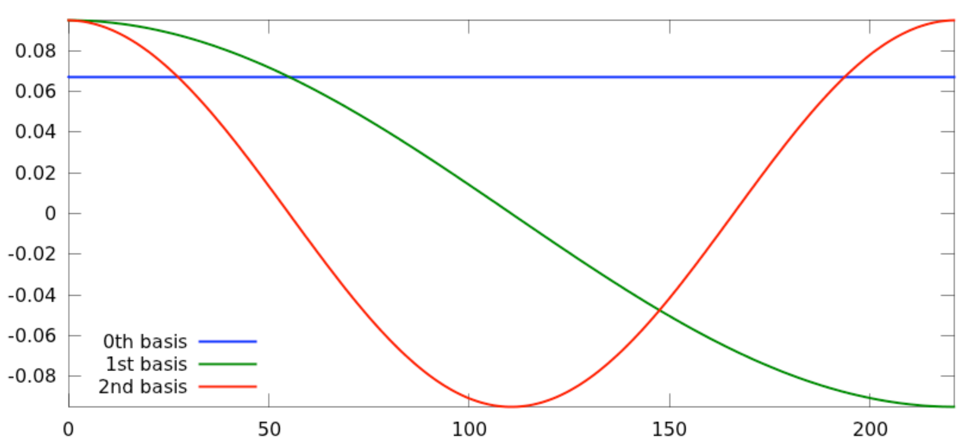
- **HMM Observations: power measurements from a constant-Q filter bank decomposition of the signal**
  - the power measurement is summed over a 3-semitone span around the fundamental of the ending and starting notes in each line in the DTW alignment
- **Improves median alignment for onsets from 118 ms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

Devaney (2014)

Score-guided performance data extraction

Monophonic

- **Timing information is available in the alignment**
- **Fundamental frequency (F_0), and amplitude can be reliably extracted**
- **Perceptually grounded models for pitch and loudness have been developed**
- **Characterizing F_0 trajectories is under-studied**



Decomposition of F_0 trace with the Discrete Cosine Transform to estimate slope and curvature

Devaney, Mandel and Fujinaga (2011)
Devaney and Wessel (2013)

Score-guided performance data extraction

Polyphonic

- Again, timing information is available in the alignment
- Fundamental frequency (F_0), and amplitude are harder to extract
- Although once extracted, perceptually grounded models for pitch and loudness and the DCT for characterizing F_0 trajectories can be used
- Currently exploring the using High Resolution methods with Roland Badeau for the task of score-guided extracting of frequency and loudness information in polyphonic audio

Experiments with Performers

Overview

- Intonation in trained singers in the Western Art Music tradition
- Solo and small ensemble (2-4 voices)
- Various aspect of the work was done in collaboration with Dan Ellis, Jason Hockman, Ichiro Fujinaga, Michael Mandel, Peter Schubert, David Wessel, and Jon Wild

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

Recording Set-Up

- **Rooms**
  - CIRMMT Labs at McGill
  - St Mathias Church, Montreal
- **Microphones**
  - Solo singers and the entire ensembles were recorded with a pair of cardioid microphone
  - Each ensemble singer was miked with a cardioid headband mic

- **Recording Equipment**
  - Lab: Mac Pro
  - Church: portable 16-track recorder

Solo Singing

Overview

- Musical Material
  - Schubert's "Ave Maria"
    - 3x a cappella & 3x accompanied
- 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 analyzed
- Singers listened to and approved their own recordings

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

Ensemble Singing

Overview

- Musical Material
  - 3-part chord progression by Giambattista Benedetti
  - 4-part piece by Praetorius (“Es ist ein Ros entsprungen”)
- Singers
  - combinations of professional SATB ensemble who performed with a conductor
- Melodic semitones and whole tones analyzed in different vertical (harmonic) contexts
- Conductor listened to and approved the recordings

Devaney, Mandel, and Fujinaga (2012)

Two-Part Singing

Overview

- Musical Material
  - Semitone pattern sung against a recorded version of the lower-line that was detuned in various ways at two pitch heights
- Singers (6 of 12 subjects)
  - 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 different vertical (harmonic) contexts

Data Analysis

Linear regression

- Dependent variable (for all experiments)
  - interval size in cents
- Independent variables (varied by experiment)
  - direction (all)
  - singer or level of experience (solo and 2-part)
  - harmonic context
  - leading tone or not (solo)
  - other contextual cues as developed by Bregman, Narumour, and Lerdahl/Krumhansl (solo)
  - vertical interval context (ensemble and 2-part)
  - accompanied versus a cappella (solo)
  - equal temperament or retuned (2-part)

Commonality between performers

Observable trends

- **General Tuning Trends**
  - No strict adherence, on average smaller than equal temperament (more so for semitones than whole tones)
  - Ascending semitones were significantly larger on average than descending semitones (in solo and 2-part singing)

Commonality between performers

Observable trends

- **Musical Context**
  - **Solo singing**
    - Non-pros' semitones were significantly smaller in leading tone contexts than non-leading tone contexts
    - Bregman (1990) prediction of smaller variance at peaks and valleys in the melody
      - true only for semitones
    - Narmour's IR model (1990)
      - semitones and whole tone size was inversely proportional to "pitch-reversal" value
    - Krumhansl and Lerdahl's melodic attraction model (2007)
      - only a few of the factors were statistically significant

Commonality between performers

Observable trends

- **Musical Context**
  - **Ensemble singing**
    - **Benedetti**: Melodic whole tones sung over a P5 were 15 cents larger on average than those sung over a M3
    - **Praetorius**: Vertical intervals in cadential contexts were significantly closer to Just Intonation than those in non-cadential contexts
    - **2-part**: Semitones sung a perfect octave above the lower voice were 7 cents larger on average than those sung above other intervals

Is there an effect of training?

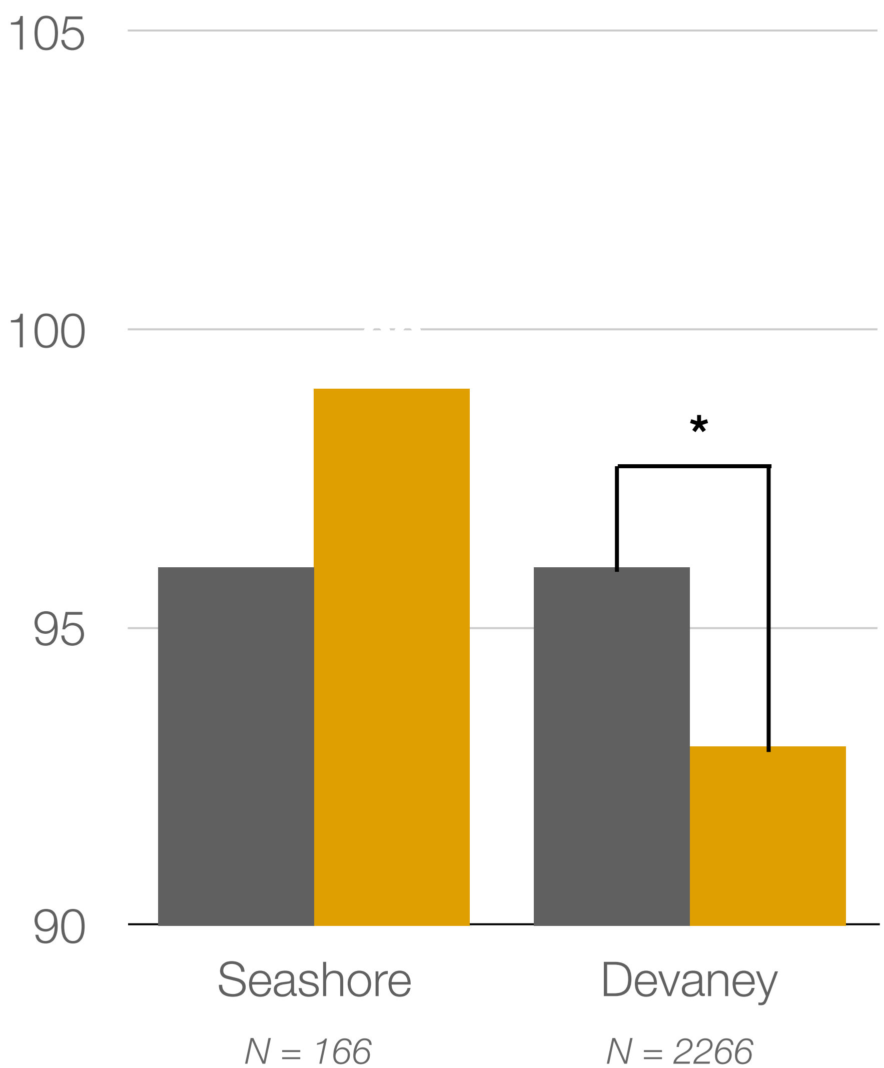
Professions versus non-professions in solo experiment

- **Effect of Training**
  - **Accompaniment**
    - Solo non-pros' accompanied semitones were 3 cents larger on average than their *a cappella* semitones
    - *The were no significant effect for detuning of the accompanying voice in the 2-part experiment*
  - **Consistency**
    - Pros were more consistent with one another
  - **Interval size**
    - Pros' semitones were significantly larger on average (closer to equal temperament)

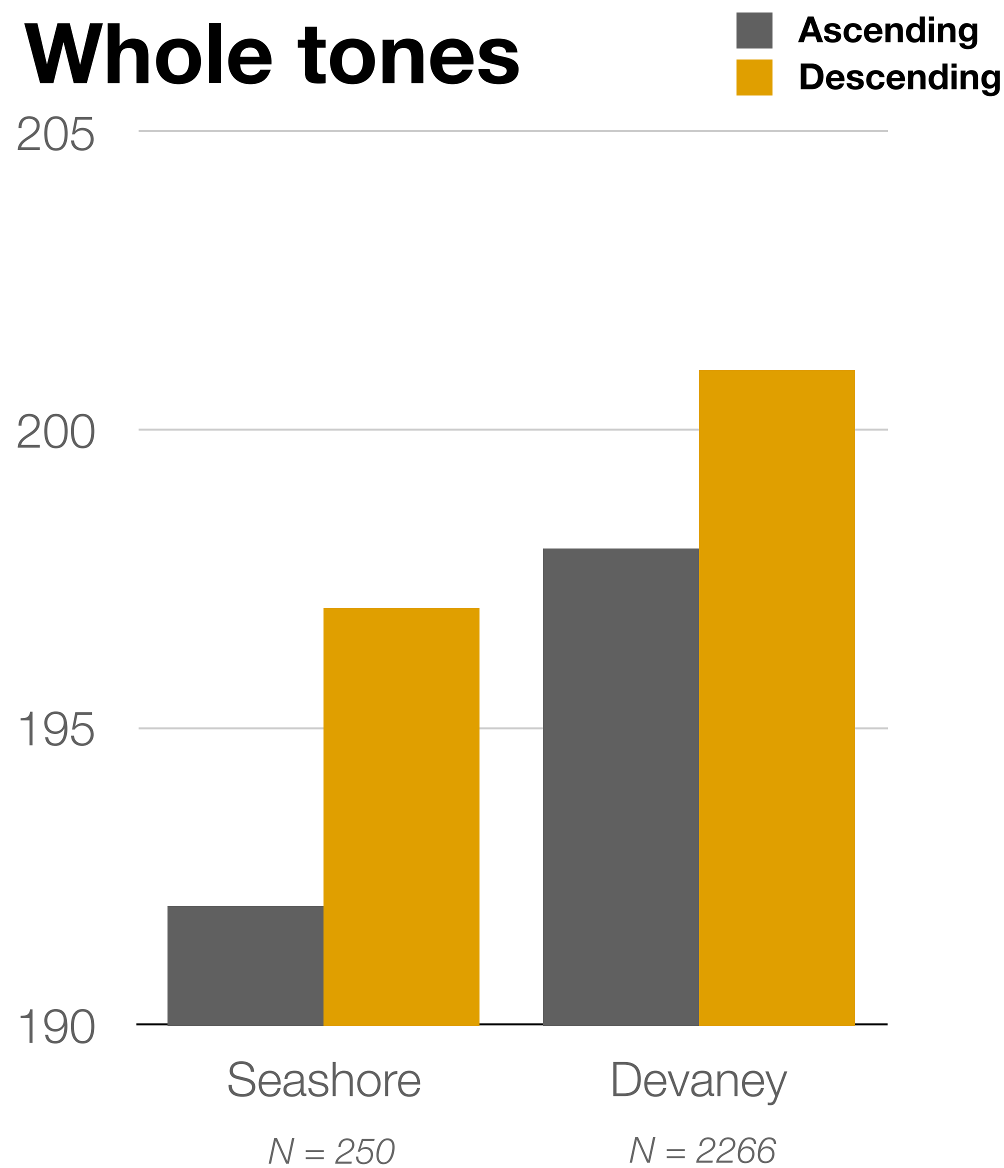
Incorporating Seashore data

Comparative analysis of Seashore and contemporary data

Semitones

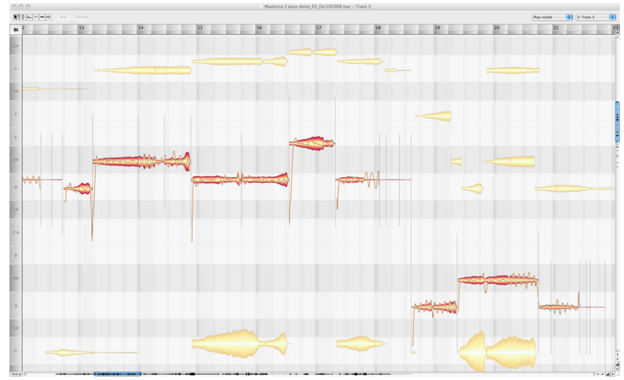


Whole tones



Some Implications for Composition

Alternative tunings



- **Precise reproduction of a prescribed tuning system requires cuing at the precise pitch level**
  - Suggested by variability in highly-trained singers and the lack of effect with detuned accompaniment
- **Singers likely have more control over smaller intervals**
  - Semitones tuning showed greater contextual effects than whole tone tuning and preliminary analysis of larger intervals shows that they are even more variable
- **Vertical tuning effects are more likely to be achieved at phrase endings**

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

Conclusions
Summary and future directions.

5

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 and considered some of the implications of my findings for composition

Current Work

Where I am going

- **Developing methods for making statistical comparison between performances**
  - examining the issue of inter- and intra-singer similarity
- **Developing more robust tools for automatic extraction of performance data from recordings**
  - making the current tools more reliable and more accessible to other researchers
- **More contextualized experiments**
  - studying existing recordings of a singer performing the same piece at different points in their career

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

Bregman, A.S. 1990. *Auditory Scene Analysis*. MIT Press, Cambridge, MA.

Devaney, J. 2014. Estimating onset and offset asynchronies in polyphonic score-audio alignment. Journal of New Music Research. 43 (3): 266–75.

Devaney, J. Under Review. Recapturing the data in Seashore's musical performance measurements. Musicae Scientiae.

Devaney J., M. Mandel, and D. Ellis. 2009. Improving MIDI-audio alignment with acoustic features. In Proceedings of the Workshop on Applications of Signal Processing to Audio and Acoustics. 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): 108–36

Devaney, J., M. Mandel, and I. Fujinaga. 2011. Characterizing Singing Voice Fundamental Frequency Trajectories. Proceedings of the Workshop on Applications of Signal Processing to Audio and Acoustics, 73–6.

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 the International Society of Music Information Retrieval conference. 511–6.

Devaney, J., and D. Wessel. 2013. Pitch perception of time-varying sung tones. *Society for Music Perception and Cognition conference*.

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.

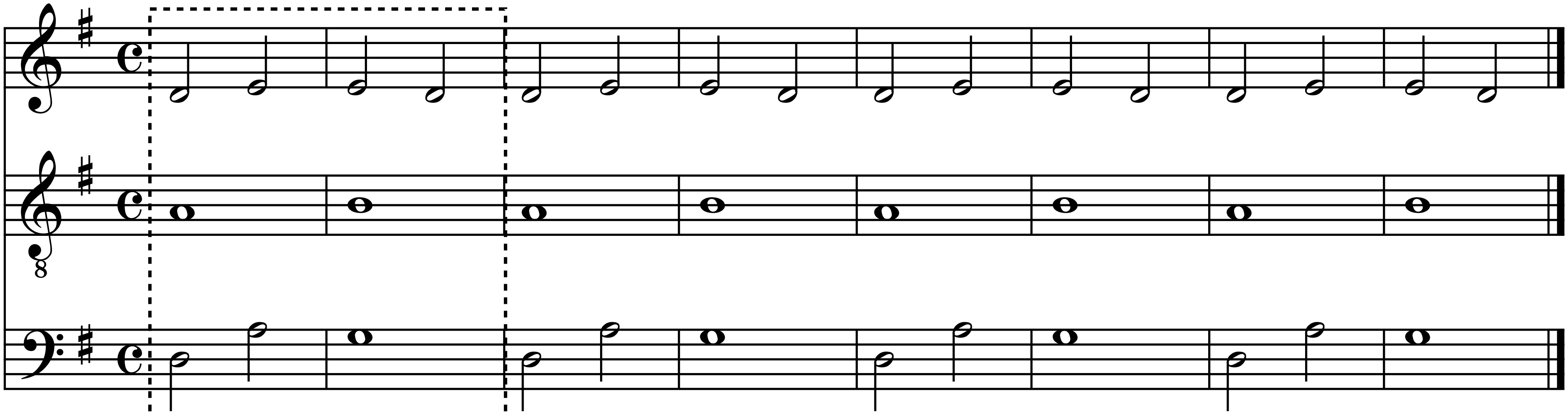
Lerdahl, F., and C. Krumhansl. 2007. Modeling tonal tension. *Music Perception* 24, 329–366

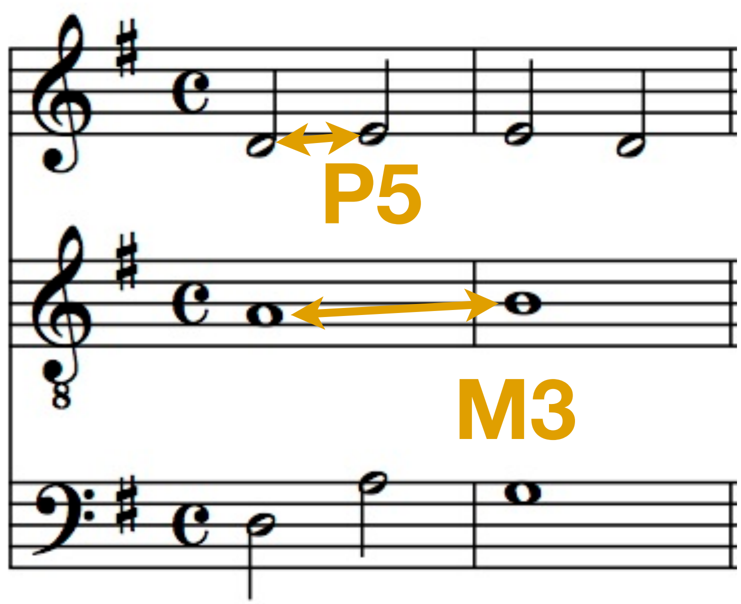
Narmour, E. 1990. *The Analysis and Cognition of Basic Musical Structures*. University of Chicago Press., Chicago, IL.

Orio, N., and D. Schwarz, D. 2001. Alignment of monophonic and polyphonic music to a score. In Proceedings of the International Computer Music Conference. 155–8.

Three-Part Singing

Exercises





Four-Part Singing

Praetorius - Es ist ein Ros’ ent sprungen

Two-Part Singing

Tuning systems (in relation to equal temperament)

Just Intonation

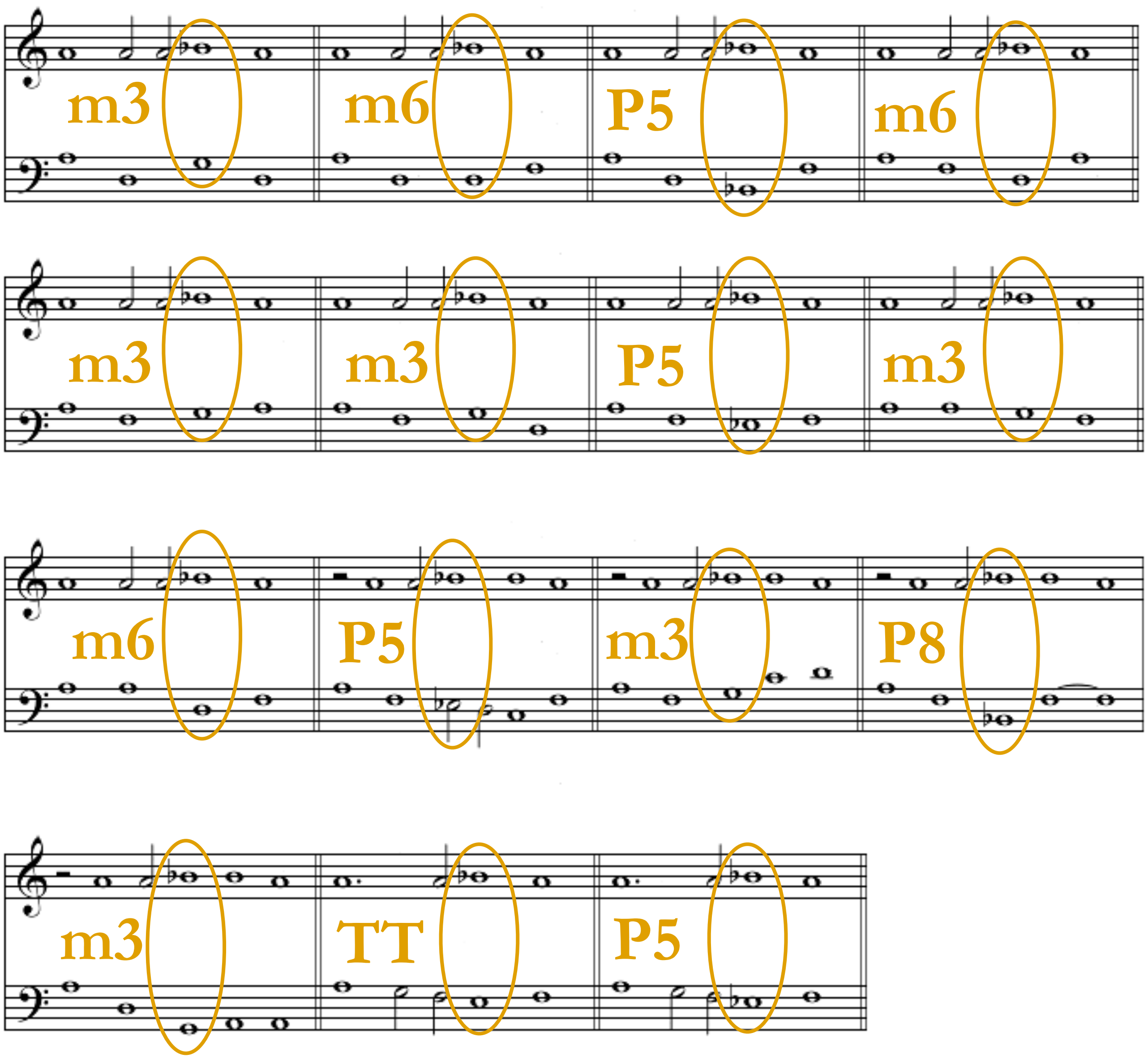
1	D -2	G -4	D -2		
2	D -2	D -2	F 14		
3	D -2	Bb 12	F 14		
4	F 14	D -2	A 0		
5	F 14	G 18	A 0		
6	F 14	G 18	D 20		
7	F -8	Eb -12	F -8		
8	A 0	G -4	F -8		
9	A 0	D -2	F -8		
10	F 14	Eb 10	D -2	C 16	F 14
11	F -8	G -4	C -6	D -2	
12	F -8	Bb -10	F -8		
13	D 20	G 18	A 22	A 22	
14	G -4	F -8	E 2	F -8	
15	G -4	F -8	Eb -12	F -8	

Modified Just Intonation

1	D -2	G 18	D 20		
2	D -2	D -25	F -8		
3	D -2	Bb -10	F -8		
4	F 14	D 20	A 22		
5	F -8	G -4	A -23		
6	F -8	G -27	D -25		
7	F 14	Eb 33	F 14		
8	A 0	G 18	F 37		
9	A 22	D 20	F 14		
10	F -8	Eb -12	D -2	C -6	F -8
11	F 14	G 18	C 16	D 20	
12	F 14	Bb 12	F 14		
13	D	G -27	A -23	A -23	
14	G 18	F 14	E 2	F 14	
15	G 18	F 14	Eb 10	F 14	

Two-Part Singing

Exercises



Prior Findings on Vocal Intonation

- **Schoen (1922) – accompanied solo singers**
  - less sharp when descending than when ascending
- **Prame (1997) – accompanied solo singers**
  - intonation deviated substantially, but not consistently, from equal temperament
- **Jers and Terström (2005) – 16-voice ensemble**
  - greater intonation dispersion at a faster tempo
  - ascending intervals were larger than descending intervals

Prior Findings on Vocal Intonation

- **Vurma and Ross (2006) – solo singers**
  - ascending/descending semitones smaller than EQT
  - ascending/descending tritones and fifths larger than than EQT
- **Howard (2007a, 2007b) – a cappella quartets**
  - used non-equal temperament with a tendency toward, though not full compliance with, Just Intonation
- **Vurma (2010) – 2-part singing against a synthesized lower voice**
  - singers' intonation did not change significantly when the synthesized voice was detuned

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 ✓