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**
  - 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 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.

Quantitative Performance Analysis

A brief history

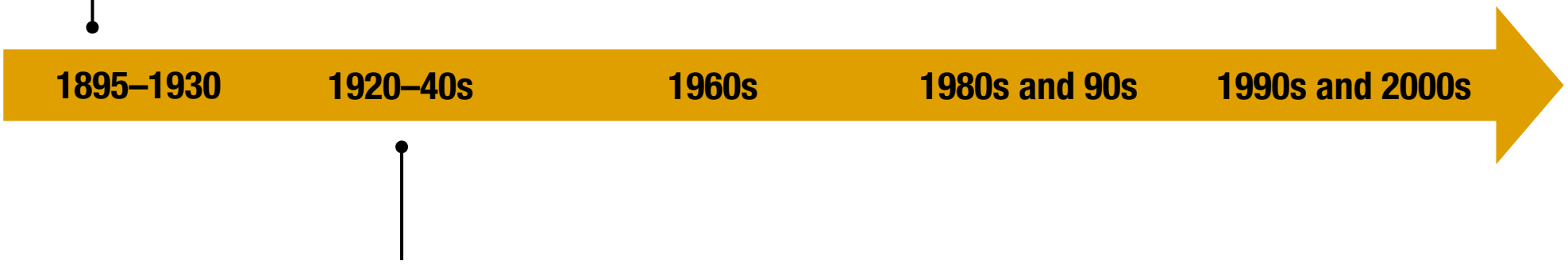
Pioneers

Binet and Courtier
Sears
Miller

Quantitative Performance Analysis

A brief history

Pioneers
Binet and Courtier
Sears
Miller

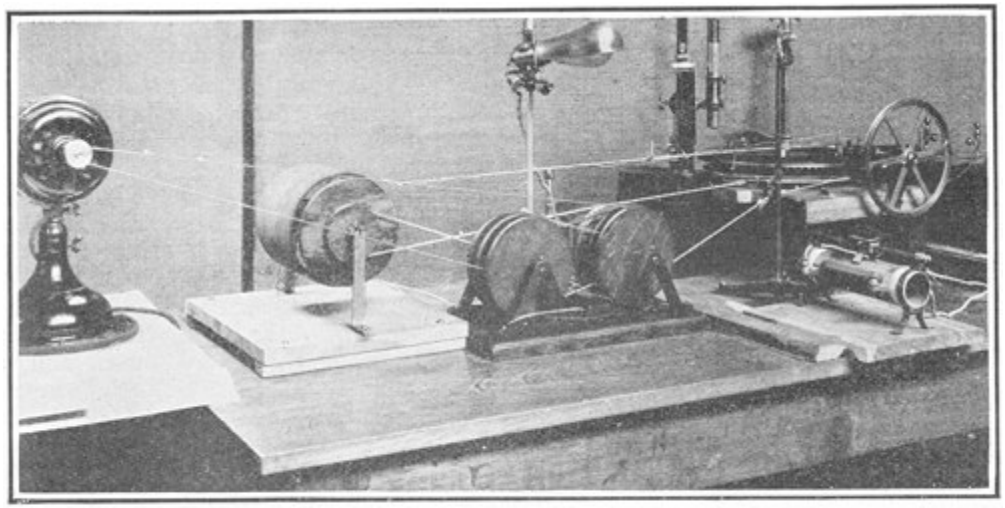


University of Iowa
Seashore and colleagues

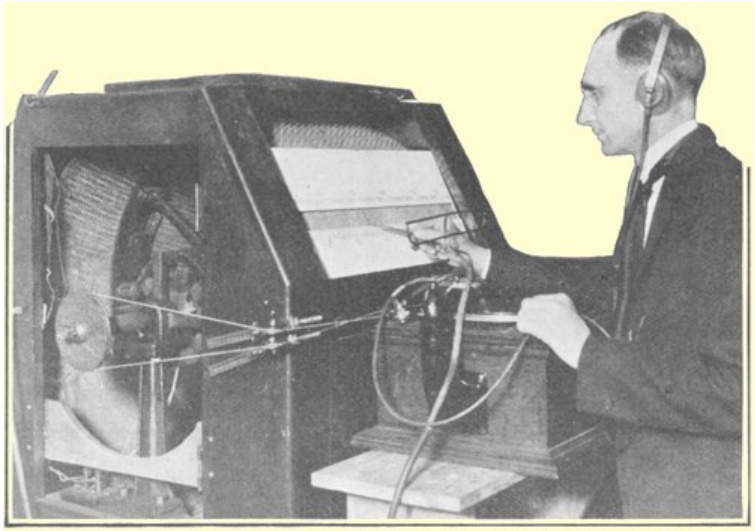
Quantitative Performance Analysis

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



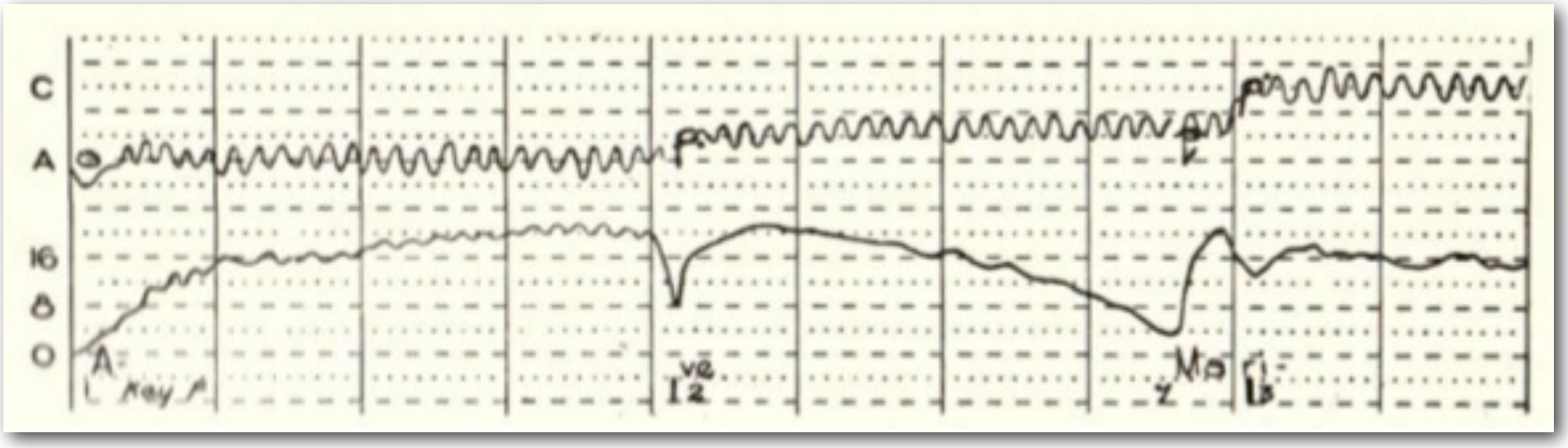
The tonoscope for analyzing the pitch of the tones on a disk phonograph record

Cary (1922)

Quantitative Performance Analysis

University of Iowa

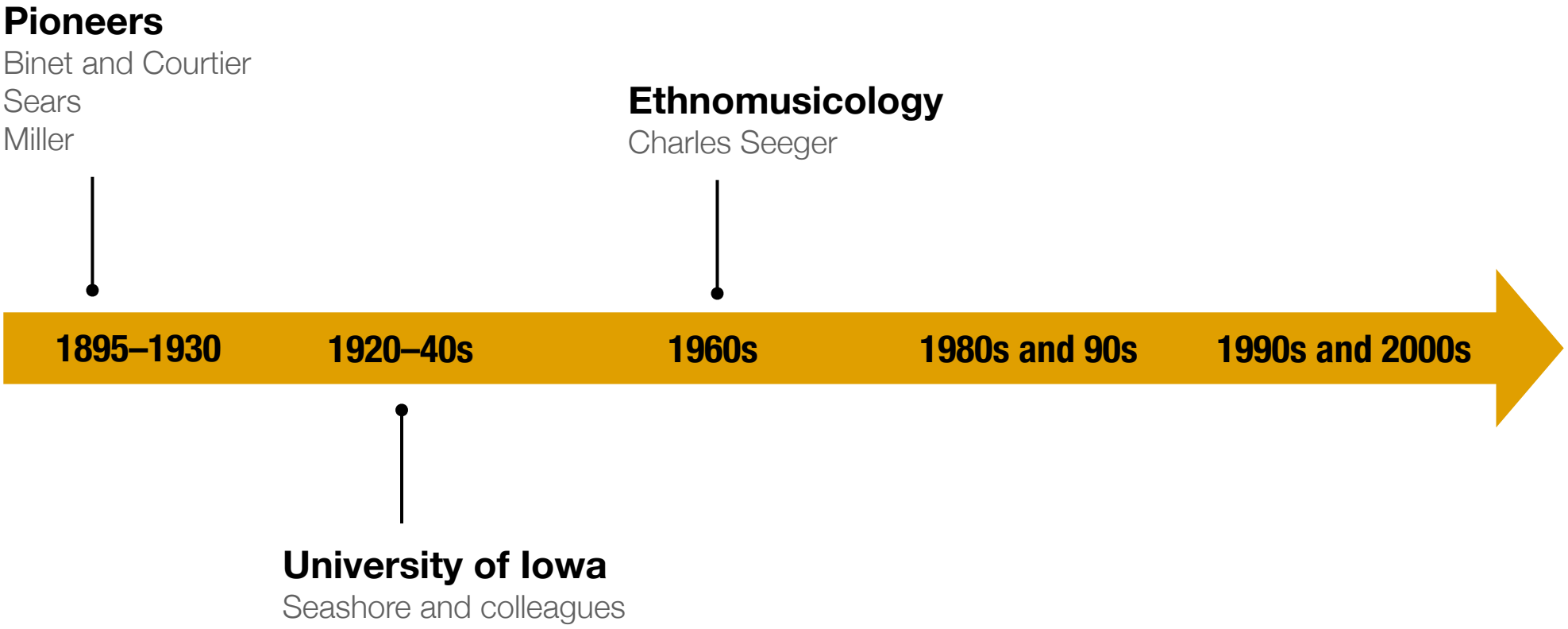
Performance scores



Seashore (1936)

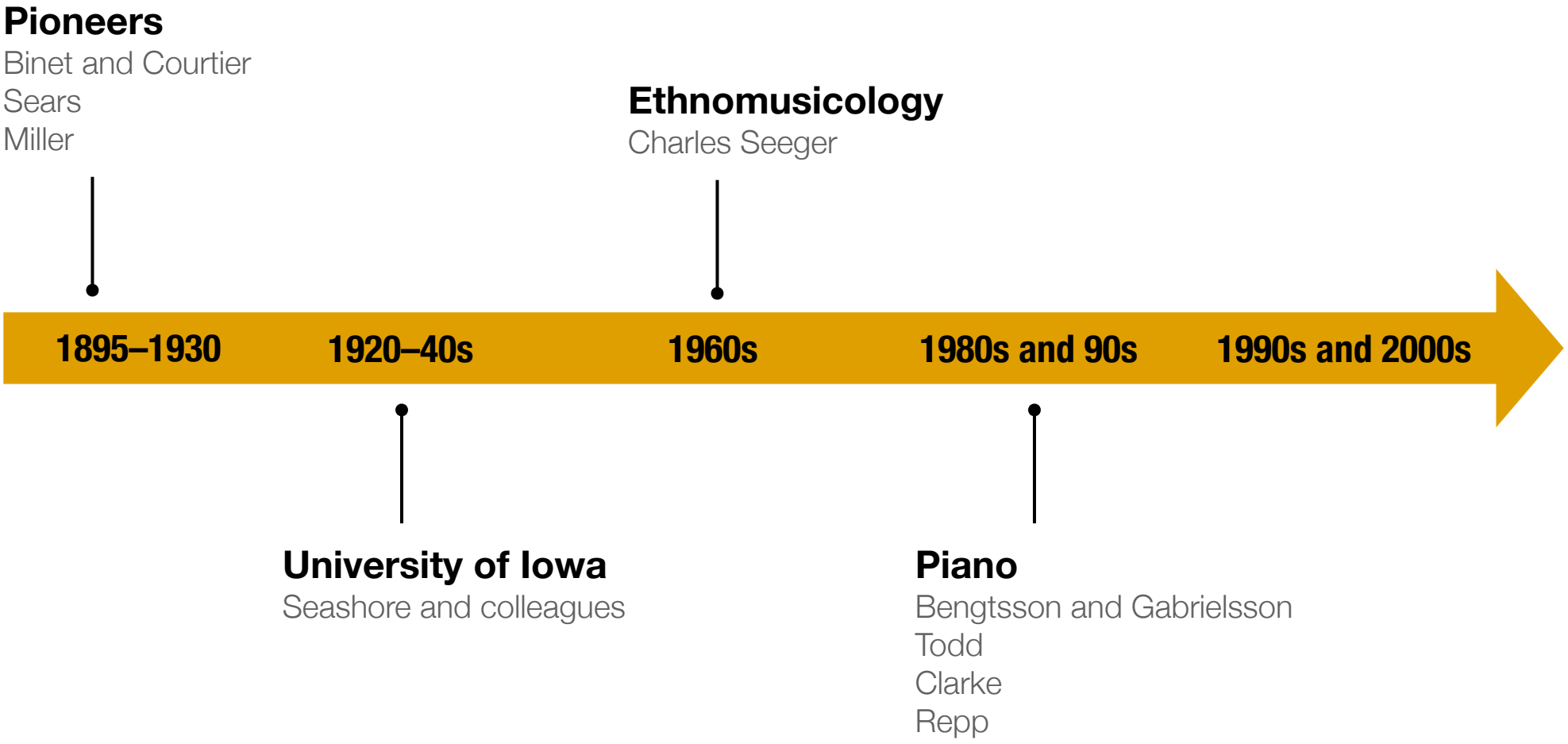
Quantitative Performance Analysis

A brief history



Quantitative Performance Analysis

A brief history



Quantitative Performance Analysis

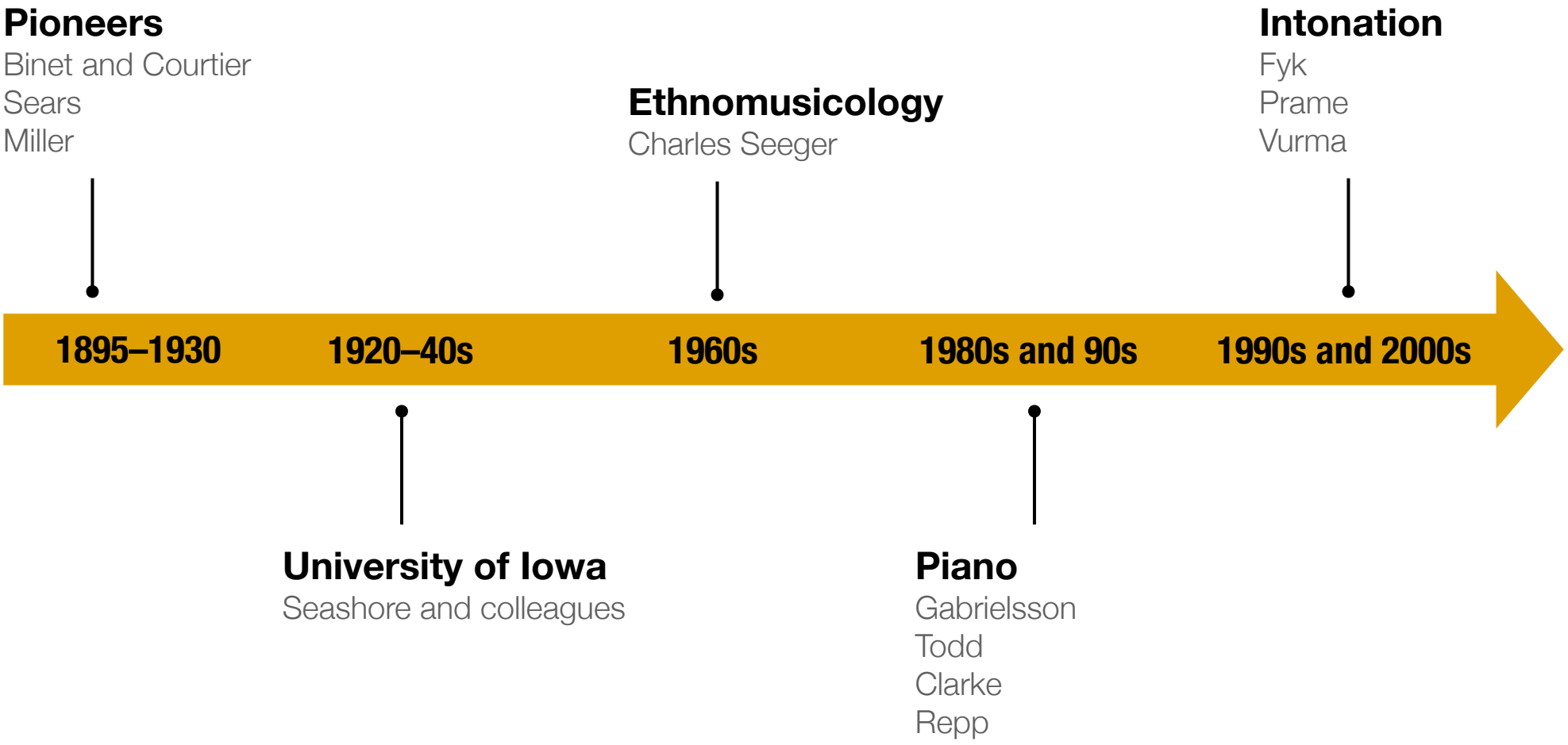
Popularity of the piano

- 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

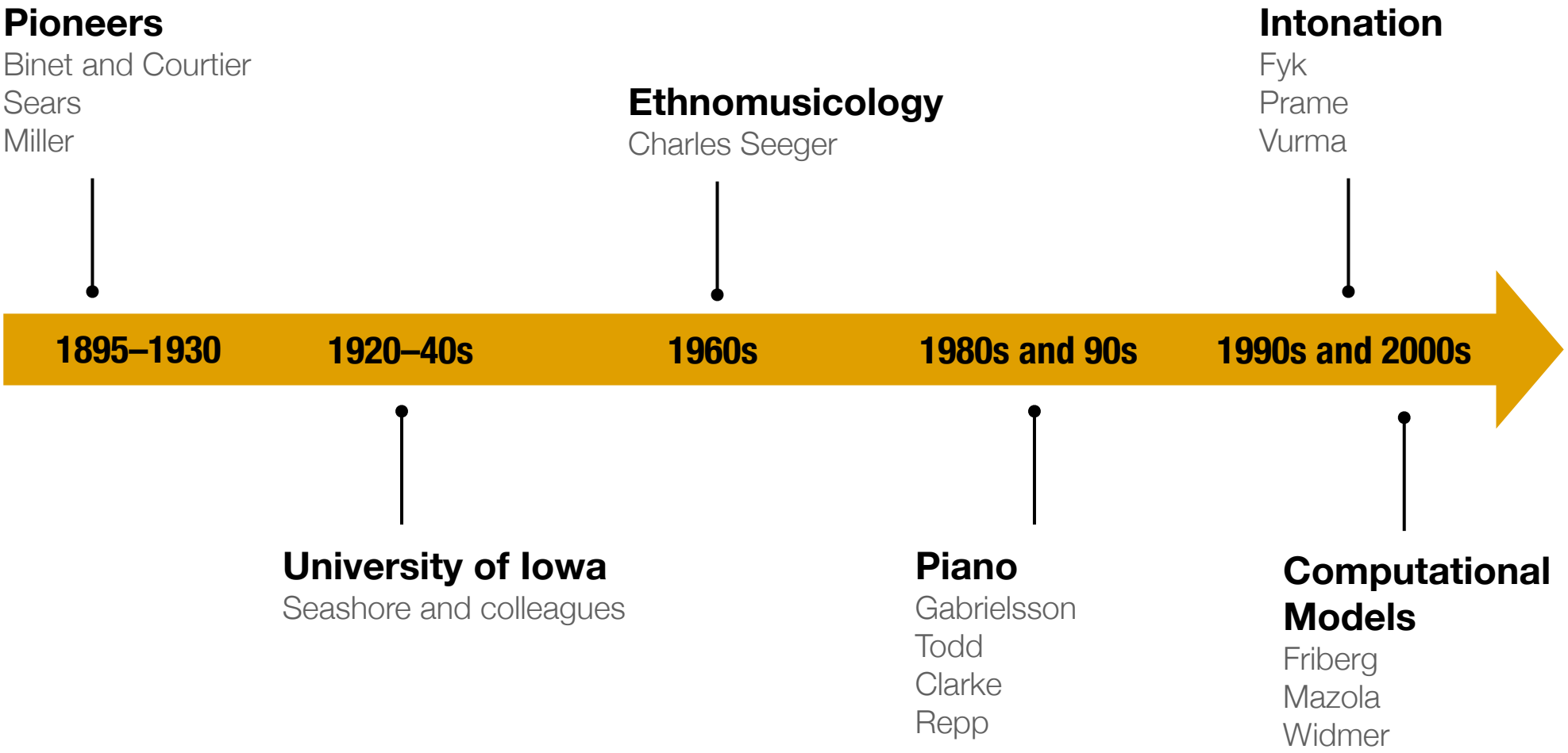
Quantitative Performance Analysis

A brief history



Quantitative Performance Analysis

A brief history



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.

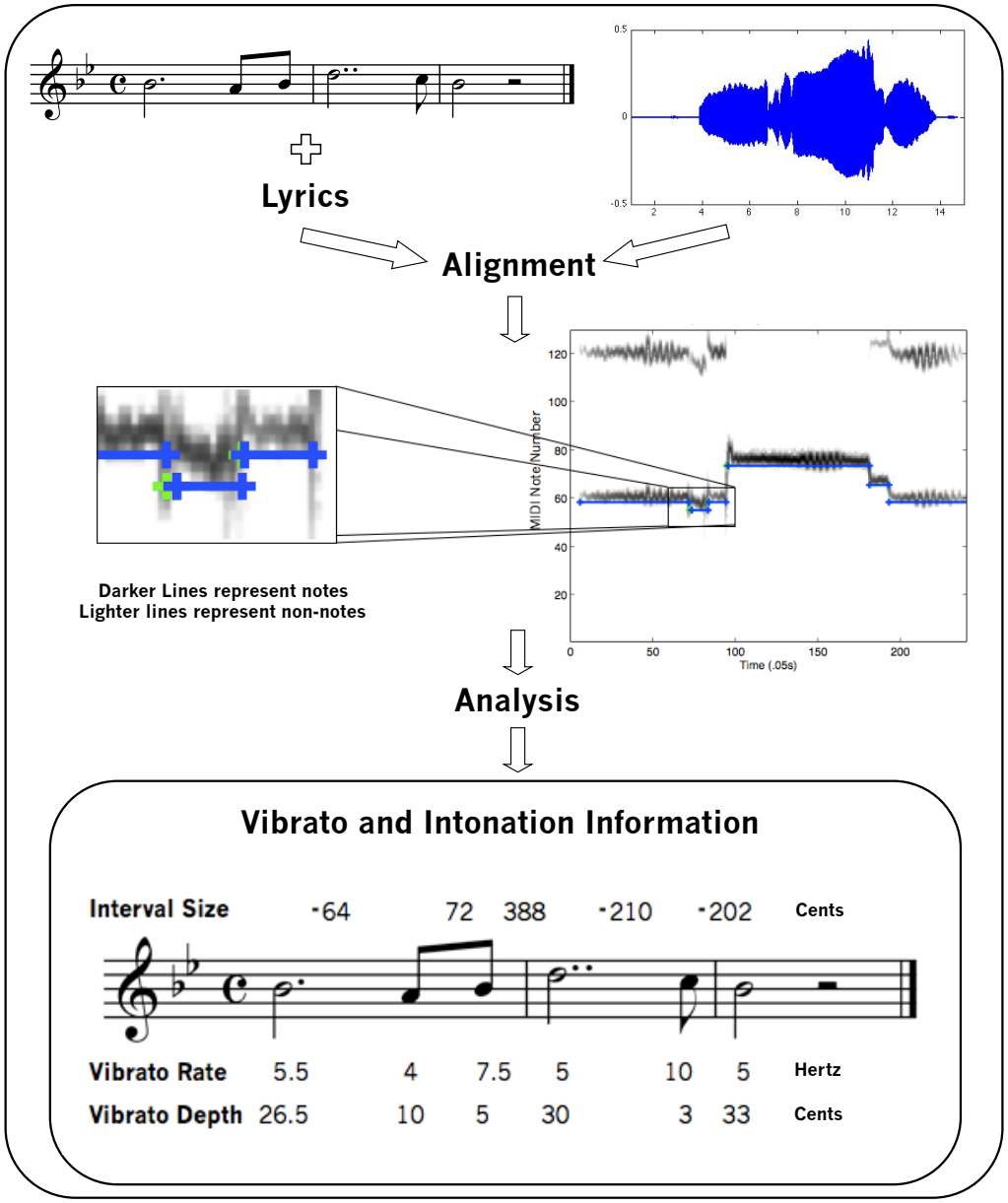
Experiments 4
Studies of intonation in the singing voice.

Conclusions 5
Summary and future directions.

AMPACT

Automatic Music Performance and Comparison Toolkit

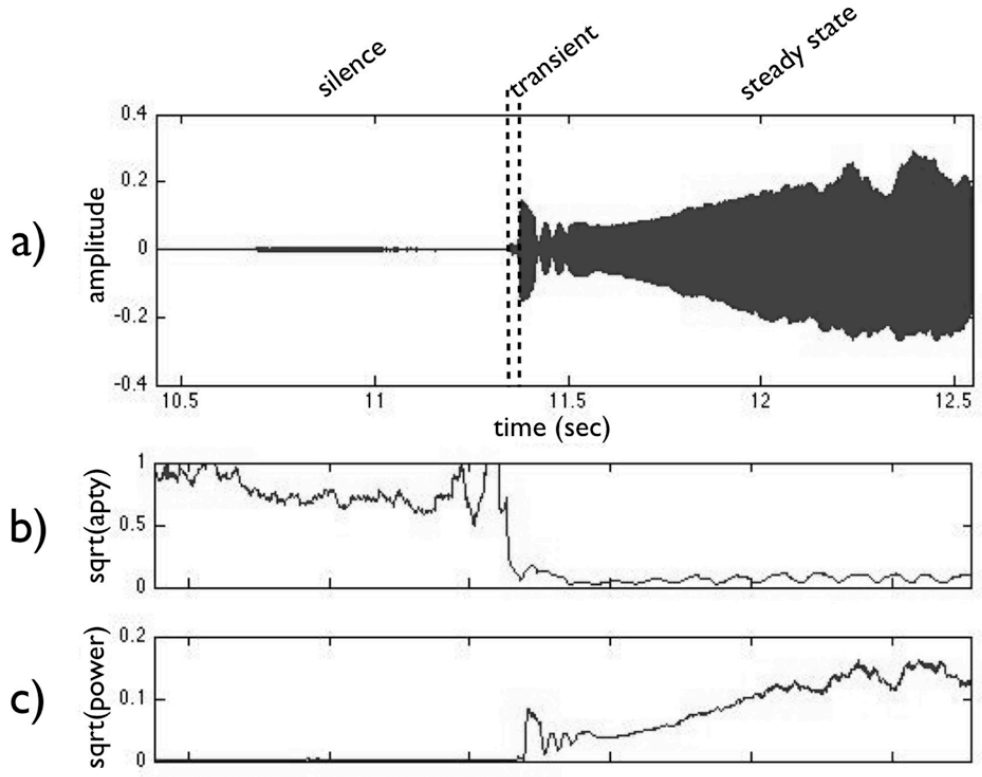
www.ampact.org



Monophonic audio

Identifying onsets and offsets

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

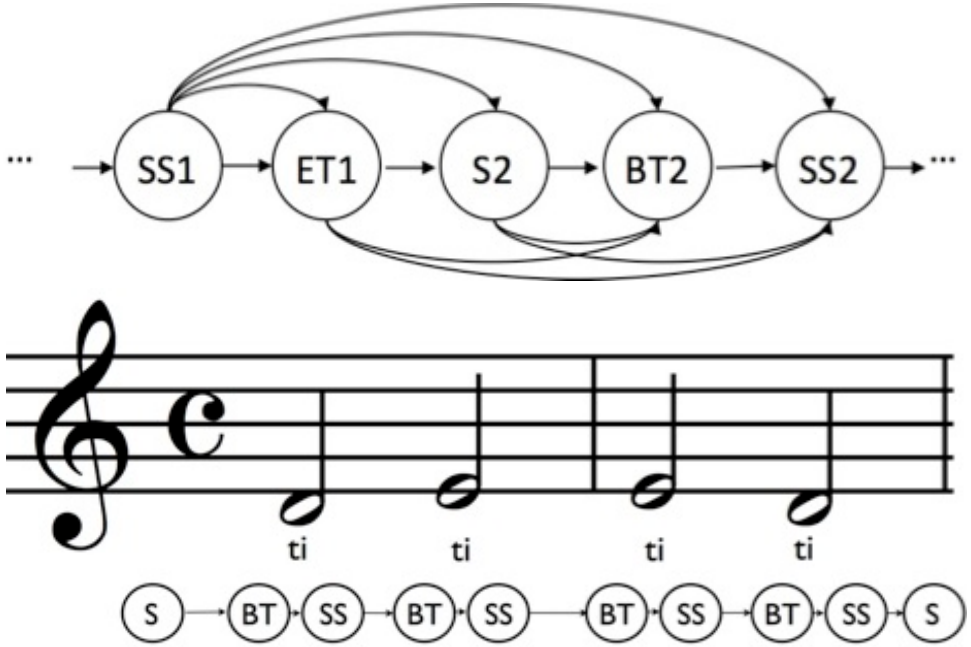


Devaney, Mandel, and Ellis (2009)

Monophonic audio

Identifying onsets and offsets

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

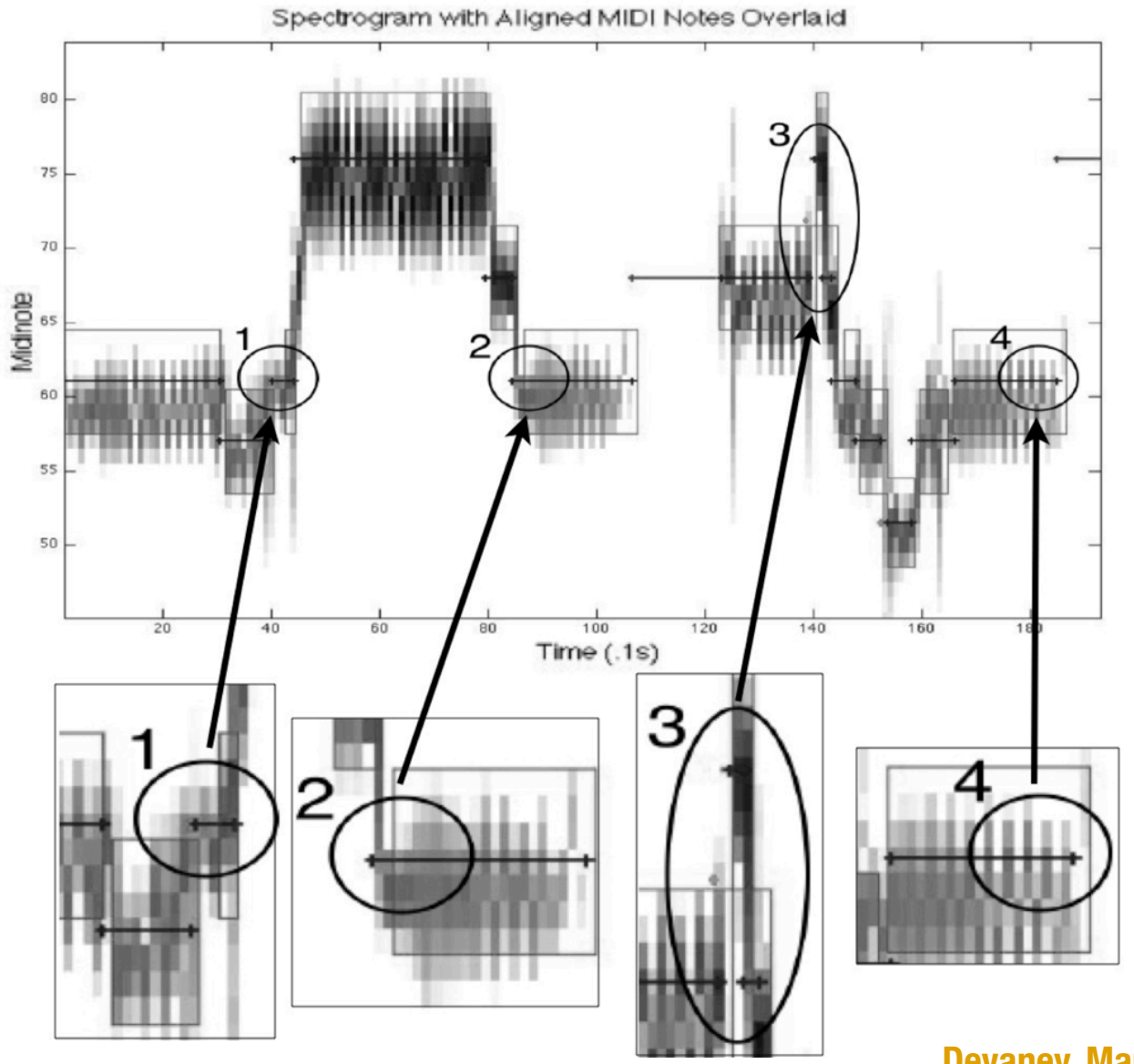


- **Improves median alignment error from 52 ms to 28 ms**

Devaney, Mandel, and Ellis (2009)

Monophonic audio

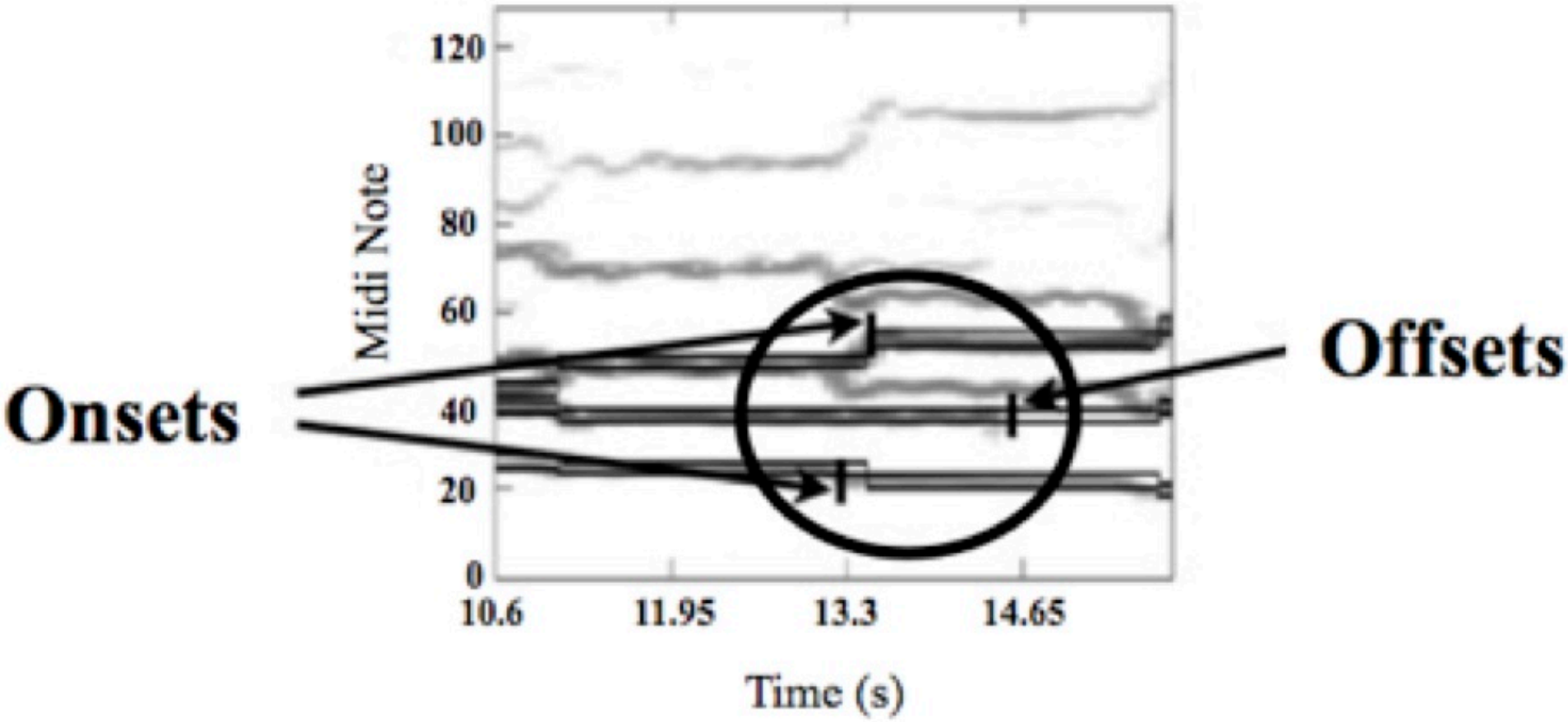
Identifying onsets and offsets



Devaney, Mandel, and Ellis (2009)

Polyphonic audio

Identifying asynchronies between voices



Devaney and Ellis (2009)

Polyphonic audio

Identifying asynchronies between voices

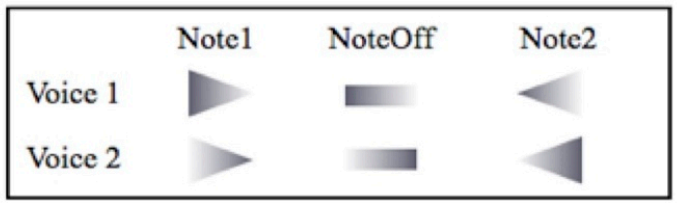
- **Multi-pass DTW/HMM algorithm**
- **DTW determines general note transitions**
  - providing a single offset/onset location for all of the musical lines
- **HMM finds the location of each line's onsets and offsets within a +/- 125 ms window around the DTW estimate**

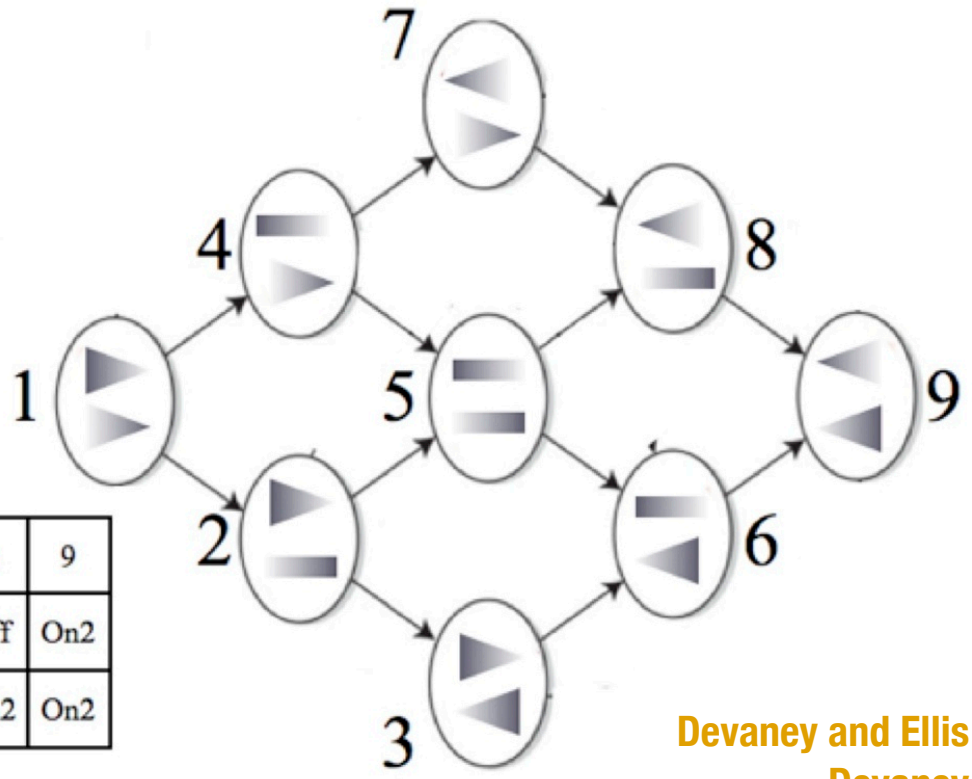
Devaney (2014)

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.

Conclusions 5
Summary and future directions.

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 (Columbia), Jason Hockman (McGill), Ichiro Fujinaga (McGill), Michael Mandel (Ohio State), Peter Schubert (McGill), and Jon Wild (McGill)

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

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

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

- **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 analyzed**

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
- **EFFECT OF TRAINING**
  - Pros were more consistent with one another
  - Pros' semitones were 6 cents larger on average
  - Non-pros tended to compress leading tones
  - Non-pros' accompanied semitones were 3 cents larger than *a cappella* semitones

Three-Part Singing

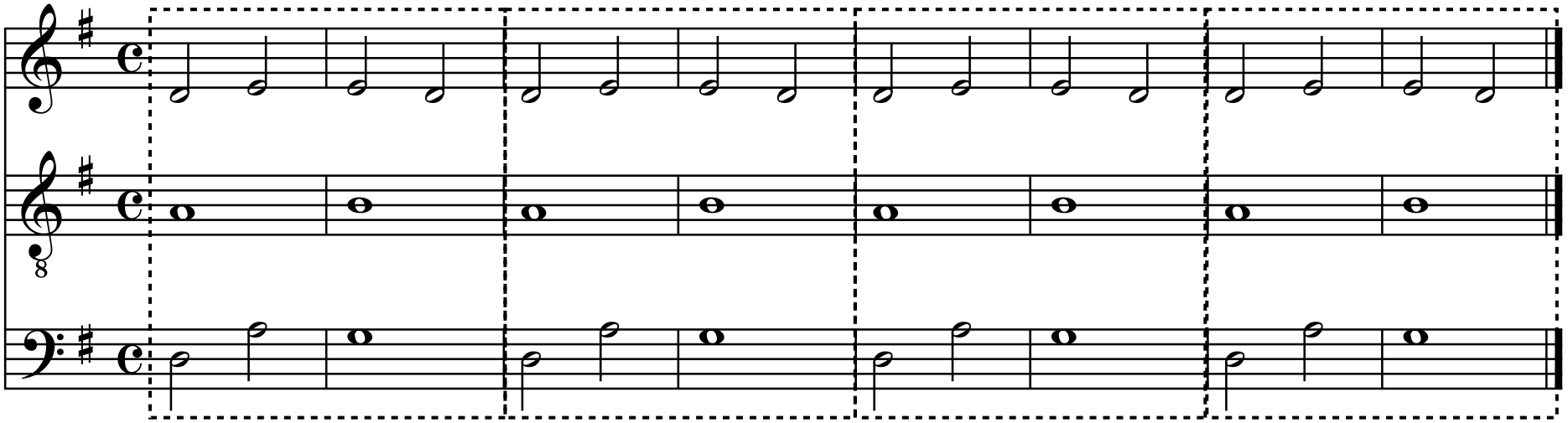
Overview

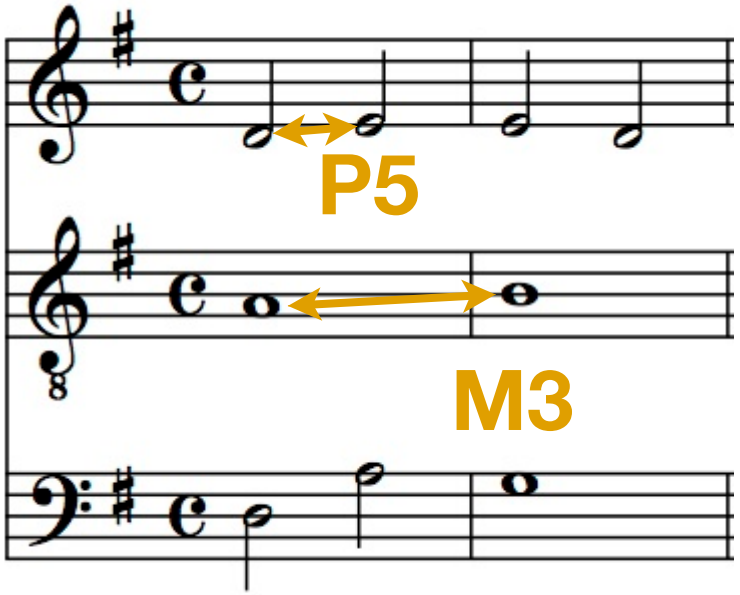
- **Chord progression by Giambattista Benedetti**
- **4 ensembles**
  - Ensemble 1 (lab): semi-professional alto, tenor, and bass singers who performed without a conductor - *pilot*
  - Ensemble 2 (lab): professional alto, tenor, and bass singers who performed with a conductor
  - Ensemble 3 (church): professional soprano, alto, and tenor singers who performed with a conductor
  - Ensemble 4 (church) professional alto, tenor, and bass singers who performed with a conductor
- **Melodic whole tones in M3 and P5 vertical contexts**

Devaney, Mandel, and Fujinaga (2012)

Three-Part Singing

Exercises

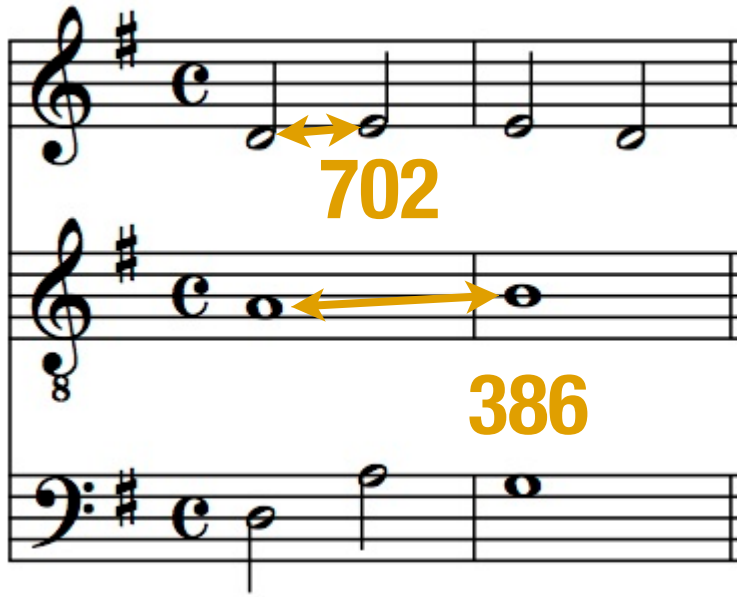




Three-Part Singing

Significant trends

- **TUNING SYSTEMS:** No strict adherence, generally closer to equal temperament
- **DIRECTION:** not significant
- **VERTICAL INTERVAL CONTEXT:** melodic whole tones sung over a P5 were 15 cents larger on average than those sung over a M3



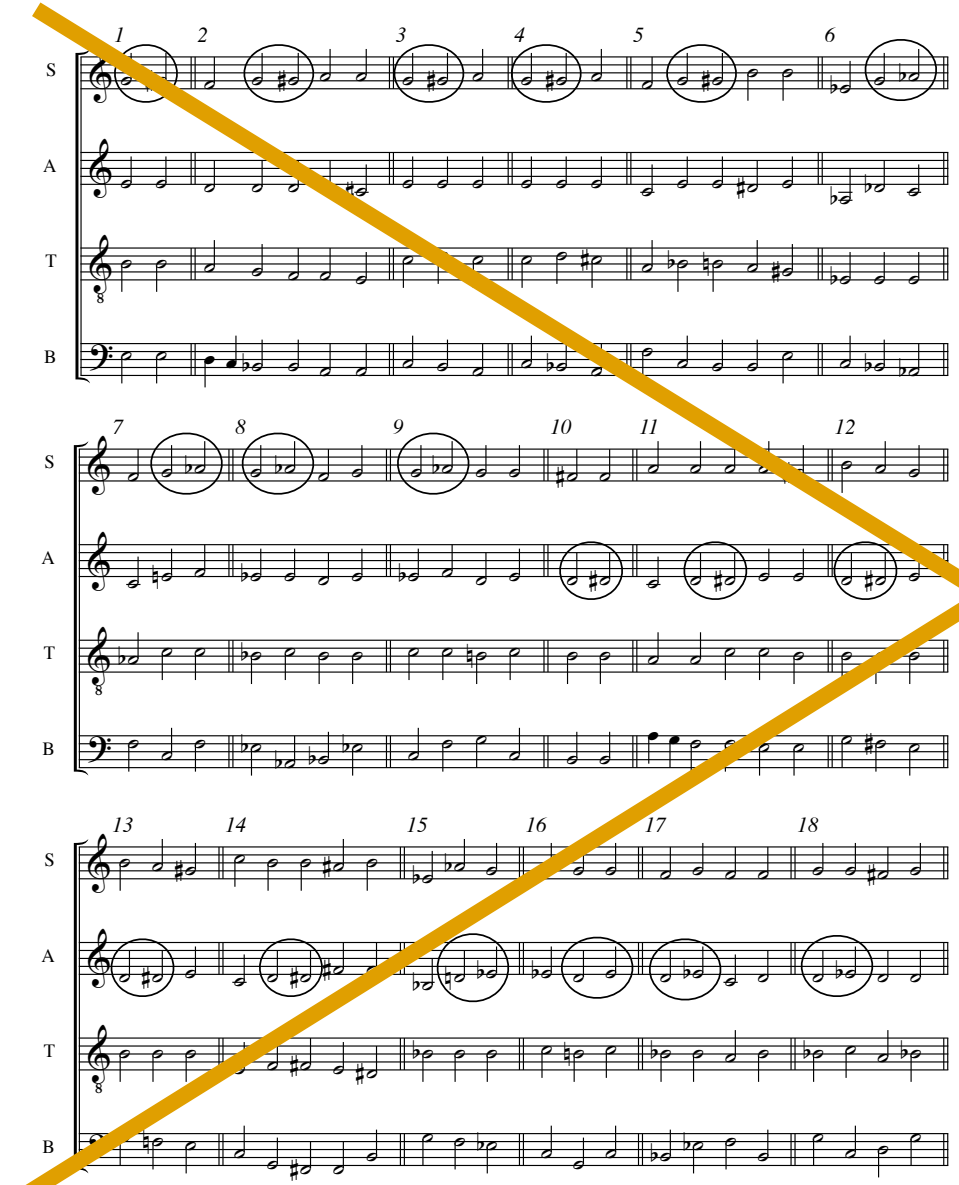
Four-Part Singing

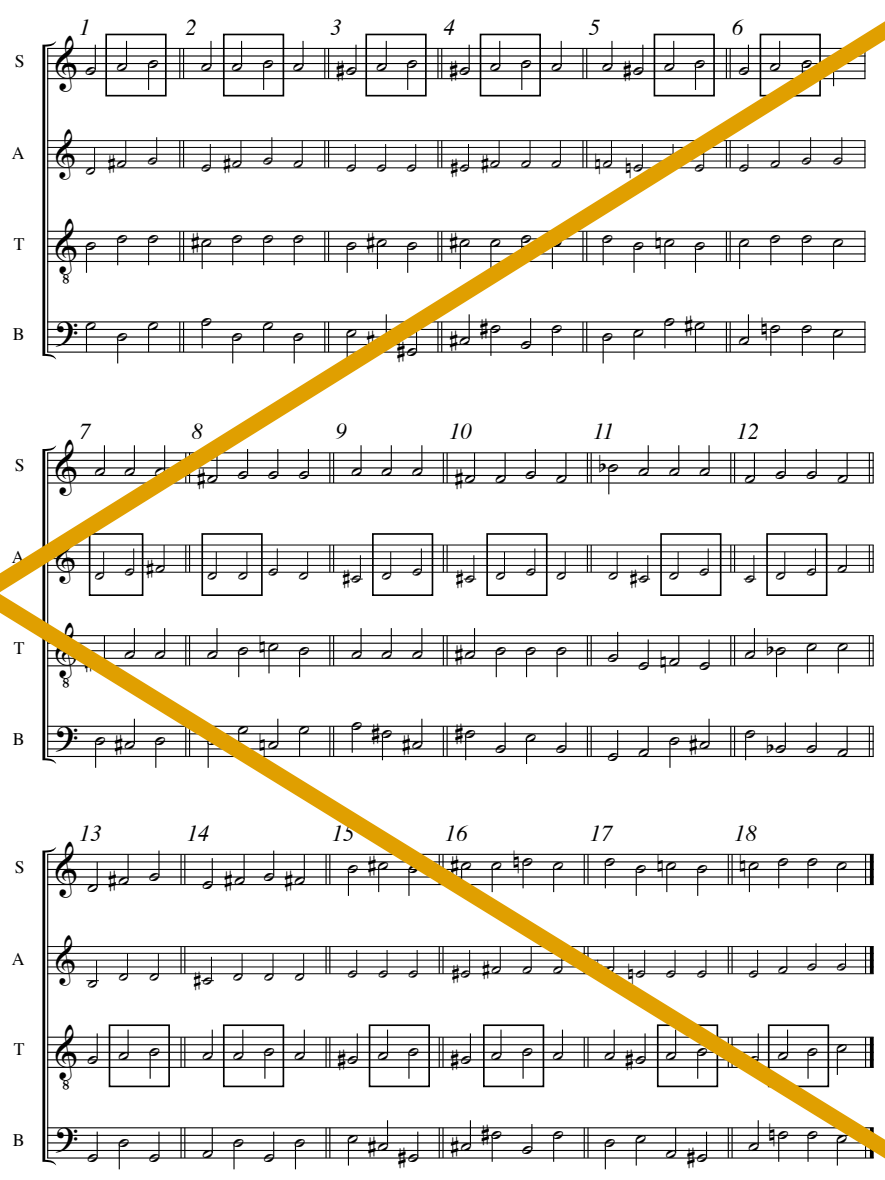
Overview

- **Exercises composed by Jonathan Wild and Peter Schubert and a piece by Praetorius**
- **3 ensembles**
  - Ensemble 1 (lab): semi-professional SATB ensemble who performed without a conductor - *pilot*
  - Ensemble 2 (lab): professional SATB ensemble who performed with a conductor
  - Ensemble 3 (church): professional SATB ensemble who performed with a conductor
- **Melodic semitones and whole tones in various vertical contexts**
- **Vertical intervals in cadential contexts**

Four-Part Singing

Exercises by Wild and Schubert





Four-Part Singing

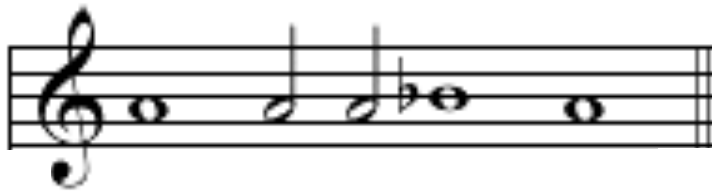
Significant trends for Praetorius

- **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 detuned in various ways 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

Tuning systems (in relation to equal temperament)

Just Intonation

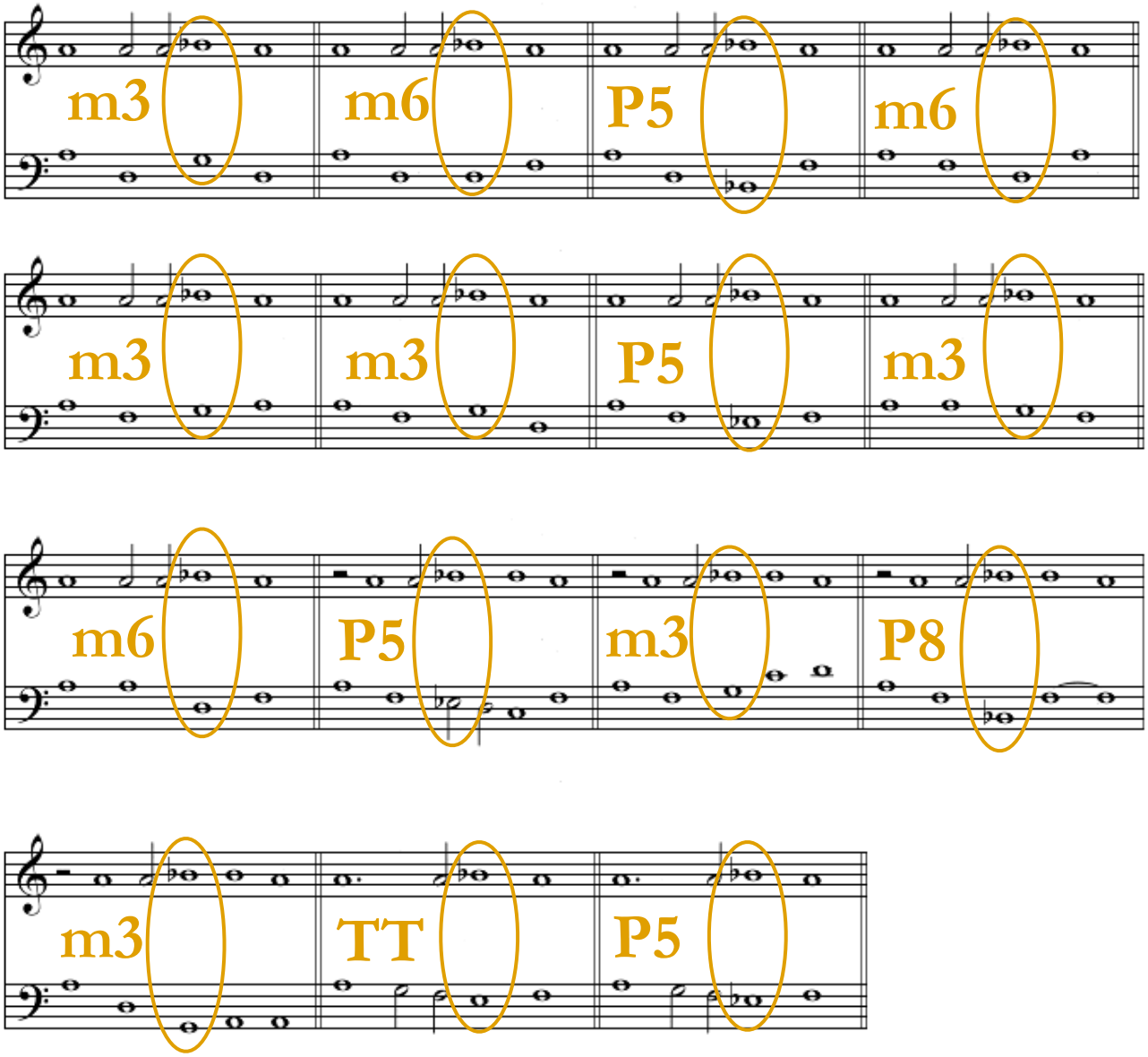
1	D	G	D		
	-2	-4	-2		
2	D	D	F		
	-2	-2	14		
3	D	Bb	F		
	-2	12	14		
4	F	D	A		
	14	-2	0		
5	F	G	A		
	14	18	0		
6	F	G	D		
	14	18	20		
7	F	Eb	F		
	-8	-12	-8		
8	A	G	F		
	0	-4	-8		
9	A	D	F		
	0	-2	-8		
10	F	Eb	D	C	F
	14	10	-2	16	14
11	F	G	C	D	
	-8	-4	-6	-2	
12	F	Bb	F		
	-8	-10	-8		
13	D	G	A	A	
	20	18	22	22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

Modified Just Intonation

D	G	D			
-2	18	20			
D	D	F			
-2	-25	-8			
D	Bb	F			
-2	-10	-8			
F	D	A			
14	20	22			
F	G	A			
-8	-4	-23			
F	G	D			
-8	-27	-25			
F	Eb	F			
14	33	14			
A	G	F			
0	18	37			
A	D	F			
22	20	14			
F	Eb	D	C	F	
-8	-12	-2	-6	-8	
F	G	C	D		
14	18	16	20		
F	Bb	F			
14	12	14			
D	G	A	A		
	-27	-23	-23		
G	F	E	F		
18	14	2	14		
G	F	Eb	F		
18	14	10	14		

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:** not significant
- **VERTICAL INTERVAL CONTEXT:** Semitones sung a perfect octave above the lower voice were 7 cents larger on average than those sung above other intervals
  - there were 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

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 ✓

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.

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

Future Work

Where I am going

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

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.

Howard, D. M. 2007a. Equal or non-equal temperament in a cappella SATB singing. *Logopedics Phoniatrics Vocology*, 32: 87–94.

Howard, D. M. 2007b. Intonation drift in a capella soprano, alto, tenor, bass quartet singing with key modulation. *Journal of Voice*, 21 (3): 300–15.

Jers, H. and S. Ternström. 2005. Intonation analysis of a multi-channel choir recording. *TMH-Quarterly Progress and Status Report* 47(1): 1–6.

Prame E. 1997. Vibrato extent and intonation in professional western lyric singing. *Journal of the Acoustical Society of America*, 102, pp. 616–21.

Schoen M. 1922. An experimental study of the pitch factor in artistic singing. *Psychological Monographs*, 31, pp. 230–59.

Vurma, A and J. Ross. 2006. Production and perception of musical intervals. *Music Perception*. 23(4): 331–44.

Vurma, A. 2010. Mistuning in two-part singing. *Logopedics Phoniatrics Vocology* 35: 24–33.

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

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**

Four-Part Singing

Praetorius - Es ist ein Ros’ ent sprungen