Studying singing voice performance through recorded audio

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

Introduction
Motivations.
1

A brief history
Quantitative approaches to performance analysis.
2

Extracting Performance Data from Recordings
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

Introduction

Why study musical performance?

- **Performances convey musicians' interpretations**
- **Performances are what listeners actually hear**
- **Studying performance can help us gain insight into**
  - commonality between performers
  - how an individual's performance practice evolves as they gain more experience
  - how performance practices evolve over time

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.
1

A brief history
Quantitative approaches to performance analysis.
2

Extracting Performance Data from Recordings
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

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

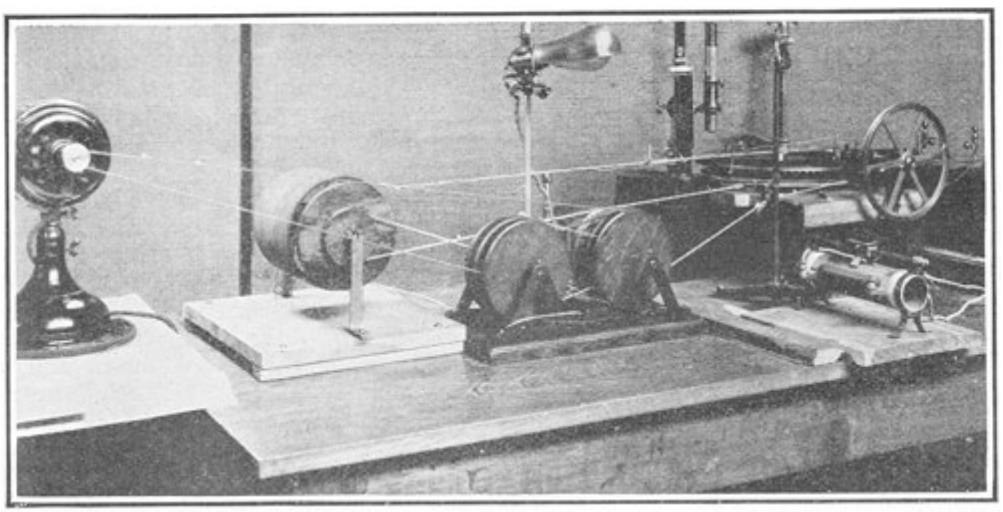
1895–1930 **1920–40s** **1960s** **1980s and 90s** **1990s and 2000s**

University of Iowa
Seashore and colleagues

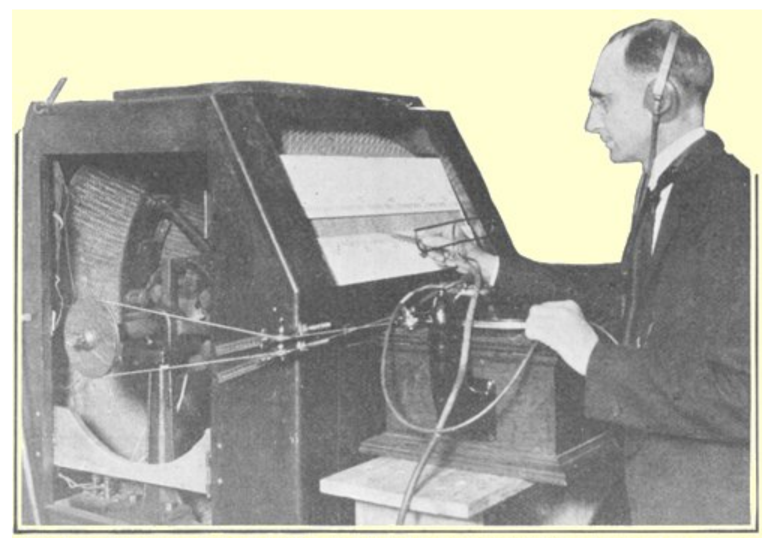
Quantitative Performance Analysis

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

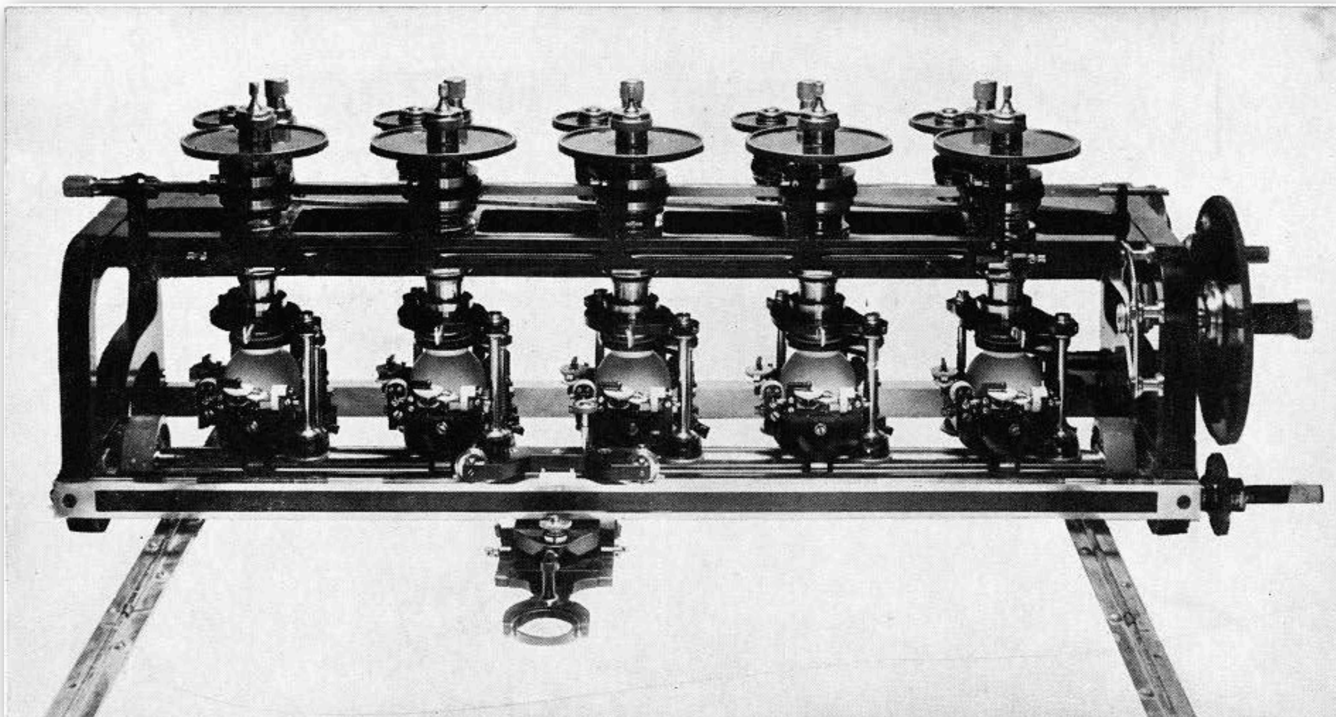


The tonoscope for analyzing the pitch of the tones on a disk phonograph record

Cary (1922)

Quantitative Performance

Phonophotography technique

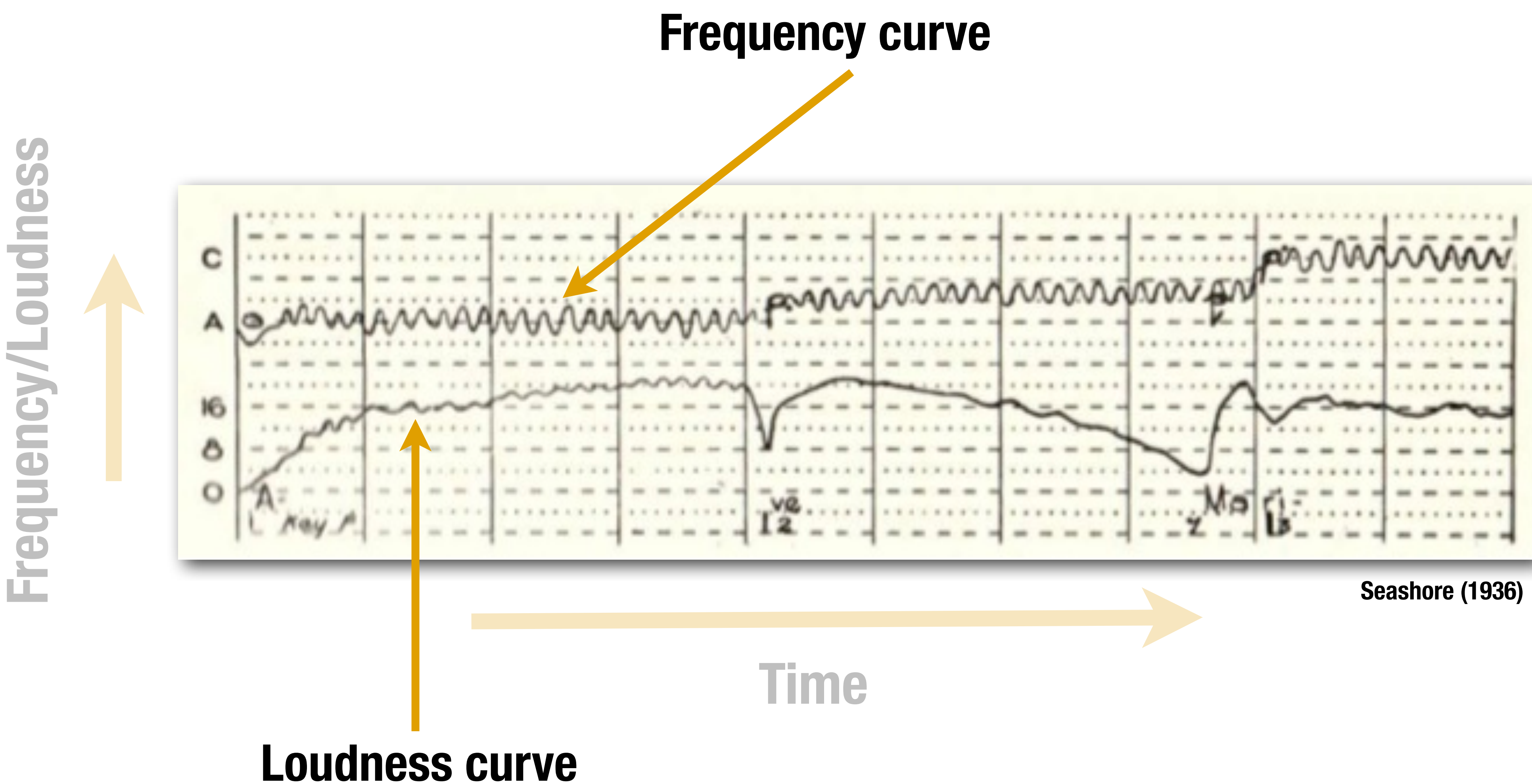


Henrici Harmonic Analyzer Seashore (1937)

- Frequency graphed in 10 cent units
- Intensity graphed in decibels
- Timing information as a function of linear space

Performance Scores

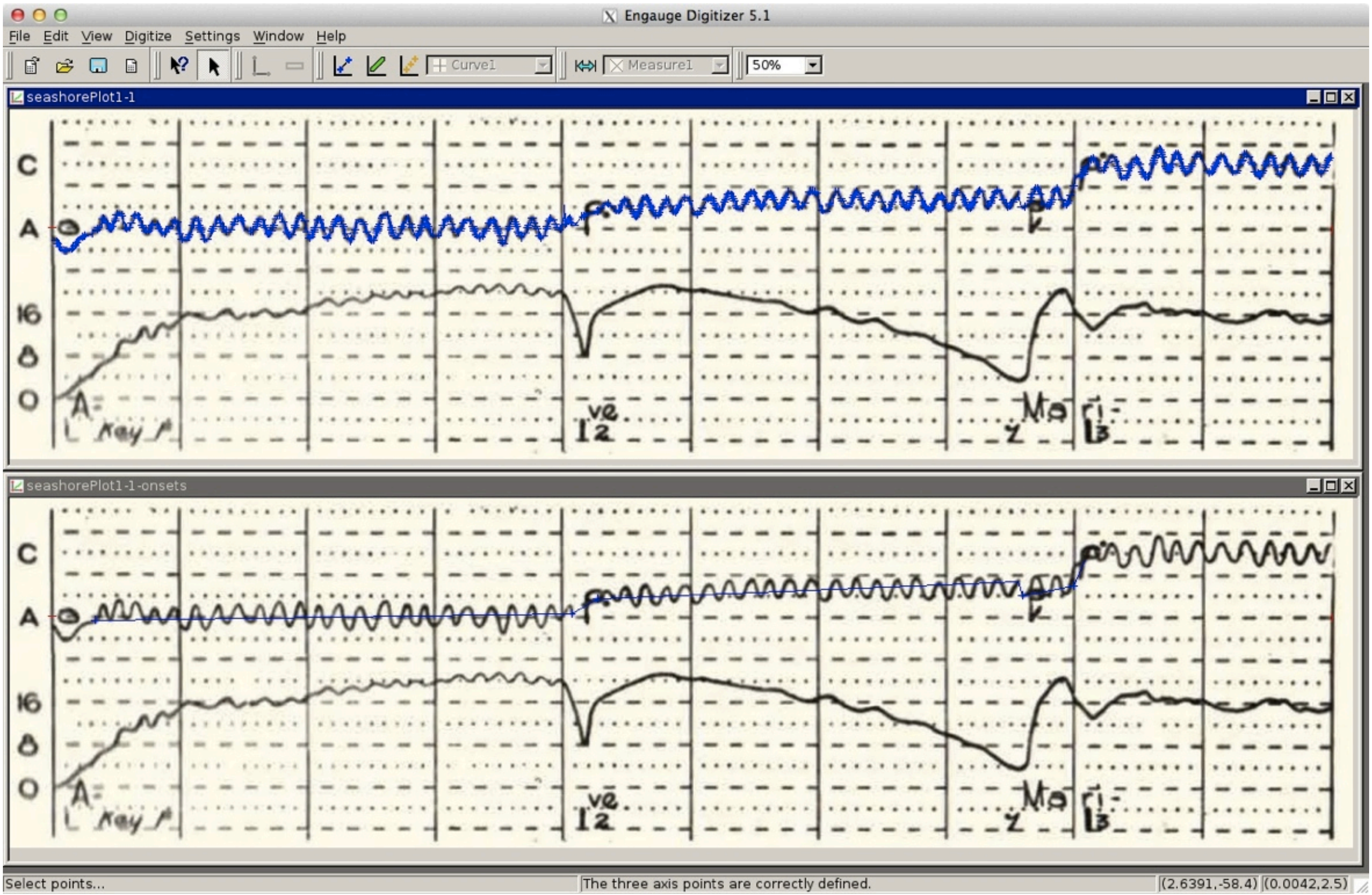
University of Iowa



Seashore (1936)

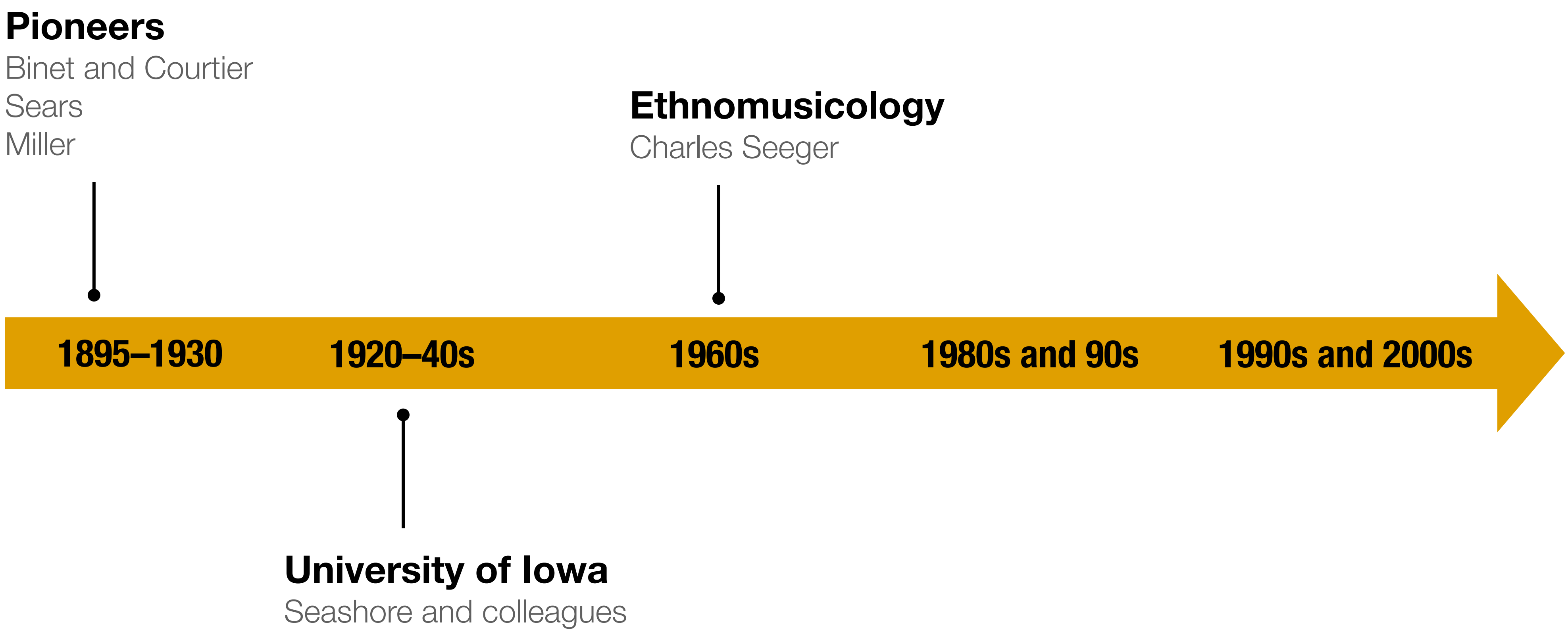
Performance Scores

Digitizing the data



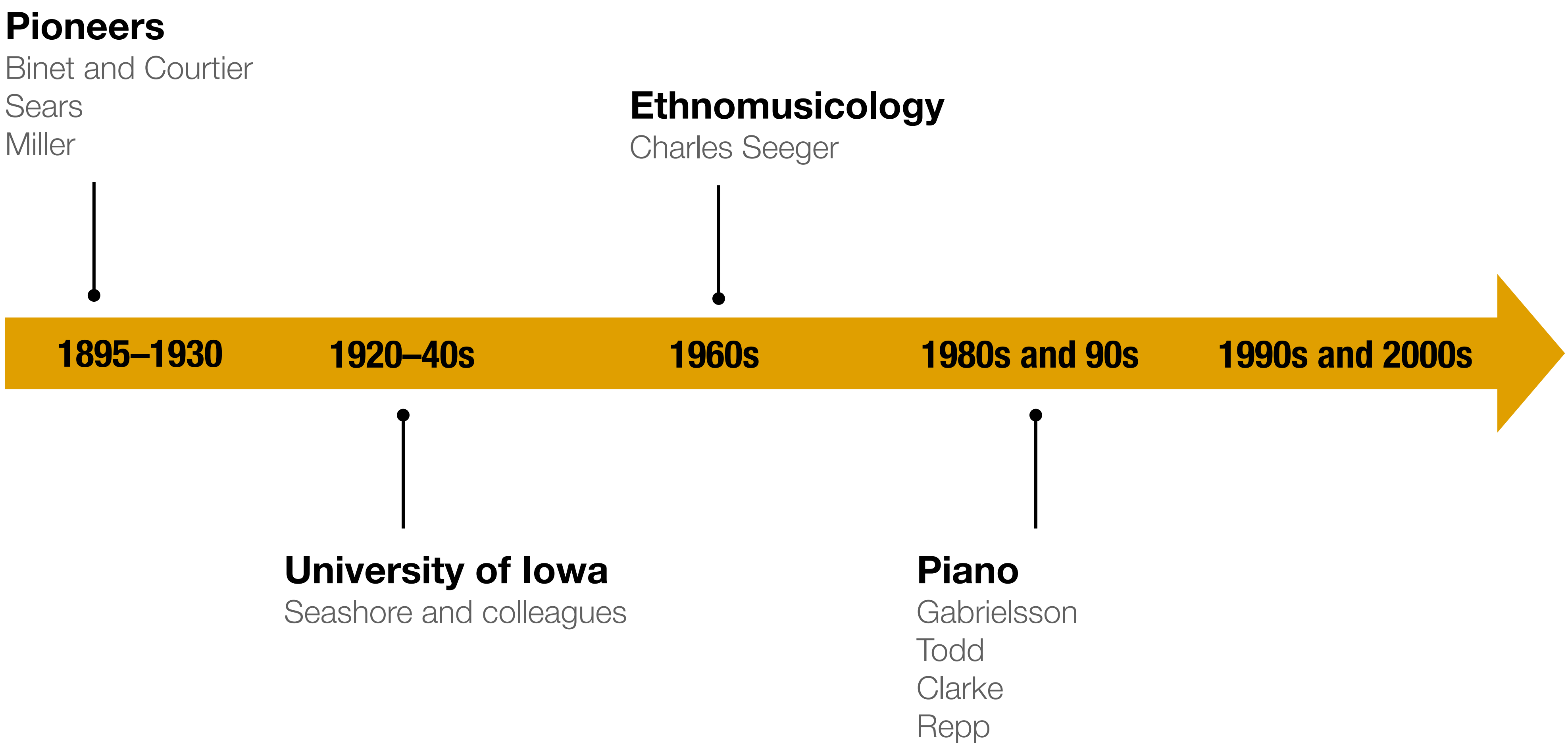
Quantitative Performance Analysis

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

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

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

Computational Models
Friberg
Mazola
Widmer

Introduction
Motivations.
1

A brief history
Quantitative approaches to performance analysis.
2

Extracting Performance Data from Recordings
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

Studying Audio Recordings

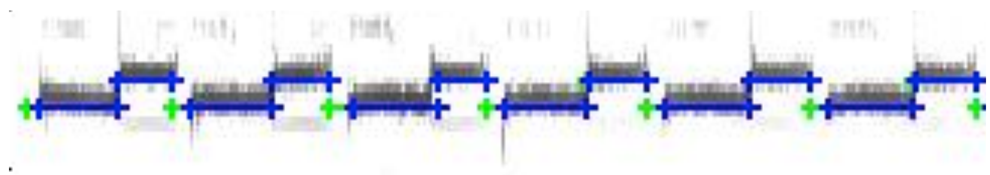
Advantages and challenges of extracting data

- Advantages
  - Allows for existing recordings to be studied
- Challenges
  - Difficulty of extracting data accurately
  - Questions of how to model the extracted data in a perceptually meaningful way

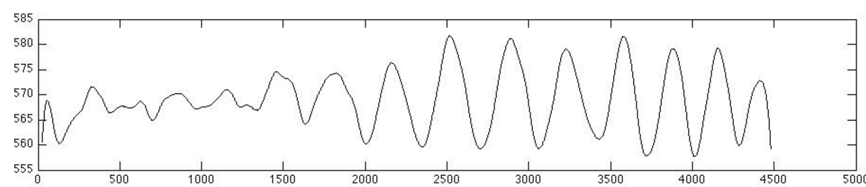
Extracting Performance Data

Example workflow

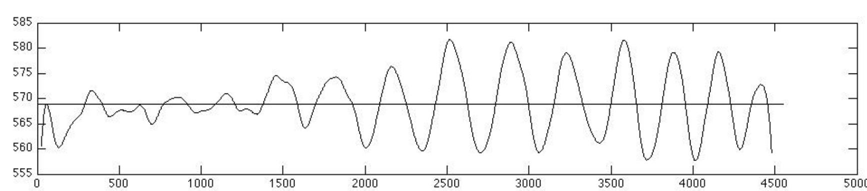
Identify Note Onsets and Offsets



Fundamental Frequency (F0) Estimation



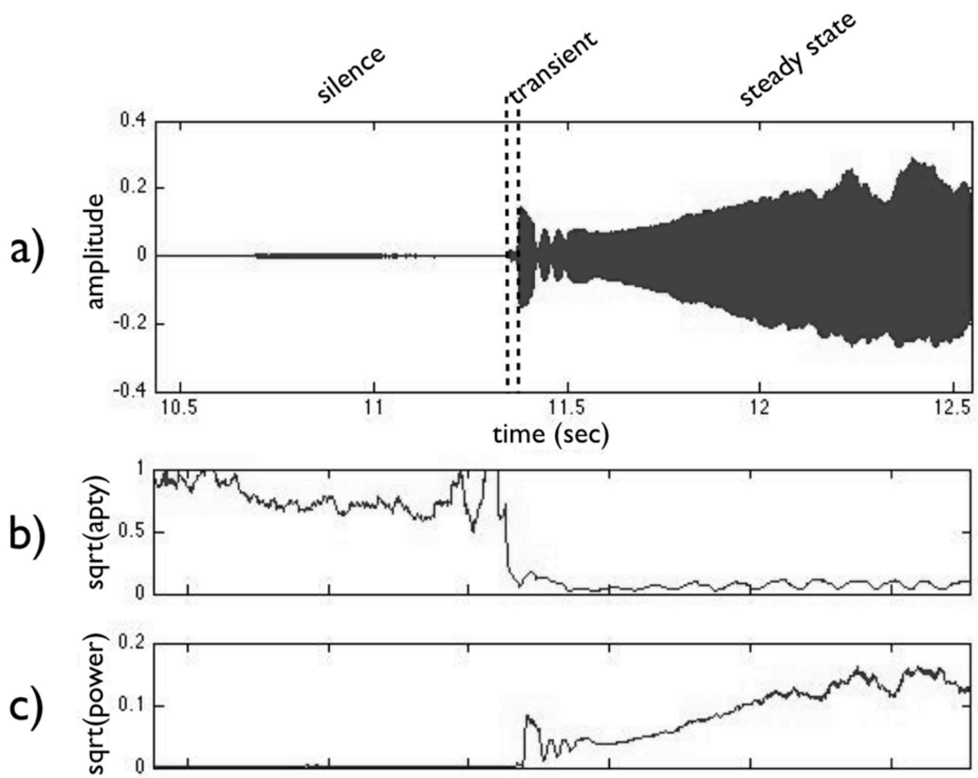
Perceived Pitch



Monophonic audio

Identifying onsets and offsets

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

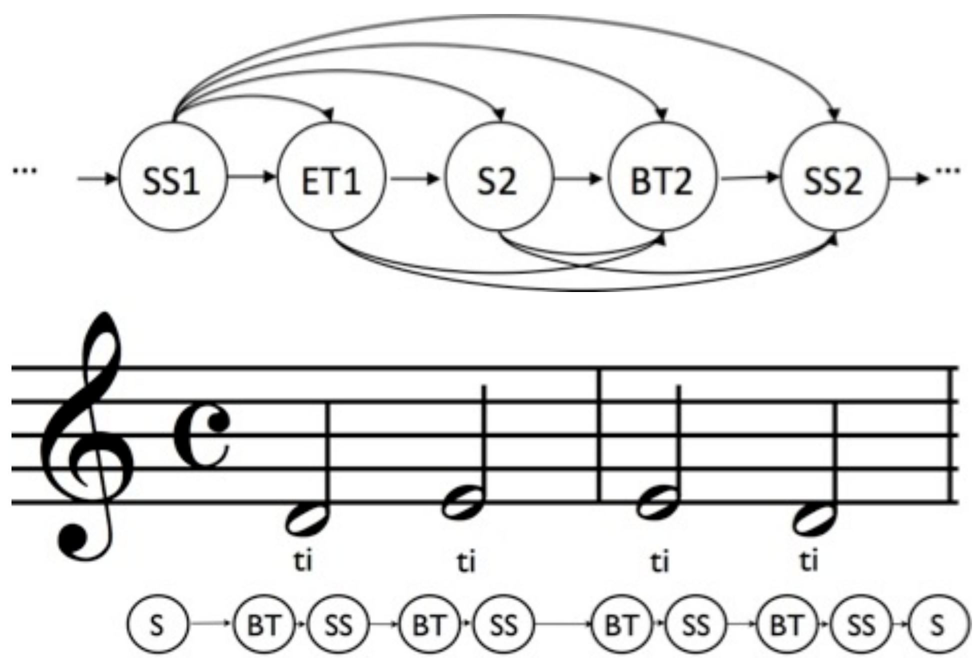


Devaney, Mandel, and Ellis (2009)

Monophonic audio

Identifying onsets and offsets

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

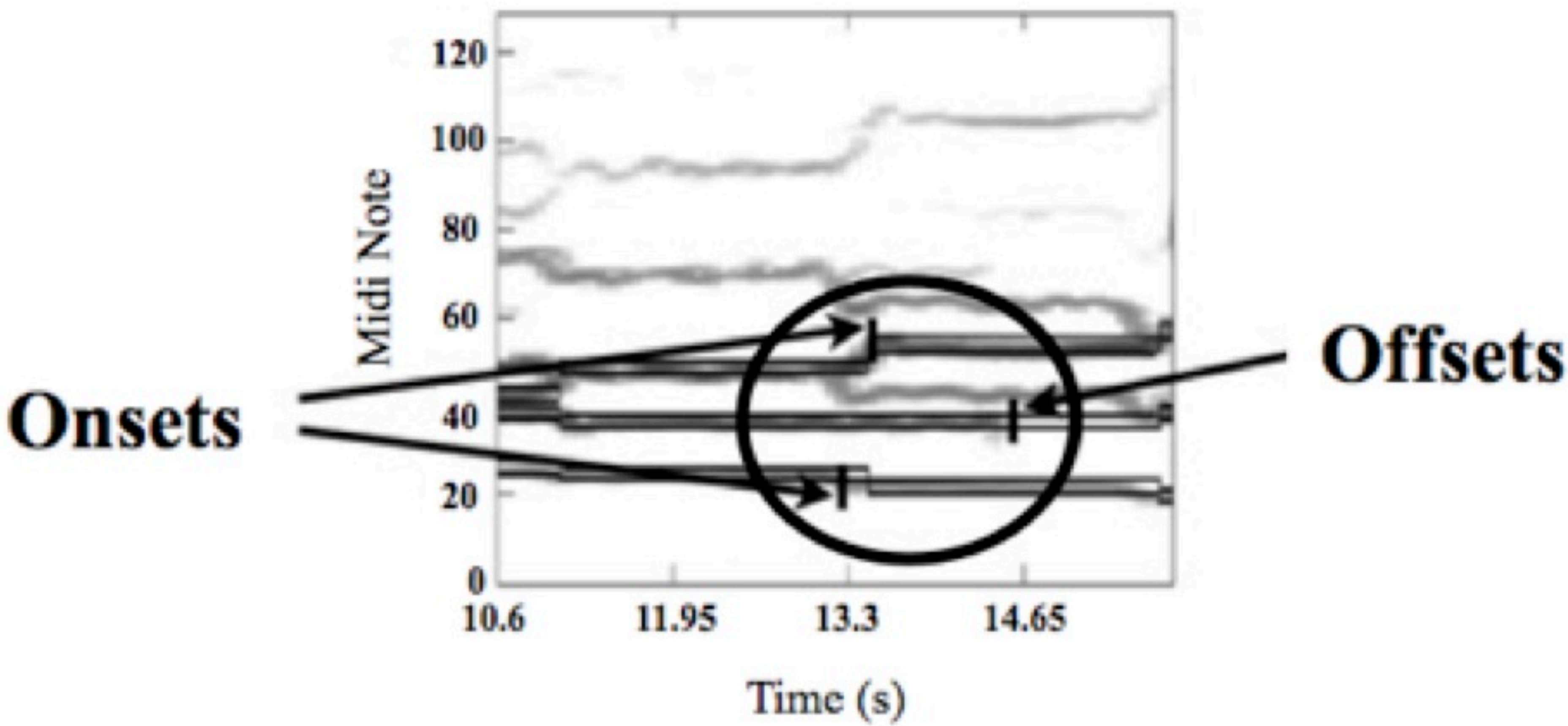


- **Improves median alignment error from 52 ms to 28 ms**

Devaney, Mandel, and Ellis (2009)

Moving towards polyphonic recordings

Still using MIDI-audio alignment



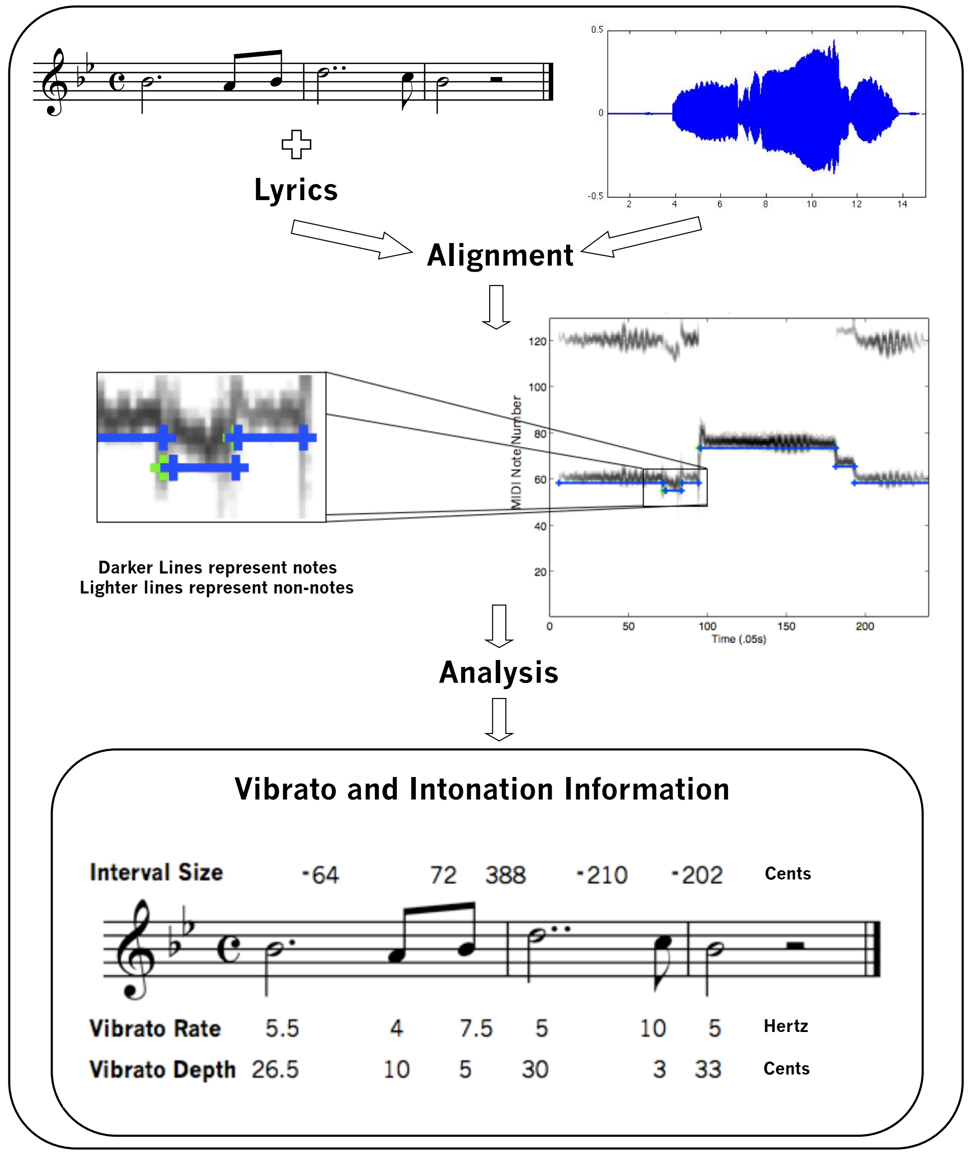
How to extract pitch and loudness data still an open question

Devaney (2014)

AMPACT

Automatic Music Performance Analysis and Comparison Toolkit

www.ampact.org



Devaney, Mandel, and Fujinaga (2012)

Introduction
Motivations.

1

A brief history
Quantitative approaches to performance analysis.

2

Extracting Performance Data from Recordings
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

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

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)

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

- 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)
    - vertical interval context (ensemble and 2-part)
  - accompaniment
    - 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

- **HARMONIC CONTEXT**
  - **Solo singing**
    - Non-pros exhibited a significant difference between semitones in leading tone and non-leading tone contexts
      - semitones in a leading context were significantly smaller on average

Commonality between performers

Observable trends

- **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)

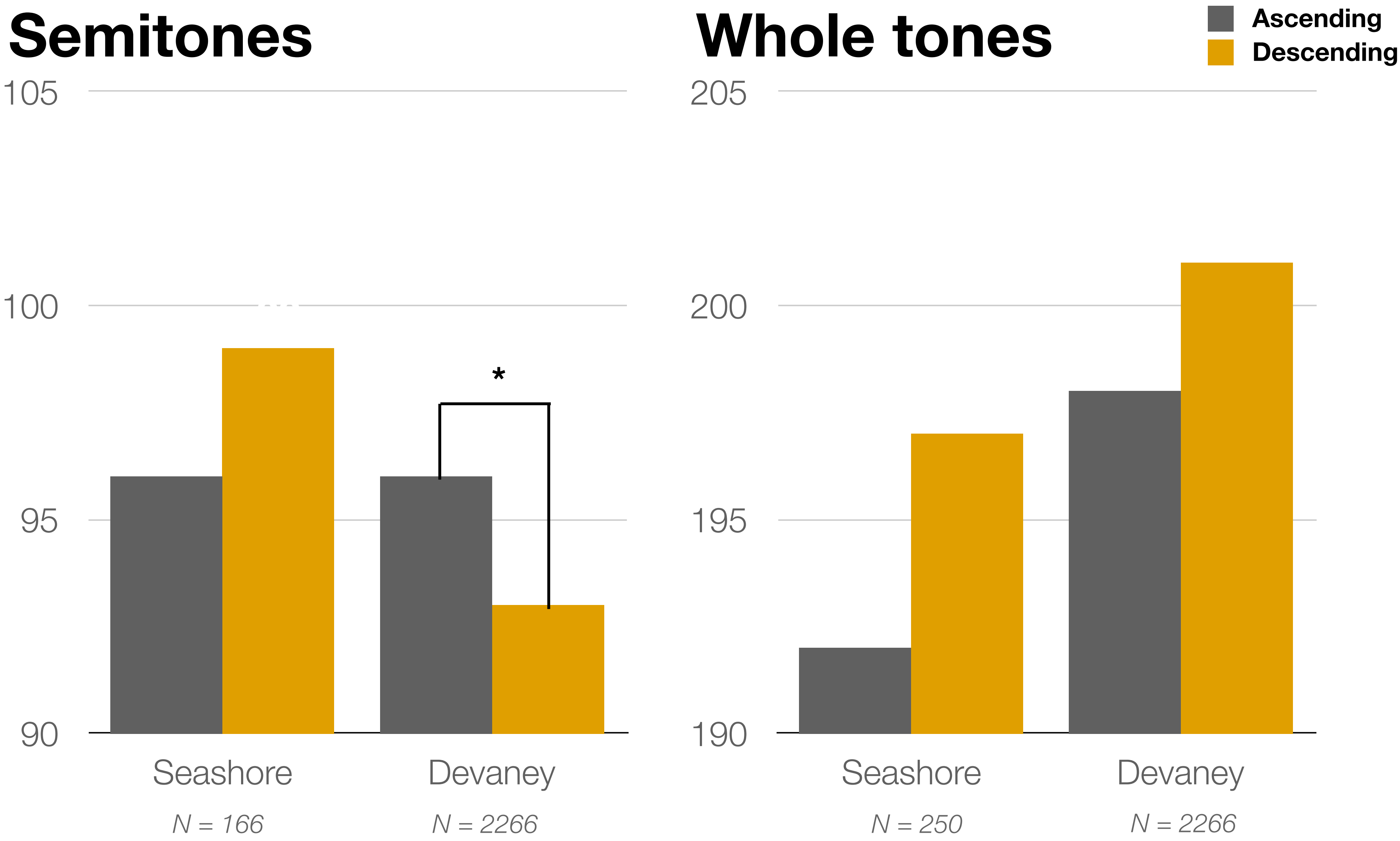
Summary of Results

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 ✓

Incorporating Seashore data

Comparative analysis of Seashore and contemporary data



Introduction
Motivations. 1

A brief history
Quantitative approaches to performance analysis. 2

Extracting Performance Data from Recordings
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
  - highlighted some of the challenges of automatically extracting performance data from recordings and how to address them
  - summarized some of my findings on vocal intonation practices in the western art music tradition

Future Work

Where I am going

- **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**
  - focused experiments about interactions in ensembles
  - studying existing recordings of a singer performing the same piece at different points in their career
- **Integrating more qualitative information**
  - performers intentionality
  - listener perception/reception

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

Cary, H. December 1922. Are you a musician? Professor Seashore's specific psychological tests for specific musical abilities. Scientific American 326–327

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

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 International Society for Music Information Retrieval conference. 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.

Seashore, C. 1938. Psychology of Music. Iowa City, IA: University of Iowa Press.

Seashore, H. G. 1936. An objective analysis of artistic singing. In University of Iowa Studies in the Psychology of Music. Vol. IV: Objective Analysis of Musical Performance, ed. C. Seashore, 12–157. Iowa City, IA: University of Iowa..

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.