Video-based Vibrato Detection and Analysis for Polyphonic String Music

Bochen Li, Karthik Dinesh, Gaurav Sharma, Zhiyao Duan

Audio Information Research Lab
University of Rochester

The 18th International Society for Music Information Retrieval

Oct 23-27, 2017

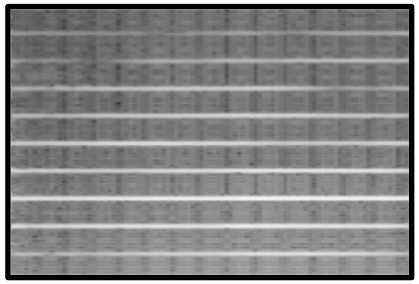
Suzhou, China

Introduction: Vibrato in Music

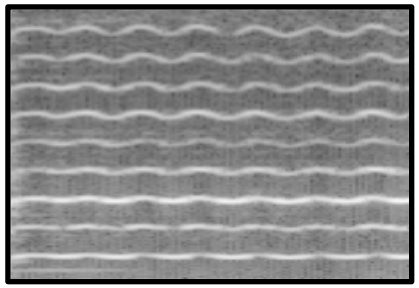
- Important artistic effect
- Pitch modulation of a note in a periodic fashion
- Characterized by Rate & Extent

Audio

Spectrogram



Non-vibrato



Vibrato

Applications of Vibrato Analysis

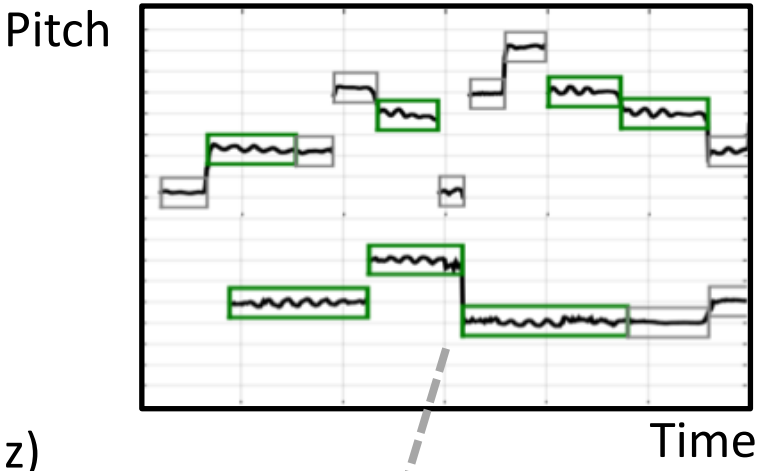
- Musicological studies
- Sound synthesis
- Voice extraction

Introduction: Problem Statement

Vibrato Detection & Analysis for polyphonic music played by string instruments

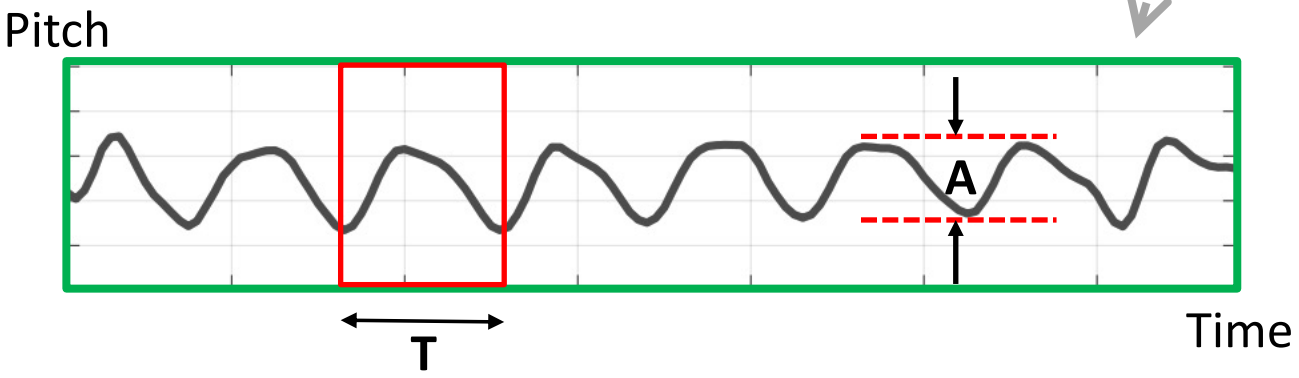
Vibrato Detection

- Note-level vibrato/non-vibrato classification



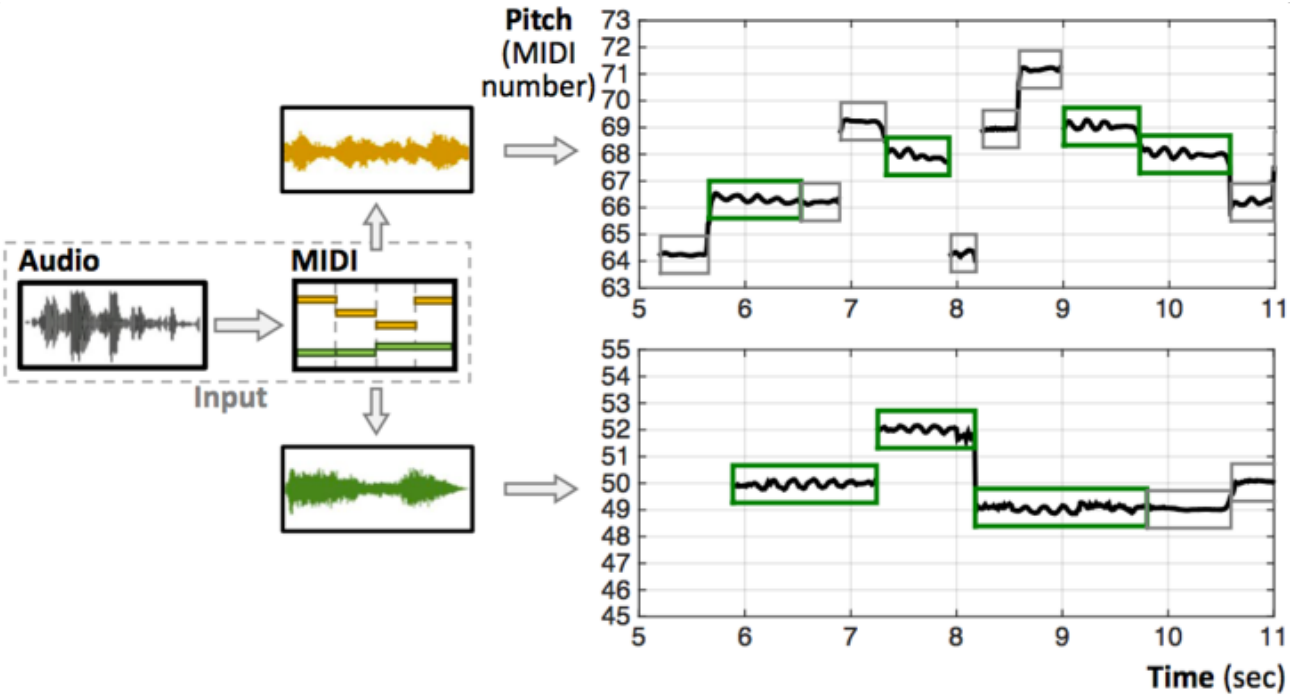
Vibrato Analysis

- Vibrato rate: speed of pitch variation (1/T Hz)
- Vibrato extent: amount of pitch variation (A cents)



Introduction: Prior Audio-based Methods

- Score-informed [Abeßer et al. 2015] (Baseline)



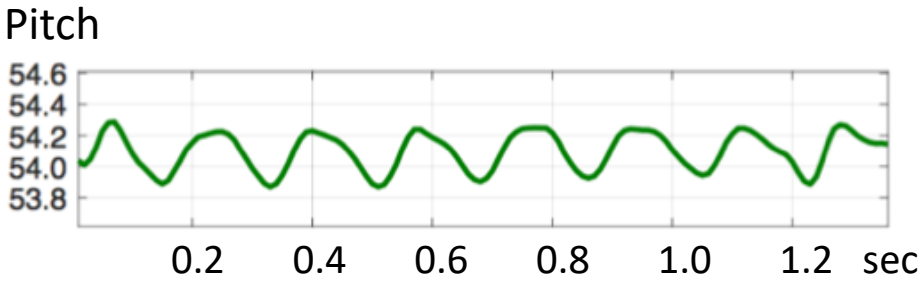
- Template-based [Driedger et al. 2016]
- Harmonic partial [Hsu et al. 2010]

Major drawbacks

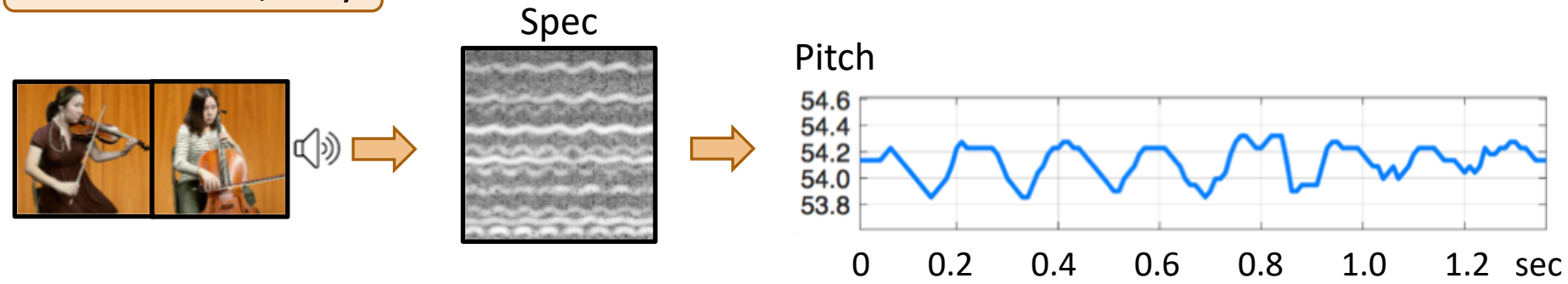
- One source from mixture
- Fails in high polyphony

Proposed Method Overview and Key Contribution

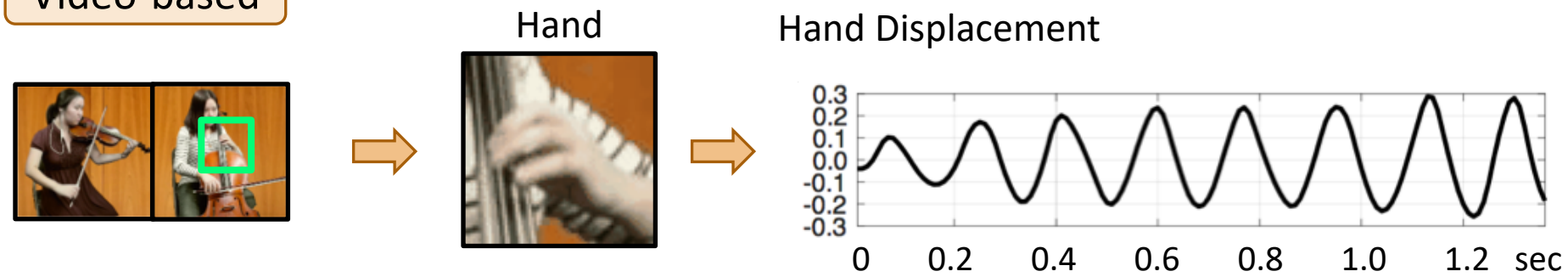
Ground-truth



Audio-based, Poly

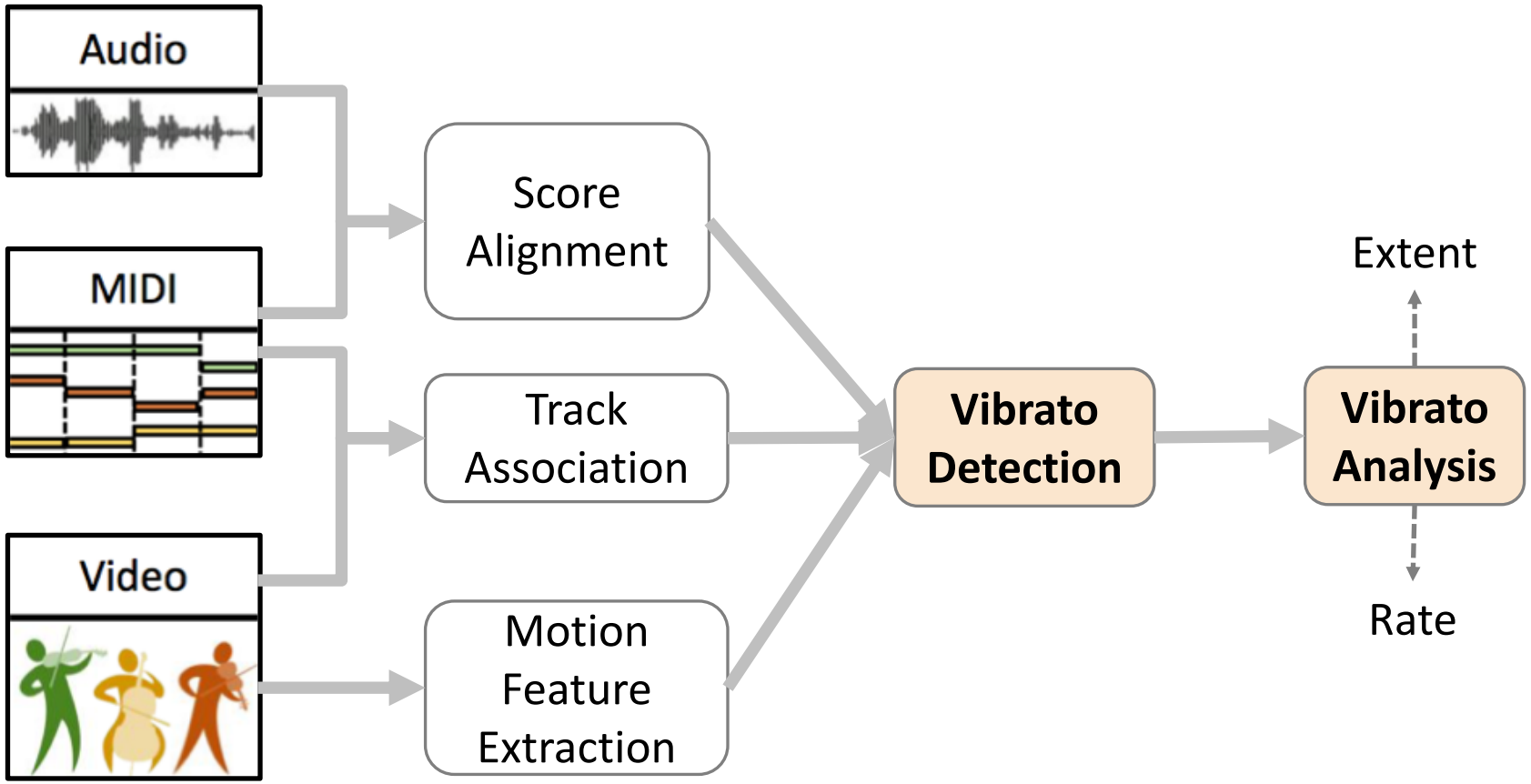


Video-based



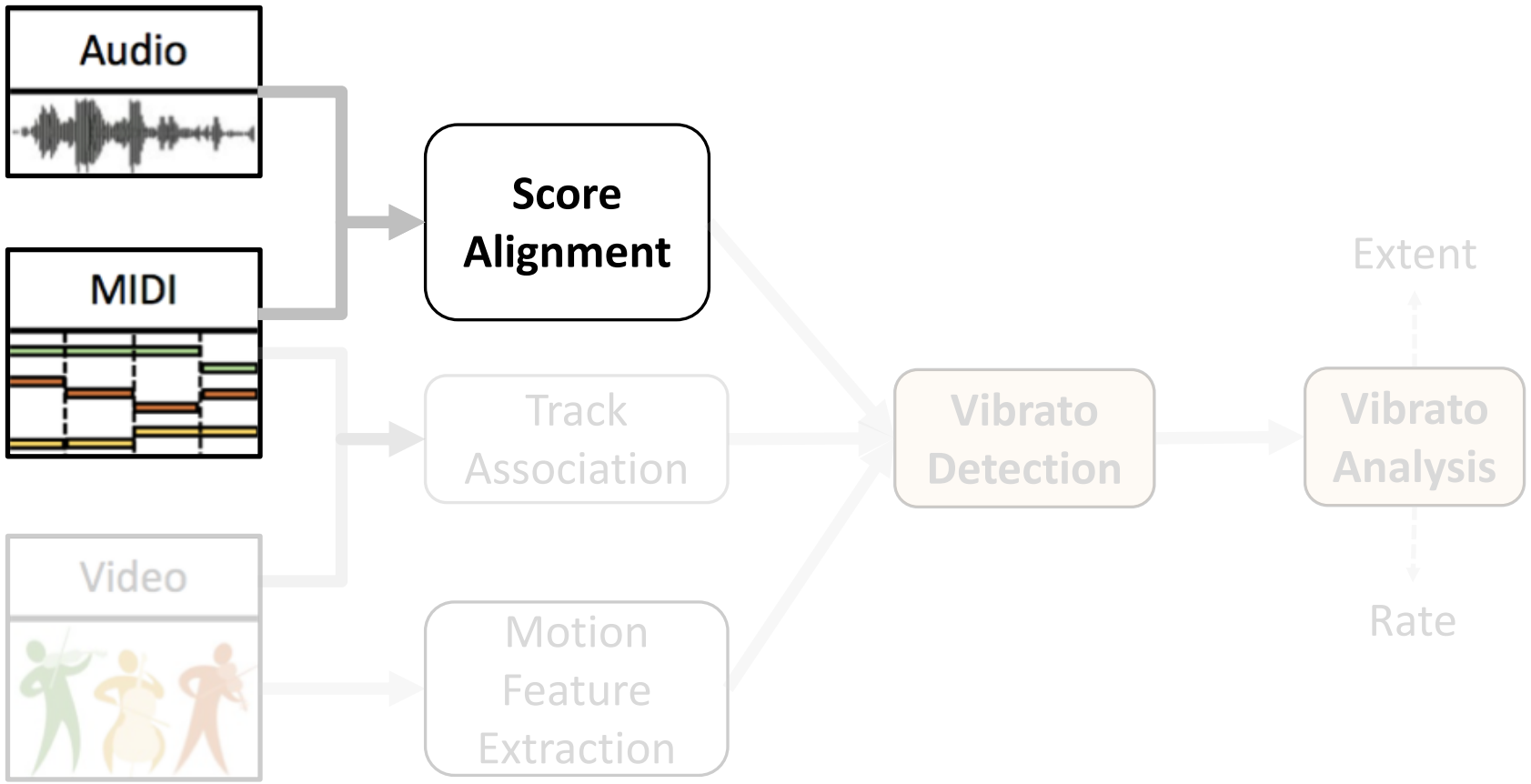
Proposed Method Overview

Video-based Method



Proposed Method

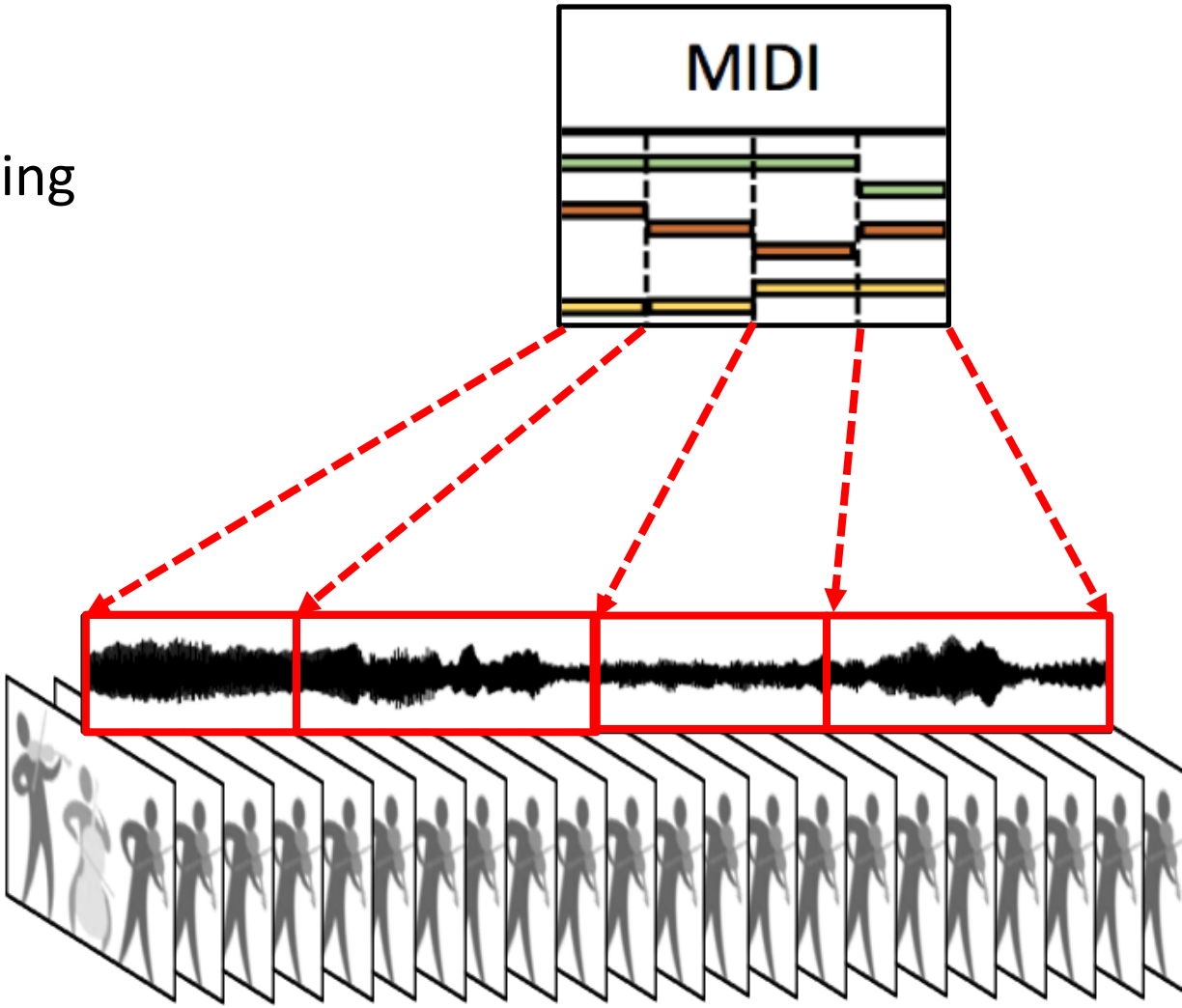
Score Alignment



Proposed Method

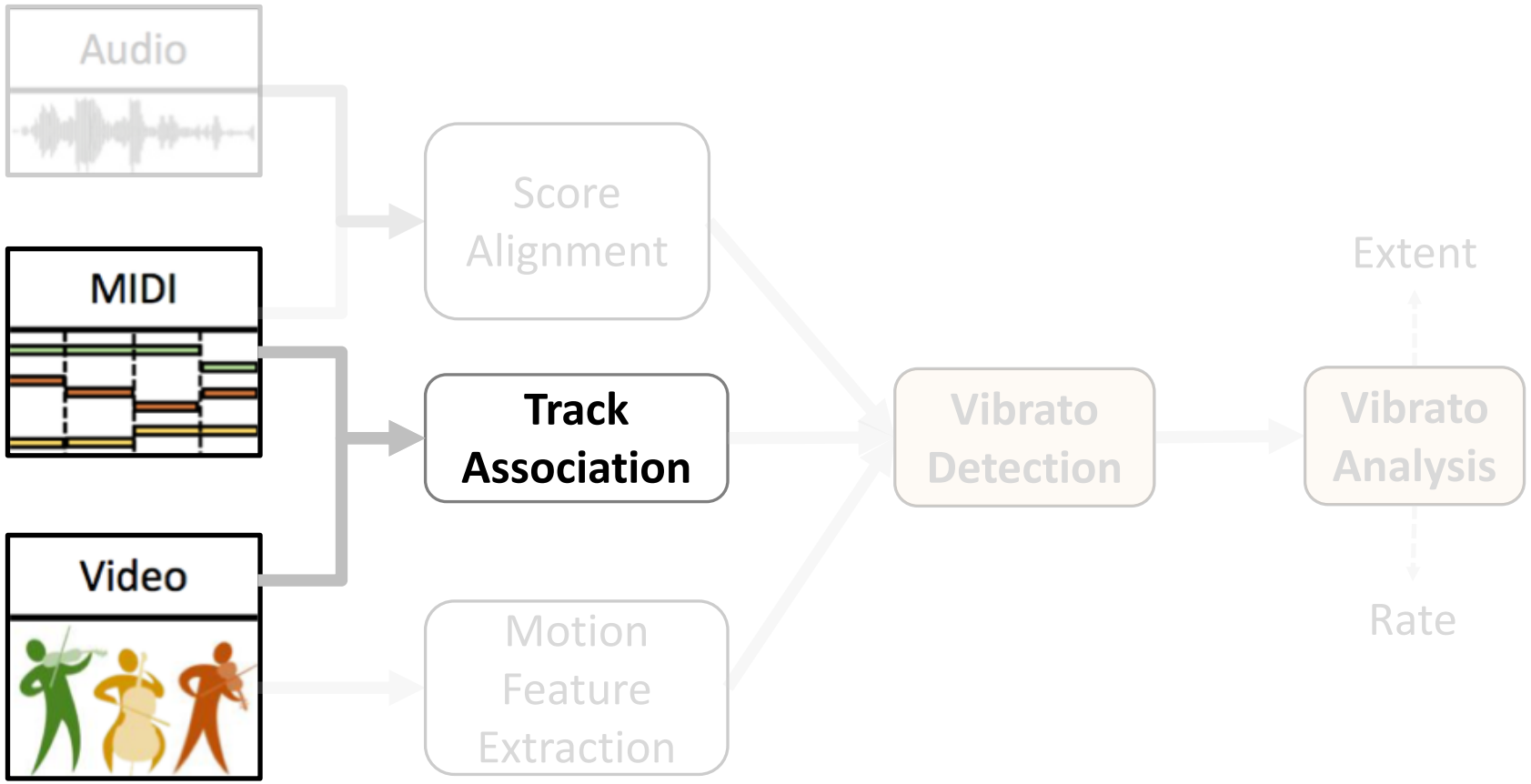
Score Alignment

- Chroma feature
- Dynamic Time Warping



Proposed Method

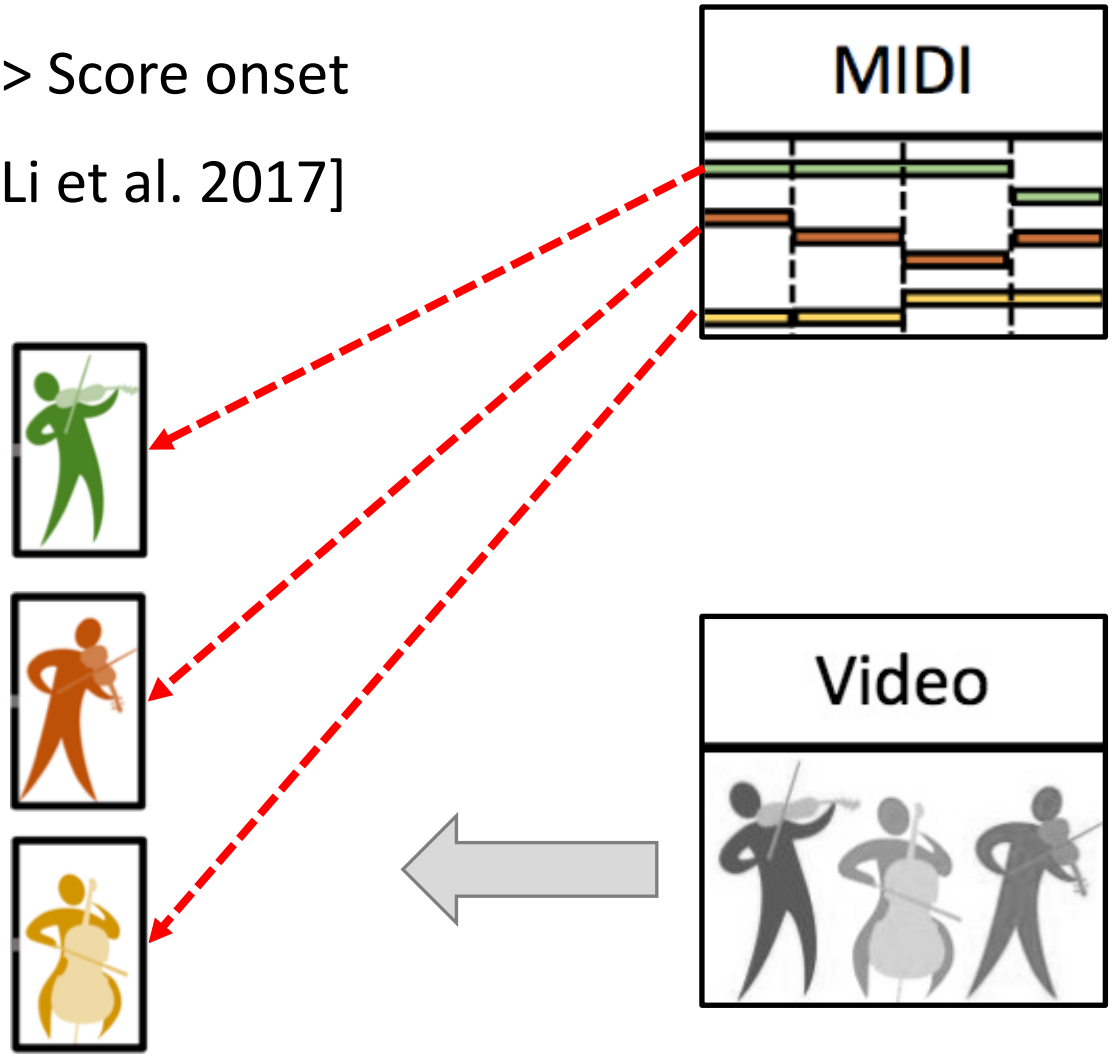
Track-player Association



Proposed Method

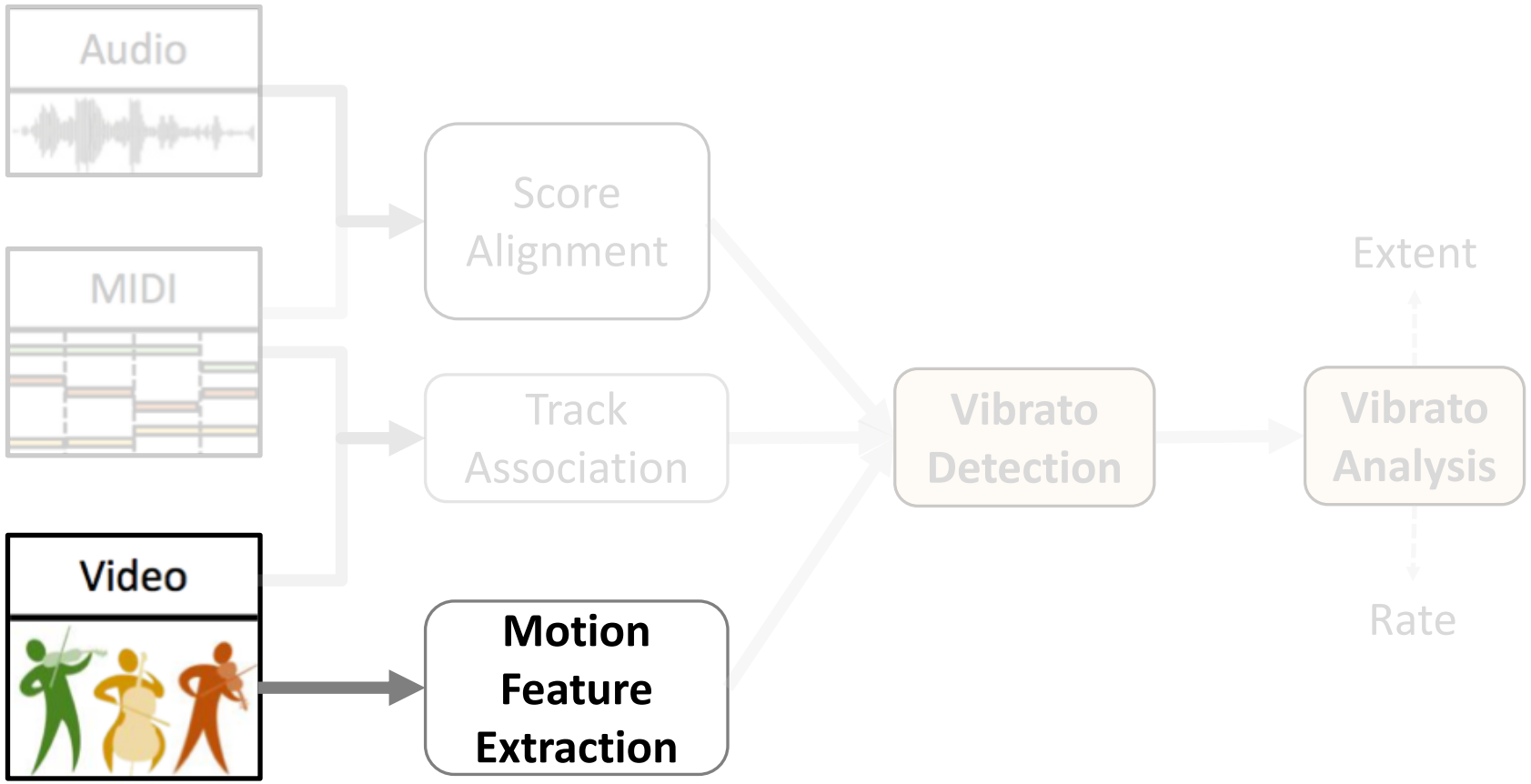
Track-player Association

- Bow motion <--> Score onset
- Previous work [Li et al. 2017]



Proposed Method

Track-player Association



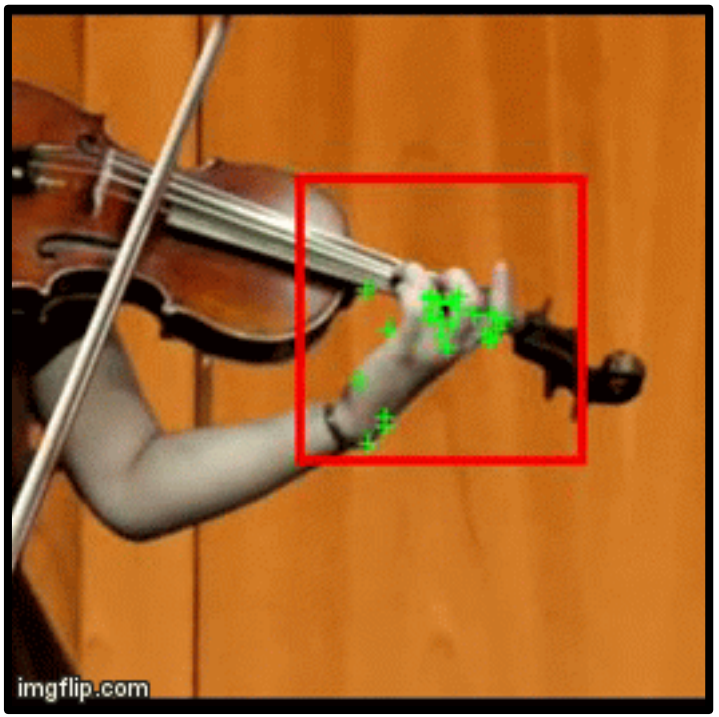
Proposed Method

Motion Feature Extraction

- Hand tracking

  - KLT tracker with 30 feature points

  - Bounding box: 70 x 70 pixels



Proposed Method

Motion Feature Extraction

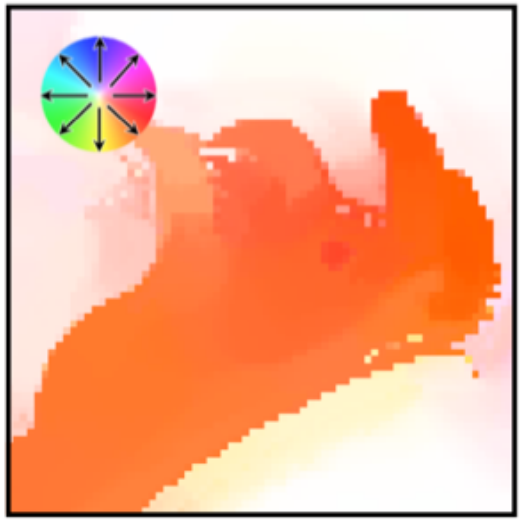
- Fine-grained motion capture

  - Optical flow estimation → pixel-level motion velocities

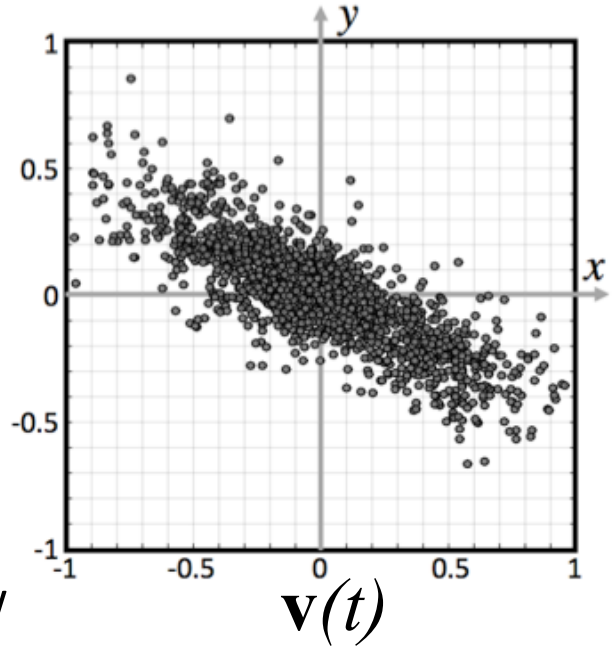
  - Frame-wise average: $\mathbf{u}(t) = [u_x(t), u_y(t)]$

  - Subtract moving mean: $\mathbf{v}(t) = \mathbf{u}(t) - \bar{\mathbf{u}}(t)$

Original Frame

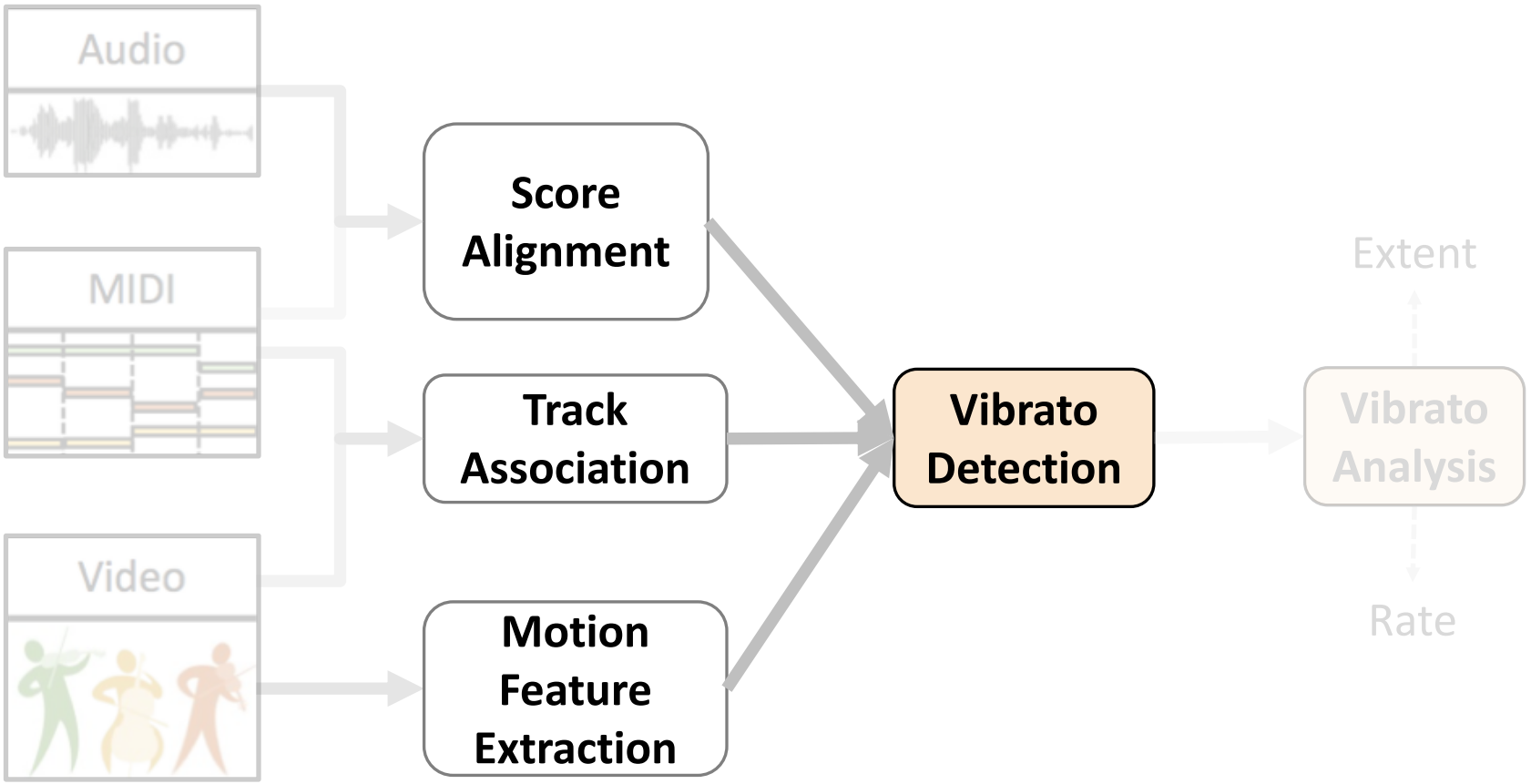


Color-encoded Optical Flow



Proposed Method

Track-player Association



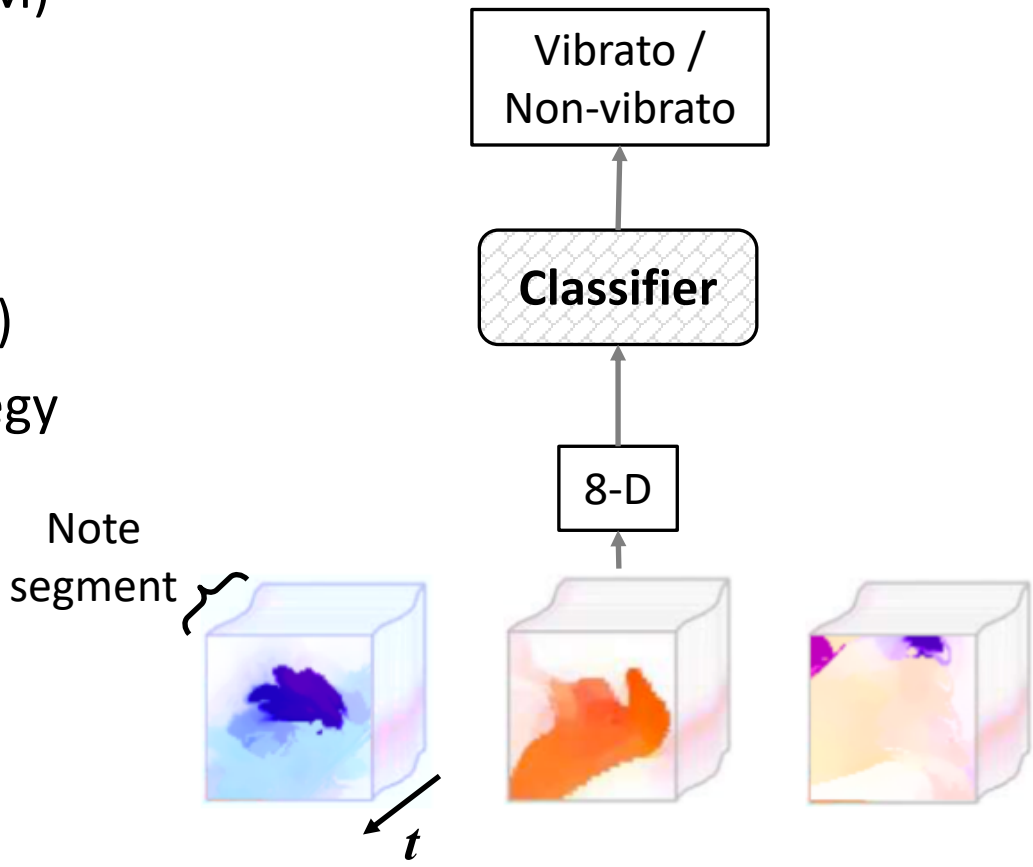
Proposed Method

Vibrato Detection

Method 1: Supervised framework

- Support Vector Machine (SVM)
- 8-D feature

  Zero-crossing rate (4-D)
  Frequency (2-D)
  Auto-correlation peaks (2-D)
- Leave-one-out training strategy



Proposed Method

Vibrato Detection

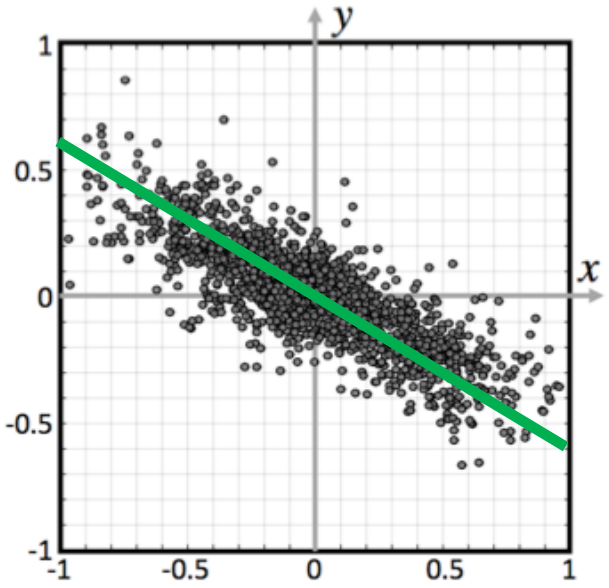
Method 2: Unsupervised framework

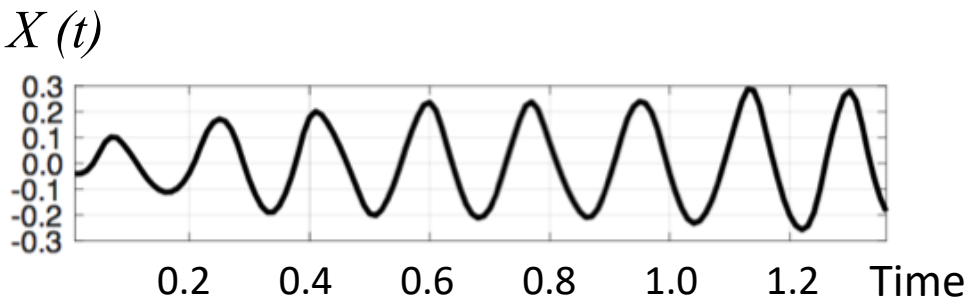
- Principal Component Analysis (PCA)
- 1-D Motion Velocity Curve:

$$V(t) = \frac{\mathbf{v}(t)^T \tilde{\mathbf{v}}}{\|\tilde{\mathbf{v}}\|}$$

- Integration → Motion Displacement Curve:

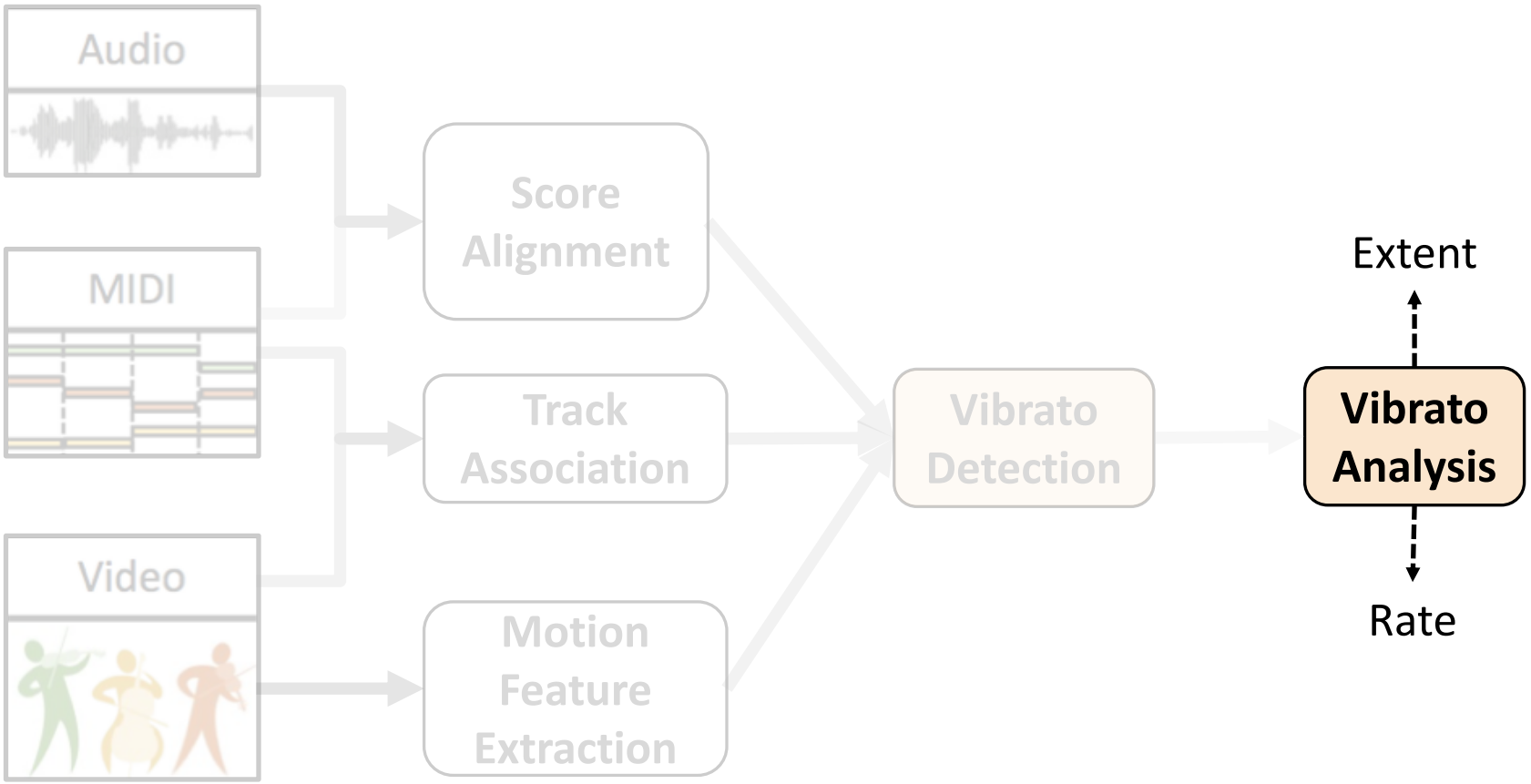
$$X(t) = \int_0^t V(\tau) d\tau$$





Proposed Method

Vibrato Analysis



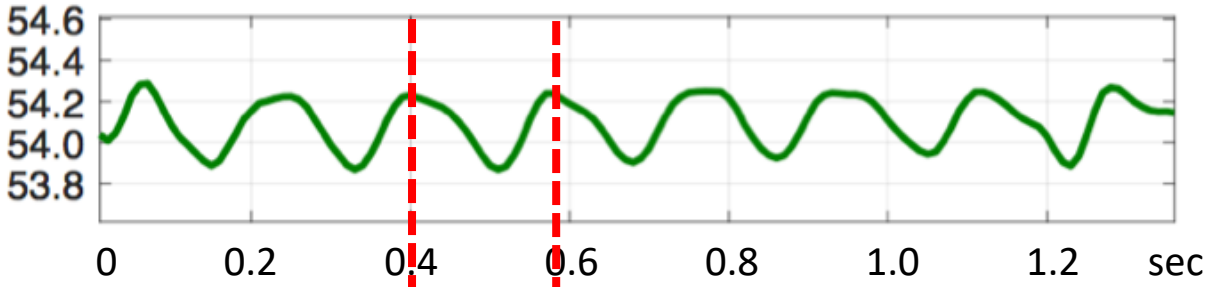
Proposed Method

Vibrato Analysis

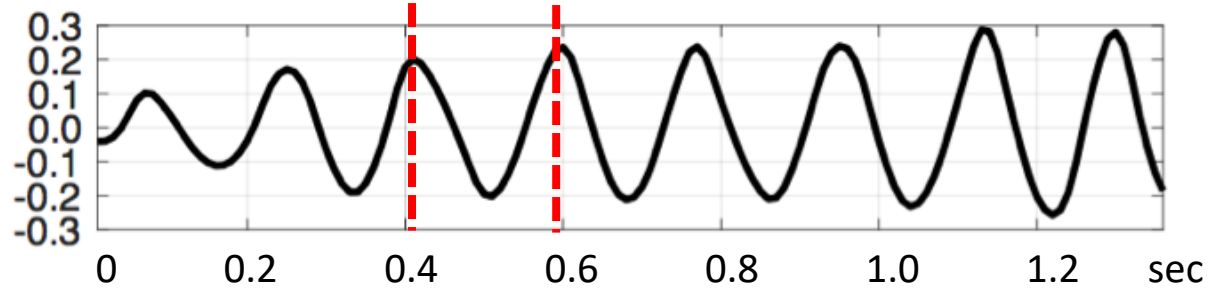
Rate

- Motion rate = Vibrato rate
- Quadratic interpolation
- Peak distance on auto-correlation of motion curve $X(t)$

Ground-truth pitch contour



Motion displacement Curve $X(t)$



Proposed Method

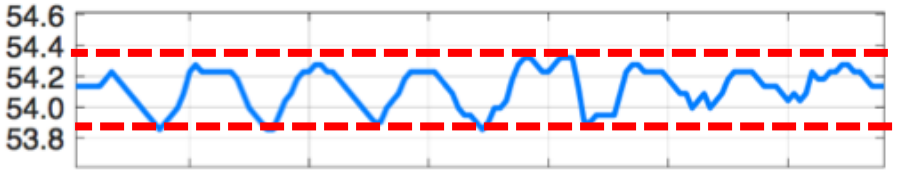
Vibrato Analysis

Extent

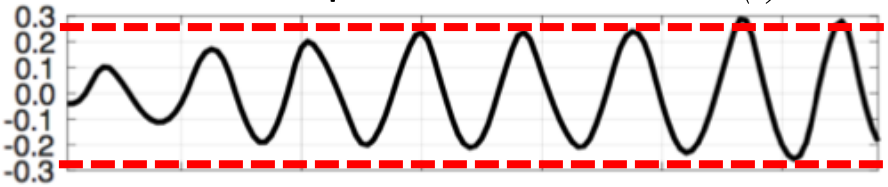
- Motion extent ≠ Vibrato extent
- Pixel → Musical cents
- Scale motion curve $X(t)$ to fit pitch contour

Ground-truth pitch contour

Estimated pitch contour



Motion displacement Curve $X(t)$



Estimated vib extent

$$\hat{v}_e = \arg\min_{v_e} \sum_{t=t^{on}}^{t^{off}} \left| 100 \cdot F(t) - v_e \frac{X(t)}{\hat{w}_e} \right|^2$$

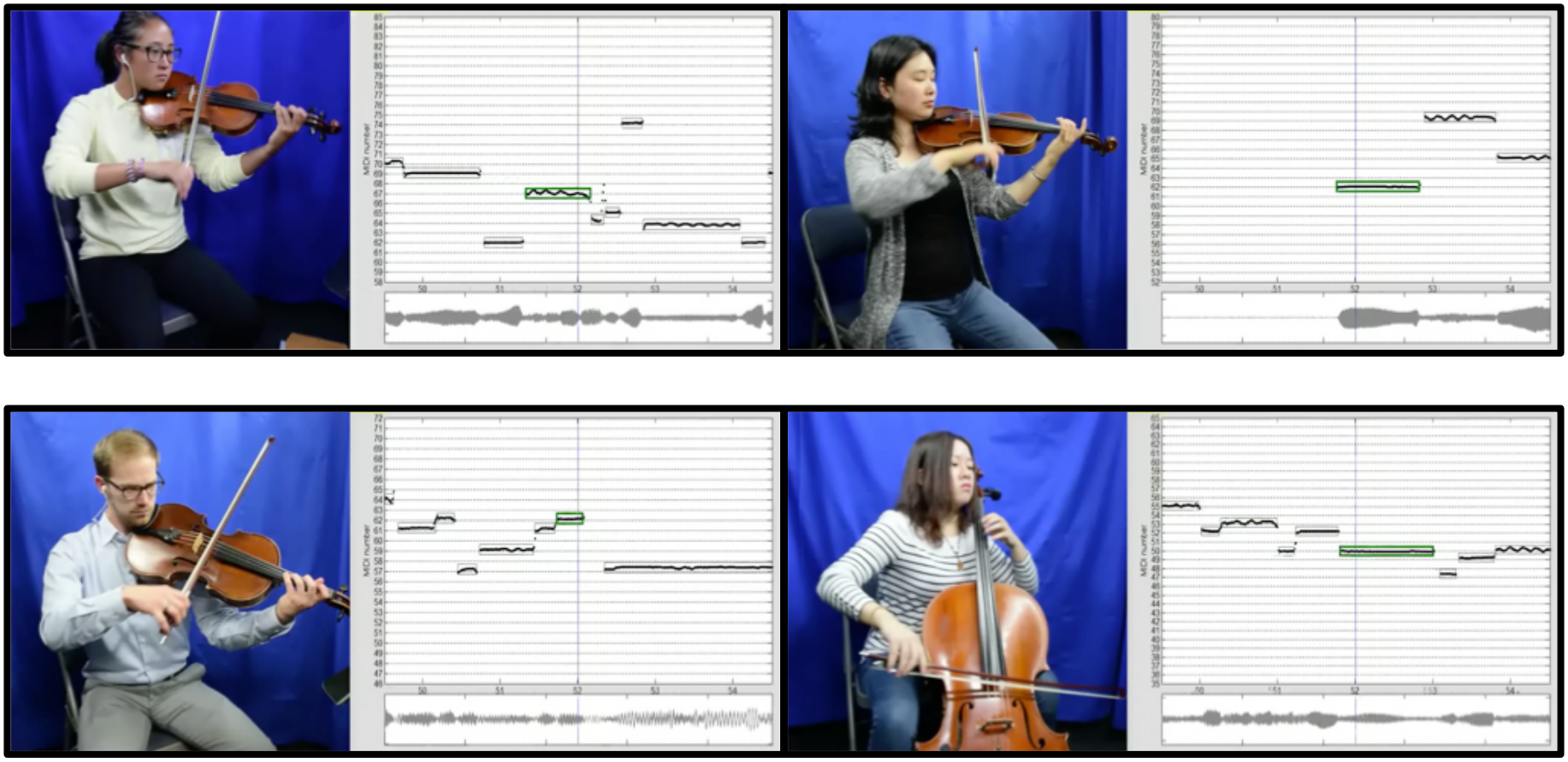
Pitch contour

Motion extent

Demo of Dataset

Dataset: URMP Dataset

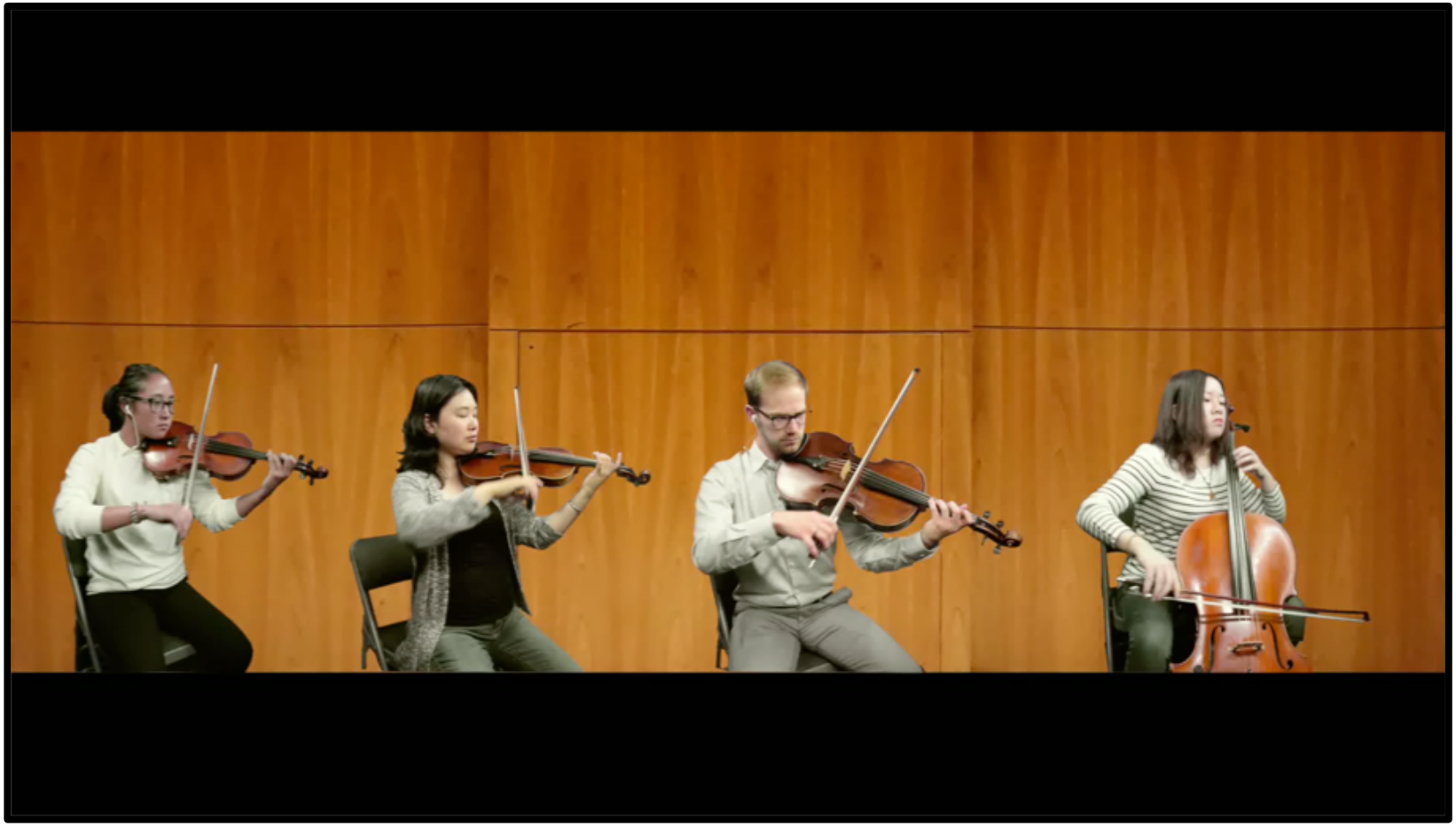
- Individually recorded in sound booth
- Annotated frame-level / note-level pitch



Demo of Dataset

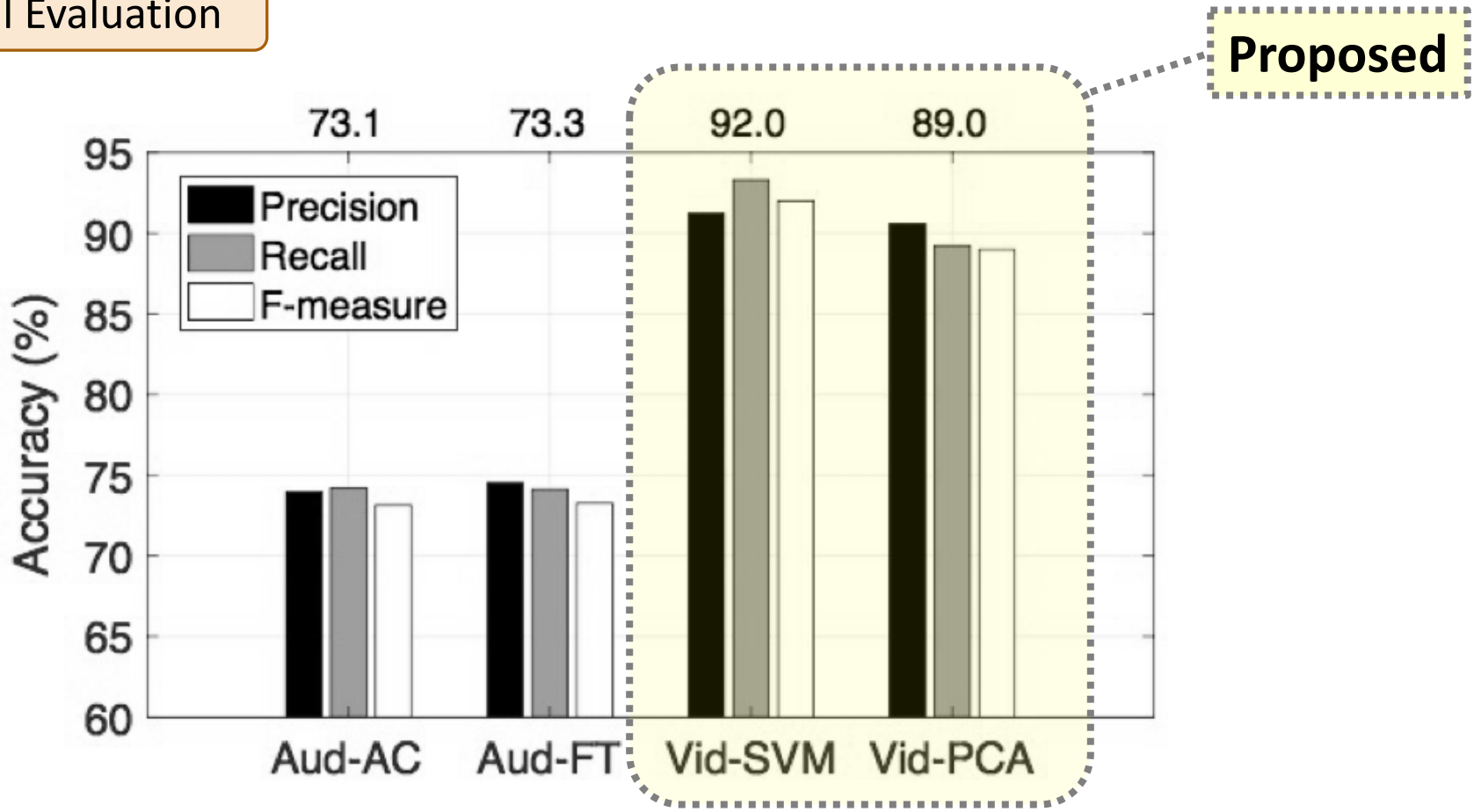
Dataset: URMP Dataset

- Assembled together with concert stage background



Experiments: Vibrato Detection Results

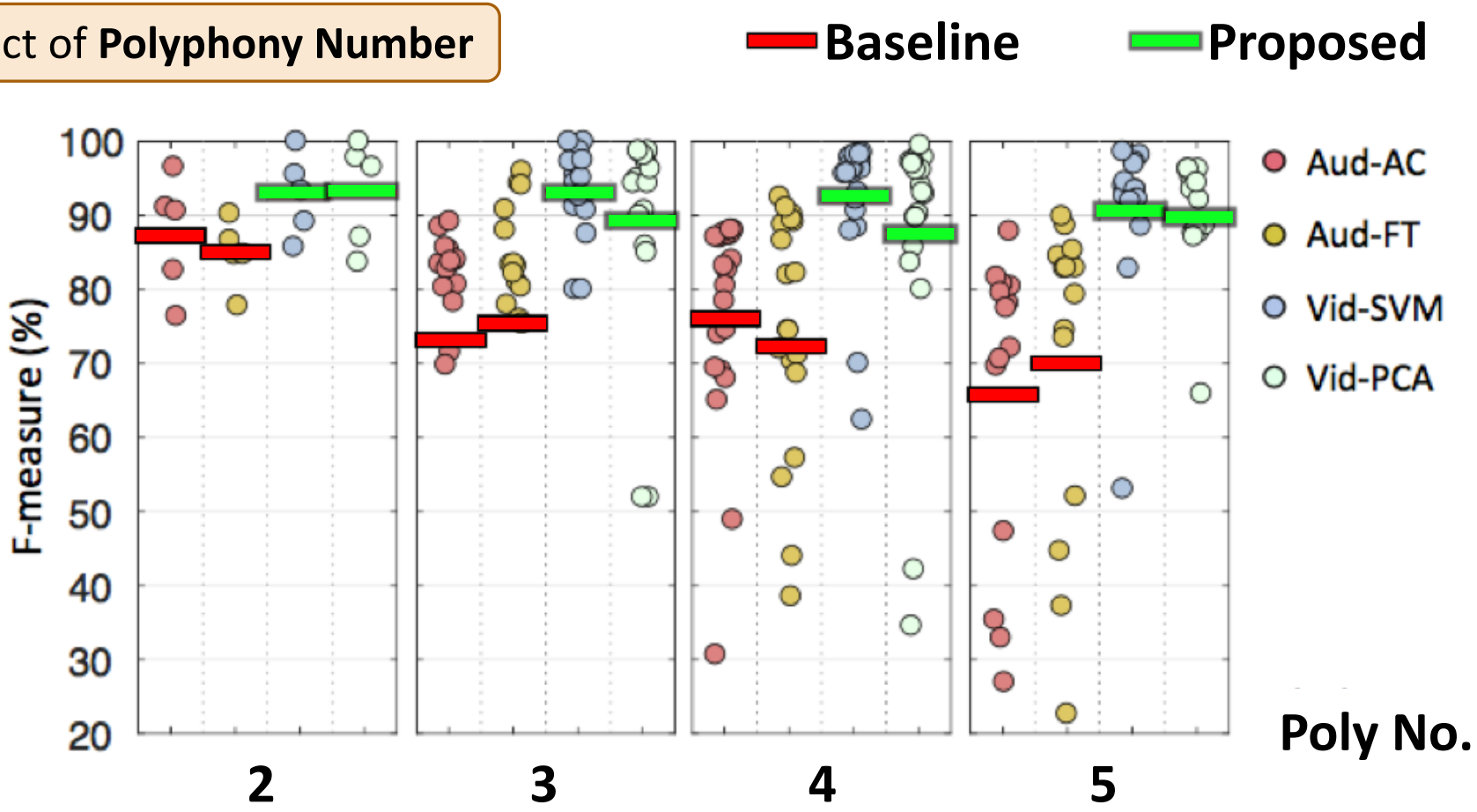
Overall Evaluation



- Video-based method → 92% F-measure
- Improvement over audio-based method
- SVM > PCA

Experiments: Vibrato Detection Results

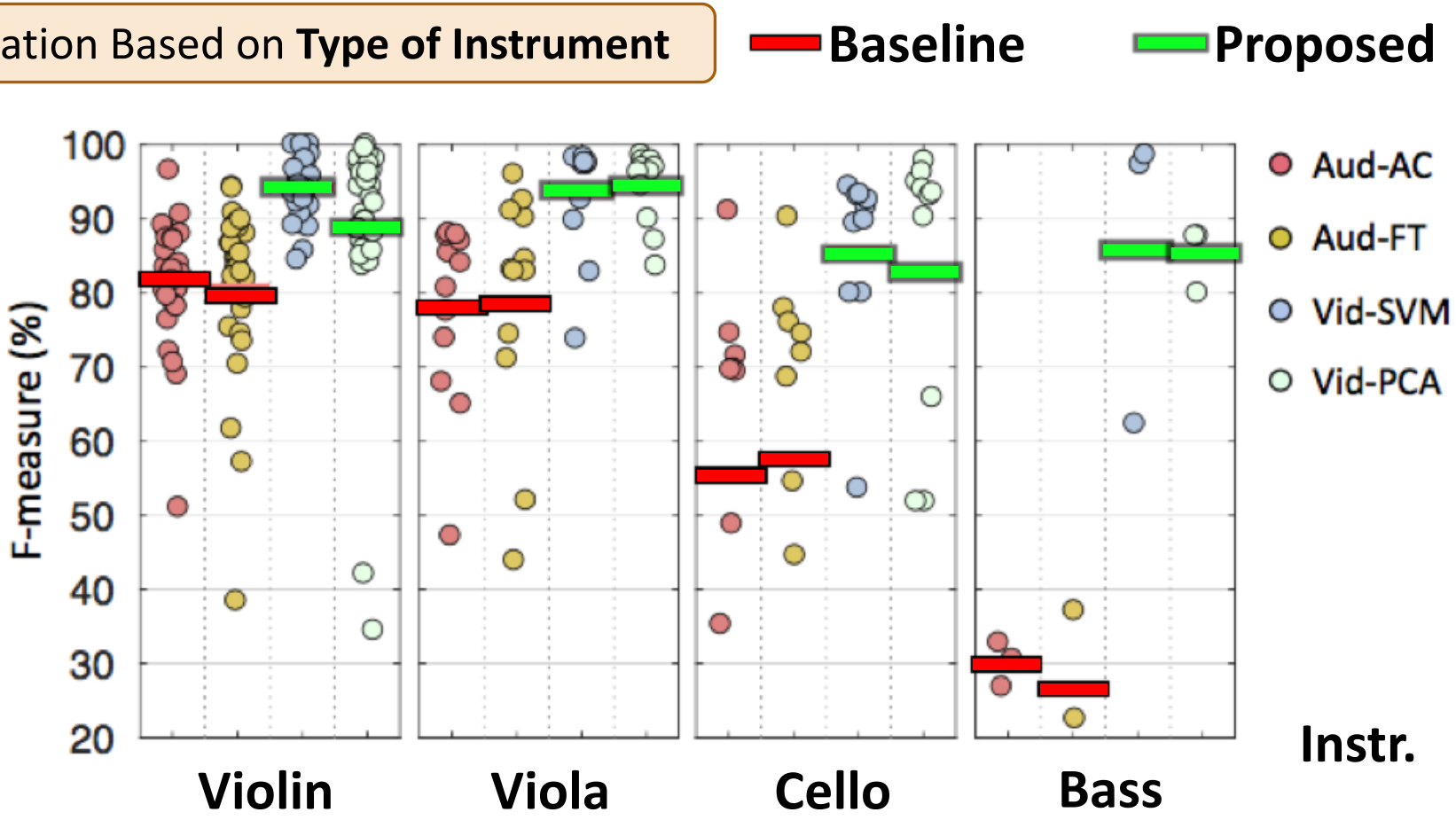
Impact of **Polyphony Number**



- Audio-based method: Poly ↗ Performance ↘
- Proposed video-based method: Robust

Experiments: Vibrato Detection Results

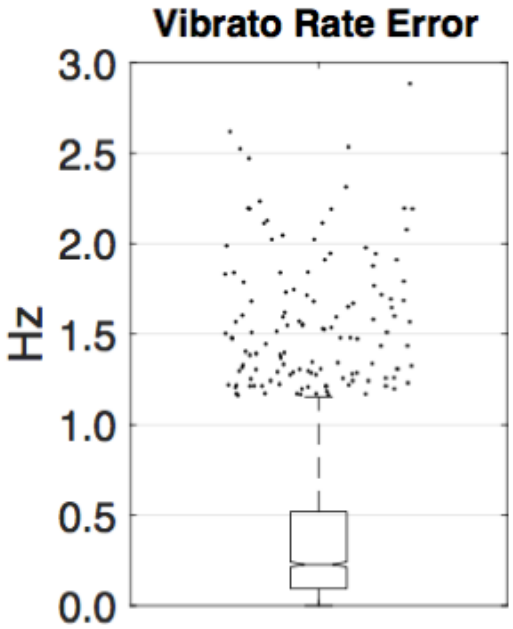
Variation Based on **Type of Instrument**

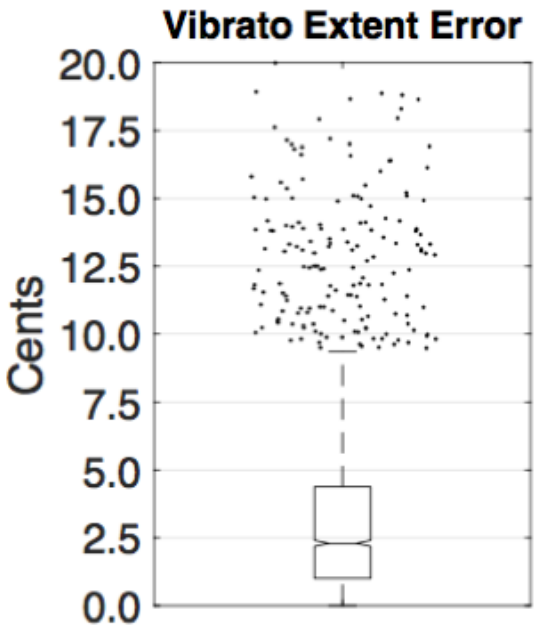


- Audio-based method: Pitch range ↘ Performance ↘
- Proposed Video-based method: Robust

Experiments: Vibrato Analysis Results

Vibrato Rate / Extent





- 2290 vibrato notes
- Rate error: 0.38 Hz
- Extent error: 3.47 cents

Conclusions

- Proposed **video-based** vibrato detection/analysis offers significant improvement over conventional audio-only analysis
- Compared to audio-based methods, proposed video-based method is
  - Robust for **polyphonic** sources
  - Robust for different types of **instruments**
- Proposed method provides good estimates for vibrato rate and extent
  - A powerful tool for analyzing string **ensembles**


air
AUDIO INFORMATION RESEARCH


MELIORA
UNIVERSITY of
ROCHESTER

Thank you!

Experiments: Dataset

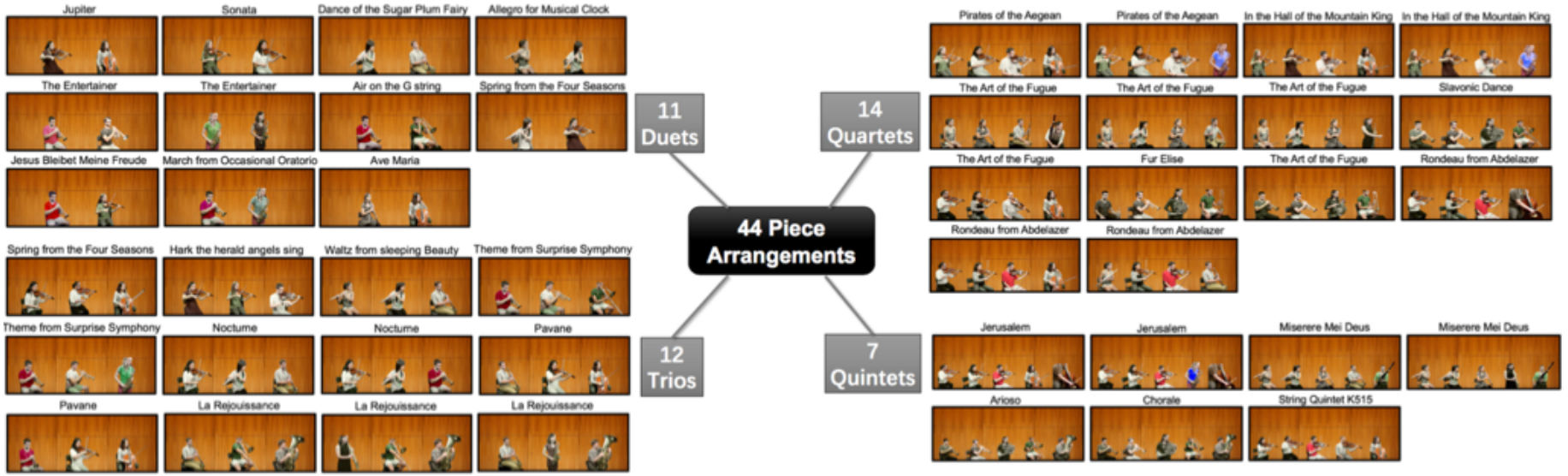
URMP Dataset

- 19 string ensembles (57 tracks)
- 5 duets, 4 trios, 7 quartets, 3 quintets
- Audio: 48k Hz
- Video: 1080P, 29.97 fps

Demo of Dataset

Dataset: URMP Dataset

- 14 instruments, 44 piece arrangements

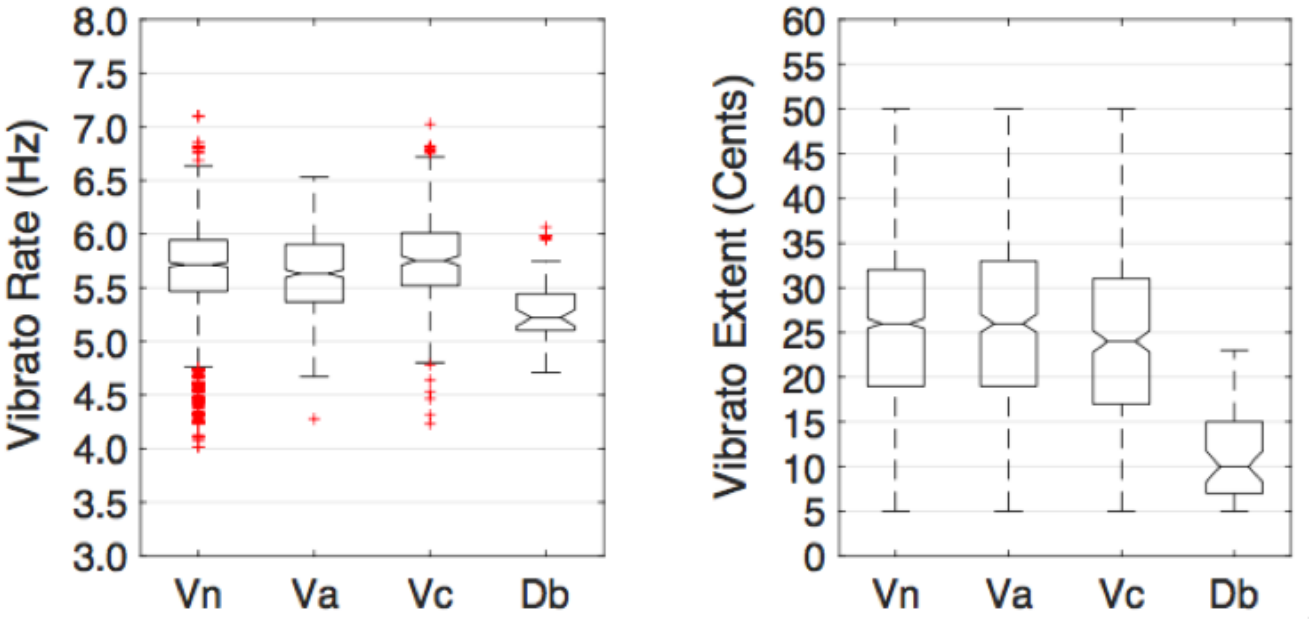


Experiments

Results

Potential Application on Musicologies

Vibrato characteristics for different **instruments**



- Test on TPs from Vid-PCA method: 2290 vibrato notes
- Average error: 0.38 Hz / 3.47 cents
- Double bass → lower rate / extent [1]

[1] James Paul Mick. *An analysis of double bass vibrato: Rates, widths, and pitches as influenced by pitch height, fingers used, and tempo*. PhD thesis, The Florida State University, 2012.