Aligning Semi-improvised Audio with Its Lead Sheet

Zhiyao DUAN & Bryan PARDO

Northwestern University, EECS Department, Evanston, IL, USA.

zhiyaoduan00@gmail.com • *pardo@northwestern.edu* • *http://music.cs.northwestern.edu*

Introduction

▶ Existing audio-score alignment methods assume that the audio performance is **faithful** to a **fully-notated** score.

▶ Semi-improvised music (e.g. jazz) strongly violates this assumption.

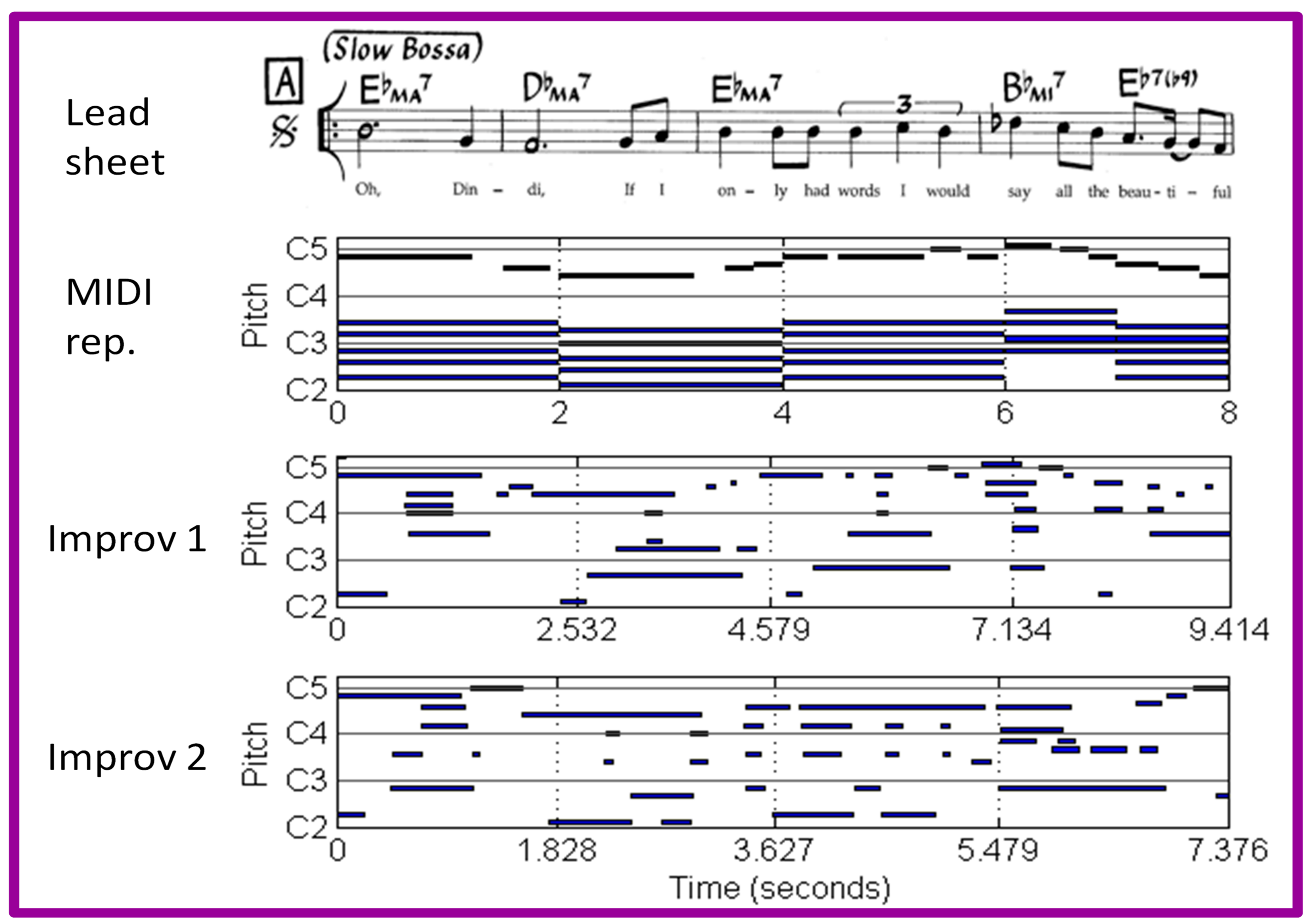
▶ We propose a system for aligning a semi-improvised music audio performance with its score, i.e. a lead sheet.

▶ Requires **no prior training** on the lead sheet to be aligned.

▶ Handles **structural changes** (e.g. jumps, repeats) in the performance.

▶ Obtains promising results on 24 piano performances and 12 full-band commercial recordings of 3 jazz lead sheets.

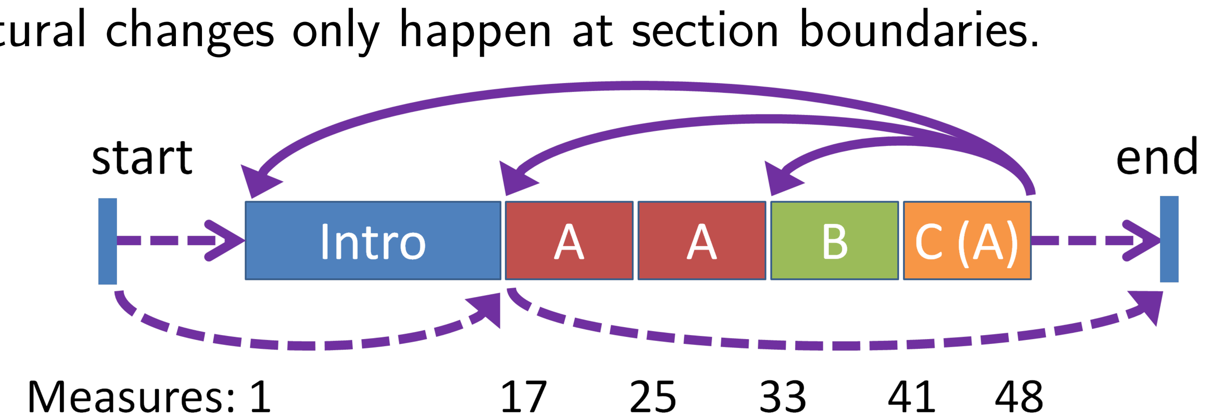
Problem Analysis



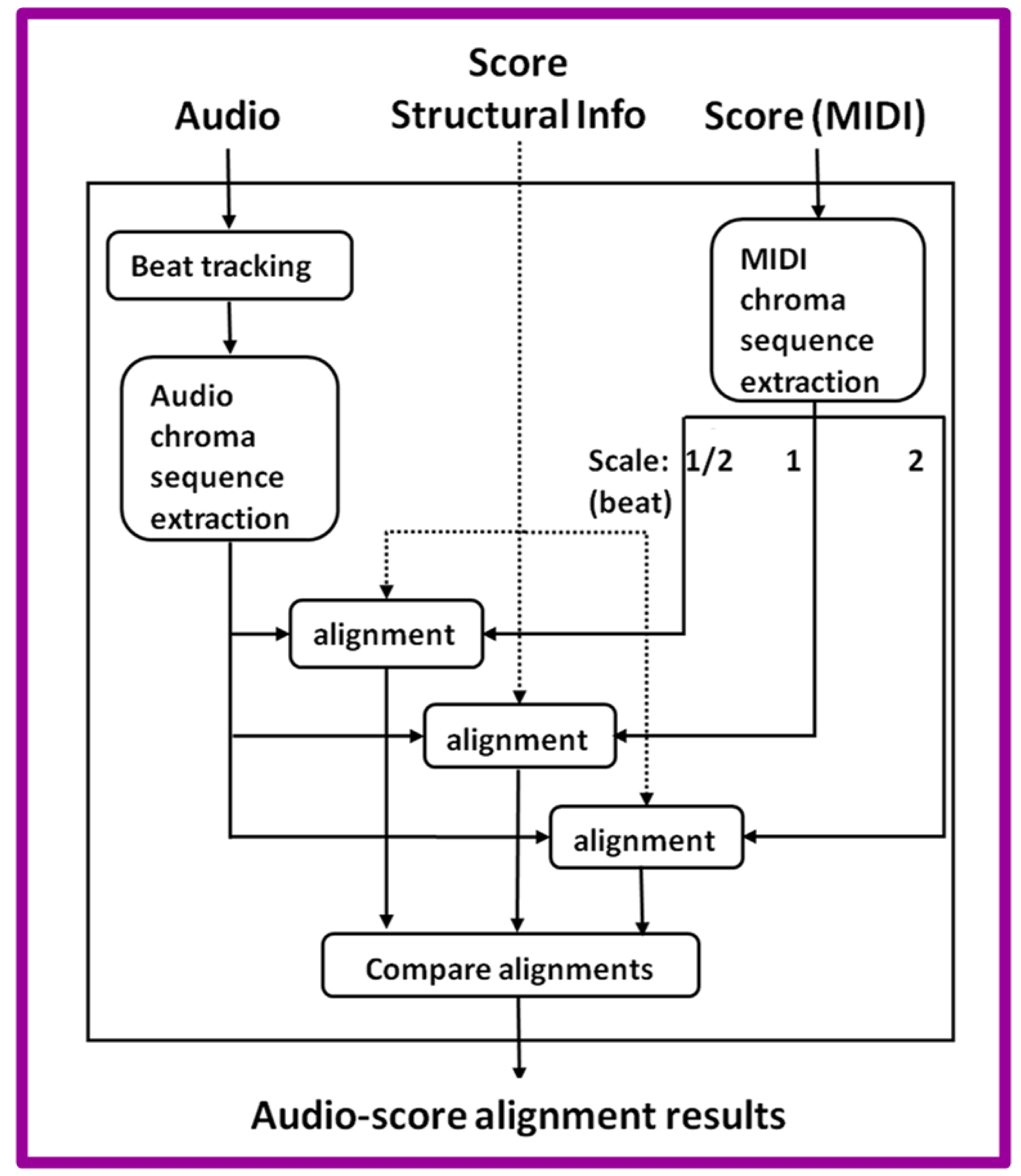
- Harmony is rendered in free rhythmic patterns.
- Melody may be significantly altered (as in Improv 2).
- Performers often make unexpected jumps and repeats.

▶ Essential Information for Alignment

- Performed notes correspond to score harmony at scale of two beats.
- Structural changes only happen at section boundaries.



Proposed System



▶ **Tracking Audio Beat:** we use the method in [1].

▶ **Extracting Chromagrams**: 1 for audio and 36 for MIDI

- Calculate an audio chromagram for **segments** of length $l = 2$ beats and hop $h = 1/4$ beats.
- Calculate a MIDI chromagram for each of the 3 scales of segments: (l, h), $(1/2l, 1/2h)$ and $(2l, 2h)$.
- Transpose MIDI chromagrams 12 times for possible key transpositions.

▶ **Aligning Chromagrams**: Let $\mathbf{A} = (\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_m)$ be the audio chromagram, $\mathbf{S} = (\mathbf{s}_1, \mathbf{s}_2, \cdots, \mathbf{s}_n)$ be the score chromagram.

- Audio may start from anywhere on the score:

$$\mathbf{C}(0,0) = 0, \mathbf{C}(i,0) = i \cdot c_1, \mathbf{C}(0,j) = 0 \qquad (1)$$

- Audio may jump at possible jumping points:

$$\mathbf{C}(i,j) = \min \begin{cases} \mathbf{C}(i, j-1) + c_1, & \text{skip audio} \\ \mathbf{C}(i-1, j) + c_2, & \text{skip score} \\ \min_{k \in \mathcal{P}(j)} \mathbf{C}(i-1, k) + d(\mathbf{a}_i, \mathbf{s}_j), & \text{transition} \end{cases} \qquad (2)$$

where $\mathcal{P}(j)$ is the set of **segments** (at the scale of 2 beats), from which a performance might transition to j. $d(\mathbf{a}_i, \mathbf{s}_j)$ is defined as:

$$d(\mathbf{a}_i, \mathbf{s}_j) = \arccos\left(\frac{\mathbf{a}_i^T \mathbf{s}_j}{\|\mathbf{a}_i\| \|\mathbf{s}_j\|}\right) \qquad (3)$$

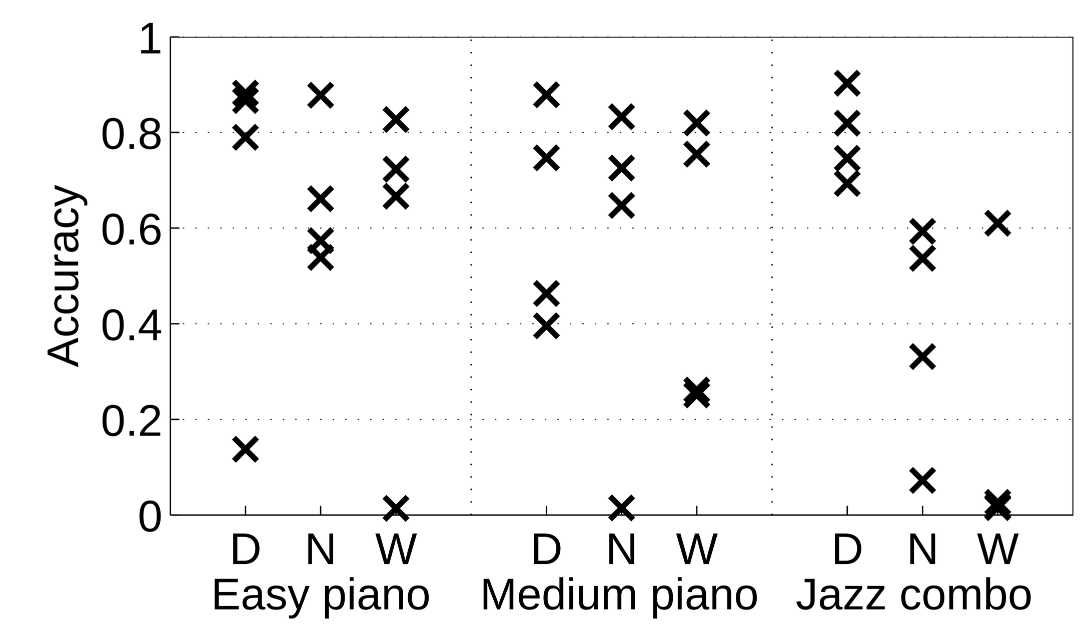
- Audio may end at anywhere on the score: Trace back from $C(m, j_1)$, where $j_1 = \arg\min_j C(m, j)$.

Experiments

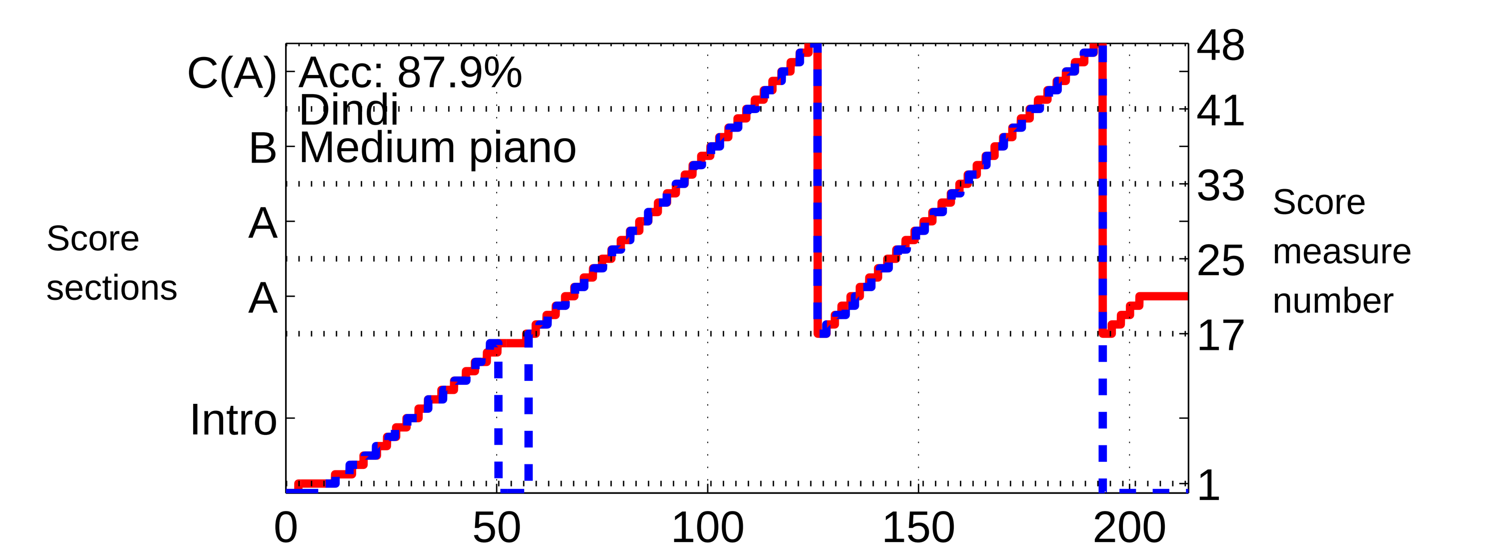
▶ Dataset and Measure

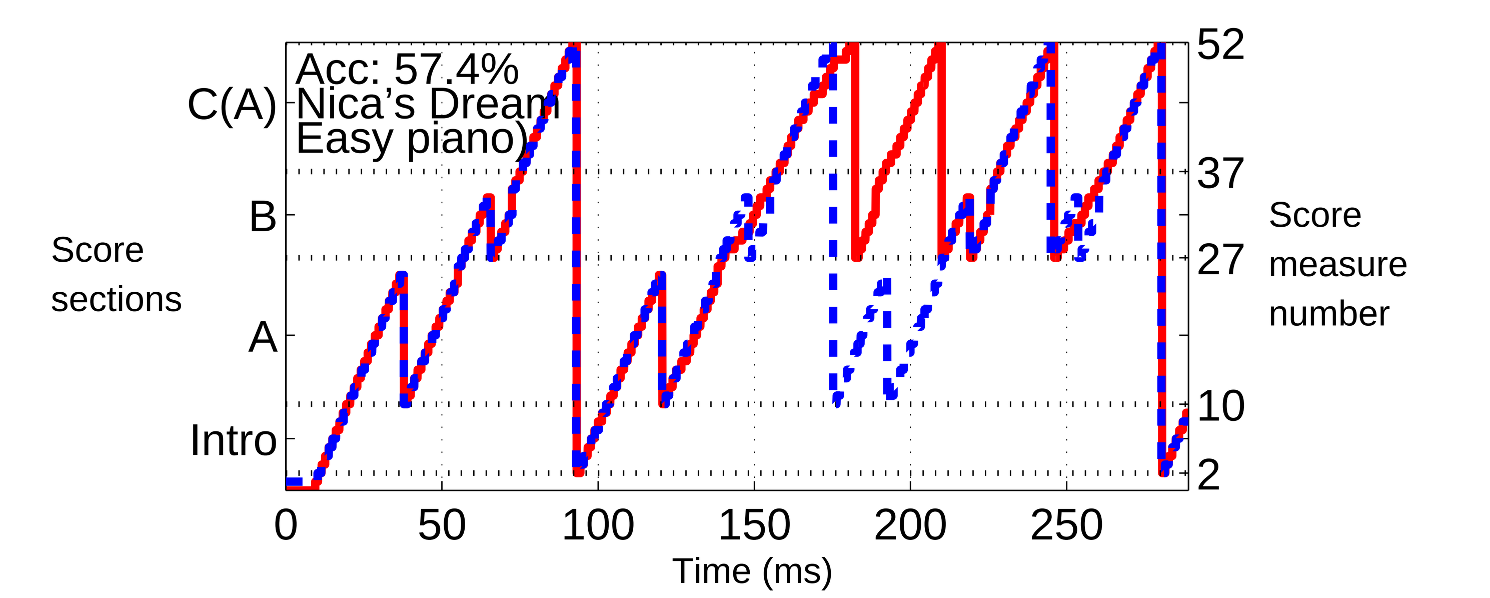
- 3 jazz lead sheets: *Dindi* by Antonio Carlos Jobim, *Nicas's Dream* by Horace Silver and *Without A Song* by Vincent Youmans.
- 12 performances for each lead sheet (4 easy piano, 4 medium piano and 4 commercial jazz band recordings).
- View alignment as a **classification** problem: assign each audio frame (46ms long) a score measure number.
- Accuracy: % of frames that are correctly assigned a measure number.
- Ground-truth measure numbers are obtained manually.

▶ Results:



▶ Examples: ground-truth (blue) vs. system output (red)





▶ **More Examples:** `http://www.cs.northwestern.edu/~zdu459/ismir2011/examples`

[1] D. Ellis, "Beat tracking by dynamic programming," *J. New Music Research, Special Issue on Beat and Tempo Extraction*, vol. 36 no. 1, pp. 51–60, 2007.

This work was supported by NSF grant IIS-0643752.