SCORE FOLLOWING WITH HIDDEN TEMPO USING A SWITCHING STATE-SPACE MODEL

Yucong Jiang (yjiang3@richmond.edu)
University of Richmond

Chris Raphael (craphael@indiana.edu)
Indiana University Bloomington

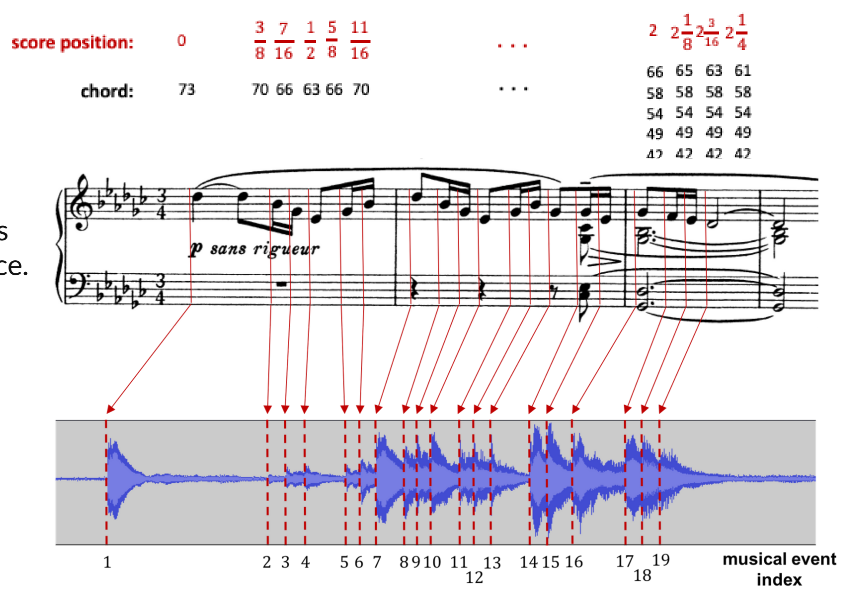
1 Introduction

What is Score Following?

The score-following problem involves building a computer program that can trace musical events in a given musical score during a live performance.

Why Score Following?

- Page turner
- Automatic accompaniment systems
- Virtual score composition
- Real-time audio enhancement/feedback



Monophonic music:

Polyphonic music:

2 Current Bottleneck

Existing score-following algorithms can still stumble on some challenging cases, especially when the data model is not reliable:

- Shared notes among neighboring chords
- Blurring effects caused by fast playing
- Pedaling

3 Research Aims

- To present a new method designed to improve the timing model—this aspect is especially meaningful in those challenging cases.
- To understand the nature of this problem better through empirical experiments.

4 The Model

1. In the first diagram, the "time step" is the **chord index**.
2. In the second diagram, the "time step" is the **audio frame index**.

1. Kalman Filter Model for Tempo

a linear dynamical system:

$$o_{k+1} = o_k + l_k t_k + \varepsilon_{k+1}$$

smooth tempo

$$\boxed{t_{k+1} = t_k + \eta_{k+1}}$$

$$o_1 \sim N(\mu_{o,1}, \sigma^2_{o,1})$$

$$t_1 \sim N(\mu_{t,1}, \sigma^2_{t,1})$$

$$\varepsilon_k \sim N(0, \sigma^2_{\varepsilon,k}),\ k = 2, \ldots, K$$

$$\eta_k \sim N(0, \sigma^2_{\eta,k}),\ k = 2, \ldots, K$$

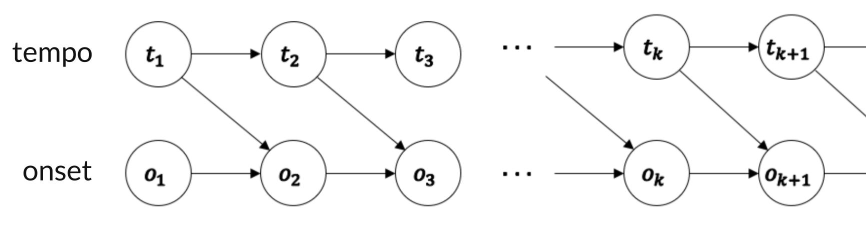
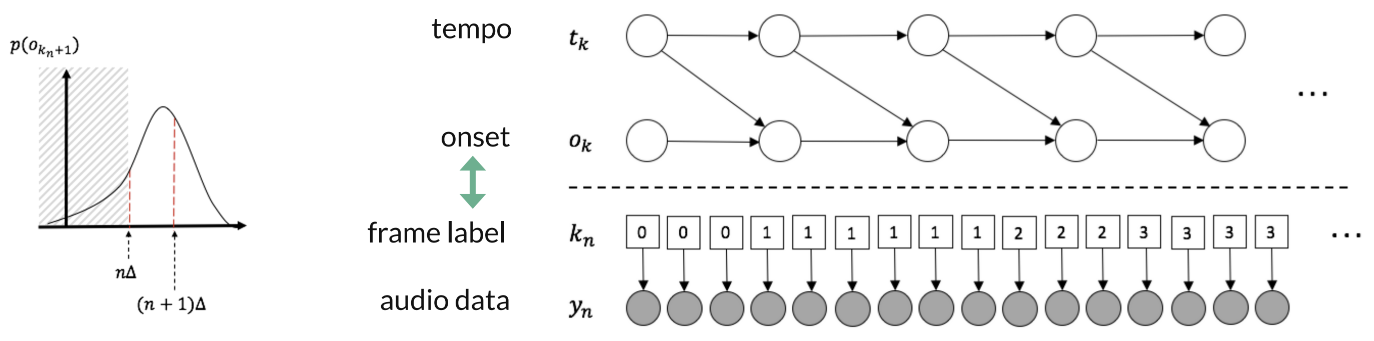


Figure 4.1: Linear dynamical system of the tempo and the onset.

$$k_n = \min\{k \in \{0, \ldots, K\} : n\Delta < o_{k+1}\}$$

$$o_k \approx \Delta \cdot \min\{n \in \{1, \ldots, N\} : k_n = k\}$$

2. Frame-wise Representation

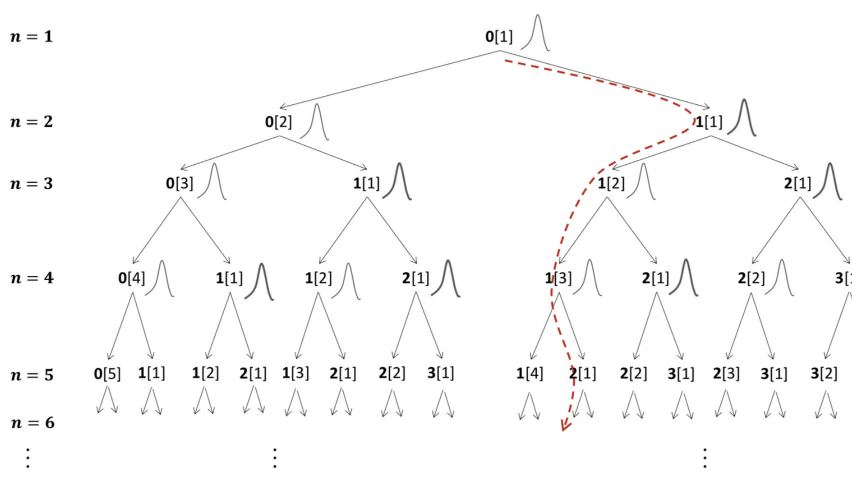


5 Computation

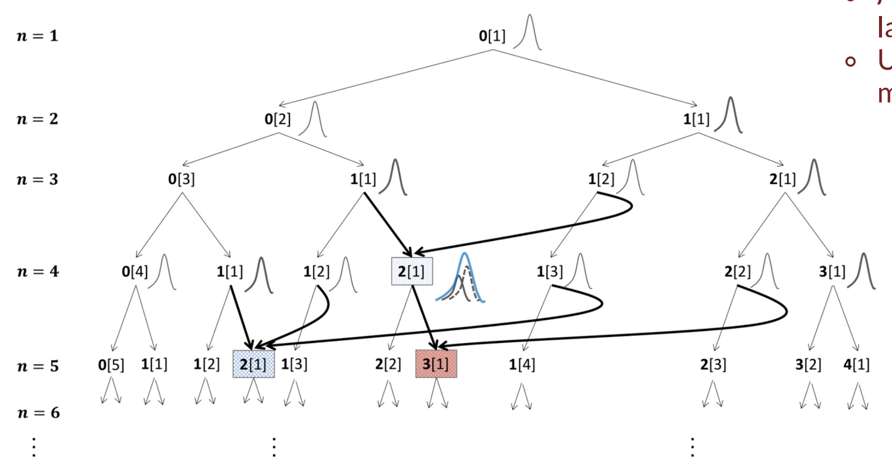
Tree Representation

- **[Chord Age]:** the number of frames a chord has lasted so far.
- **Left Branch:** staying in the same chord.
- **Right Branch:** moving on to the next chord.

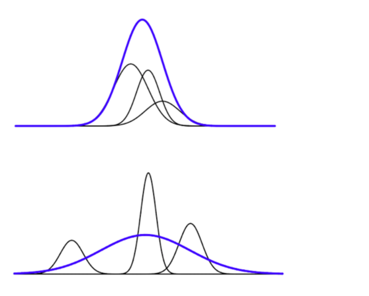
The tree grows exponentially with time (or with n)!



Approximation



- At each frame, merge nodes with the same label and age.
- Use a single Gaussian to approximate a mixture of Gaussians.



6 Experiments

Composer	Piece	Abbreviation	#Excerpts
Mozart	Piano Concerto No. 17 in G major, mvmt1	Mozart	3
Schumann	Piano Concerto in A minor, mvmt1	Schumann	3
Chopin	Barcarolle, Op. 60	Chopin_barcarolle	2
Chopin	Prelude, Op. 28 No. 4	Chopin_prelude	2
Chopin	Ballade No. 1	Chopin_ballade	8
Liszt	Paganini S.141, No. 3 (*La campanella*)	Liszt	5
Rachmaninoff	Prelude, Op. 3, No. 2	Rachmaninoff	5
Schubert	Six Moments, D. 780 No. 2	Schubert_780	1
Schubert	Ständchen, D 957 No. 4 from Schwanengesang	Schubert_ständchen	4
Debussy	Prelude, No. 2 (*Violes*)	Debussy_violes	1
Debussy	Prelude, No. 8 (*La fille aux cheveux de lin*)	Debussy_fille	3
Beethoven	Piano Sonata No. 8 (*Sonata Pathétique*)	Beethoven_pathétique	1
Beethoven	Piano Sonata No. 31	Beethoven_31	8
Haydn	Piano Sonata No. 24 in D major, mvmt1	Haydn_mvmt1	1
Haydn	Piano Sonata No. 24 in D major, mvmt2&3	Haydn_mvmt23	3

- 15 solo piano pieces
- 50 excerpts

- Typical length: 40~90 seconds
- 48 minutes in total

- Sampling rate: 8k Hz
- Hop size: 16 ms

- **Baseline: Music Plus One** (hidden Markov model)

Evaluation Method

Frame-wise accuracy:

$$Acc_n = \underbrace{\sum_{\substack{k_n = \kappa_n \\ 1 \le a_n \le n - \kappa_n + 1}}}_{\text{correct hypotheses}} p(k_n, a_n | y_1^n)$$

$$Acc = \sum_n Acc_n / N$$

Results

12 excerpts failed (accuracy < 40%) by either program:

- fatal error
- high uncertainty

	baseline	tempo tracking
# failed excerpts	11	9
average accuracy	15.1%	22.1%

The proposed method is measurably better than the baseline:

coefficient	value
$\bar{d}$	-0.04
t	-2.7307
df	37
p-value	0.009623

	baseline	tempo tracking
average accuracy	65.0%	69.1%

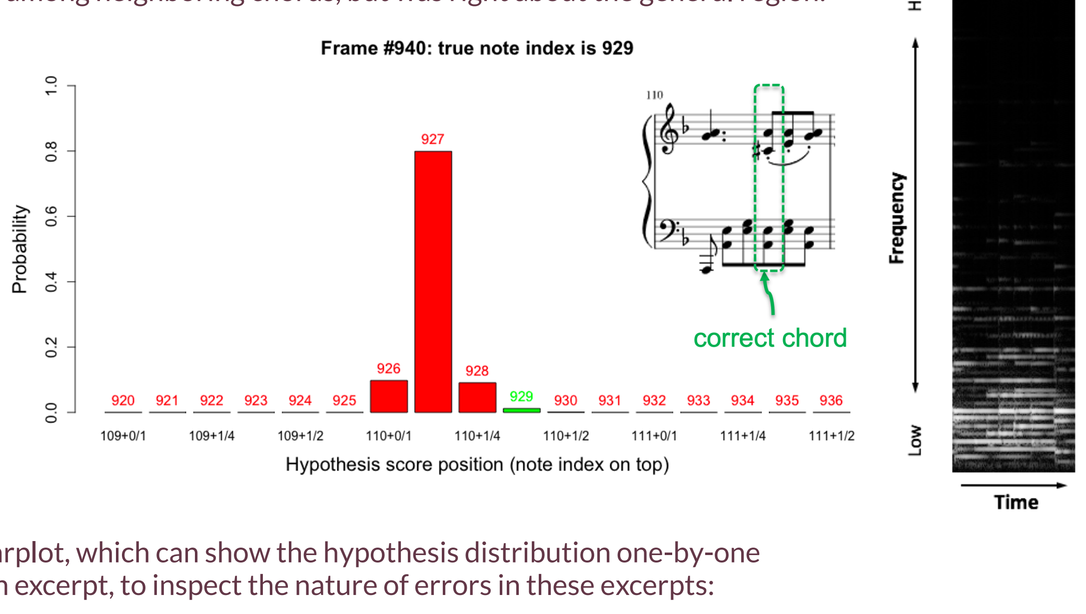
Table 5.10: Average accuracies of 38 excerpts.

Case-by-case Analysis

An example of a typical scenario where the program was "confused" among neighboring chords, but was right about the general region:

Proposed method:

- is robust even with incorrect default tempo;
- recovers more easily from mistakes;
- however, can suffer at fast playing places.



Use a "movie" version of the barplot, which can show the hypothesis distribution one-by-one quickly through all frames of an excerpt, to inspect the nature of errors in these excerpts:

Excerpt	Piece	Baseline	Tempo	Observation
#23	Liszt	25.7%	51.2%	Baseline got lost half-way through after a section of 14 repeated chords in a row.
#43	Beethoven_31	31.8%	46.5%	Incorrect default tempo. 14.4% higher accuracy among the other six excerpts (successfully followed).
#46	Beethoven_31	34.0%	62.7%	
#15	Chopin_ballade	0.4%	27.2%	Baseline: completely lost near the beginning when the sound was blurring. Tempo: followed the region.
#16	Chopin_ballade	0.3%	36.2%	
#13	Chopin_ballade	11.8%	27.5%	Baseline: got lost starting around 1/3 through when the sound started to blur. Tempo: sometimes "confused," but always recovered.
#19	Liszt	2.1%	18.6%	Both programs got lost near the beginning (repeated chords and patterns), but only the proposed method recovered.
#21	Liszt	0.5%	13.4%	

7 Conclusion

We can speculate that treating the tempo as a variable helps the program adapt to unpredictable performance variations, and that modeling the tempo as smooth helps discriminate among hypotheses.

In conclusion, this paper presents an innovative new method for improved score-following, and suggests a promising direction for future research endeavors.