

Spectral Learning for Expressive Interactive Ensemble Music Performance

Gus (Guangyu) Xia

Yun Wang

Roger Dannenberg

Geoffrey Gordon

School of Computer Science

Carnegie Mellon University

Introduction: Musical background

Interaction

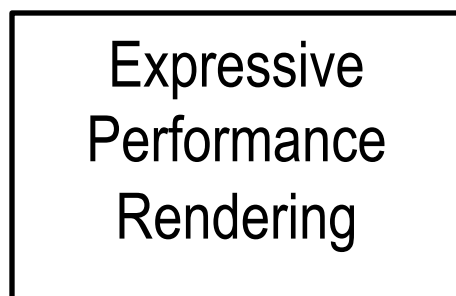
Expression

Rehearsal

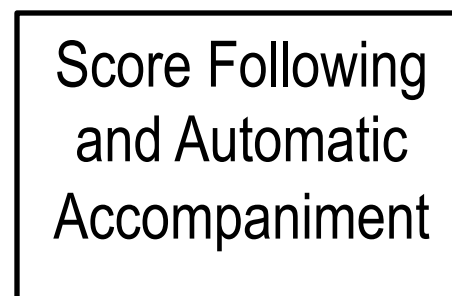


Introduction: Technical background

Expression



Interaction



Expressive Interactive Performance

Introduction: Problem Overview

For interactive ensemble music performance, how can we build artificial performers that automatically improve their ability to sense and respond to human musicians' expression with rehearsal experience?



- We start from piano duets
- Focusing on **expressive timing** and **dynamics**
- Model expressions as co-evolving time series
- Use **Spectral Learning**

Outline

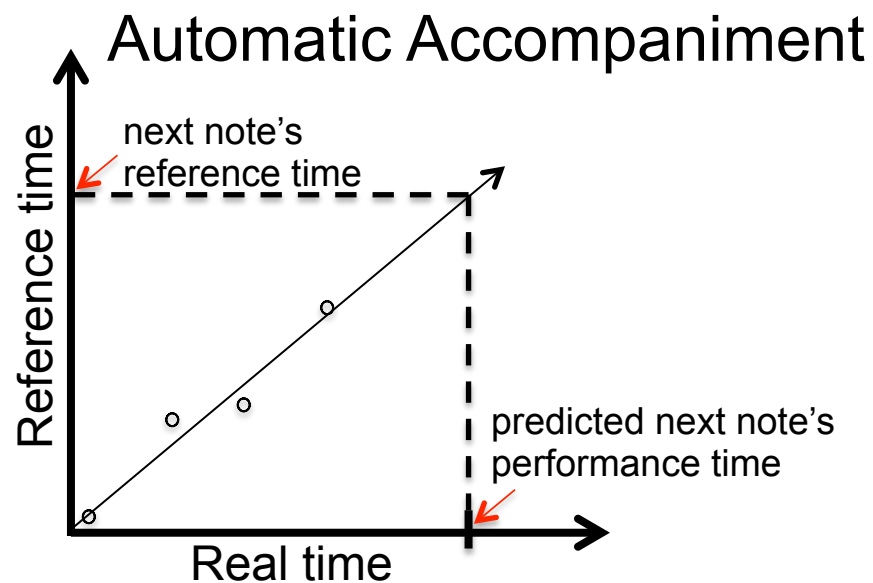
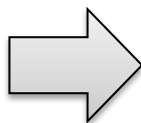
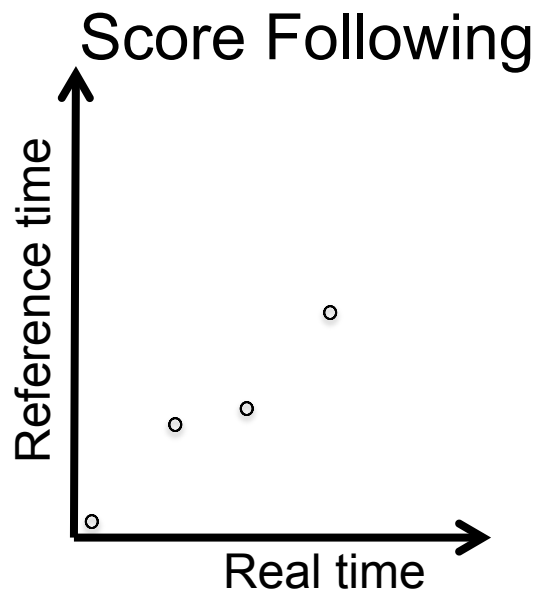
- Introduction
- **Data Collection**
- **Method**
- **Demos**
- **Future Work & Conclusion**

Data Collection

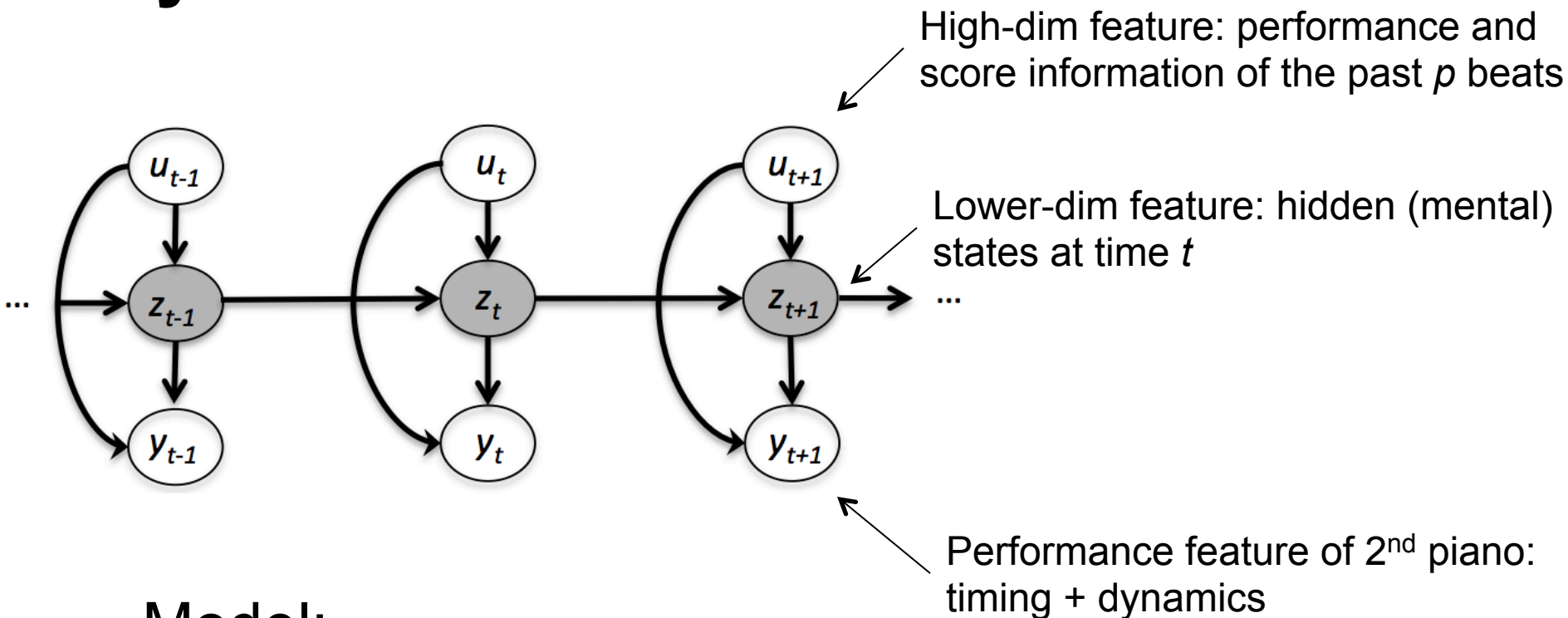
- Musicians:
 - 10 music graduate students play duet pieces in 5 pairs.
- Music pieces:
 - 3 pieces of music are selected, *Danny boy*, *Serenade* (by *Schubert*), and *Ashokan Farewell*.
 - Each pair performs every piece of music 7 times.
- Recording settings:
 - Recorded by electronic pianos with MIDI output.

Base line

- Linear extrapolation based on previous notes :



Spectral Learning for Linear Dynamic System

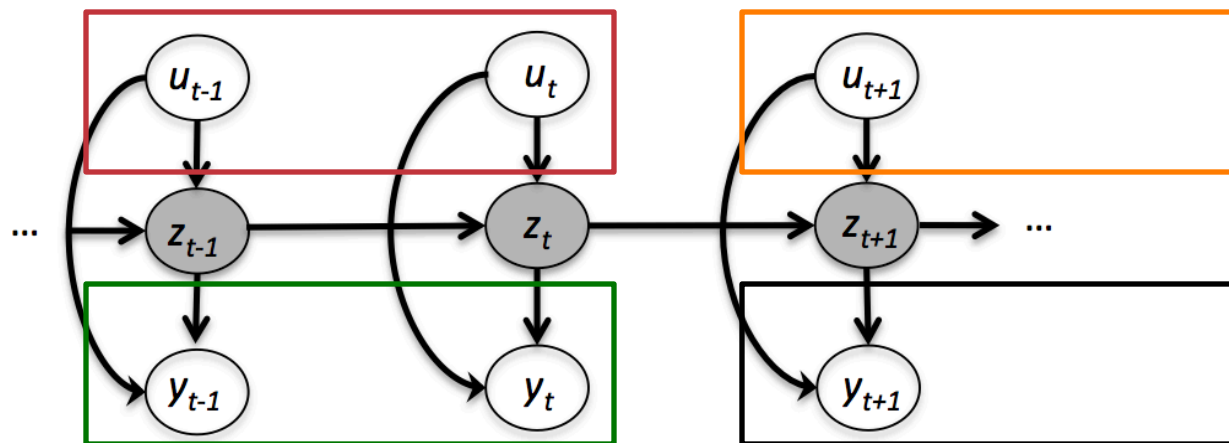


■ Model:

$$z_t = Az_{t-1} + Bu_t + w_t \quad w_t \sim \mathcal{N}(0, Q)$$

$$y_t = Cz_t + Du_t + v_t \quad v_t \sim \mathcal{N}(0, R)$$

Spectral Learning(1): Oblique projections

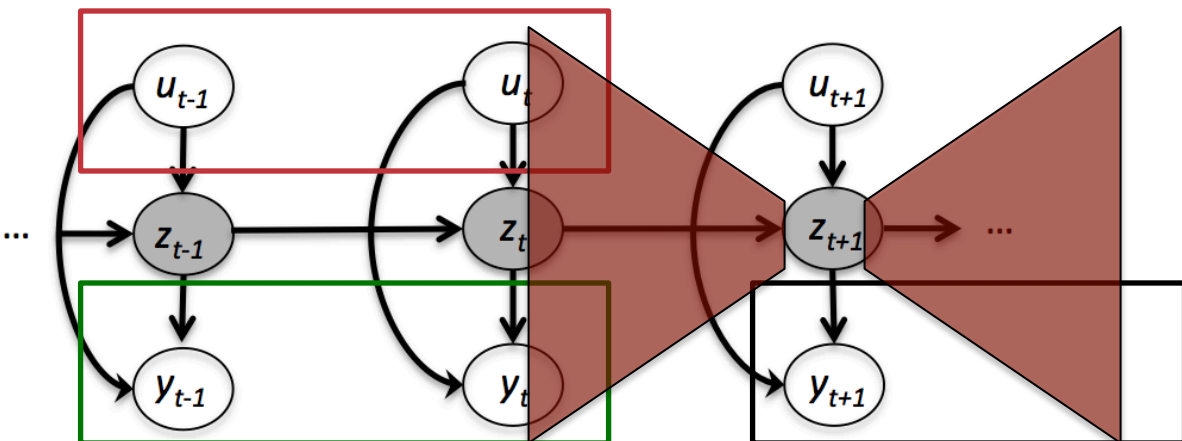


$$\mathbb{E}(Y_F) = [\beta_{Y_H} \beta_{U_H} \beta_{U_F}] \begin{bmatrix} Y_H \\ U_H \\ U_F \end{bmatrix}$$

- We don't know the future.
- Partially explain future observations based on the history

$$\tilde{Y}_F \stackrel{\text{def}}{=} [\hat{\beta}_{Y_H} \hat{\beta}_{U_H} 0] \begin{bmatrix} Y_H \\ U_H \\ 0 \end{bmatrix}$$

Spectral Learning(2): State estimation



$$z_t = Az_{t-1} + Bu_t + w_t$$

$$y_t = Cz_t + Du_t + v_t$$

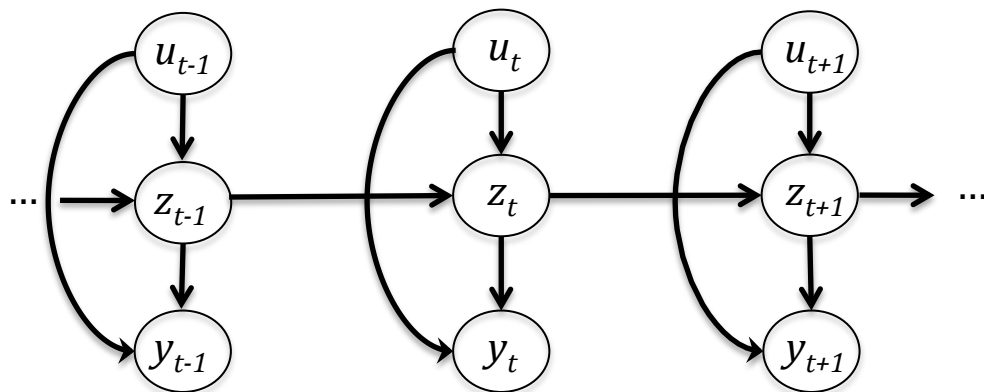
$$\tilde{Y}_F = \begin{bmatrix} C \\ CA \\ CA^2 \\ \vdots \end{bmatrix} z_{t+1}$$

- States estimation by SVD

$$\tilde{Y}_F = U\Sigma V^T = (\overset{1}{U\Sigma\bar{2}})(\overset{1}{\Sigma\bar{2} V^T})$$

- Moreover, enforce a bottleneck by throwing out near-zero singular values and corresponding columns in U and V.

Spectral Learning(3): Estimate parameter



$$z_t = Az_{t-1} + Bu_t + w_t$$

$$w_t \sim \mathcal{N}(0, Q)$$

$$y_t = Cz_t + Du_t + v_t$$

$$v_t \sim \mathcal{N}(0, R)$$

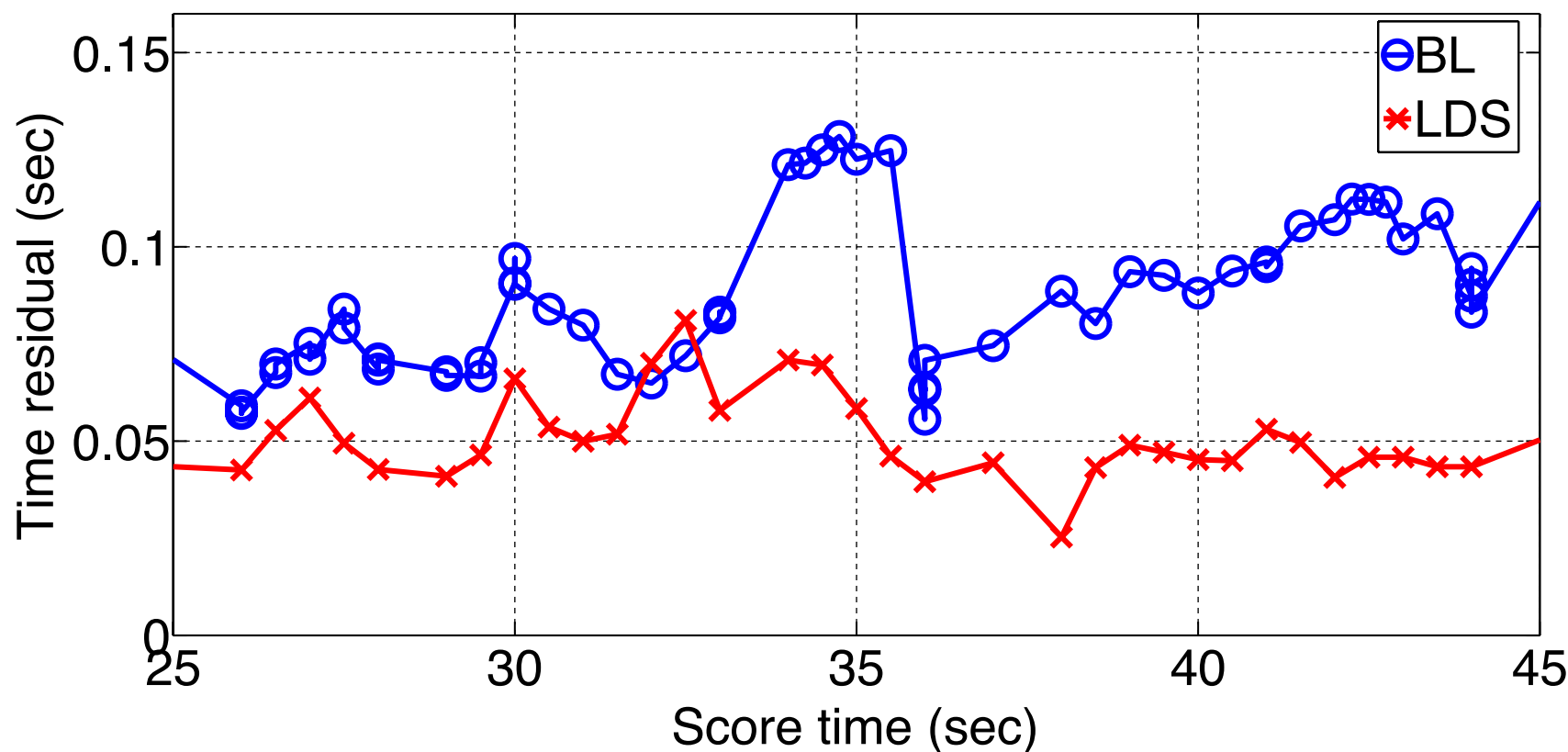
- Based on estimated hidden states, the parameters could be estimated from the following equation:

$$\begin{bmatrix} \hat{Z}_f^- \\ Y_f \end{bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{bmatrix} \hat{Z}_f \\ U_f \end{bmatrix} + \begin{bmatrix} e_w \\ e_v \end{bmatrix}$$

Result: An example

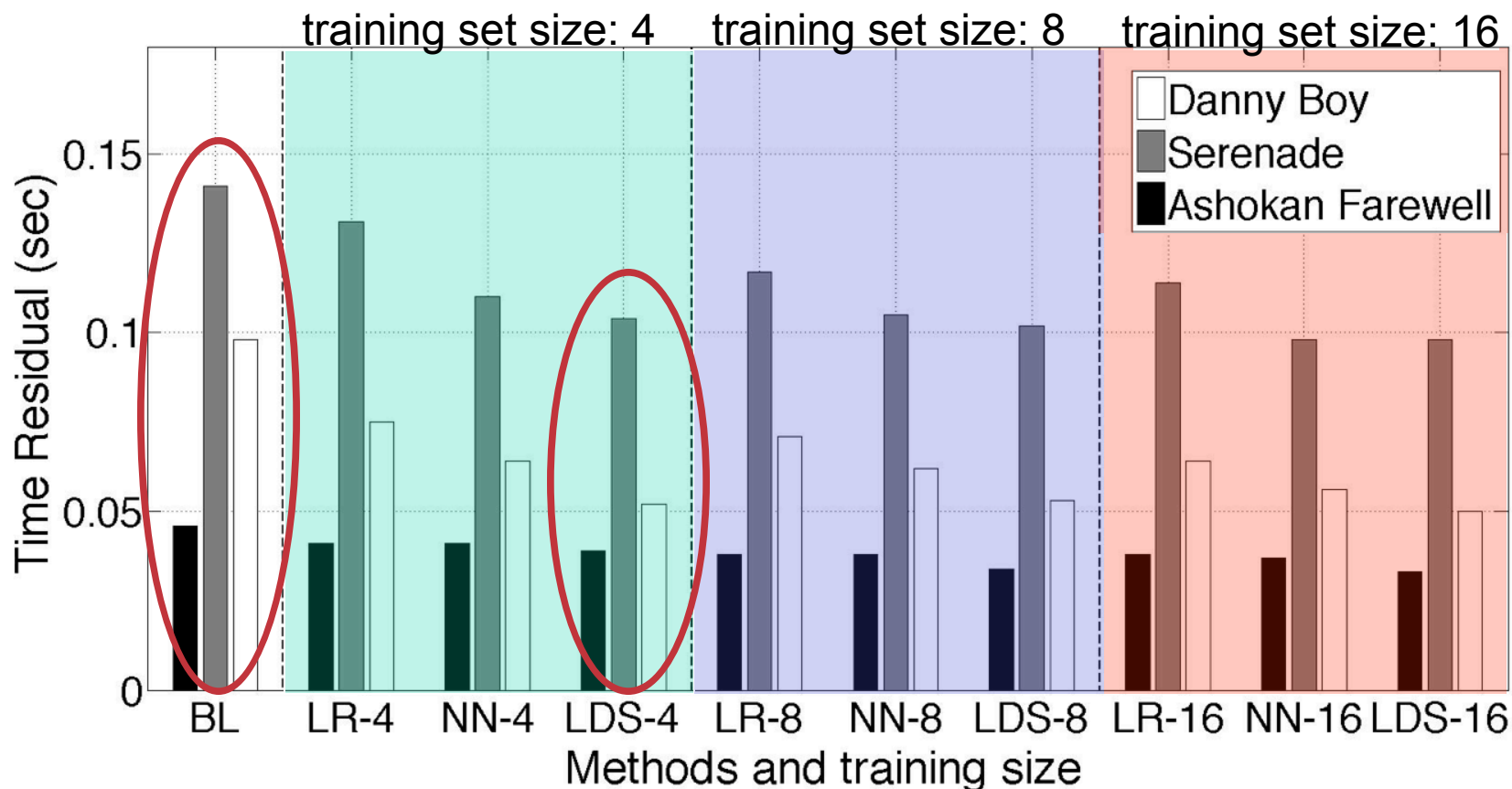
BL: Base-line Linear extrapolation

LDS: Spectral Learning (ONLY 4 rehearsals!)



Result: Overall

LR: Linear regression NN: Neural network



Audio Demo

- Base Line:



- Spectral Learning: 4 training examples



Conclusion

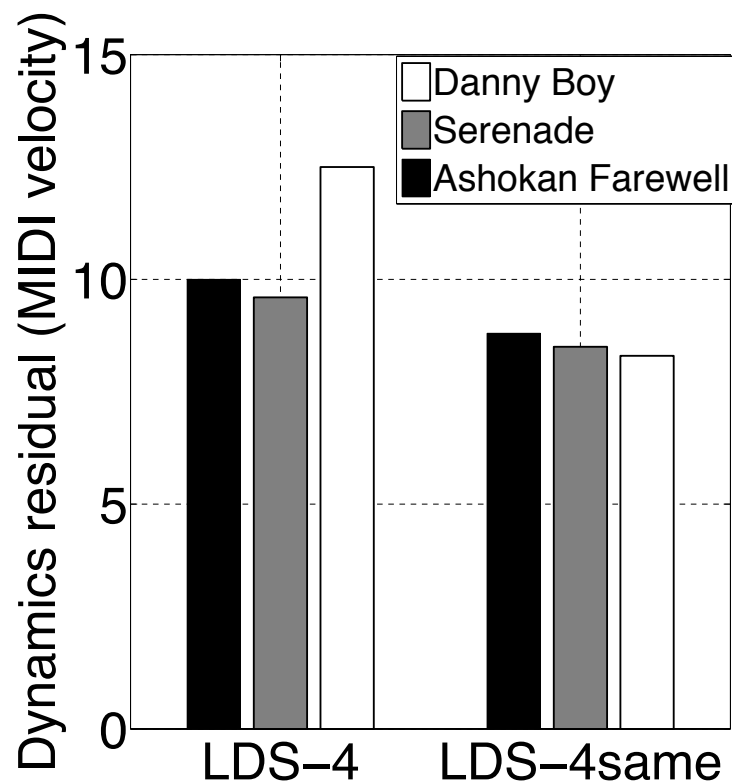
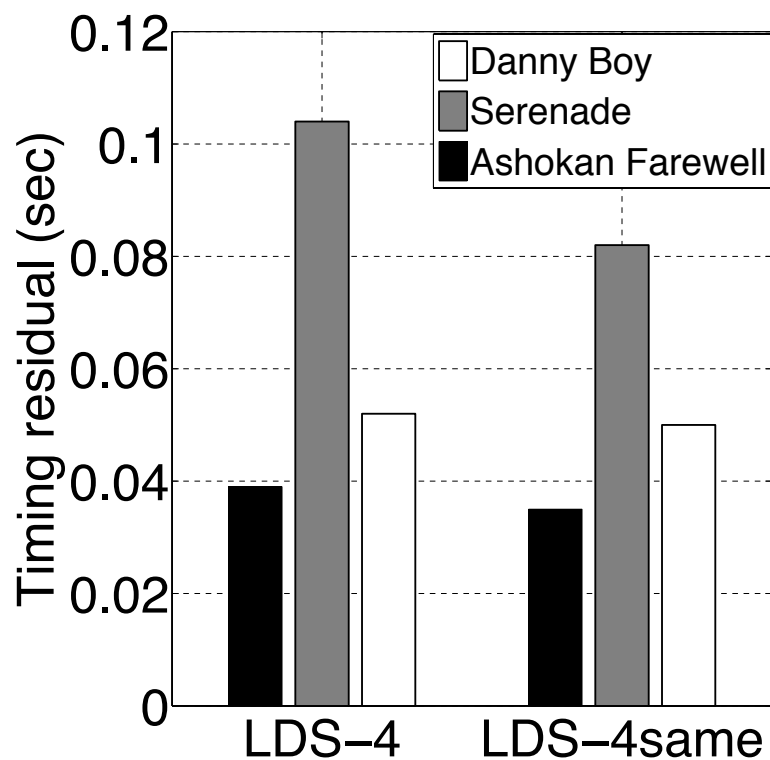
- An artificial performer for interactive performance using spectral learning
- A combination of expressive performance and automatic accompaniment
- Generate more human-like interactive performance just based on 4 rehearsals

Future Work

- Cross-piece models (in progress)
- Plugin with music robots (in progress)
- Online learning and decoding

Thanks!

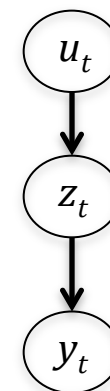
Result: Performers effect



Neural Network

- Model:

$$z_t = f(W_1 u_t + b_1) \quad f(x) = \begin{cases} x, & (x > 0) \\ 0, & (x \leq 0) \end{cases}$$
$$y_t = W_2 z_t + b_2$$



- 10-dimensional hidden layer
- Objective function: mean absolute error
- 30 epochs of SGD training
- Learning rate decays from 0.1 to 0.05 exponentially

Audio Demo

- Base Line:



- Note-specific approach: 34 training examples



- General feature approach: 4 training examples



- LDS approach: 4 training examples

