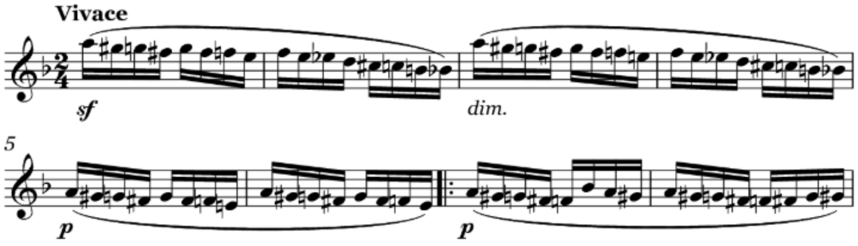
Musical Expressiveness

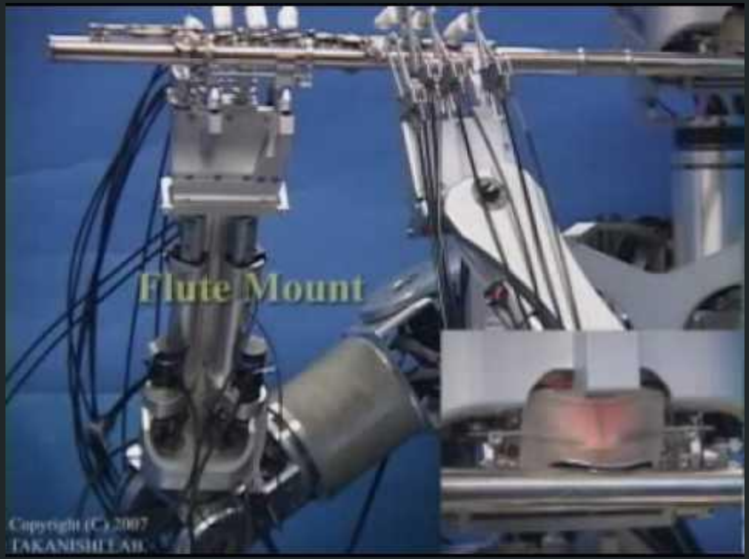
Gavin Baker, Claire Wenner

Outline

I. Background
   A. Robotic Musicians
   B. Expressiveness definition
II. Approaches
   A. User Defined
   B. Models
III. Main Methods
   A. Cancino-Chacón et. al 2016
   B. Xia et. al 2015
IV. Comparison
V. References

Robotic Musicians





The robot is good - but is the performance as good as a human performer?

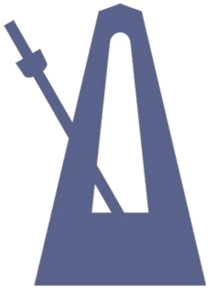
The difference is the expressiveness of the performer!

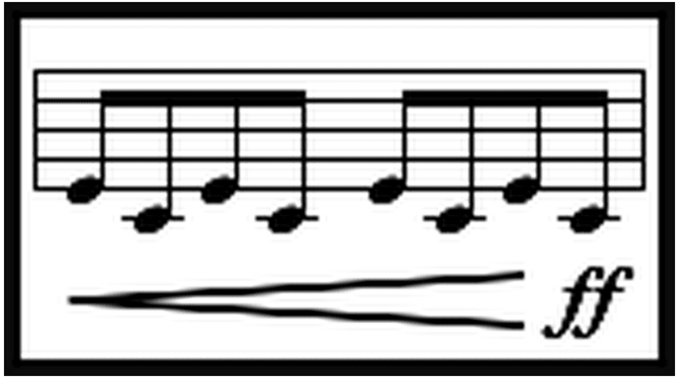
What is Musical Expressiveness?

- Wikipedia: “Musical expression is the art of playing or singing with a personal response to the music”
- “the micro-deviations from the notated dictates of the score a performer executes while playing”

Musical Expressiveness Parameters

- Tempo
- Pitch
- Duration
- Dynamics
- Onset times





User Defined Expressiveness

- ForMuLa- programming language for specifying music
  - Includes dynamic tempo, dynamics, etc.
  - User specified information- same output every time
- Robotic instruments (as seen previously)
  - High accuracy necessary for expressive performance

```
:ap  rhb        (6 pitches - - - ; do 1 beat of right hand)
   dup $ 12 + $ $ $ $ $ dup $ 12 + $
;ap

:ap  rh         (right-hand process)
   ::sh1        (volume shape within each beat)
      begin
         0 1|32 ocon -20 1|32 ocon
         -40 1|16 ocon
         -20 1|32 ocon -40 1|32 ocon
         -10 1|32 ocon -20 1|32 ocon
      again
   ;;sh
   ::ash        (articulation control)
      absolute
      begin
         6(32 1|32 ocon 5(32 1|32 ocon
         1(32 1|8 ocon
         1(16 1|32 ocon 1(32 1|32 ocon
      again
   ;;sh
   3 oct /32
(measure 1)
   2 0 do
      d f+ a b g+ c+ rhb f+ g+ a b g+ c+ rhb
      e+ g+ b +c+ a+ f+ rhb g+ b +c+ +d b+ a rhb
   loop
(measure 3)
   +d+ +f+ +a +b +g+ +c+ rhb b+ +d+ +f+ +g+ +e+ +c+ rhb
   +c+ +e +g +a +f+ b rhb a+ +c+ +e +f+ +d+ b rhb
   b +c +d+ +e+ +d a rhb g+ b +d +e +c+ a rhb
   e+ g+ b +c+ a+ f+ rhb c+ e+ g+ a g d rhb
(measure 5)
```

Models for Expressiveness

- Kim et. al 2011 uses statistical models of structure-expression relations in polyphonic piano performance
- Flossman et. al uses a linear Gaussian model for musical expression of every note
- Basis functions for linear and non-linear models (Cancino-Chacón et. al 2016)
- Spectral learning with linear dynamic system models (Xia et. al 2015)

No real consensus on a “good” method!

Basis Functions

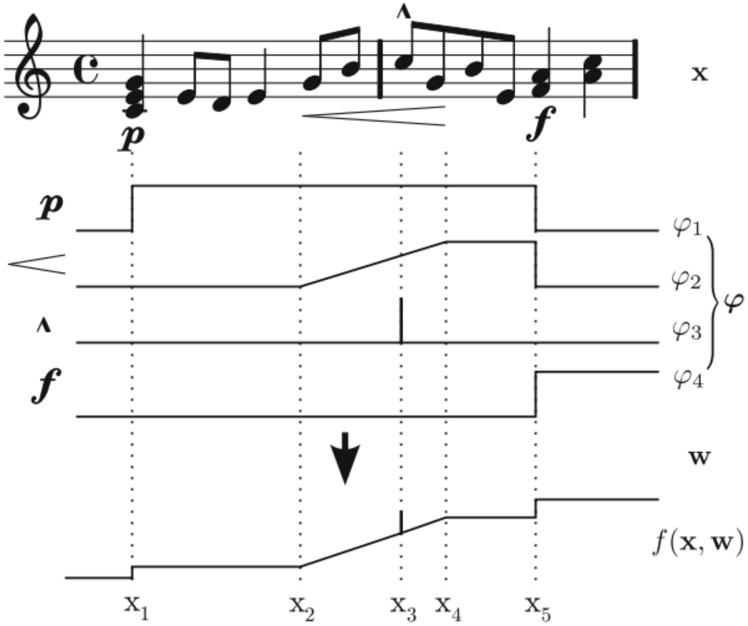
Basis Functions for Musical Expressiveness

- Basis function maps values into different spaces
  - Notes marked Forte
  - Notes on beat 1
  - Notes with crescendo over them
- Deterministic
- Used as classifiers to parameterize musical expressiveness features
- Can be interpolated in a number of different ways

Linear Model

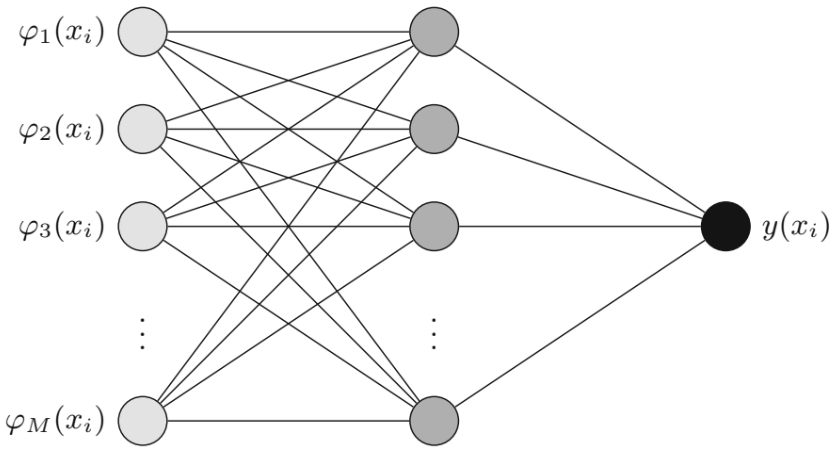
- Expressiveness as a linear combination of bases.
- Creates larger functions determining pieces of expressiveness
- Assumes no interplay between functions

$$y_i = \boldsymbol{\varphi}(x_i)^{\mathrm{T}}\mathbf{w},$$



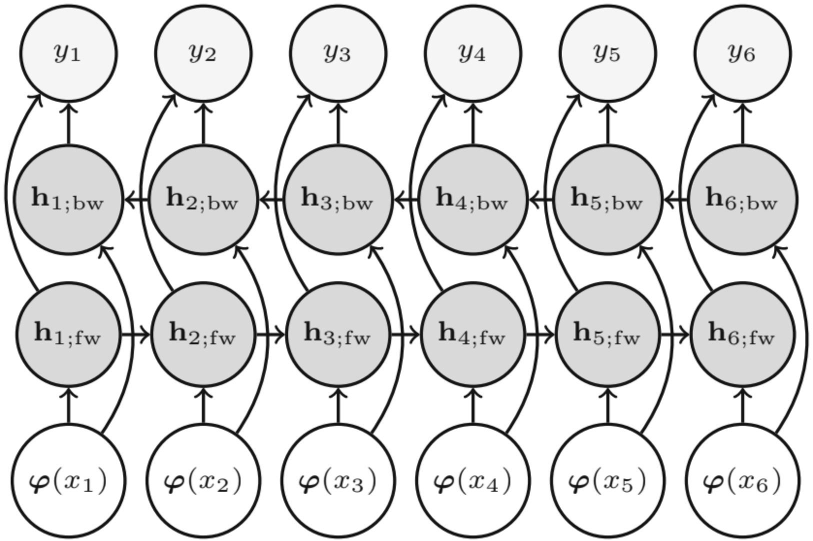
Nonlinear Model

- Allows for interplay between different facets of expressiveness (pitch, dynamics, tempo, pitch durations)
- Uses feed-forward NN to allow nonlinear combinations of basis functions
- Generally small network size is sufficient.



Temporal Factors

- Additionally to Feed-Forward NNs, use LSTM architecture with Forward-Backward.
- Allows for noticeable broader changes in expression as compared to nonlinear combinations by themselves



Solo Performance vs Ensemble Performance

- Each instrument can produce its own set of basis functions for its own dynamics
- Potential for huge number of basis functions in large ensembles
  - Same piece played with different instrumentation should have the same dynamics
- Merging operation on instrument classes
  - Reduces number of functions
  - Ensures consistency across different instrumentations

Contrasting Results of Different Models

- Magaloff Corpus
  - Missed notes, but only expressiveness matters, not notes themselves
  - Human musicians are inherently ground truth for expressiveness parameters
- Slight universal improvement from Linear -> Nonlinear
  - Does that mean anything?
- Some bases have more sway on ending expressiveness (Dynamics):
  - Beat 1
  - fff, ff, pp, ppp
  - accent

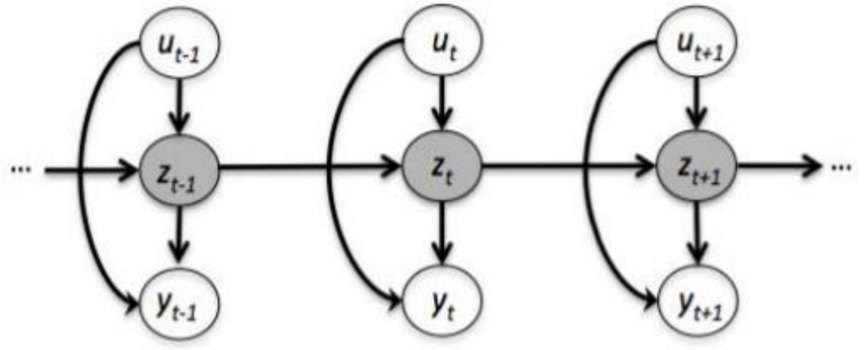
Spectral Learning

Interactive expressiveness

- Performance is rarely one musician alone
  - Even “solo” performances often have accompaniment
- Musicians have to agree on expressiveness for a cohesive performance
  - Rehearsals allow for communication
- Xia et. al, 2015 considers a piano duet where the artificial pianist’s expressiveness considers the other pianist

Model: Linear Dynamic System

- **U**: both the 1st pianist's musical expression and score information
- **Z**: hidden mental states of the 2nd pianist that influence the performance
- **Y**: 2nd pianist's musical expression



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

Input Features

Score information

- High pitch contour
- Low pitch contour
- Beat phase

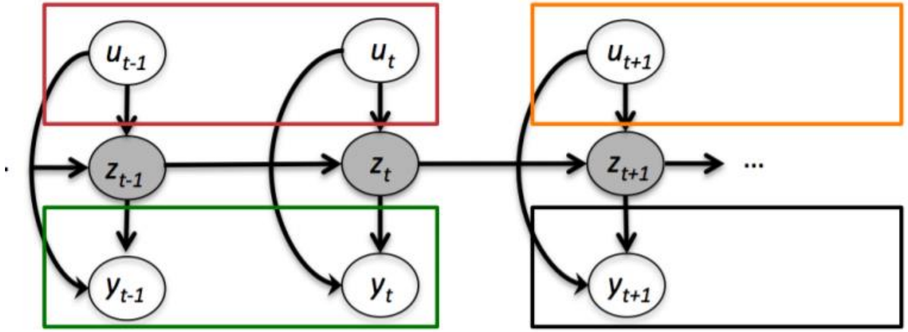
1st Pianist Expression

- Tempo
- Onset deviation
- Duration
- Dynamics

Training: Spectral Learning

- Learns hidden states by predicting future performance from past features
- Forces prediction to go through low-rank bottleneck
- Three regressions
  - First two estimate hidden states by oblique projections and SVD
  - Third estimates parameters

Step1: Oblique Projections



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

- Since the future U is unknown in real-time, use the oblique projection of future Y
- Partial explanation of future based on past

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

Step 2: State Estimation

$$\tilde{Y}_F = \Gamma_f Z_f \stackrel{\text{def}}{=} \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{\frac{d}{2}-1} \end{bmatrix} \left[Z_{\frac{d}{2}+1}, Z_{\frac{d}{2}+2}, \dots, Z_{T-\frac{d}{2}} \right]$$

$$\tilde{Y}_F = \mathcal{U}\Lambda\mathcal{V}^T = (\mathcal{U}\Lambda^{\frac{1}{2}})(\Lambda^{\frac{1}{2}}\mathcal{V}^T)$$

- Use Singular Value Decomposition (SVD) to estimate states
- Throw out non-zero singular values

Step 3: Parameter Estimation

- With the estimated hidden states, we can estimate parameters

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

$$\hat{Z}_f^S = A\hat{Z}_f + BU_f^S + e_w$$
$$Y_f = C\hat{Z}_f + DU_f + e_v$$

Training

- Learn LDS parameters for a specific song from 5-6 pairs of human pianists playing duet
- Machine pianist has 4 rehearsals with human soloist based on these parameters

Results of Spectral Learning Method

- Methods for comparison:
  - Linear Regression
  - Neural Network
  - Timing estimation used in automatic accompaniment systems (baseline)
- Compare dynamics residual and timing residual
  - How different is each method from the human performance?
- LDS performs best, but not always by very much
- Demos: Human soloist with machine accompaniment

Rule-based approach

Spectral learning with 4 rehearsals

Which Method is Superior? (Pros/Cons)

Spectral Learning

- Good results
- Considers expressiveness of other pianist

- Assumes linear function
- Evaluation is limited

Basis Functions

- Good Results
- Extensible via ML techniques
- Easily extensible beyond dynamics into other areas of expressiveness

- Prone to one-off errors
- Not useful for real-time performance/reading

References

[1] C. E. Cancino-Chacón and M. E. C. Grachten, “An Evaluation of Score Descriptors Combined with Non-linear Models of Expressive Dynamics in Music,” *Discovery Science Lecture Notes in Computer Science*, pp. 48–62, 2015.

[2] C. E. Cancino-Chacón, T. Gadermaier, G. Widmer, and M. Grachten, “An evaluation of linear and non-linear models of expressive dynamics in classical piano and symphonic music,” *Machine Learning*, vol. 106, no. 6, pp. 887–909, Sep. 2017.

[3] D. Anderson and R. Kiuvila, “Formula: a programming language for expressive computer music,” *Computer*, vol. 24, no. 7, pp. 12–21, 1991.

[4] E. Schubert, S. Canazza, G. D. Poli, and A. Rodà, “Algorithms can Mimic Human Piano Performance: The Deep Blues of Music,” *Journal of New Music Research*, vol. 46, no. 2, pp. 175–186, May 2017.

[5] G. Grindlay and D. Helmbold, “Modeling, analyzing, and synthesizing expressive piano performance with graphical models,” *Machine Learning*, vol. 65, no. 2-3, pp. 361–387, 2006.

[6] J. Murphy, J. Mcvay, P. Mathews, D. A. Carnegie, and A. Kapur, “Expressive Robotic Guitars: Developments in Musical Robotics for Chordophones,” *Computer Music Journal*, vol. 39, no. 1, pp. 59–73, 2015.

[7] J. Solis and A. Takanishi, “Anthropomorphic Musical Robots Designed to Produce Physically Embodied Expressive Performances of Music,” *Guide to Computing for Expressive Music Performance*, pp. 235–255, 2012.

[8] G. Xia, Y. Wang, R. B. Dannenberg, an G. Gordon, “Spectral Learning for Expressive Interactive Ensemble Music Performance,” *In Proceedings of the 16th International Society for Music Information Retrieval Conference* (pp. 816–822), Málaga, Spain: ISMIR.

[9] S. Flossmann, M. Grachten, and G. Widmer, “Expressive Performance Rendering with Probabilistic Models,” *Guide to Computing for Expressive Music Performance*, pp. 75–98, 2012.

[10] T. H. Kim, S. Fukayama, T. Nishimoto, and S. Sagayama. 2011. Polyhymnia : An Automatic Piano Performance System with Statistical Modeling of Polyphonic Expression and Musical Symbol Interpretation. *Proceedings of the International Conference on New Interfaces for Musical Expression*, pp. 96–99.