MUSICAL EXPRESSIVENESS

Tessa Kettelberger

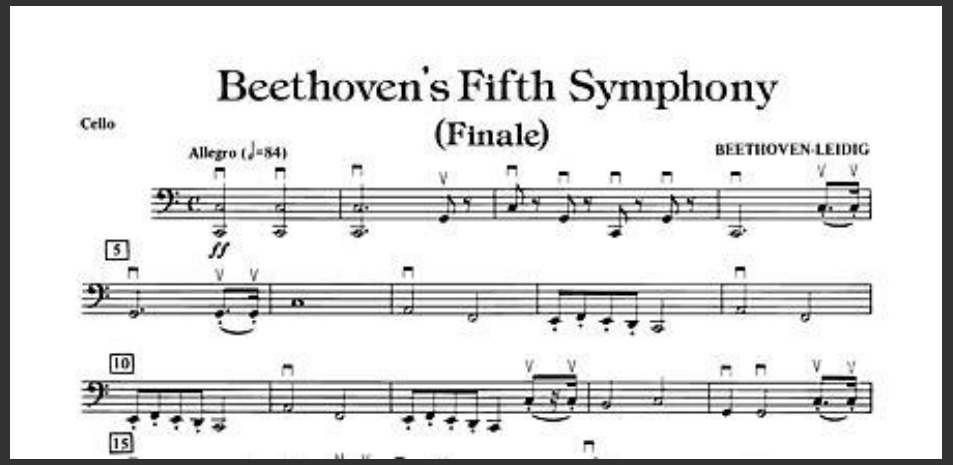
THE PROBLEM OF MUSICAL EXPRESSIVENESS

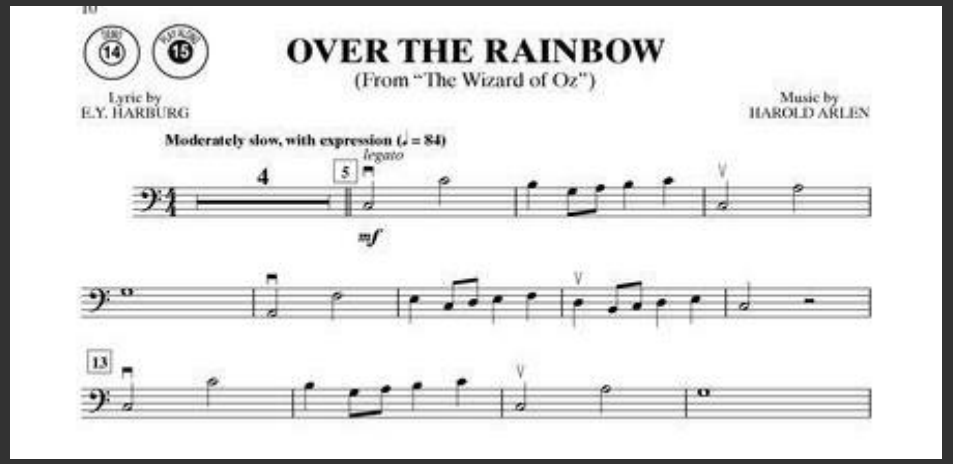
How do we account for differences between musical performance and the literal score?

Part of humans' understanding of the music they listen to, create, and perform is the idea of interpretation and stylistic choice.

THE PROBLEM OF MUSICAL EXPRESSIVENESS

- Performances very rarely adhere exactly to the score
- How can we model this relationship between the performance and the score?
  - Expression in addition to score
  - **Expression as variation from the score**
- How can we model the performance of variable or interpretation-based cues :
  - Dynamics
  - Articulation
  - Tempo
- How do we model the influences that lead to these interpretations?





THE PROBLEM OF MUSICAL EXPRESSIVENESS: CHALLENGES

- If we are trying to reach “human level”, what is that level?
- Approaches must assume that the task is done in a common way by humans
- How much information do we need to take into account?
- Does high-scoring expressiveness sound good?

THE PROBLEM OF MUSICAL EXPRESSIVENESS: APPLICATIONS

- Computer performance of music
  - Music generation
- Computer listening and audio feature extraction
- Computer creativity
- Player identification

APPROACHES

- Psychological
- Explicit models
- Data Mining and Machine Learning
  - Linear models
  - Nonlinear models

PSYCHOLOGY-BASED AND EXPLICIT MODELS

(old models)

PSYCHOLOGY-BASED AND EXPLICIT MODELS

- Timmers, R. & Honing H. (2002), Friberg, A. & Sundberg, J. (1999), Todd, N. (1992)
  - Annotate scores explicitly to capture a specific performance
  - Represent timing-based expressive choices by note-length ratios
  - Treat tempo as velocity
  - Tempo-dynamic coupling
  - Use the body and movement as a basis for measurement

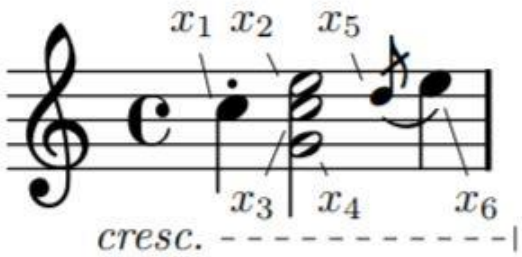
LINEAR REPRESENTATIONS

BASIS FUNCTION MODELING

- LBM (Linear Basis Models) are common
- Treat the score as an effect on the performance
- Represent the meaning of the score through a discrete basis functions
- Learn weights for each function's influence on performance outcome

BASIS FUNCTION MODELING: THE FUNCTIONS

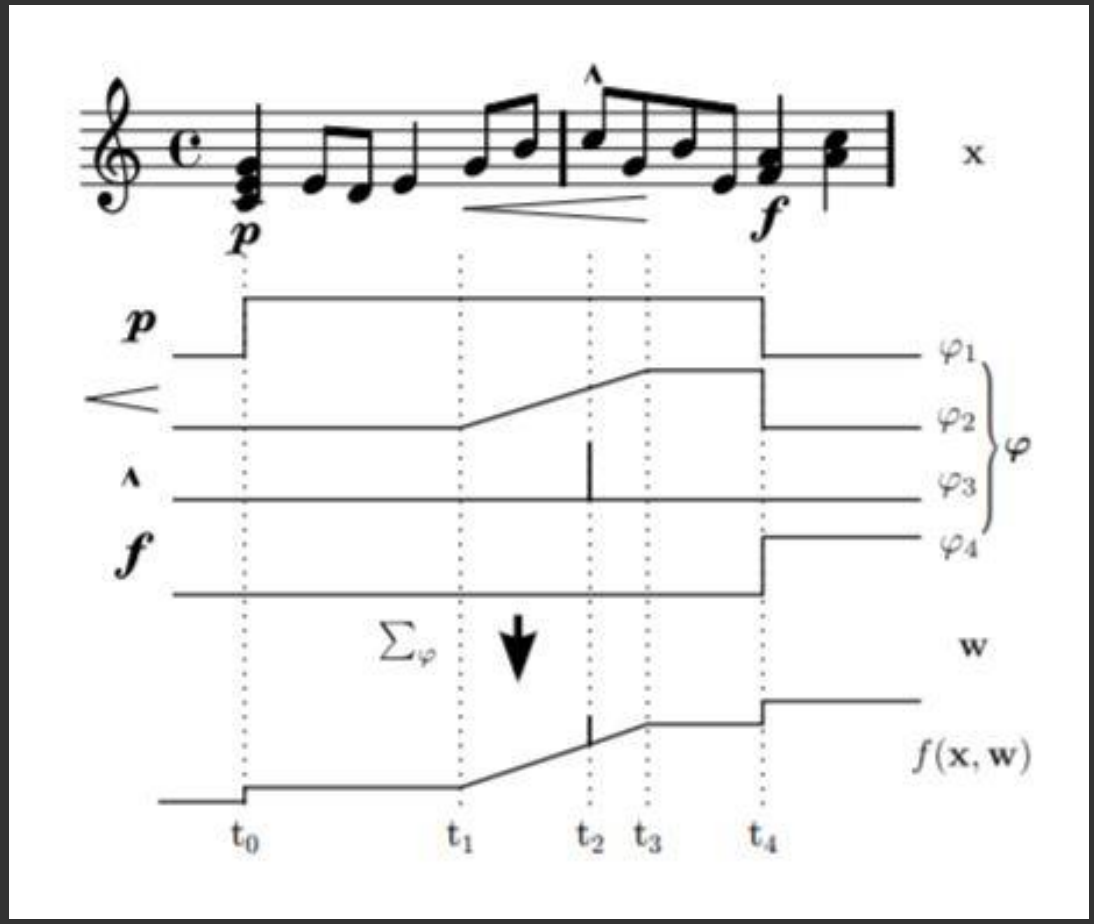
$\varphi : X \longrightarrow [0,1]$ for note elements X



note	$\varphi_{staccato}(.)$	$\varphi_{grace}(.)$	$\varphi_{pitch}(.)$	$\varphi_{pitch^2}(.)$	$\varphi_{crescendo}(.)$
x_1	1	0	$\frac{72}{127}$	$(\frac{72}{127})^2$	0.00
x_2	0	0	$\frac{76}{127}$	$(\frac{76}{127})^2$	0.25
x_3	0	0	$\frac{72}{127}$	$(\frac{72}{127})^2$	0.25
x_4	0	0	$\frac{67}{127}$	$(\frac{67}{127})^2$	0.25
x_5	0	1	$\frac{74}{127}$	$(\frac{74}{127})^2$	0.75
x_6	0	0	$\frac{76}{127}$	$(\frac{76}{127})^2$	0.75

BASIS FUNCTION MODELING

- Each aspect of performance = a sum of relevant, weighted basis functions
- Functions may be global or local
- Multiple types of basis functions

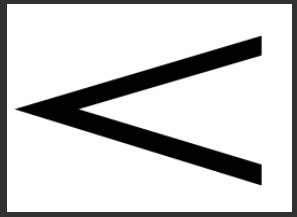


BASIS FUNCTION MODELING

- Indicator basis functions
- Polynomial pitch-specific basis functions
- Expressive dynamics basis functions
  - Constant, impulsive, and gradual dynamics
- Implication-realization basis functions







BASIS FUNCTION MODEL: THE MODEL

- Each φ, with an applied weight, corresponds to a **y**, its *expressive parameter* or *target*
- LBMs determine the optimal level of influence of **y** with *weights* **w.**
- The target **y** can be represented through the equation below, where **x** is the vector of notes and E is noise

$$\mathbf{y} = f(\mathbf{x}, \mathbf{w}) + \epsilon = \boldsymbol{\varphi}(\mathbf{x})\mathbf{w} + \epsilon$$

- Learn **w** from a set of performances given to us in the form (**x**,**y**) where the optimal **w** is given by:

$$\hat{\mathbf{w}} = \operatorname{argmin}_{\mathbf{w}} \| \mathbf{y} - \boldsymbol{\varphi}(\mathbf{x})\mathbf{w} \|$$

BASIS FUNCTION MODEL

- Evaluated accuracy with r, R2 goodness of fit:
  - How strongly the predicted model and the actual performance correlate
  - Local models overfit with all parameters included
  - Global models do best with all parameters included:
  - R2 of 0.162 and r of 0.462
  - Differs significantly across pieces

FREQUENT CORRELATED TRENDS

- Uses *Frequent Correlated Trends* to represent 'imperfect' behaviors ie. Imprecise score interpretations
- A behavioral model based on representing mined repeated patterns in the structural (score) dimension and the other (performance) dimension
- Step 1: Learn frequently repeating patterns or trends
- Step 2: Treat these patterns as musical units and learn what score patterns co-occur with what performance patterns

FREQUENT CORRELATED TRENDS

Automatic pattern extraction (Similar Sequence Miner) to extract trends:

1. Do a scan of the symbolically represented song counting instances of each note
2. Scan again on a transposed version of the song
3. Eliminate infrequent notes/patterns
4. Merge commonly occurring strings of size k that have k-1 elements in common to make patterns of size k+1
5. Eliminate infrequent patterns (Go back to 3 and repeat)

STEPS OF PATTERN RECOGNITION

ABCABCD

1. K = 1
2. Eliminate infrequent patterns

ABCABC

Merge strings of size k with k-1 common elements

K=1

AB BC CA AB BC

AB BC CA AB BC

- Eliminate infrequent patterns

AB BC AB BC

AB BC AB BC

- K=2
- Merge patterns with k-1 common elements

ABC ABC

PATTERN EXTRACTION AND PROBABILITY

- Transform pattern observations into only a few abstract values with a parser
  - Melodic surfact
- Create from this abstract, general representation a smaller alphabet of patterns

Higher, Lower, or Same : Only 3 ‘Notes’

PATTERN EXTRACTION AND PROBABILITY

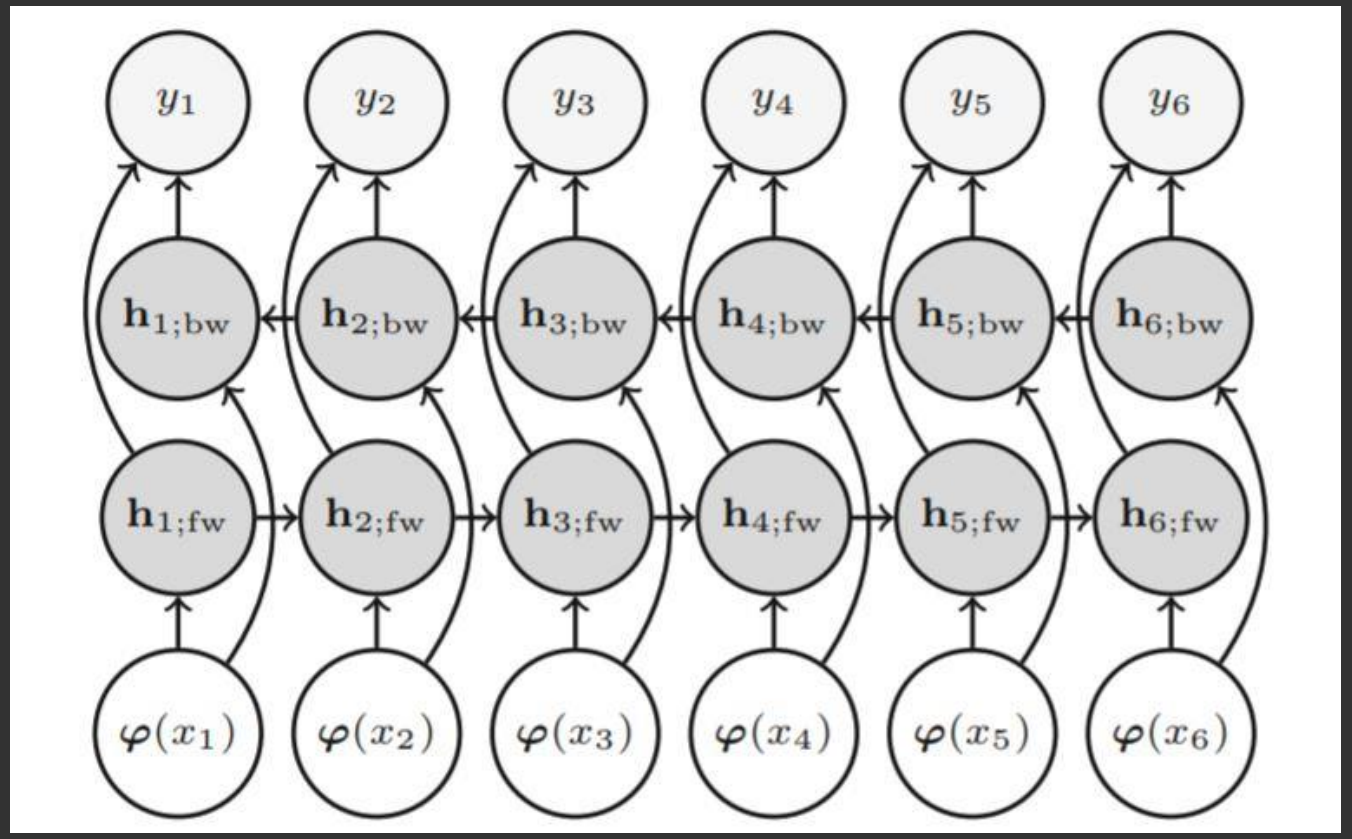
- Create probability distributions based on patterns occurring at the same time in the score and performance dimension
- Use these distributions for classification with an appropriate distance function
- Able to learn the interpretation styles of 23 performers and guess correctly 90% of the time

NON-LINEAR MODELING

RECURRENT NBMS

- Nonlinear Basis-Function Models
- A suggested improvement to LBMs which takes backward temporal data
- Operates in units of time, not notes
- Proposed RNN with one bidirectional layer and LSTM

RECURRENT NBMS



$$\mathbf{h}_t = f_h\left(g_\varphi\left(\boldsymbol{\varphi}(x_t)\right) + g_h\left(\mathbf{h}_{t^*}\right)\right)$$

Input: Matrix of basis functions

Output: Weights

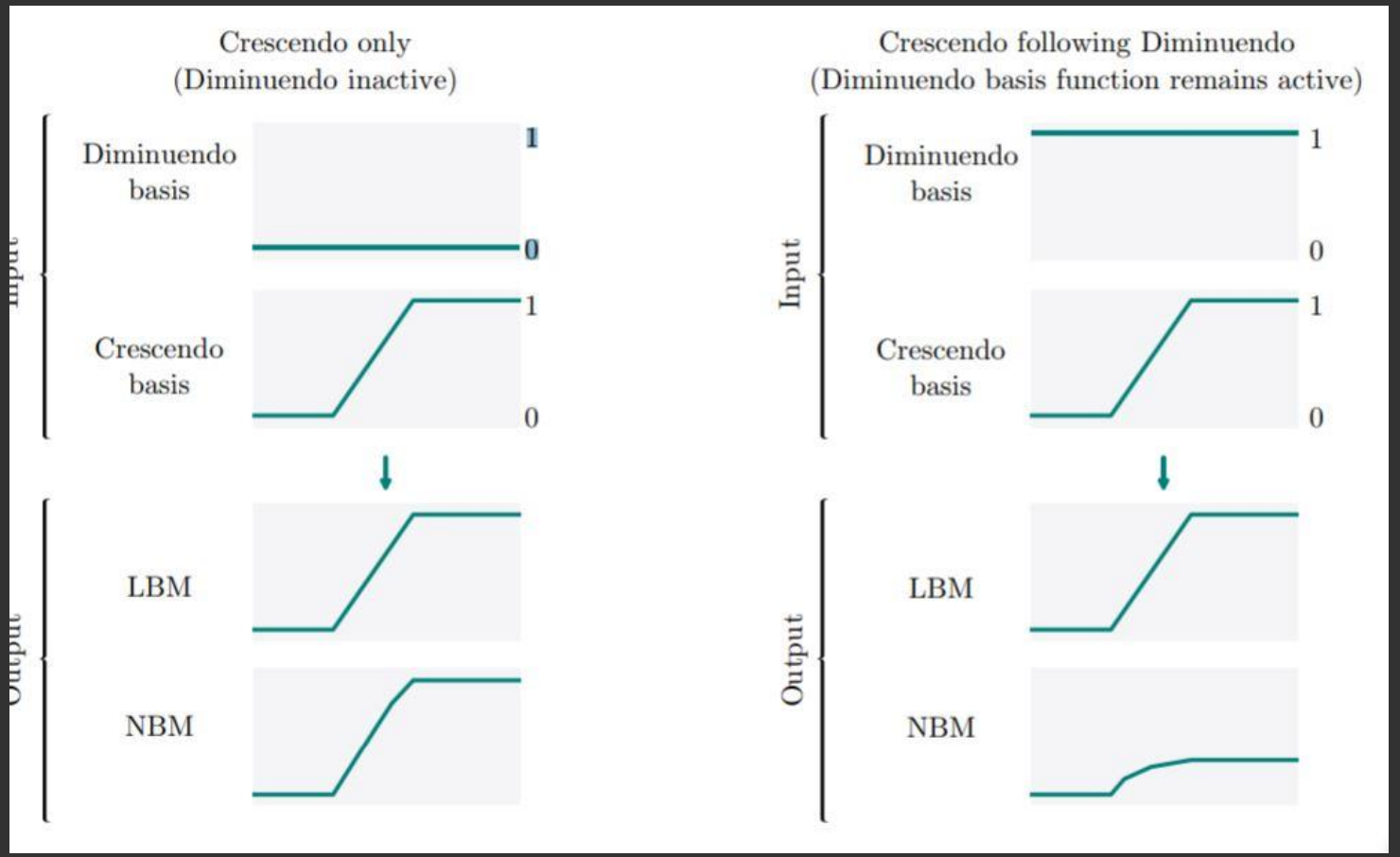
RECURRENT NBMS

- Given training pairs of scores **Φ** and targets **t**
- Target for training: expressive performance
- Minimize the mean squared error between predictions and the targets

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmin}} \frac{1}{K} \sum_{k} \| f(\boldsymbol{\Phi}_k; \mathbf{w}) - \mathbf{t}_k \|^2.$$

RECURRENT NBMS

- Significant accuracy improvement over LBMs (r=0.312 vs. r=0.609)
- Approach also validated by the discovery of examples of the interpretation of musical markings being influenced by markings that follow
  - This affect is not captured in linear approach



REFERENCES

Cancino-Chacon et al.: An evaluation of linear and non-linear models of expressive dynamics in classical piano and symphonic music.

Molina Solana, Miguel Jose. Representation of uncertain and imprecise behaviors and its application to music performance. Universidad de Granada. 2012.

Grachten, Maarten and Widmer, Gerhard. Linear Basis Models for Prediction and Analysis of Musical Expression.

Timmers, R. & Honing, H. On music performance, theories, measurement and diversity. Cognitive Processing (International Quarterly of Cognitive Sciences).

Friberg, A. & Sundberg, J. Does music performance allude to locomotion? A model of final ritardandi derived from measurements of stopping runners.

Todd, N. (1992) The dynamics of dynamics: A model of musical epression. Department of Psychology, University of Exeter.

Van Herwaarden, S., Grachten, M, & Bas de Haas, W. (2014). Predicting Expressive Dynamics in Piano Performances Using Neural Networks

Palmer, C.(1996). Anatomy of a performance: Sources of musical expression. Music Perception, 13(3).

Grindlay, G.,& Helmbold, D. (2006). Modeling, analyzing, and synthesizing expressive piano performance with graphical models. Machine Learning, 65.

Bishop, C. M. (2006). Pattern recognition and machine learning. New York, NY: Springer Science.