Music Expressiveness

Joseph Jaeger and Yiyang Wang

What is musical expression?

- Related to musical interpretation
- Key parameters affecting expression:
  - Dynamics
  - Tempo
  - Articulation
  - Timbre

Musical expression in the context of MIR

Goals:

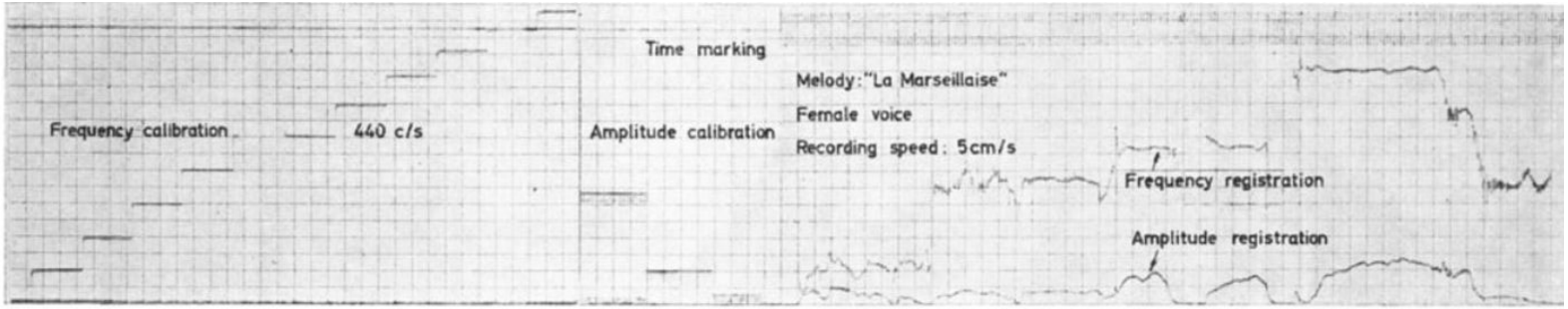
- To predict, analyze, and generate musical expression

Motivation and applications:

- Improve our understanding of musical expression
- Objective evaluation and comparison of musical performances
- Music tutoring software
- Improved music generation
- Improved MIDI rendering

Historical analytical methods

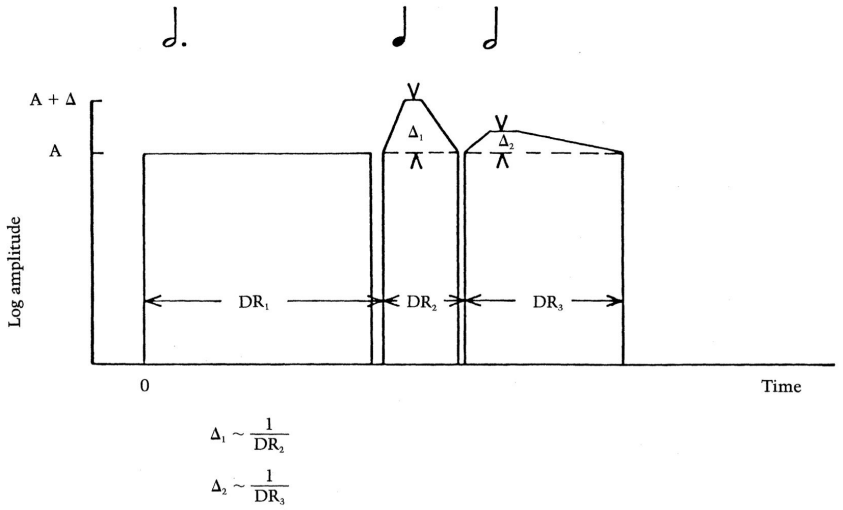
- Carl E. Seashore (1866-1949)
  - Made physical measurements of musical performances throughout his career using various instruments
  - “The Measure of a Singer” (1915), “The Vibrato” (1932), and many more
- Revival of music performance measurements in the 1960s
- Tove et al. (1965) designed new equipment and obtained the data below



Rule-Based Models

Synthesis-by-Rule (Sundberg et al., 1983)

- Realization: vowel synthesizer
- Controller:
  - A notation program
  - A text-to-speech program
- 7 sets of rules for synthesizing the (e.g. amplitude vs. pitch, accents vs. note-value contrasts, amplitude continuity)
- Evaluated by the jury (experts) and listening tests
- Iterative analysis-by-synthesis approach



(accents rule illustration)

Sundberg, J., Askenfelt, A., and Fryde´n, L. Musical performance: A synthesis-by-rule approach. Computer Music Journal, 7(1):37–43, 1983.

Data-Driven Models

Analyze expressive timing: Li et al. (2017)

- Tempo curves can represent expressive timing
- Clustering expressive timing into a pre-defined unit can be used for analysis of timing
  - Clustering by musical phrase is a logical choice
  - But, phrase length is not constant
- Li et al. propose clustering the coefficients of the regressed tempo curve rather than clustering the timing directly
- Used GMMs to fit expressive timing distribution within a phrase and regressed polynomial coefficients
- The model fitting regressed polynomial coefficients performed best

The Basis Mixer: Chacón and Grachten (2016)

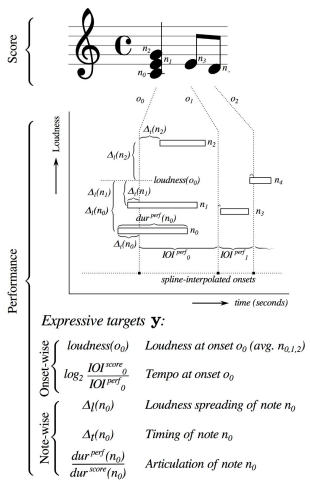
- Used to synthesize expressive performances of scores
- Basis functions numerically encode score content
  - Originally, only encoded explicit score markings
  - Now, it also encodes non-explicit information like tension features

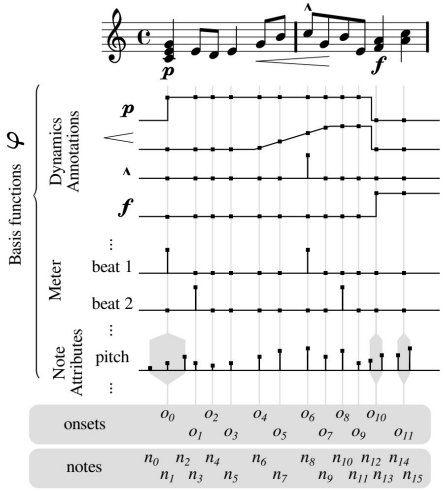
Advantages:

- Linear model is easy to interpret
- Easy to find optimal weight parameters

Disadvantages:

- Model cannot learn non-linear relationships
- Only works for score-based music

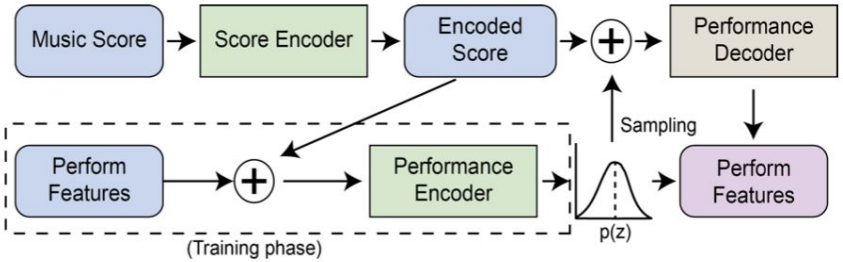


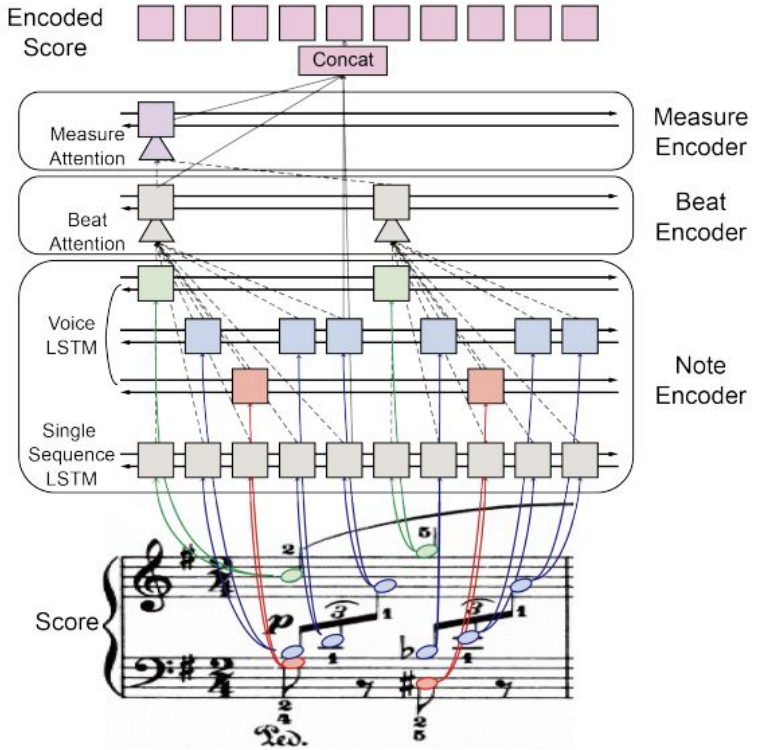


http://carloscancinochacon.com/documents/online_extras/basis_mixer/basis_mixer.html

VirtuosoNet (Jeong et al., 2019)

- MusicXML → MIDI
- Hierarchical RNN - hierarchical attention networks (HAN) for score encoding
- Style-informed performance encoding with conditional variational autoencoder (CVAE)

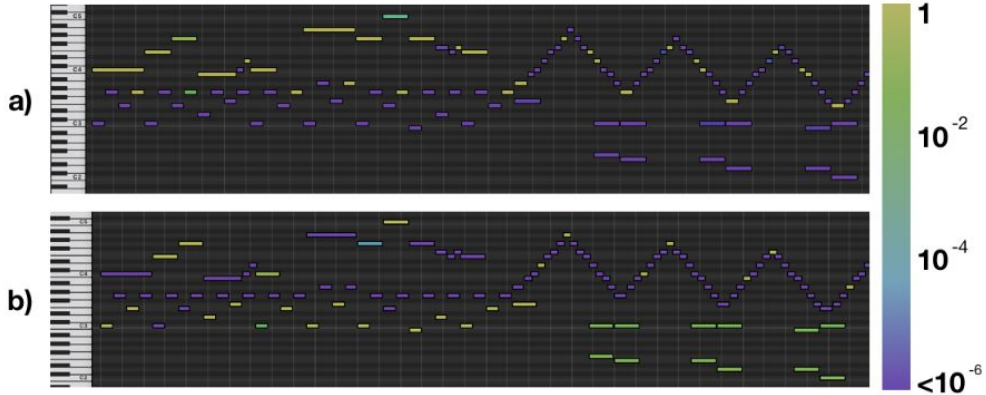




Jeong, Dasaem, Taegyun Kwon, Yoojin Kim, Kyogu Lee and Juhan Nam. "VirtuosoNet: A Hierarchical RNN-based System for Modeling Expressive Piano Performance." International Society for Music Information Retrieval Conference (2019).

VirtuosoNet (Jeong et al., 2019) (cont.)

- Multi-head attention
- An optional measure-level module with nearly identical architecture for a broad picture. Predictions are then fed into the note-level module for refinement
- Realization: remapped MIDI to a Disklavier (A.I. Pianist)



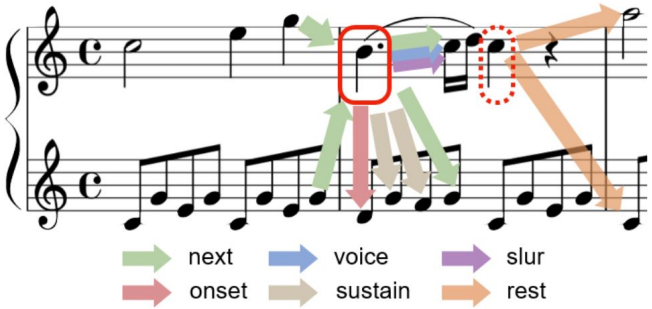
Compare with Basis Mixer:
https://youtu.be/EMdaoXxPkMY?t=14

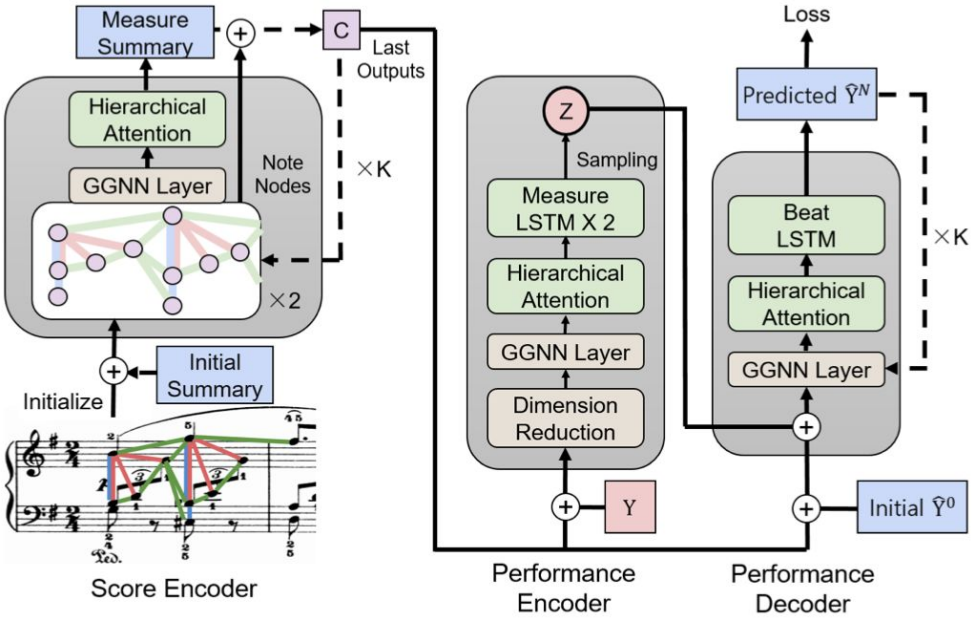
Different styles:

https://youtu.be/EMdaoXxPkMY?t=170

Follow-up? Graph Neural Network (GNN) (Jeong et al., 2019)

- Encoding score input as a graph:
  - Nodes: music notes (w/ set of features)
  - Edges: relations of neighboring notes





Dasaem Jeong, Taegyun Kwon, Yoojin Kim, and Juhan Nam. Score and performance features for rendering expressive music performances. In Proc. of Music Encoding Conf., 2019.

Discussion

Challenges:

- Availability and scope of datasets (Western Classical Music centric)
- Evaluation of model and unbiased assessment of performances
- Parameters for adequate modeling not well understood
- The nuance of deviation from the reference is hard to capture

Limitations (esp. generative models):

- MIDI-domain instructional set as output
- Data hungry

Discussion (cont.)

Trends:

- Data-driven models
- Performance-driven human-computer interactive expressive systems (conductor simulator, auto-accompaniment)
- Music psychology and cognition inspired models (rather than rooting in music theory alone)

Questions?