Modelling Pitch Expectation in Melody: A Comparison of Variable-Order Markov and Transformer-Based Approaches

Ioannis Emmanouilidis
Queen Mary University of London
yiannisem@gmail.com

Marcus Pearce
Queen Mary University of London
marcus.pearce@qmul.ac.uk

centre for digital music

Introduction

- Expectation is widely regarded as a key mechanism underlying the emotional impact of music.
- Some models of expectation capture **statistical learning** (the identification of statistical regularities from a lifetime of musical listening, and from the piece at hand).
- We compare a **probabilistic model** (IDyOM) with a **deep-neural** approach (Music Transformer) to assess how well they capture pitch expectation in melody.

Models

- **IDyOM**: estimates the conditional probability of each possible next note given the preceding musical context, using n-gram modelling.
- Trained on corpus of monophonic melodies.
- Different IDyOM models can be created using various **viewpoints** (enables the tracking of various pitch and rhythm characteristics).
- **Music Transformer**: originally purposed for generative music.
- **Decoder-only** architecture, uses **relative self-attention** to predict the next musical event from a sequence of symbolic tokens.
- Trained on MAESTRO and fine-tuned on MCCC.

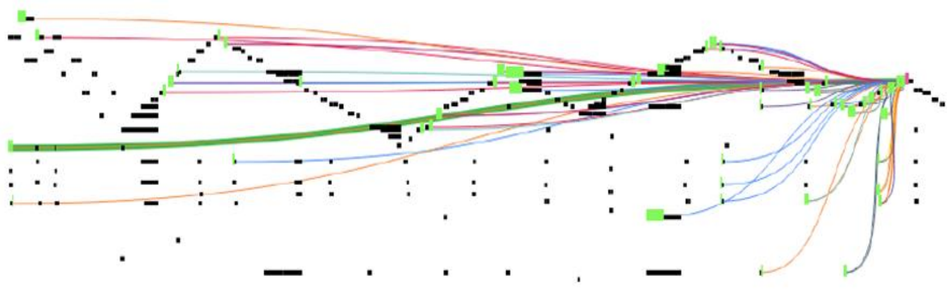

FIG 1. A piano-roll visualisation showing which musical events most strongly affect a prediction made by the Music Transformer

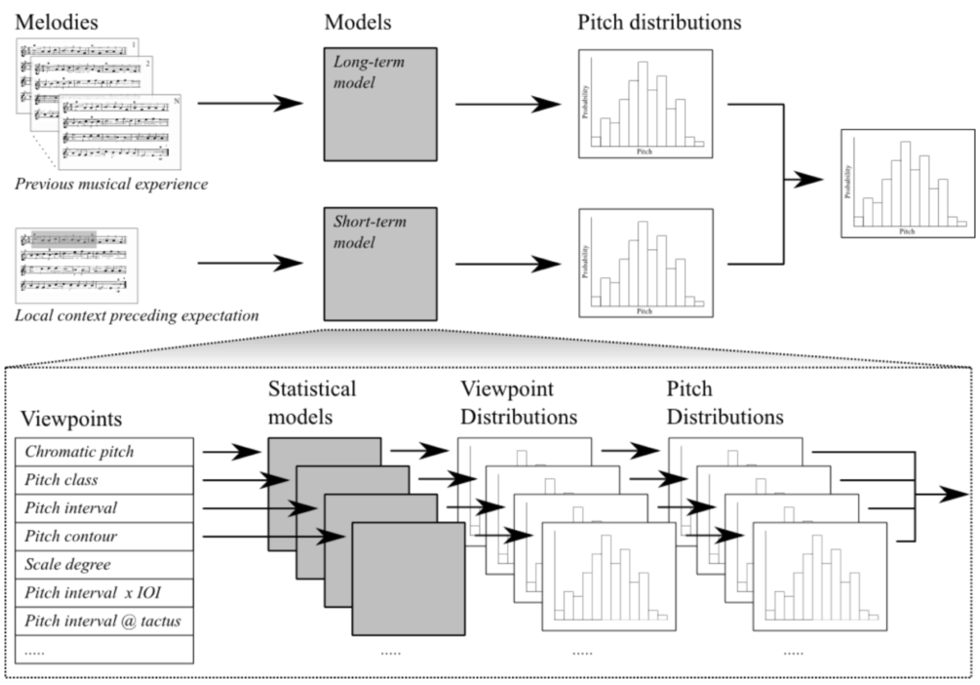

FIG 2. A visualisation of how IDyOM combines various viewpoints in a long-term and short-term model to generate predictions

Methodology

- **OLS regression** to quantify how closely the models' predictions aligned with human ratings of the likelihood of different pitch continuations to **two-note and longer melodic stimuli.**
- **Unique variance partitioning** based on two-predictor regression.
- **Comparison of correlations** (Steiger's test).

Model	Viewpoints
IDyOM 1	`(cpitch)`
IDyOM 2	`((cpint dur) cpintfip cpcint)`
IDyOM 3	`((cpint cpintfref))`
IDyOM 4	`(cpintfib cpintfip (cpint cpintfref) (cpitch ioi))`

FIG 2. Viewpoints chosen in each IDyOM model used

Results

- **Two-note stimuli**: the Music Transformer significantly outperformed IDyOM 1; the optimised IDyOM 2 significantly outperformed the Transformer.
- **Longer stimuli**: IDyOM 4 (optimised) slightly outperformed the Transformer, but the difference was not statistically significant.
- **Mostly overlapping variance** in all cases.

Stimulus	IDyOM model	R^2 (IDyOM)	R^2 (Transformer)	R^2 (both)	Unique IDyOM	Unique Transformer	Z	p
Two-note	1	0.367	0.559	0.581	0.022	0.214	3.5796	0.003
Two-note	2	0.687	0.559	0.701	0.142	0.014	-3.3742	0.0007
Long	1	0.679	0.664	0.782	0.118	0.103	0.2685	0.7883
Long	3	0.554	0.664	0.714	0.050	0.160	-1.8755	0.0607
Long	4	0.729	0.664	0.786	0.122	0.057	1.3385	0.1807

TABLE 1. Results table showing the coefficient of determination (R^2) for each regression, the unique contributions of each model, and the Z and p values from the Steiger's test

Conclusions

- The Music Transformer modelled melodic expectation competently, but without outperforming the optimised IDyOM models.
- The low unique variances suggest that the two models tend to predict the same context-probe pairs well across many stimuli.
- Could indicate that they use similar representations, but further work is needed to render the Transformer's predictions more interpretable.
- Such speculation can be guided by examining which notes in the context receive higher softmax probability.

Queen Mary
University of London