Predictability and Statistical Memory in Classical Sonatas and Quartets

Can mathematics reveal what makes one composer sound different from another?

I used information theory and probabilistic modelling to measure how much "memory" is built into classical music – how strongly each note depends on the notes that come before it in a melody.[1]

To measure musical memory, I developed three complementary mathematical definitions and applied them to the piano sonatas and string quartets of Haydn, Mozart, Beethoven, and Schubert.

My dataset consisted of **316 sonata movements** and **289 quartet movements** from the publicly available MIDI database kunstderfuge.[2]

Each test asked the same basic question in a different way:
How far in the past does a note "remember"?

Haydn **Mozart** **Beethoven** **Schubert**

Public-domain images of composers from Wikimedia Commons, 2025.

- *Joseph Haydn painted by Thomas Hardy (1791)*
  https://commons.wikimedia.org/wiki/File:Joseph_Haydn.jpg
- *Wolfgang Amadeus Mozart painted by Barbara Krafft (1819)*
  https://commons.wikimedia.org/wiki/File:Wolfgang-amadeus-mozart_1.jpg
- *Ludwig van Beethoven painted by Joseph Karl Stieleer (1820)*
  https://commons.wikimedia.org/wiki/File:Joseph_Karl_Stieler%27s_Beethoven_mit_dem_Manuskript_der_Missa_solemnis.jpg
- *Franz Schubert painted by Wilhelm August Rieder (1875)*
  https://commons.wikimedia.org/wiki/File:Franz_Schubert_by_Wilhelm_August_Rieder_1875.jpg

Methods and Results

Time-delayed mutual information: Rate of information[3] decay between distributions of notes separated by different lags. *How much information about a note is contained in the choice of an earlier note; how rapidly does this decay with intervening lag?*

$$I(\tau) = I(X_t; X_{t+\tau}) = \sum_{x_t} \sum_{x_{t+\tau}} \mathbb{P}(x_t, x_{t+\tau}) \log \left(\frac{\mathbb{P}(x_t, x_{t+\tau})}{\mathbb{P}(x_t)\mathbb{P}(x_{t+\tau})} \right)$$

Markov model goodness of fit: Log likelihood ratio[4] of best-fit first- and second-order Markov chains, and associated Bayesian Information Criterion (BIC). *How much better is a melody described by a generative model that predicts each note from the two previous notes, versus from one previous note?*

$$\mathrm{BIC} = -2\log(L) + d\log(n)$$

$$\log(L) = \sum_{t=1}^{n-1} \log\left(P(X_t, X_{t+1})\right)$$

$$P = \begin{array}{cc} X_{t-2} & X_{t-1} \\ 1 & 1 \\ 1 & \vdots \\ 1 & m \\ 2 & 1 \\ 2 & \vdots \\ 2 & m \\ \vdots & \vdots \\ m & m \end{array} \begin{array}{c} X_t \\ \begin{array}{ccc} 1 & \cdots & m \end{array} \\ \begin{bmatrix} {}_1p_{11} & \cdots & {}_1p_{1m} \\ \vdots & \ddots & \vdots \\ {}_1p_{m1} & \cdots & {}_1p_{mm} \\ {}_2p_{11} & \cdots & {}_2p_{1m} \\ \vdots & \ddots & \vdots \\ {}_2p_{m1} & \cdots & {}_2p_{mm} \\ \vdots & \ddots & \vdots \\ {}_mp_{m1} & \cdots & {}_mp_{mm} \end{bmatrix} \end{array}$$

Mixture transition distributions (MTD): A low-dimensional family of higher-order Markov chains, with weighted contributions across orders.[5] *How much weight is associated with different lags, in a generative model that is additive across lags?*

$$\mathbb{P}(X_t = x_0 \mid X_{t-1} = x_1, \ldots, X_{t-k} = x_k) = \sum_{i=1}^{k} \beta_i P(x_i, x_0)$$

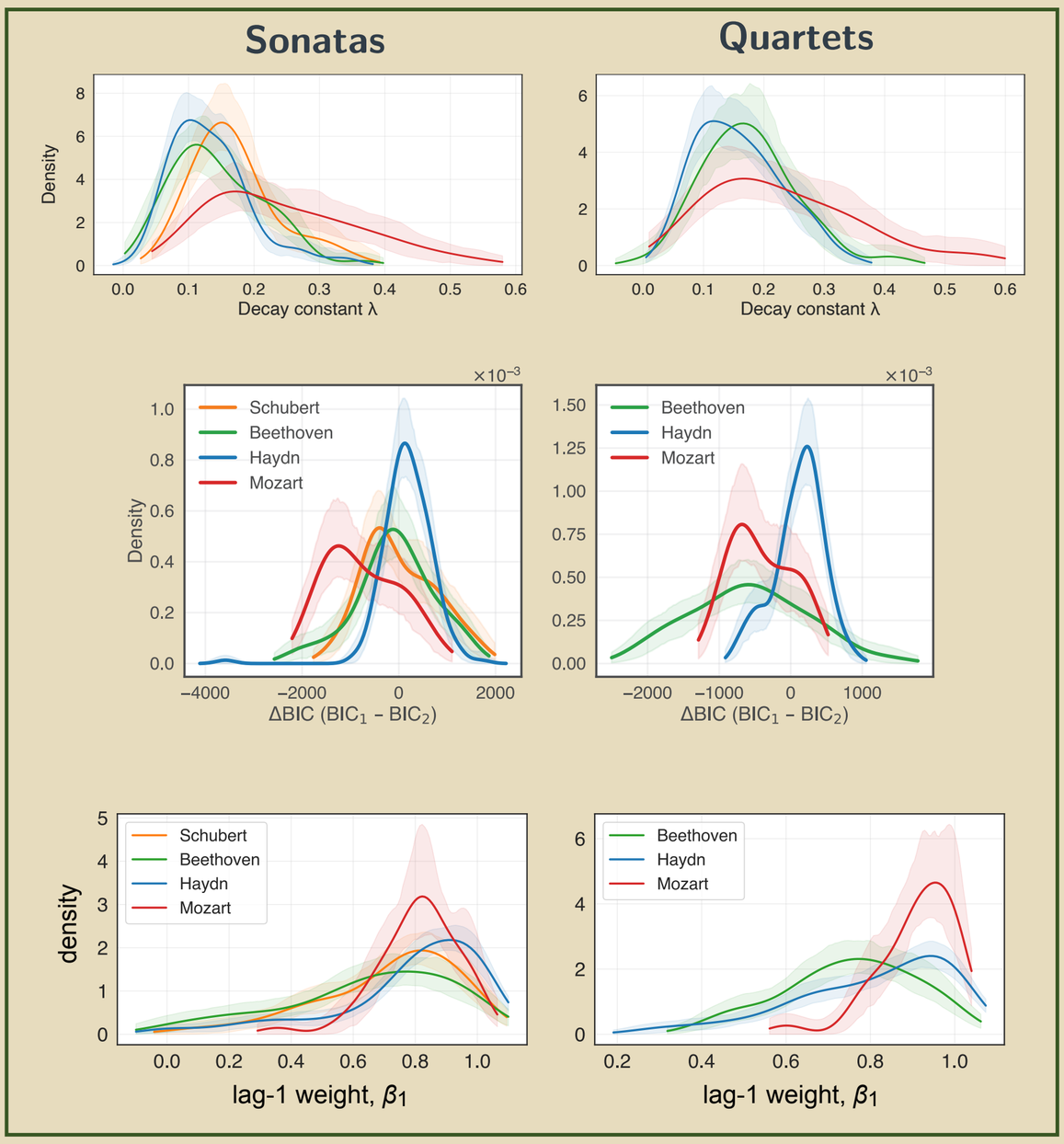

Figure created by author and adapted from Chen-Plotkin L, Kulkarni SS, and Bassett DS. Predictability and statistical memory in classical sonatas and quartets, arXiv (physics.soc-ph) 2509.24172 (2025)

Across all three methods, Mozart's melodies show less "memory" or more surprise.
Mutual information between lagged notes decays more rapidly in Mozart's melodies than those of other composers.
Order-one models are favored over order-two by BIC more often for Mozart than for other composers.
MTD models of Mozart's melodies place more weight on short lags, compared to other composers.

Connections to other fields

How do these **mathematical results** compare to **musicological accounts**[6] of these composers' styles?

What are the tradeoffs between these approaches and **LLM-based models** of music?[7]

What role can science and mathematics play in understanding **art, culture, and creativity**?

Can we **quantify** aesthetic experience? Perhaps more importantly: ***should*** we?

References

1. Chen-Plotkin L, Kulkarni SS, Bassett DS. Predictability and Statistical Memory in Classical Sonatas and Quartets. arXiv: 2509.24172 (2025).
2. Kunst der Fuge. (n.d.). *Kunst der Fuge: MIDI files database*. http://www.kunstderfuge.com/
3. Shannon CE. A mathematical theory of communication. The Bell system technical journal. 2:379-423 (1948).
4. Wilks SS. The large-sample distribution of the likelihood ratio for testing composite hypotheses. The annals of mathematical statistics. 9:60-2 (1938).
5. Raftery AE. A model for high-order Markov chains. Journal of the Royal Statistical Society Series B: Statistical Methodology.47:528-39 (1985).
6. Caplin WE. Classical form: A theory of formal functions for the instrumental music of Haydn, Mozart, and Beethoven. Oxford University Press; 1998.
7. Huang CZ et al. Music transformer. arXiv:1809.04281 (2018)