Application of Hidden Markov Models to music performance style classification via timing and loudness features

C. Vaquero Patricio
Universiteit van Amsterdam,
Music Cognition Group
e-mail: c.vaqueropatricio@uva.nl

E.Chew
Queen Mary University of London,
Centre for Digital Music
e-mail: elaine.chew@qmul.ac.uk

Hidden Markov Models (HMMs) have been widely used in modeling time series data and, especially, in speech recognition [5]. In expressive music performance modeling, HMMs have served varied purposes such as generating expressive performances [2] and score following through machine listening [6]. We explore a novel use of HMMs for stylistic classification of performances of the same music piece. In Western art music it is common to distinguish performers and performance styles by their historically informed approaches, or lack thereof, as well as by the characteristics of the instrument being played. In Baroque music performance, we may differentiate recordings into those played on a Baroque period instrument (tuning around 415 Hz for concert A) and on a modern one (around 440 Hz). Aesthetic performing choices are often based on the performers' artistic and musicological approaches as well their previous exposure to other performances. These choices are commonly manifested in performers' use loudness and timing patterns, which are the elements of musical phrasing. Due to the availability of recordings and quantitative methodologies we may differentiate between the use of these expressive features between modern and Baroque performances. The questions we ask include: How well-defined is the division of Baroque and modern performances? Can quantitative methods capture if a performer plays in one style vs. the other? And, ultimately, can listeners perceive these differences? In this study we focus on quantifying performance style differences using HMMs with ground truth being the aforementioned division between Baroque and modern performances commonly found in Classical music culture.

The method is tested on performances of the first half (343 notes in total) of the Prelude of Bach's Suite in G Major (BWV1007). The excerpt was chosen for its isochronous rhythm and the lack of tempo or dynamic markings in the score. A dataset of twelve recordings performed by Anner Bylsma, Jaap Ter Linden and Peter Wispelwey on Baroque cello and by Mstislav Rostropovich, Jean-Guihen Queyras and Yo-Yo Ma on modern cello were collected. These performers were chosen because they have each recorded the piece at two different periods of their careers. Onsets for every note were manually annotated (twice and averaged) using Sonic Visualiser [7]. The timing ($log_2(tempo)$) is calculated per note duration and smoothed over each beat (four consecutive sixteenth notes) [1]. The loudness values (in sones) for each note played were extracted using a

Short Time-Varying Loudness model proposed by Moore [3]. The set of features used therefore consist of the timing and loudness values separately, the timing and loudness combined, and the 1st and 2nd derivatives of each of these features, in order to study the "acceleration" (phrasing trajectory in our case) and jerk (rate of change of acceleration), respectively.

With the goal of automatically classifying the performances we train our models using the HMM implementation on scikit-learn [4]. We carry different experiments building classifiers models for baroque and modern styles according to the possible combinations of features. Our aim is to model the transition of expressive trends as possible expressive trajectories. We initially represent these trajectories as two different states within the hidden layer. We run experiments gradually incrementing the number of hidden states to 15. Once we have trained our baroque and modern models we test each performance by leave-one-run-out cross-validation and obtain the likelihood of each performance belonging to one or the other style model. We validate these models based on a randomized permutation test with 1000 iterations.

The randomized permutation test results show that the HMM models obtained are significantly different ($p < 0.05$) when the included features are loudness and the first and second derivatives of loudness (with 2 and 3 hidden states) or timing and the first derivative of timing (with 2 hidden states). Models based on loudness and timing combined did not discriminate significantly. A main challenge faced in this study is the limited training data available. Further research should be carried with a larger dataset in order to validate the strength of this methodology. We show how by modeling performance styles using HMMs we may quantify and compare how well a performance fits one aesthetic style vs. another using different combinations of features, and how distant performances are from their respective styles. Further research could use this methodology to analyze the evolution of aesthetic trends by sequentially adding different performances to the models.

References

[1] E. Chew and C. Callender. Conceptual and experiential representations of tempo: Effects on expressive performance comparisons. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 7937 LNAI, pages 76–87. 2013.

[2] G. Grindlay and D. Helmbold. Modeling, analyzing, and synthesizing expressive piano performance with graphical models. *Machine Learning*, 65:361–387, 2006.

[3] B. Moore, Brian R. Glasberg, and T. Baer. A model for the prediction of thresholds, loudness, and partial loudness. *Journal of the Audio Engineering Society*, 45(4):224–240, 1997.

[4] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine Learning in {P}ython. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[5] L.R Rabiner and B.H. Juang. An introduction to hidden Markov models. *IEEE ASSP Magazine*, 3(January):4–16, 1986.

[6] C. Raphael. Music Plus One and Machine Learning. In *International Conference on Machine Learning*, pages 21–28, 2010.

[7] C. Cannam, C. Landone, and M. Sandler. Sonic visualiser: An open source application for viewing, analysing, and annotating music audio files. In *Proceedings of the ACM Multimedia 2010 International Conference*, pages 1467–1468, Firenze, Italy, October 2010.