Composer Recognition based on 2D-Filtered Piano-Rolls

Gissel Velarde[1] **Tillman Weyde**[2] **Carlos Cancino Chacón**[3]
David Meredith[1] **Maarten Grachten**[3]

[1] Department of Architecture Design & Media Technology, Aalborg University, Denmark
[2] Department of Computer Science, City University London, UK
[3] Austrian Research Institute for Artificial Intelligence, Vienna, Austria

{gv, dave}@create.aau.dk, t.e.weyde@city.ac.uk,
{carlos.cancino, maarten.grachten}@ofai.at

Summary

We propose a method for music classification based on the use of convolutional models on symbolic pitch–time representations (i.e. piano-rolls).

Background:

- Similar principles of perceptual organization operate in both vision and hearing [3].
- Studies suggest direct interaction between visual and auditory processing in the brain[4, 10, 6].
- Convolutional models have been used to model the physiology and neurology of visual perception [2], [9]. Filters perform tasks like contrast enhancement or edge detection.
- Visually motivated features generated from spectrograms have been successfully used for music classification (see [12, 1]).

The proposed method: is based on the analysis of texture in 2D pitch–time representations.

- Parsing of the music into separate voices is not required,
- Extraction of any other predefined features is not required.

Findings: We show that:

- Filtering significantly improves recognition.
- The results of the experiments suggest that the method is robust to encoding, transposition and amount of information.
- Our best classifier reaches state-of-the-art performance on discriminating between Haydn and Mozart string quartet movements.

Applications: Recommendation systems, music database indexing, music generation and systems as an aid in resolving issues of spurious authorship attribution.

Method

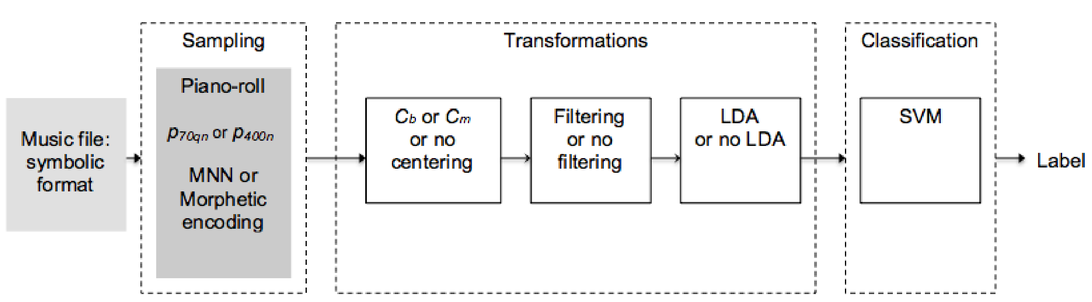

Figure 1: Overview of the method. Music, represented symbolically, is first sampled to 2D images of piano-rolls. Then, various transformations or processing steps are applied to the images, including convolution with predefined filters (Gaussian and Morlet). The order of applying these transformations is from left to right. These transformations are applied in order to find a suitable normalization (i.e., alignment between the images) before classification, and to test the robustness of the method to transformations. Finally, the images are classified with an SVM.

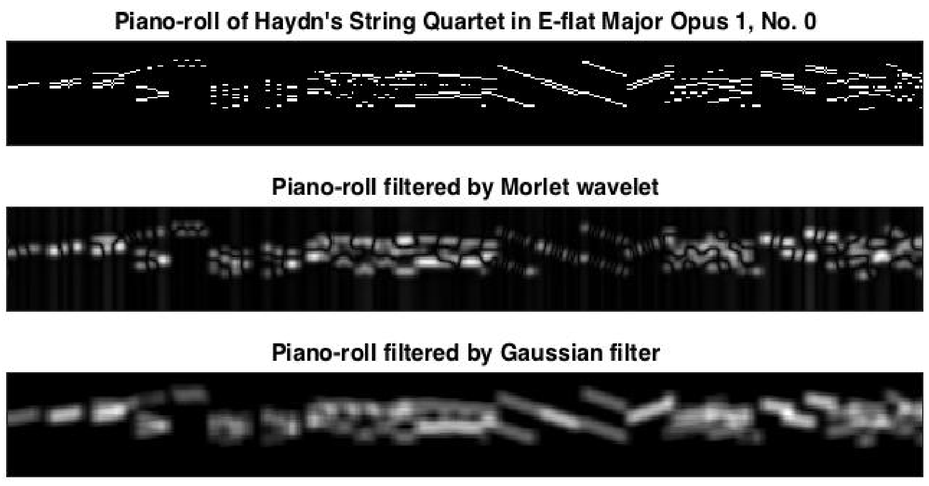

Figure 2: Piano-roll (p_{400n}) morphetic pitch representation (top) of Haydn's String Quartet in E-flat Major Opus 1, No. 0 and its transformations filtered by the Morlet wavelet at a scale of 2 pixels oriented of 90 degrees (second image), and by a Gaussian filter of size 9×9 pixels with $\sigma = 3$ (third image). p_{400n} and its filtered versions are each 56×560 pixels

Experiments

Task: Composer recognition.
Dataset: 54 string quartet movements by Haydn and 53 movements by Mozart, encoded as **kern files, same dataset as in [11].
Evaluation: Method's classification accuracy under different transformations in leave-one-out cross-validation:

- encoding (MIDI Note Numbers (MNN) vs. morphetic pitch),
- transposition (not centering vs. centering with C_b (Pitch range centering) or C_m (Center of mass centering)) and
- amount of information (p_{70qn} (first 70 qn of each piece) vs. P_{400n} (first 400 notes of each piece)).

Results

- At 5% significance level, filtering significantly improves recognition (Wilcoxon rank sum = 194.5, $p = 0.0107$, $n = 12$, with Morlet wavelet), (Wilcoxon rank sum = 203, $p = 0.0024$, $n = 12$, with Gaussian filter) (see Table 1).
- Excluding results with C_m, the performance of the method under different transformations is not significantly affected:
  - MNN vs. morphetic pitch (Wilcoxon rank sum = 269.5, $p = 0.8502$, $n = 16$),
  - not centering vs. pitch range centering C_b (Wilcoxon rank sum = 311.5, $p = 0.0758$, $n = 16$),
  - p_{70qn} vs. P_{400n} (Wilcoxon rank sum = 242, $p = 0.4166$, $n = 16$) (see Table 1)
- There is no significant difference between the results obtained by van Kranenburg and Backer [11] and our best classifier (Wilcoxon rank sum = 11449, $p = 0.8661$, $n = 107$) (see Table 2).

	Pitch–time representation	LDA Morlet	LDA Gauss	LDA NF	Morlet	Gauss	NF
Morphetic pitch	p_{70qn}	65.4	58.9	57.9	53.3	68.2	58.9
	$C_b(p_{70qn})$	65.4	60.7	47.7	57.9	63.6	51.4
	$C_m(p_{70qn})$	53.3	60.7	52.3	64.5	59.8	56.1
	p_{400n}	67.3	**80.4**	57.0	63.6	72.9	55.1
	$C_b(p_{400n})$	62.6	72.9	54.2	61.7	66.4	53.3
	$C_m(p_{400n})$	65.4	65.4	55.1	66.4	70.1	53.3
MNN	p_{70qn}	64.5	67.3	66.4	62.6	66.4	64.5
	$C_b(p_{70qn})$	**70.1**	61.7	63.6	67.3	61.7	61.7
	$C_m(p_{70qn})$	63.6	57.9	57.0	66.4	56.1	54.2
	p_{400n}	66.4	69.2	64.5	65.4	63.6	64.5
	$C_b(p_{400n})$	54.2	64.5	52.3	58.9	58.9	49.5
	$C_m(p_{400n})$	53.3	62.6	42.1	56.1	63.6	44.9

Table 1: Haydn and Mozart String Quartet classification accuracies in leave-one-out cross validation for different configurations of classifiers (NF = no filtering).

Method	Accuracy
Proposed best classifier	**80.4**
Van Kranenburg and Backer (2004) [11]	79.4
Herlands et al. (2014) [5]*	80.0
Hillewaere et al. (2010) [7]*	75.4
Hontanilla et al. (2013) [8]*	74.7

Table 2: Classification accuracies achieved by previous computational approaches on the Haydn/Mozart discrimination task. * indicates that a different dataset was used from that used in the experiments reported here.

Future work

- In preliminary experiments, we have seen that diverse configurations of classifiers (i.e. different filter types, orientations, centering, etc.) seem to provide complementary information, potentially for ensembling
- Also in preliminary experiments, we have observed that the method can be applied to synthetic audio files and audio recordings. In this case, audio files are sampled to spectrograms.
- We are optimistic that our proposed method can perform similarly on symbolic and audio data, and might be used successfully for other style discrimination tasks such as genre, period, origin, or performer recognition.

References

[1] Yandre MG Costa, LS Oliveira, Alessandro L Koerich, Fabien Gouyon, and JG Martins. Music genre classification using lbp textural features. *Signal Processing*, 92(11):2723–2737, 2012.

[2] John G Daugman. Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Research*, 20(10):847–856, 1980.

[3] Diana Deutsch. *Psychology of Music*. Academic Press, San Diego, 3rd edition, 2013.

[4] Marc O Ernst and Heinrich H Bülthoff. Merging the senses into a robust percept. *Trends in Cognitive Sciences*, 8(4):162–169, 2004.

[5] William Herlands, Ricky Der, Yoel Greenberg, and Simon Levin. A machine learning approach to musically meaningful homogeneous style classification. In *Proceedings of the 28th AAAI Conference on Artificial Intelligence (AAAI)*, pages 276–282, 2014.

[6] Souta Hidaka, Wataru Teramoto, Yoichi Sugita, Yuko Manaka, Shuichi Sakamoto, Yôiti Suzuki, and Melissa Coleman. Auditory motion information drives visual motion perception. *PLoS One*, 6(3):e17499, 2011.

[7] Ruben Hillewaere, Bernard Manderick, and Darrell Conklin. String quartet classification with monophonic models. In *Proceedings of the 11th International Society for Music Information Retrieval Conference (ISMIR 2010)*, pages 537–542, Utrecht, The Netherlands, 2010.

[8] María Hontanilla, Carlos Pérez-Sancho, and Jose M Iñesta. Modeling musical style with language models for composer recognition. In *Pattern Recognition and Image Analysis*, pages 740–748. Springer, 2013.

[9] S Marčelja. Mathematical description of the responses of simple cortical cells. *Journal of Neuophysiology*, 70(11):1297–1300, 1980.

[10] Daniele Schön and Mireille Besson. Visually induced auditory expectancy in music reading: a behavioral and electrophysiological study. *Journal of Cognitive Neuroscience*, 17(4):694–705, 2005.

[11] Peter Van Kranenburg and Eric Backer. Musical style recognition—a quantitative approach. In *Proceedings of the Conference on Interdisciplinary Musicology (CIM)*, pages 106–107, 2004.

[12] Ming-Ju Wu, Zhi-Sheng Chen, Jyh-Shing Roger Jang, Jia-Min Ren, Yi-Hsung Li, and Chun-Hung Lu. Combining visual and acoustic features for music genre classification. In *Machine Learning and Applications and Workshops (ICMLA), 2011 10th International Conference on*, volume 2, pages 124–129. IEEE, 2011.

Acknowledgements

The work for this paper carried out by G. Velarde, C. Cancino Chacón, D. Meredith, and M. Grachten was done as part of the EC-funded collaborative project, "Learning to Create" (Lrn2Cre8). The project Lrn2Cre8 acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET grant number 610859. G. Velarde is also supported by a PhD fellowship from the Department of Architecture, Design and Media Technology, Aalborg University. The authors would like to thank Peter van Kranenburg, William Herlands, Yoel Greenberg, Jordi Gonzalez and the anonymous reviewers.