Improving Tokenization Expressiveness With Pitch Intervals

Mathieu Kermarec[1], Louis Bigo[1], Mikaela Keller[1,2]
[1] Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France | [2] Inria

Music *tokenization* methods

- Using NLP based models such as *transformers* on musical data requires to represent music as sequences of atomic elements called *tokens*
- Existing tokenization strategies: *Midi-Like*, *REMI*, *Compound Words*
  → explicit representation of pitch values
  → generalizing musical knowledge to all keys requires duplicating training data by applying transpositions, resulting in large datasets and expensive training procedures.

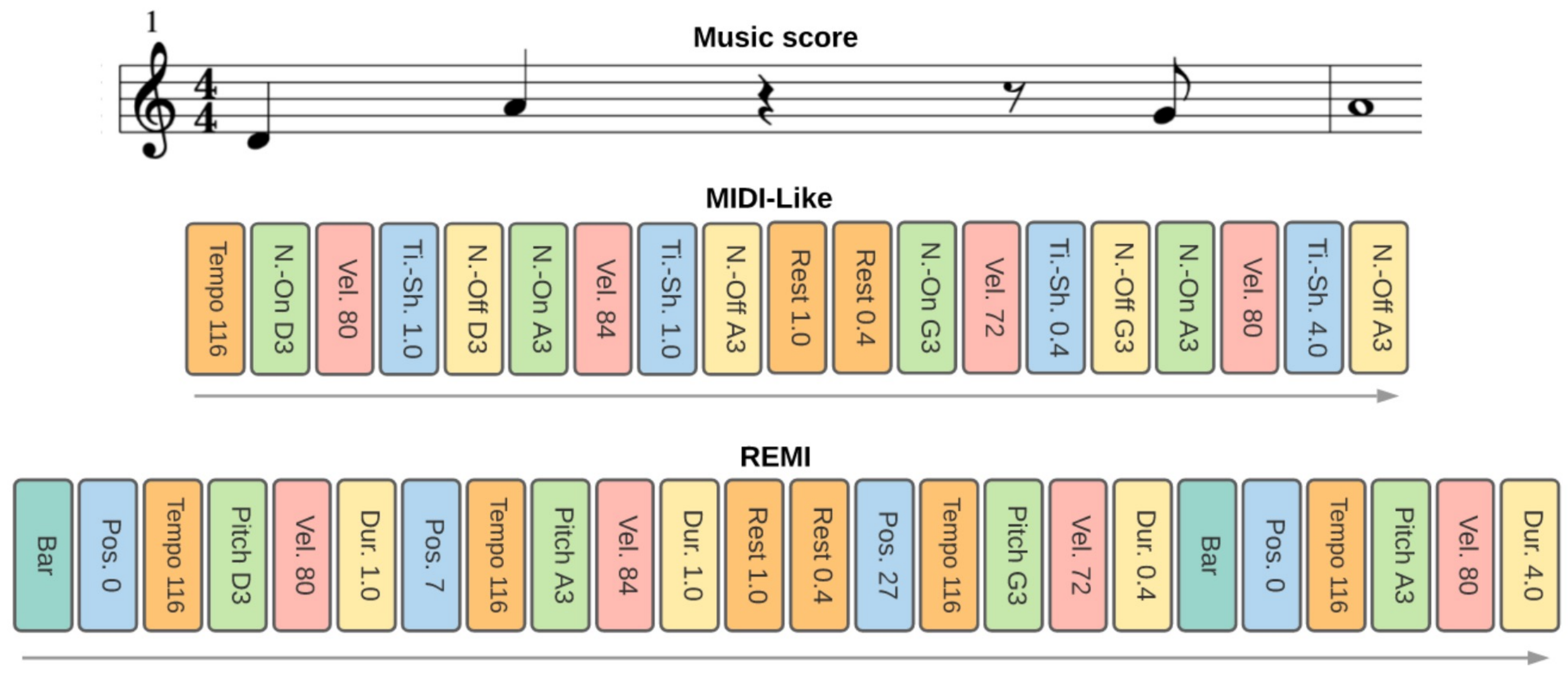
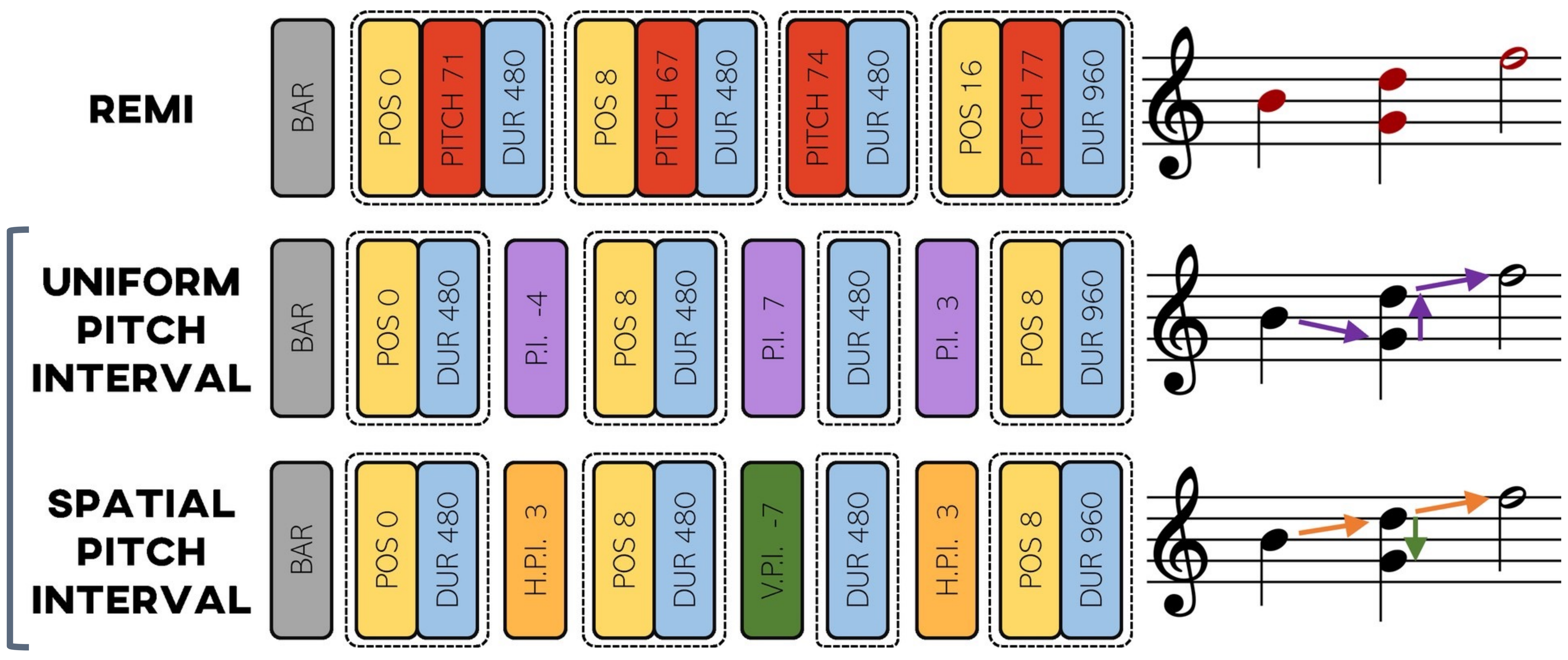


Illustration of the *MIDI-Like* and REMI tokenizations
Figure extracted from [1]

Towards a transposition-invariant *tokenization*

- Goal: helping models to grasp musical knowledge at any key without using transposition-based data augmentation
- ***Uniform Pitch Interval*** tokenization: substituting pitch tokens with interval tokens (P.I)
- ***Spatial Pitch Interval*** tokenization:
  → Vertical interval tokens (V.P.I): (descending) intervals between simultaneous notes
  → Horizontal interval tokens (H.P.I): intervals between consecutive (top) notes

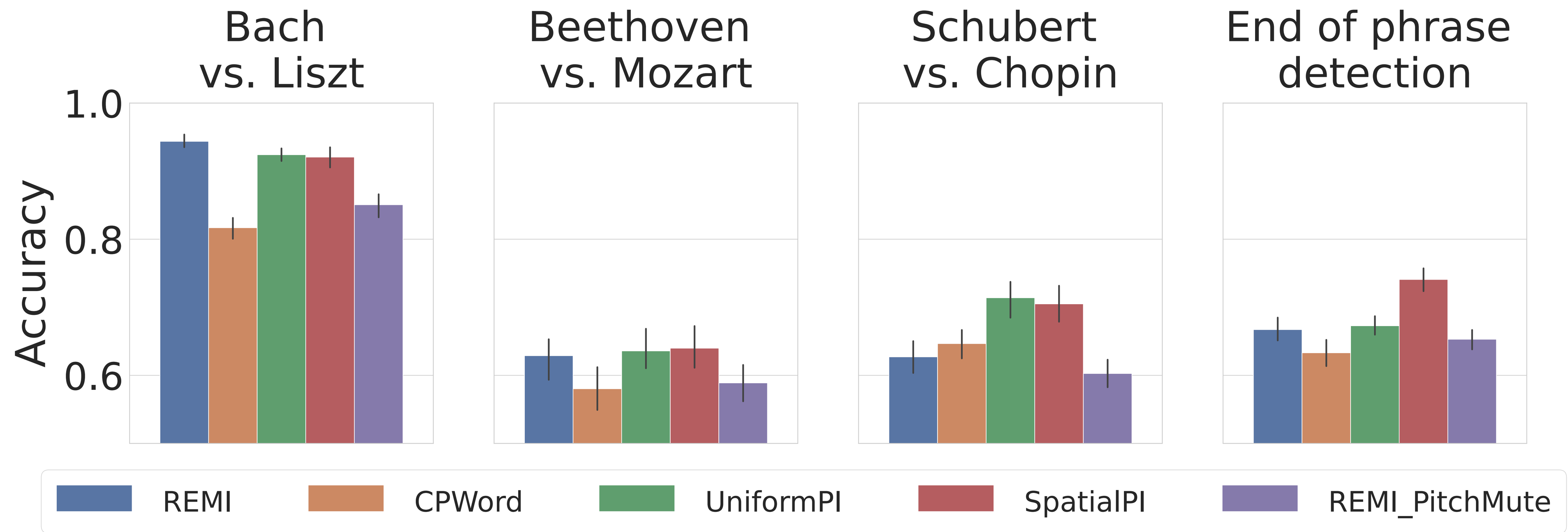


Experiments: comparing the expressiveness of various *tokenizations*

- Sequence lengths and vocabulary sizes vary significantly across tasks and tokenization strategies
- Training + evaluation of classifiers for two MIR tasks:
  → binary composer classification (*GiantMIDI-Piano* dataset)
  → end-of-phrase detection (*TAVERN* dataset)
- Classification of musical sequences as *bags-of-tokens* (TF-IDF weighted) using logistic regression models

			REMI		CP Word		Uniform P. I.		Spatial P. I.		PitchMute	
Evaluation	Nb Pieces	Dataset	Tokens	Vocab	Tokens	Vocab	Tokens	Vocab	Tokens	Vocab	Tokens	Vocab
Composer Classification *(GiantMIDI)*	740	Bach & Liszt	6.1 M	211	2.7 M	78 K	6.1 M	282	6.1 M	348	6.1 M	118
		Mozart & Beethoven	3.8 M	207	1.7 M	49 K	3.8 M	274	3.8 M	333	3.8 M	113
		Chopin & Schubert	4.9 M	210	2.1 M	59 K	4.9 M	263	4.9 M	320	4.9 M	114
End of Phrase Detection	1060	TAVERN	226 K	136	110 K	771	225 K	147	214 K	164	226 K	75

- Tokenization choices have a significant impact on the classifier performance
- Pitch Interval tokenizations perform equally or better than REMI, even in cases where absolute pitch is presumably discriminant due to the use of contrasting pitch ranges (*e.g* J.-S. Bach and F. Liszt)



- **Perspectives**
  → Experiment hybrid tokenization *e.g* interval tokens only for simultaneous notes
  → Compare tokenizations on wider tasks involving the training of *transformer* models

[1] N. Fradet et al. "MidiTok: A Python Package for MIDI File Tokenization". *ISMIR, Late-Breaking Demo*. 2021
[2] S. Oore et al. "This time with feeling: Learning expressive musical performance". *Neural Computing and Appl.* 2020
[3] Y.-S. Huang et al. "Pop music transformer: Beat-based modelling and generation of expressive pop piano compositions". *28th ACM International Conference on Multimedia.* 2020
[4] W.-Y. Hsiao et al. "Compound word transformer: Learning to compose full-song music over dynamic directed hypergraphs". *AAAI Conference on Artificial Intelligence*, 2021.