ISMIR 2017 - International Society for Music Information Retrieval Conference

A Metric for Music Notation Transcription Accuracy

Andrea Cogliati and Zhiyao Duan

<andrea.cogliati@rochester.edu> Dept. of Electrical and Computer Engineering

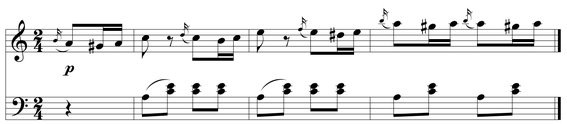
University of Rochester

Abstract

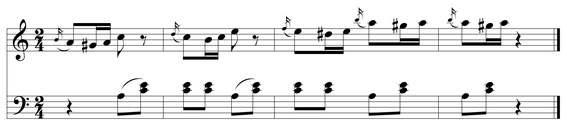
- Automatic music transcription (AMT) aims at transcribing musical performances into music notation
- Most existing AMT systems only focus on parametric transcription
- Lack of **objective** metric to evaluate music notation transcription
- The proposed edit metric counts differences between a transcription and the ground-truth music score in **twelve different musical features**
- The metric can be used to **predict human evaluations** of music notation transcription with an average R^2 of 0.564

Examples

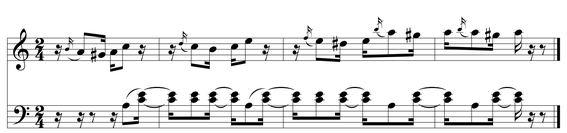
Comparison of two transcriptions of the same piece containing similar errors but with different readability:



(a) Ground truth



(b) Transcription with a wrong pickup measure

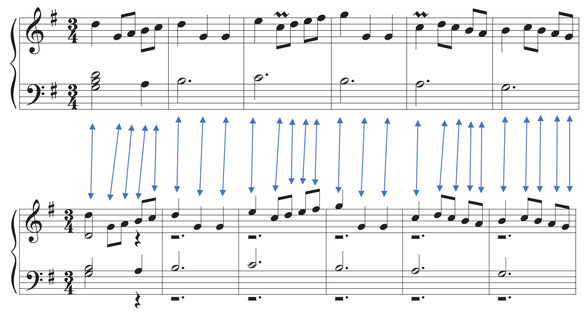


(c) Transcription off by a 16th note

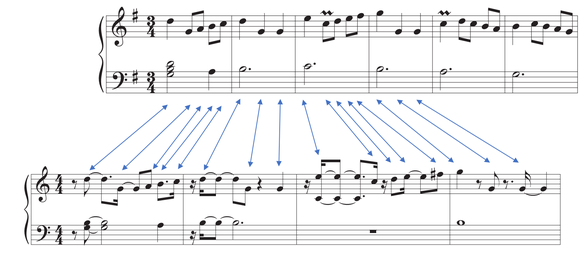
Proposed Method

- Align the transcription to the ground truth based on the pitch content only
  - Pitch content is arguably the most salient feature of a transcription
  - Invariant to meter and key mistakes
    - Increased robustness of the alignment
- Compare musical objects at aligned portions between the scores and count differences on the following features:
  - *Binary matching:* barlines, clefs, key signatures, time signatures
  - *Rests:* duration, staff assignment
  - *Notes:* spelling, duration, stem direction, staff assignment, grouping into chords
- Normalize error counts by the total number of musical objects
- Translate normalized error counts to musically relevant evaluation with a linear regression to fit human ratings
  - Human ratings of three musical aspects taken from [1]: *pitch content, rhythm notation, note positioning*
  - For each aspect, linear regression learns twelve weights, one for each normalized error count
- Human evaluators in [1] were graduate students in Music Theory
- The dataset shows a low inter-evaluator agreement
  - Average standard deviation for (score range is 1 to 10)
    - Pitch notation: 1.64
    - Rhythm notation: 1.52
    - Note positioning: 1.84

Alignment Stage



Alignment between the ground truth (top) and a transcription (bottom) of Bach's Minuet in G. Arrows indicate aligned beats.

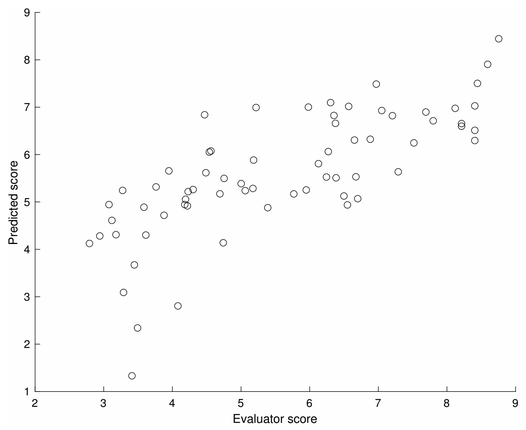


Alignment between the ground truth (top) and another transcription (bottom) of Bach's Minuet in G. Arrows indicate aligned beats.

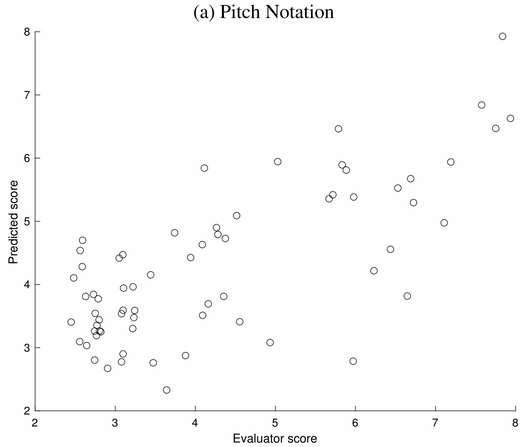
Conclusions

- Clear correlation between predicted ratings and average human ratings
  - Pitch notation R^2=0.558
  - Rhythm notation R^2=0.534
  - Note positioning R^2=0.601
- The twelve proposed error count categories capture musically relevant features of music notation transcription
- High variance between evaluator scores may reduce performance
- Full code is available at [2]

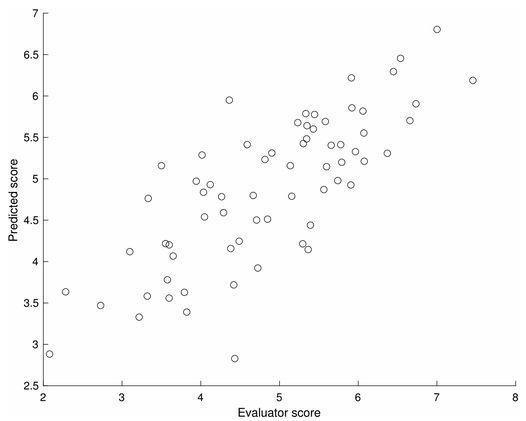
Results



(a) Pitch Notation



(b) Rhythm Notation



(c) Note Positioning

Correlation between the predicted ratings and the average human ratings.

References

[1] Andrea Cogliati, Zhiyao Duan, and David Temperley, "Transcribing human piano performances into music notation," in *Proc. International Society for Music Information Retrieval Conference (ISMIR),* 2016.

[2] http://www.ece.rochester.edu/~acogliat/repository.html