HEWS Data Set:
Human-Played Exercises & Warm-Ups Set

Amelia Murray
Jake Sandakly
Fall 2019
Music Information Retrieval
Professor: Dr. Rachel Bittner

1. INTRODUCTION

MIDI is a crucial and common part of music production as it has the power to create sounds that were previously impossible to produce. MIDI often works in tandem with VSTs (Virtual Studio Technology) to create instruments within DAWs. Musical MIDI & VSTs typically aim to match the timbre of a pre-existing real instrument or create an original timbre that does not exist.

Unfortunately, MIDI is often disregarded within the machine learning community since computer generated MIDI tends to struggle to capture the velocity that humans naturally invoke.

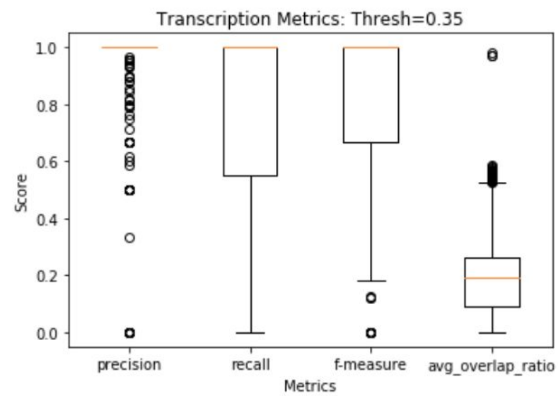
Our research seeks to expose the specific areas that current pitch detection algorithms struggle with when detecting pitch of MIDI instruments.

2. BASELINE PERFORMANCE

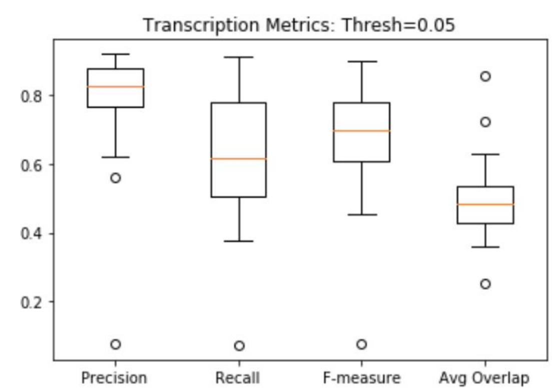
We used the piano transcription algorithm in the Python library, Madmom [1]. We tested the algorithm with the MAPS dataset [2], the LabROSA Automatic Piano Transcription (APT) dataset [3] and the Maestro dataset [4].

The following plots show the metrics for precision, recall, f-measure and average overlap ratio for the three datasets when tested with Madmom.

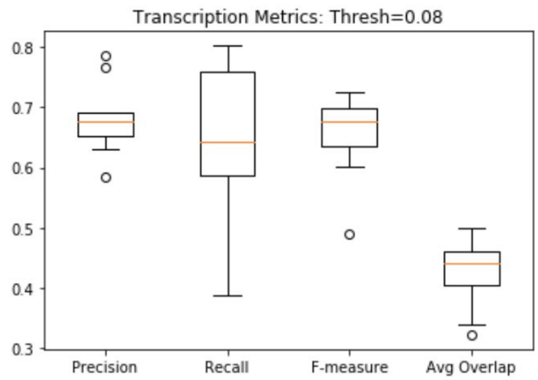
MAPS Dataset Metrics:



LabROSA APT Dataset Metrics:



Maestro Dataset Metrics:



3. ANALYSIS

Each dataset has a different composition of MIDI tracks and audio tracks.

MAPS consists of about 65 hours of Audio. It has 4 subcategories of tracks: isolated notes and monophonic excerpts, chords with random pitch notes, usual chords from Western music, and pieces of piano music.

LabROSA APT collected MIDI data from [5]. The creators made piano recordings from a subset of the MIDI files using a Yamaha Disklavier piano.

Maestro contains over 200 hours of paired audio and MIDI recordings from ten years of International Piano-e-Competition. The Repertoire is mostly classical, including composers from the 17th to early 20th century [6].

The metrics for MAPS are unusually high because MAPS was used to train Madmom. MAPS consists of many small short audio files which Madmom had difficulty processing and guessing correctly. The algorithm struggled to detect this dataset's low notes and notes/chords with overtones. Some of the files were single notes. This increased the accuracy because the algorithm was able to guess many of the tracks at 100% accuracy and precision due to the singularity of the file. In other cases where there were low notes, the algorithm struggled with transcription as it would score 0 since it would miss 1/1 low notes in certain audio files where there was one single note.

LabROSA APT was also used to train Madmom, but it was the most difficult for Madmom to correctly estimate, according to [7]. The LabROSA dataset is small and has songs as opposed to smaller audio files like MAPS.

Maestro performed well as it performed as well as LabROSA when Maestro had a higher threshold of 0.8 and LabROSA had a threshold of 0.5. Maestro had similar weaknesses to the other two datasets but specifically struggled with trills, multi-octave chords, low notes, and low note clusters.

4. CREATION OF HEWS DATASET

We hypothesized that by creating a dataset composed of MIDI files played by humans, we could further determine the strengths and weaknesses of Madmom's piano pitch transcription algorithm. The HEWS acronym stands for "Human-played Exercises & Warm-up data Set." We used common warm-ups so that we could target pain points of the algorithm more readily. The idea was to emulate warm ups and exercises that a real student playing piano might learn. Each MIDI file has 4 corresponding wave files that have matching audio of Bosendorfer Grand Piano, Yamaha Grand Piano, Steinway Grand Piano and Classic Electric Piano. The exercises include arpeggios, chords, low notes, melodies with chords, octave noodles (short melodies spanning multiple octaves), octave practice, overtone practice, random chord clusters, single random notes, various scales, trills played solo, trills with chords, and short songs.

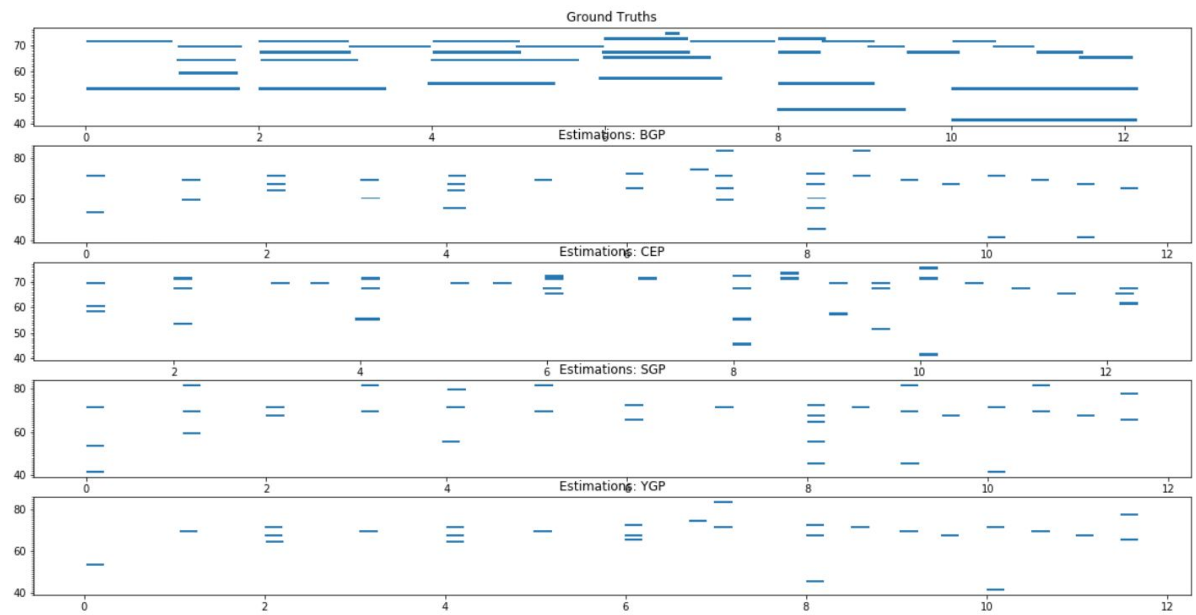
The MIDI was played on an M-Audio Keystation 88 by Murray and Sandakly and was recorded into Logic Pro. The

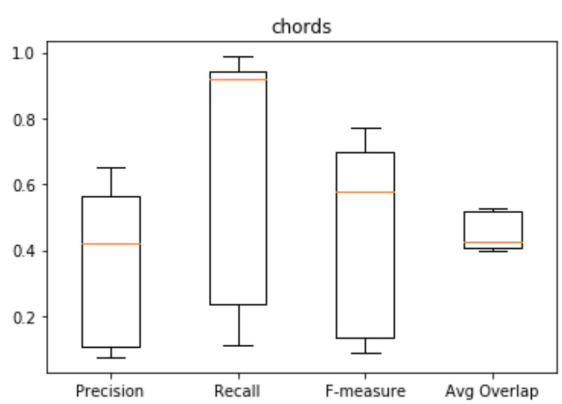
5. STATISTICS

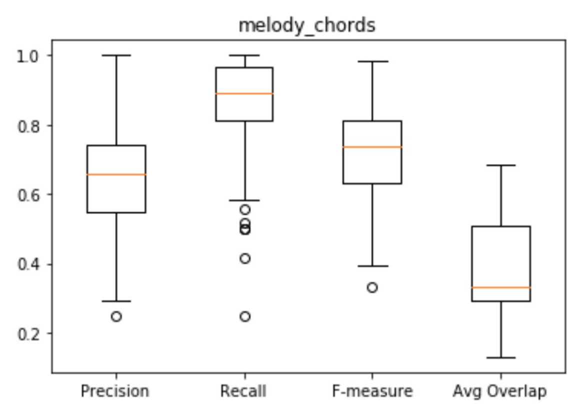
The statistics are broken down by exercise to illustrate in greater detail the specific pitfalls of the Madmom piano pitch transcription algorithm. Following the statistics is the breakdown of the statstics by instrument.

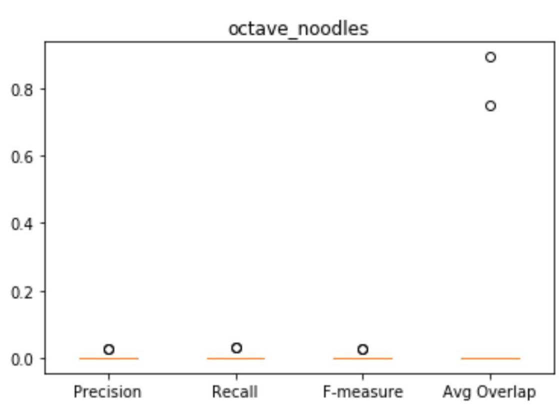
A sample transcription for ‘songs #12’:

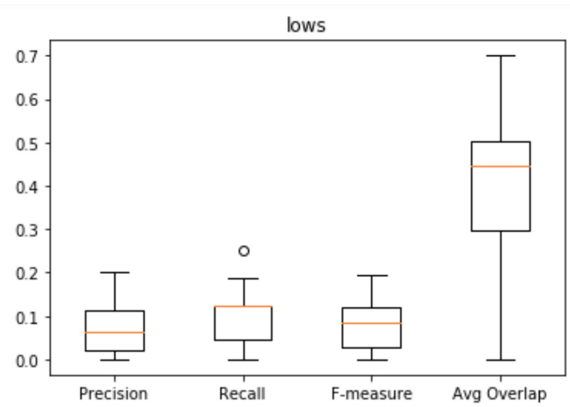
Songs 12

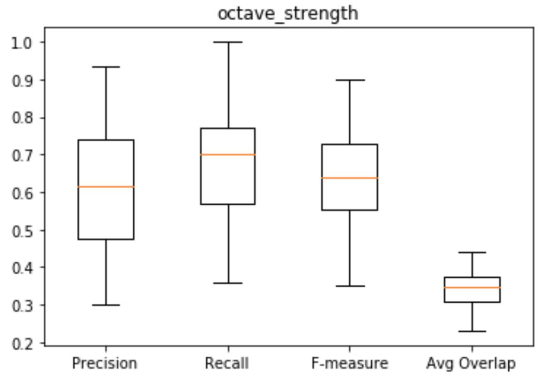


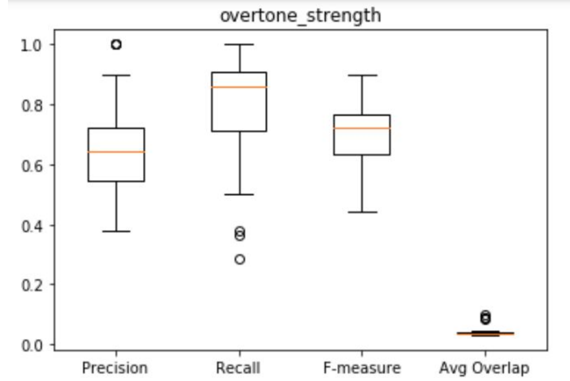


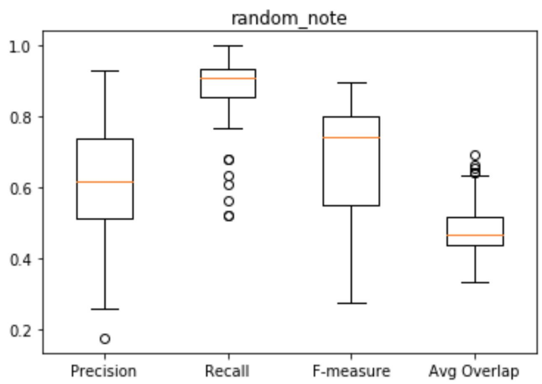


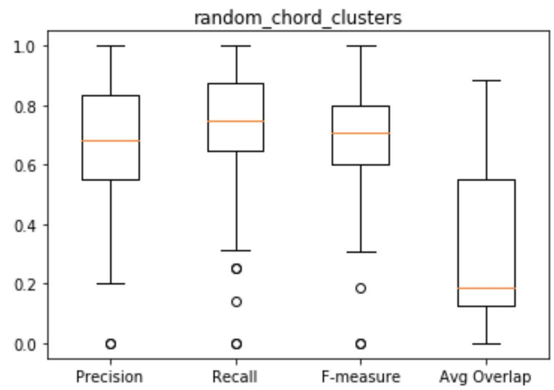


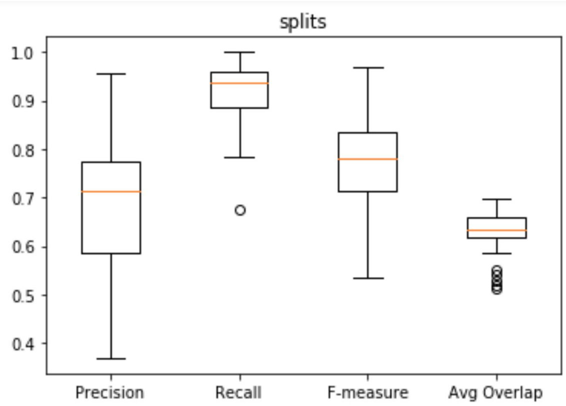


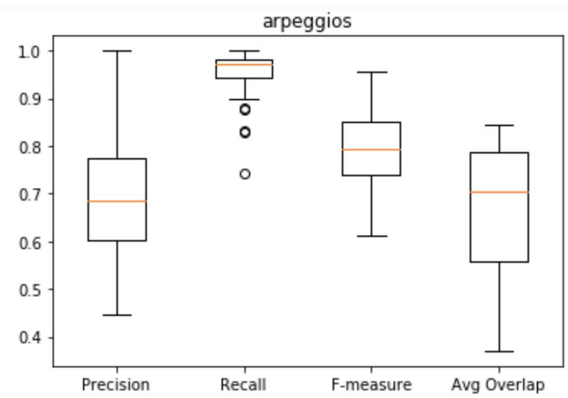

octave_strength
Precision
Recall
F-measure
Avg Overlap

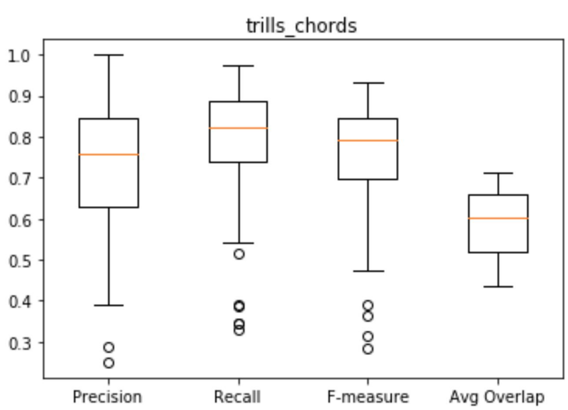

overtone_strength
Precision
Recall
F-measure
Avg Overlap

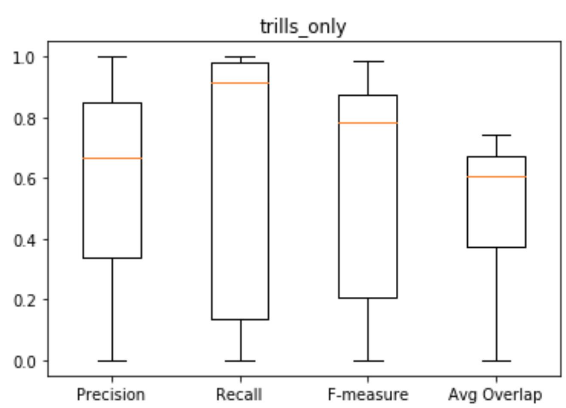

random_note
Precision
Recall
F-measure
Avg Overlap

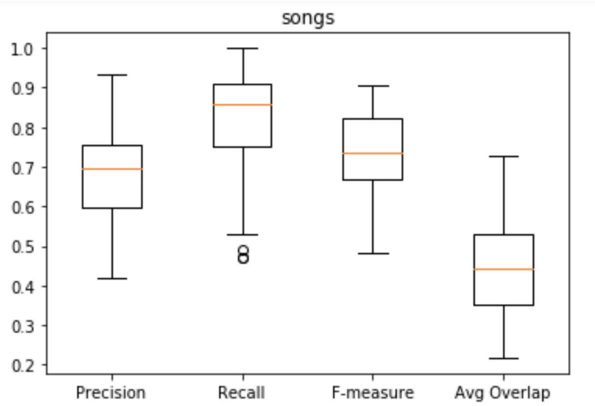

random_chord_clusters
Precision
Recall
F-measure
Avg Overlap

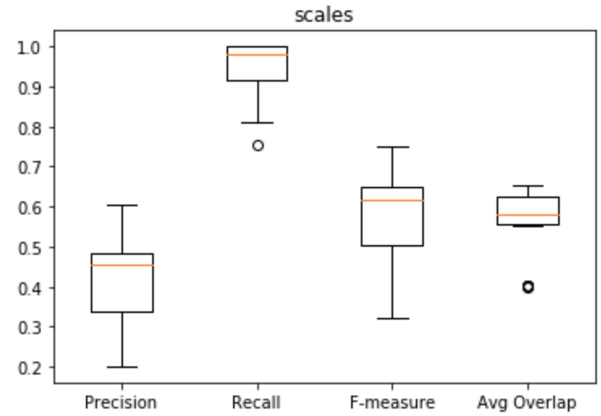

splits
Precision
Recall
F-measure
Avg Overlap

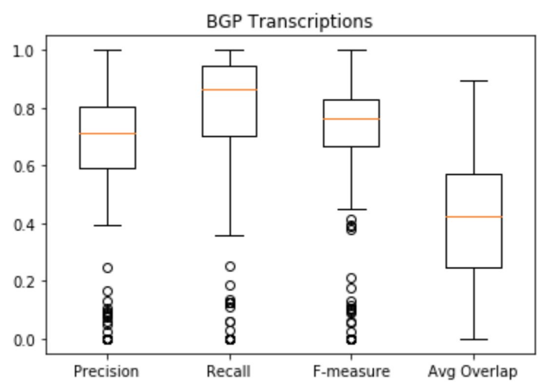

arpeggios
Precision
Recall
F-measure
Avg Overlap

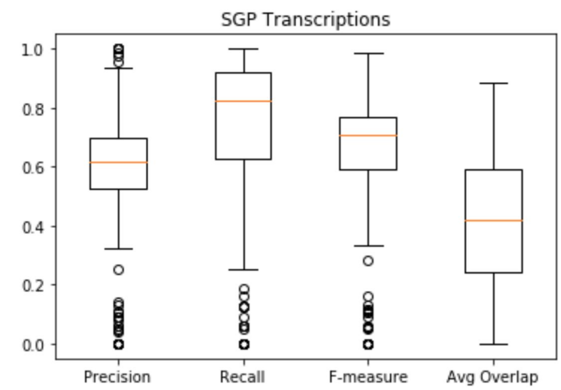

trills_chords
Precision
Recall
F-measure
Avg Overlap

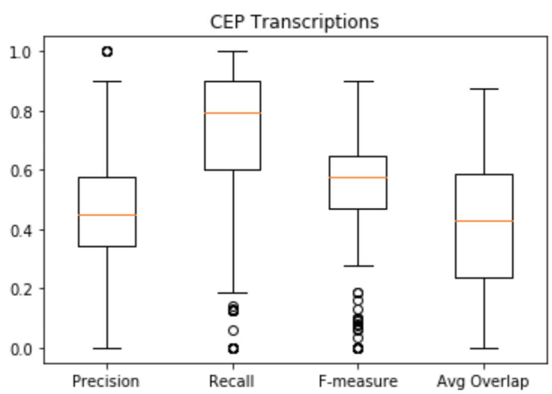

trills_only
Precision
Recall
F-measure
Avg Overlap

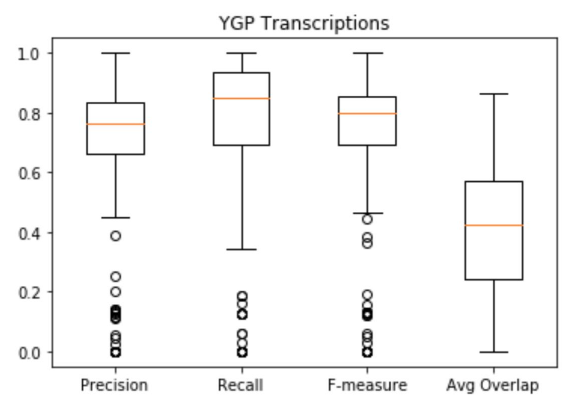

songs
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
Precision
Recall
F-measure
Avg Overlap


scales
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
Precision
Recall
F-measure
Avg Overlap


BGP Transcriptions
1.0
0.8
0.6
0.4
0.2
0.0
Precision
Recall
F-measure
Avg Overlap


SGP Transcriptions
1.0
0.8
0.6
0.4
0.2
0.0
Precision
Recall
F-measure
Avg Overlap


CEP Transcriptions
1.0
0.8
0.6
0.4
0.2
0.0
Precision
Recall
F-measure
Avg Overlap


YGP Transcriptions
1.0
0.8
0.6
0.4
0.2
0.0
Precision
Recall
F-measure
Avg Overlap

6. DATASET PERFORMANCE

The dataset performed well on any exercises that had majority mid and upper range exercises, but it did very poorly on any exercises with low notes. The ‘octave noodles’ category scored exceptionally low, meaning that Madmom is not adept at detecting and transcribing low notes.

The threshold was set at .25, which is comparable to MAPS as that dataset had a .35 threshold. HEWS is most similar to MAPS in terms of content as both datasets are composed of many small MIDI files. However, HEWS is also similar to Maestro because both datasets contain data from real humans playing, so the velocities are different. This human velocity is troubling for the Madmom algorithm because of the datasets it was trained on. Since all of the MAPS and LabROSA MIDI velocities are computer generated, and since those are two of the datasets used to train Madmom, the ability for Madmom to detect notes with a lower velocity is much less likely.

The dataset shows that the algorithm does well with songs and melodies, probably due to the fact that the low notes are less prominent. The octave strength and overtone strength exercises demonstrate that the algorithm does great with overtones but struggles with detecting notes in different octaves.

7. CONCLUSION

The HEWS dataset was created as a tool to expose the weaknesses within existing MIDI pitch tracking algorithms and learn about possible opportunities for improvement in future MIDI pitch tracking algorithms. Through this research project, we sought to create a set of human played MIDI tracks that could be used by other members of the MIR field in training similar algorithms. By using this dataset to train a MIDI pitch tracking algorithm, future algorithms could avoid having the same weaknesses that Madmom does. It is our hope that when used in tandem with other similar MIDI datasets, more robust models could be created to better detect pitch.

REFERENCES

[1] Block, Sebastian and Korzeniowski, Filip and Schluter, Jan and Krebs, Florian and Widmer, Gerhard., “Madmom: a new Python Audio and Music Signal Processing Library” in *Proceedings of the 24th ACM International Conference on Multimedia*. October 2016.

[2] V. Emiya, R. Badeu and B. David, *Multipitch estimation of piano sounds using a new probabilistic spectral smoothness principle,* IEEE Transactions on Audio, Speech and Language Processing (to be published);

[3] Poliner, Graham. *LabROSA Automatic Piano Transcription,* under The National Science Foundation, LabROSA, 2008.

[4] Curtis Hawthorne, Andriy Stasyuk, Adam Roberts, Ian Simon, Cheng-Zhi Anna Huang, Sander Dieleman, Erich Elsen, Jesse Engel, and Douglas Eck. *Enabling Factorized Piano Music Modeling and Generation with the MAESTRO Dataset*. In International Conference on Learning Representations, 2019.

[5] *http://www.piano-midi.de/*

[6] *https://magenta.tensorflow.org/datasets/maestro#dataset*

[7] Böck, Sebastian and Schedl, Markus, *Polyphonic Piano Note Transcription with Recurrent Neural Networks*, Proceedings of

the 37th International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2012.