The potential for automatic assessment of trumpet tone quality

Trevor Knight
TrevorKnight@gmail.com

Finn Upham
Finn.Upham@gmail.com

Ichiro Fujinaga
Ich@music.mcgill.ca

CIRMMT Centre for Interdisciplinary Research in Music Media and Technology

McGill

1. Examining trumpet tone quality

The goal of this study was to examine the possibility of training a classifier to determine good from bad tone quality. It is a first step to determining if computers could give automatic feedback to student musicians regarding their tone quality. It also serves to create a labelled data set that could be used to research audio features indicative of tone quality.

The recording setup

2. Recording and rating notes

Recordings:

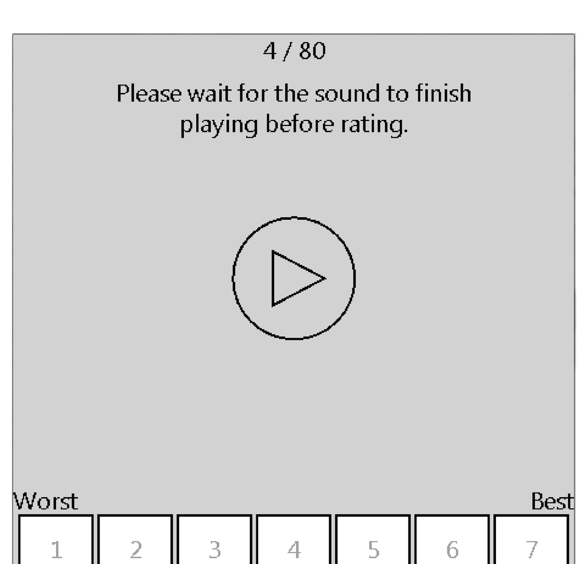
- Four trumpet players
  - with range of experience levels
  - each playing with both personal and control trumpets and mouthpieces
- 12 pitches
- Half notes (minims) at 60 bpm
- All recorded at three dynamic levels (*p*, *mf*, *ff*)

Labellings:

Averaged from the subjective ratings of 5 brass players on a 7-point Likert scale

Software:

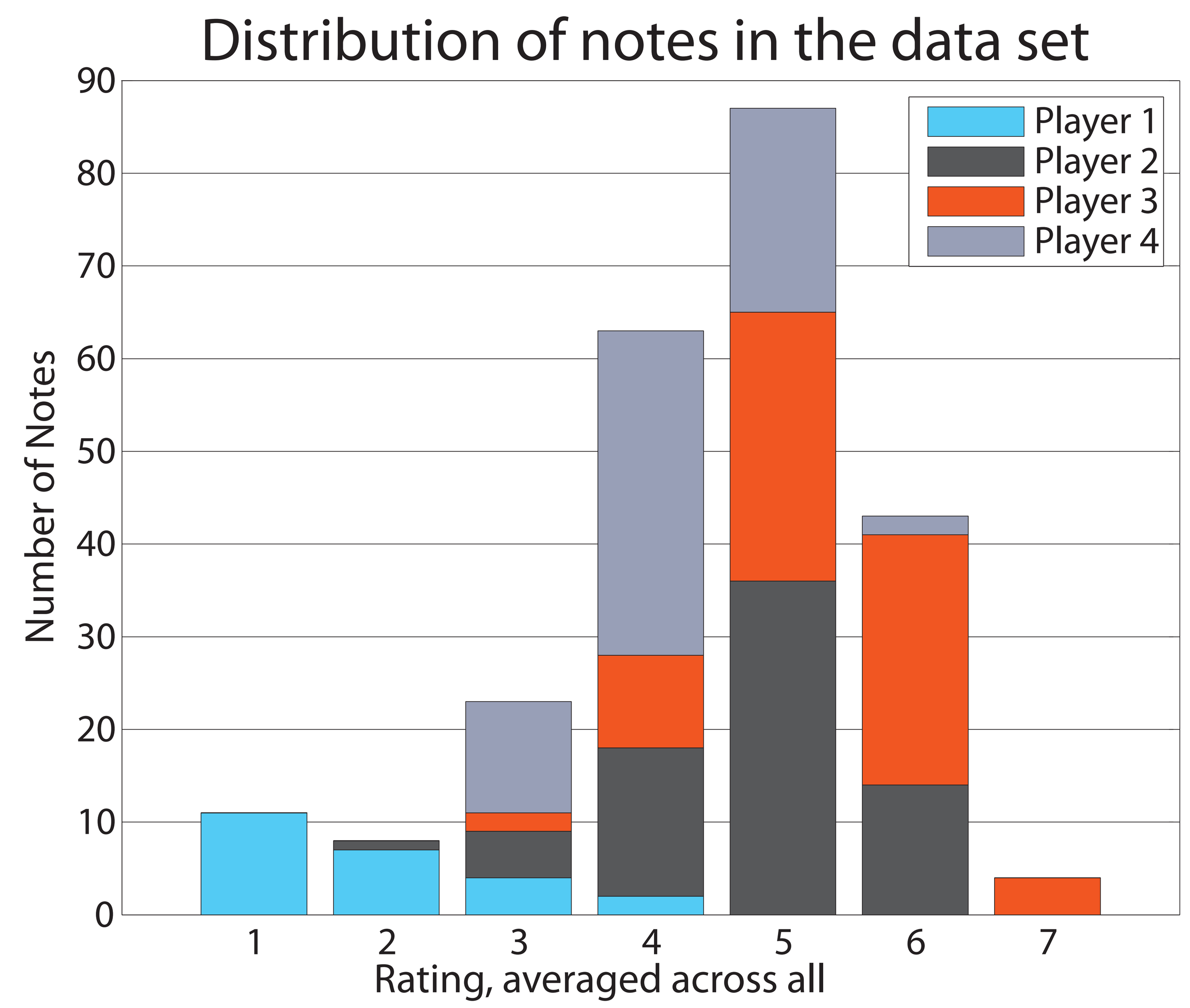
- jAudio for feature extraction
- Autonomous Classification Engine (ACE) 2.0



The rating interface

3. Data set

The notes in the data set and the contribution of each performer. Each note is only represented by its average rating. There is a total of 239 notes.



4. Subsets tested

In order to test the efficacy of a classifier, we trained and tested it using several subsets of the data.

The subsets were then tested using both five-fold cross-validation and leave-one-performer-out methods. That is to say, by training using the data from three of the four players and testing using the data from the unseen player.

Two groups: "good" and "bad", using:

a. Just the extremes (less than 2.5 and greater than 5.5)
b. More inclusive (less than 3.5 and greater than 5.5)
c. Splitting on the median value (5.4)

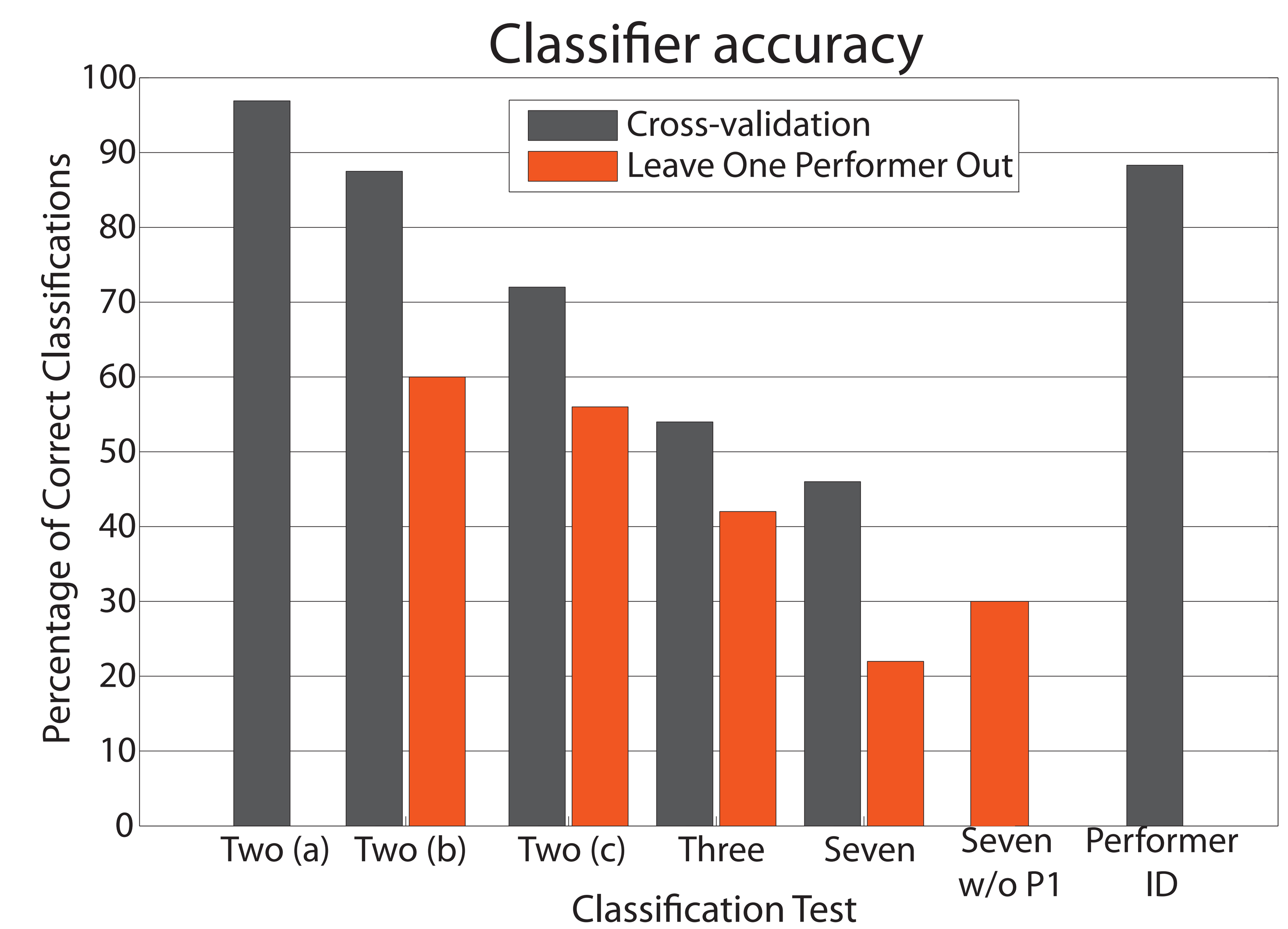
Three groups: "good," "medium," and "bad"
Seven groups: using the averaged ratings

Because Player 1 is such a large part of the ratings with an average less than 2.5, we could not test the leave-player-out method with the Two (a) class distribution and also tried testing the seven classes with and without Player 1.

Performer ID: One last test was to check for accuracy with notes labelled just with the player number.

5. Classification results

The graph below shows the classifiers accuracy with two, three, and seven levels of tone quality for both the cross-validation and leave-performer-out tests.



6. Promising potential

With just two classes, the classifier shows above 72% accuracy, up to 97% when using data from the two extremes (subset Two a).

As expected, increasing the number of classes to three or seven decreased classifier accuracy, but the accuracy still remains well above the levels expected from chance.

Classifier accuracy dropped markedly when using the leave-player-out test, suggesting the classifier may be benefiting from some performer-specific characteristics.

Further work is being conducted to verify these results with a greater number of performers and to determine the most meaningful features for tone quality.

Acknowledgements

This research would have been impossible without financial support from CIRMMT and the amazing help and advice of the CIRMMT staff: Harold Kilianski, Yves Méthot, and Julien Boissinot.