A dataset and baseline for automated assessment of timbre quality in trumpet sound

Alberto Acquilino, Ninad Puranik, Ichiro Fujinaga, Gary Scavone

Schulich School of Music, McGill University, Montreal, Québec
Centre for Interdisciplinary Research in Music Media and Technology
alberto.acquilino@mail.mcgill.ca, ninad.puranik@mail.mcgill.ca

The relationship between trumpet timbre and sound efficiency for brass pedagogues

- The musician's efficiency in trumpet playing is reflected in the timbre of the produced sound. Excessive muscle tension and inefficiencies in playing can alter the timbre quality.
- Efficient sounds are described as rich and round, while less efficient tones sound strained and shrill.

Main research questions

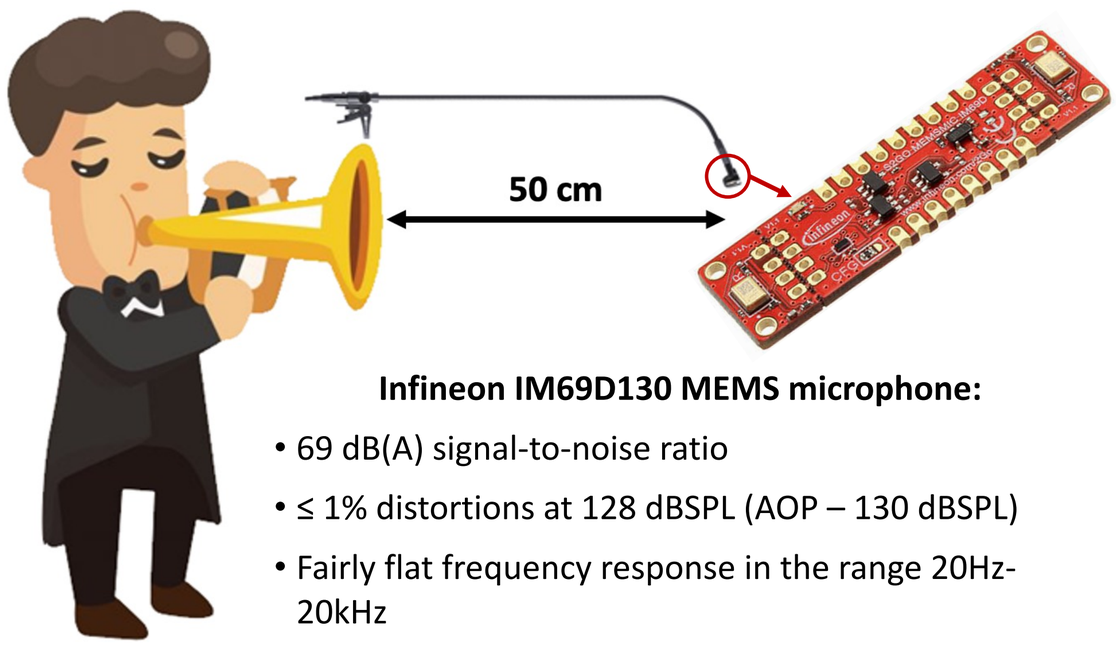
i. Can brass players and teachers consistently rate trumpet sound efficiency based solely on auditory information, as claimed in the pedagogic literature?

ii. If the above hypothesis is verified, can a machine learning model be used to assess sound production efficiency in trumpet playing with accuracy comparable to human experts?

Dataset & Methodology

1. Dataset recording

- 110 trumpet players from early beginners to international soloists
- Same microphone at the same position and distance from the trumpet bell
- Players were instructed to play isolated tones:
  - Over a specified chromatic scale:
  - At three different dynamics: *p* *mf* *f*



Infineon IM69D130 MEMS microphone:

- 69 dB(A) signal-to-noise ratio
- ≤ 1% distortions at 128 dBSPL (AOP – 130 dBSPL)
- Fairly flat frequency response in the range 20Hz-20kHz

2. Dataset preparation

- The dataset was segmented into discrete trumpet tones using the *pyin* vamp plugin yielding to over 19,000 tones.
- A representative subset of 1,481 notes at different pitch, dynamics and sound efficiency levels was selected.

3. Labels assignment

- A web interface was developed for blind grading of the trumpet tones
- The selected subset was so labeled by 8 expert players and teachers based on perceived efficiency of sound production
- The labeled dataset is now accessible on Zenodo:

Analysis & Results

- The inter-rater reliability was assessed using a subset of 100 tones graded by all experts
- A Random Forest classifier was trained using the mode of the ratings. Its accuracy and variability were assessed in comparison to the variability in human graders
- Analysis of the top discriminatory features indicates that the change in the spectrum across time is a crucial factor in the perception of trumpet sound production efficiency

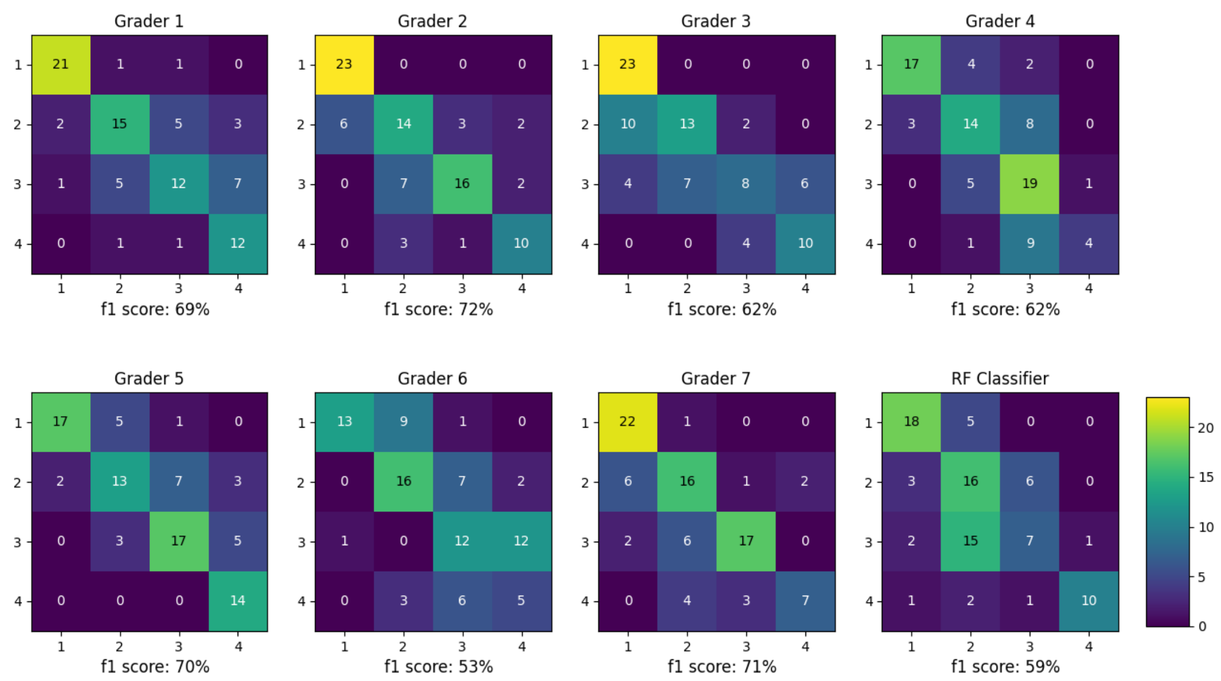


Figure 1: Confusion matrices with the predicted labels of each grader and of the trained RF classifier (horizontal axis) vs the true label as the mode of the assigned grade (vertical axis) and the corresponding f1 scores

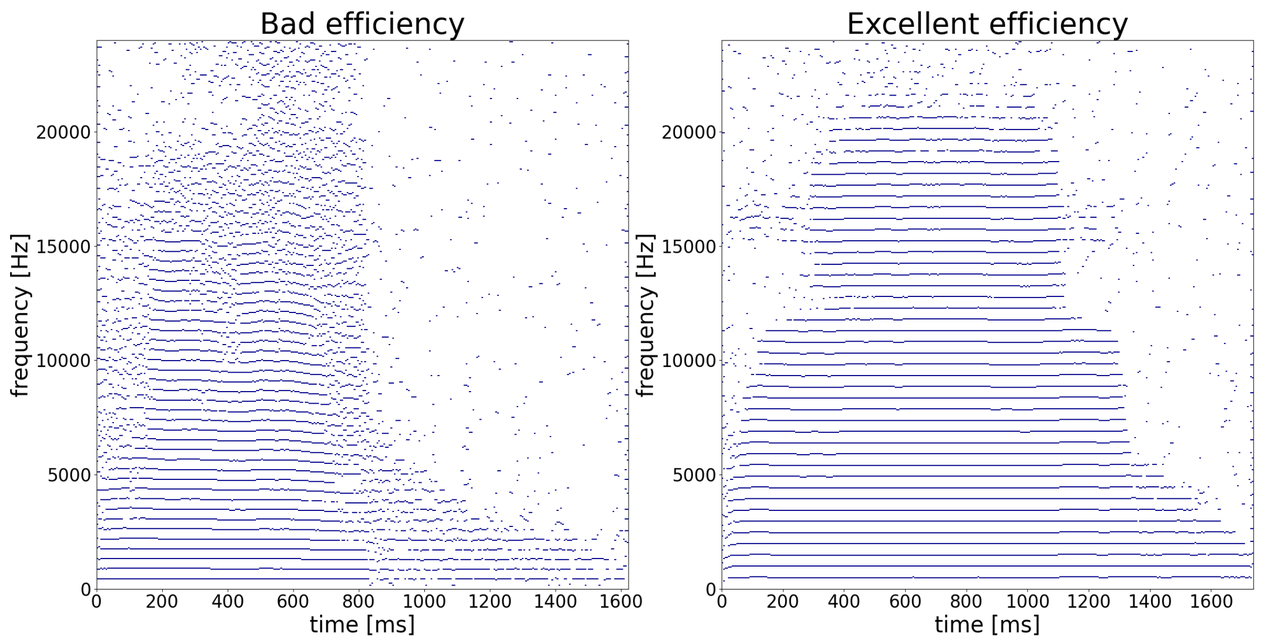


Figure 2: Visualization of the temporal evolution of spectral peaks for trumpet sounds rated as bad efficiency (left) and excellent efficiency (right)

Conclusions

- The Random Forest Classifier models trained using extracted audio features showed accuracy scores comparable to human experts.
- Features based on spectral complexity were observed to be highly important for this task of trumpet timbre discrimination.
- A representation based on the evolution of spectral peaks was developed to visualize trumpet sound efficiency

Acknowledgments: Dr. Richard H. Tomlinson Doctoral Fellowship, Fonds de Recherche du Québec

Social Sciences and Humanities Research Council of Canada | Conseil de recherches en sciences humaines du Canada | Canada

Presented at the 24th ISMIR conference, Milan, Italy, November 8th, 2023

NSERC CRSNG