Using Machine Learning Techniques to Improve Computer-Based Assessment of Musical Performances

Sumanth Sura
Advisor: Mr. Nicholas Medeiros

Engineering Need

Recently, musicians have been using online algorithms to analyze their playing. However, existing technologies for analyzing musical performance do not account for more human factors of playing, such as tone or musical technique.

Engineering Goal

The goal of the project is to develop a computer algorithm that correctly recognizes the categorizes the tone category (e.g. "feathery tone"), the extremity of that tone as a percentage accuracy (e.g. "75% match") based on how prominent the computer predicts the tone is and returns to the user when the tone is occurring and what change needs to occur in the playing to improve.

Methodology

METHODOLOGY GOAL: ANALYZE TONE + TECHNIQUE

1. Recorded original audio (violin, recorder) on iPhone in Voice Memos
2. Used Python to generate FFTs and spectrograms from the recorded data
3. Manually observe the qualifications the audio must contain
4. Manually observe the qualifications the audio must contain
5. Search for trends in the data by finding patterns in loud frequencies
6. Cross-check the trends with the expected qualities of the sound
7. Return feedback to user about accuracy and timestamps of mistakes

Background

Recently, there have emerged technologies used for analyzing musicians' performance, by looking at their note accuracy and rhythm.

The use of these technologies has increased due to COVID-19, which brought limited access to travel and fewer available music teachers.

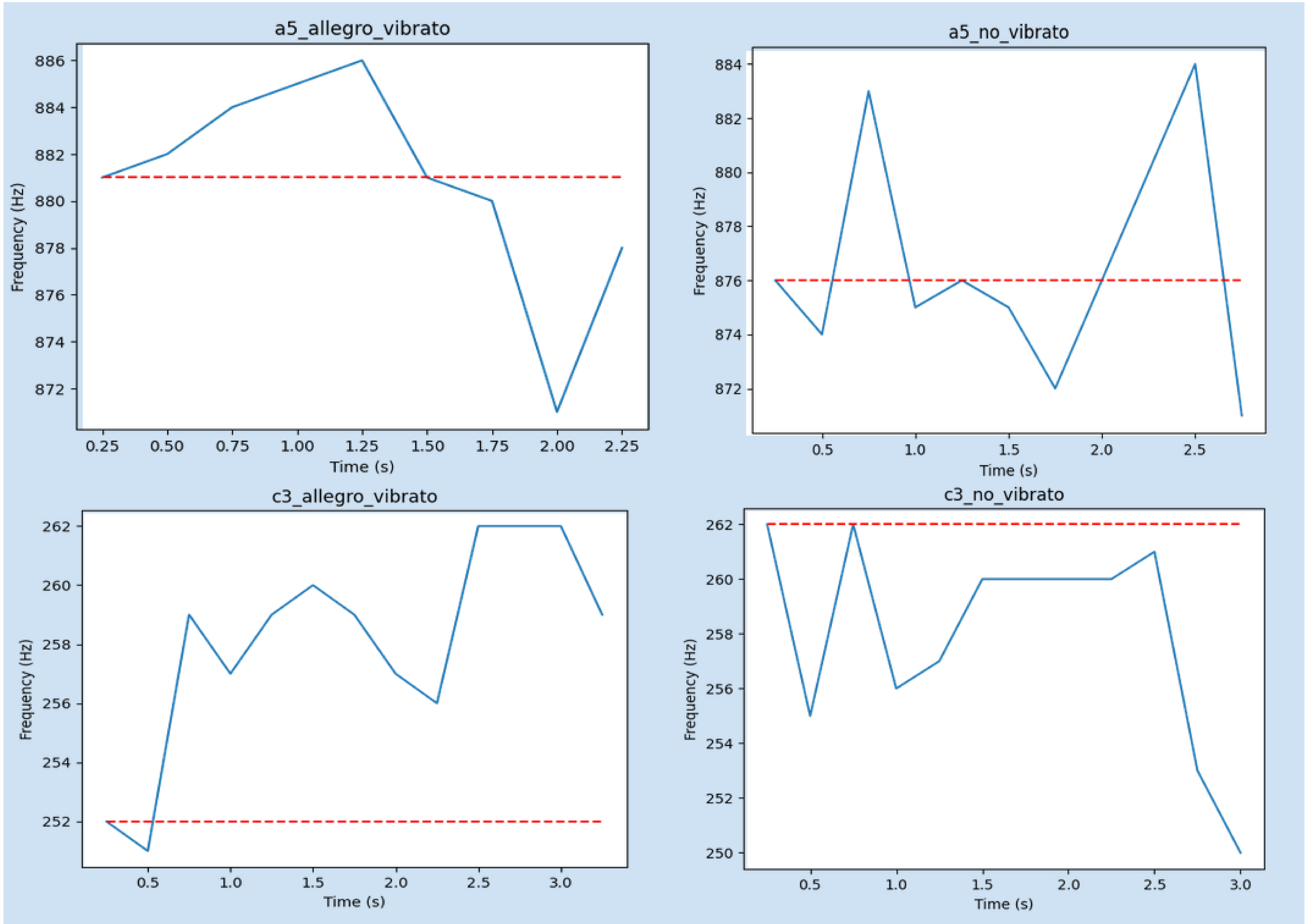
However, such technologies have limited capabilities at recognizing tone and technique.

Therefore, there exists a need for a more accurate computer algorithm for musicians.

Main Takeaway

Analyzing the trends in certain collections of frequencies simultaneously throughout time provides accurate feedback on musical technique. Graphs containing more crossing-points reveals a significant difference in recordings with better tone.

Results



Figs. 1 & 2 – Average highest frequency plots of an A5 on a violin with (left) and without vibrato (right)
Figs. 3 & 4 – Average highest frequency plots of a C4 on a violin with (left) and without vibrato (right)

The plots of the average highest frequencies throughout the recording will fluctuate greatly around the starting frequency if the recording contains vibrato. Therefore, the data has a greater standard deviation.

Analysis

Overall model achieved ~81% accuracy
Recordings without vibrato contained significant differences:

I. Had a greater number of crossing points
II. Contained higher standard deviation
III. Area under the curve with respect to the primary frequency was greater

This model suggests that computers can be trained to provide tailored feeback on tone quality and other musical techniques. It may be possible to expand the model to different instruments as well.

Conclusion

Analyzing multiple different ranges of frequencies over time, as shown in the graphs above, was proven to be successful since these overtones provide information on the timbre of the sound.

Researchers can build off this project by developing image analysis tools on frequency data from a given recording (by sampling chunks of a reasonable size of ~0.15 - 0.25s each), to determine the shape of the resulting frequency-amplitude time graph (a 2D spectrogram). This data can be cross-checked with the magnitude of the integral (the overall fluctuation) and number of crossing points.

Relevance

The rise of COVID-19: Musicians have had limited access to human teachers since the arrival of the pandemic. Teachers have also faced increased burnout.
Economic impact: The success of this project would lead to a decrease in spending for teachers. Pursuing music will become more economically viable. This would result in a greater population of musicians, increasing the overall profit of the music industry.
Efficiency.: An algorithm produces results faster than humans. The project would also be comparable to several different master recordings to give a more well-rounded assessment of performance. This mitigates human bias while giving accurate data.

Future Research

One application of this research is to evaluate its impact in the real world.

Investigate long-term effectiveness of the model
I. Collect "improvement" data from people using existing technologies
II. Repeat the process for those using tone algorithm and for those w/ human teacher
III. Compare data. Tells us how close model represents a human.
IV. Check how long it takes each group to reach a certain proficiency level

Criteria

The algorithm must accurately detect tone/technique from recordings with ≥ 90% accuracy. - **8**	The algorithm must be cheap to produce and not take up lots of RAM/CPU. - **10**	The user can input multiple recordings at once.- **10**	Lvl. 1 (Priority)
The algorithm must calculate significance of the tone and technique results.- **9**	The UI is easy to understand - **5**	The algorithm has customizable levels of difficulty for musicians with different skill level. - **8**	Lvl. 2 (Important)
The algorithm must output time of mistakes and such feedback - **7**	The algorithm weights the different elements it has analyzed - **7**	The runtime for analyzing a single input is ≤ 20 seconds - **10**	Lvl. 3 (Nice to have)

References

[1] Giraldo, S., Ortiga, A., Perez, A., Ramirez, R., Waddell, G., & Williamon, A. (n.d.). Automatic assessment of violin performance using Dynamic Time Warping Classification. IEEE Xplore. Retrieved September 27, 2022, from https://ieeexplore.ieee.org/document/8404556

[2]Senevirathna, E. D. N. W., & Jayaratne, L. (2015, November 26). Audio Music Monitoring: Analyzing current techniques for song recognition and identification - GSTF Journal on Computing (JOC). SpringerLink. Retrieved October 17, 2022, from https://link.springer.com/article/10.7603/s40601-014-0015-7

[3] Pinquier, J., Rouas, J.-L., & André-Obrecht, R. (2002). Robust speech / music classification in Audio Documents. 7th International Conference on Spoken Language Processing (ICSLP 2002). https://doi.org/10.21437/icslp.2002-551