A Signal Processing Model and Feedback System For Stringed Instrument Musicians

Introduction

What is Collé?
- Collé is an advanced bow technique that enables the violinist to play more expressively
- Collé consists of the violinist “biting” the string (i.e. pinching the bow with the index finger) then immediately “releasing” (lifting the bow off the string)

Problem
- The sound characteristics of good collé are very subtle: there are many details to listen for
- Of the 25+ named violin techniques, collé is rarely studied as it is not found in music notation. However, the precise bow control and psychomotor coordination required in collé also are the building blocks for all of the accented bowing strokes

Design Goal:
- My project goal was to develop a signal processing system which can analyze a user’s bowing technique through sounds collected by microphone and give dynamically generated feedback for effective practice. The scope of the project was limited to a specific subset of bowing techniques: the collé stroke

Methods

My approach was to record the collé of violinists of varying expertise, then to analyze the differences to determine critical features and metrics, then finally to implement these metrics in software

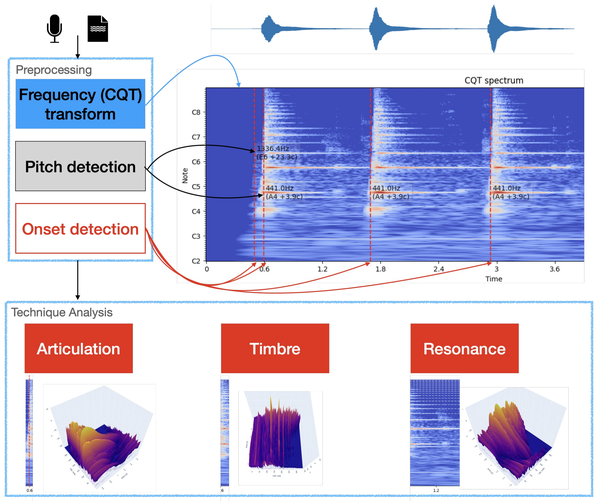
Phase 1: Data Collection
- Violinists with a wide range of experiences were recorded playing Collé, including
  - Beginners with 1-3 years of experience
  - Youth orchestra violinists with 6-9 years of experience
  - Violin teachers with 15-30+ years of experience

Phase 2: Data Analysis and critical feature extraction
- Traditional pedagogy indicates three critical features of collé:
  - **Articulation** – the rapid onset “bite” and release of the bow from the string
  - **Tone Color** – the relative warmth vs brightness during the bowed portion of the sound
  - **Resonance** – the duration of audible sound after the bow is released from the string
- A wide variety of common signal processing methods using the python librosa toolkit were applied to the recordings to try to measure each critical feature
  - Recordings of experienced teachers were used as references against which others were compared
  - In some cases, novel variations of common signal processing methods were developed to better isolate and measure the unique features of Collé

Phase 3:
- Detailed spectrogram visualizations and metrics determined from Phase 2 were implemented in python using the librosa toolkit for signal processing and the plotly toolkit for interactive 3D visualization

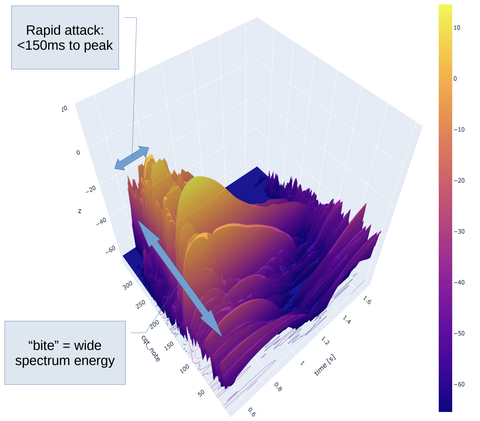
Design

The feedback system was designed as a cascade of functions to handle the following, as illustrated below:
- **Signal acquisition** from microphone/audio file
- **Preprocessing**
  - **Constant-Q transform** (CQT), similar to a short time Fourier transform, but in which frequencies are logarithmically stepped (12*4 frequency bins per musical octave)
  - **Onset detection** to identify start times of notes
    - This also identifies bounced or double-touch bow landings, a common beginner error in the Collé
  - **Pitch detection** on time intervals between onsets, based on autocorrelation
    - detected pitches are used to compute partial frequencies used in downstream metrics
    - Pitch offsets in cents based on A440 tuning are also reported. This also identifies out-of-tune playing
- **Technique analysis and metrics**
  - Metrics detailed in the “Critical Features” section are computed
- **Visualization**
  - A 3D interactive plot of the CQT of the Collé is rendered using plotly
  - Additional 2D plots to illustrate metrics are also rendered

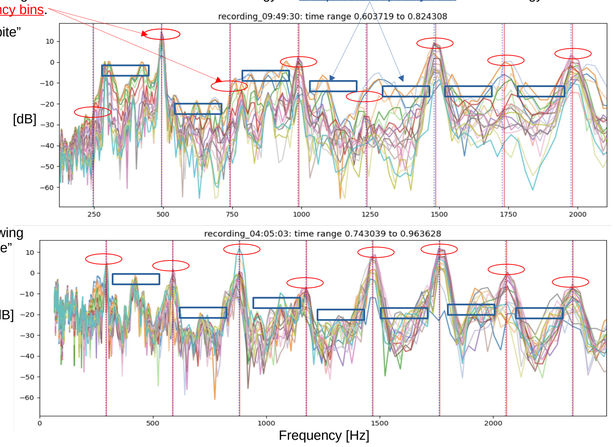


Critical features: Analysis, modeling and design of metrics

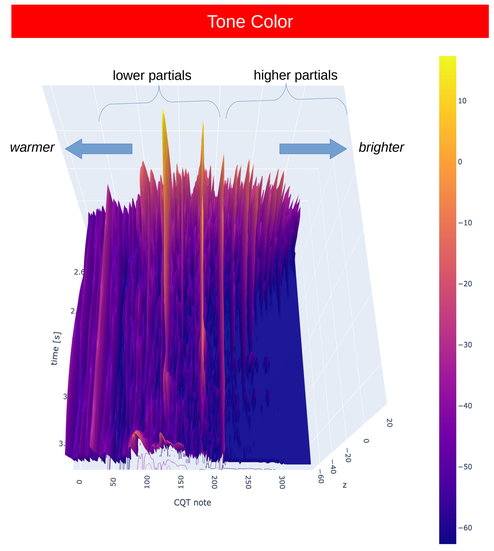
Articulation



The **Articulation** phase includes two important aspects:
1) The reference **attack time duration** is short (less than 150ms to reach peak amplitude, and 300ms to release)
2) The “**bite**” or “catch” character of the sound has no known published research on its spectral analysis.
   - By comparing the reference Collé and non-articulated bowing recordings, I observed the Collé tended to have more wide-spectrum energy at onset. This is consistent with the white noise-like “scratch” sound that it is often described as.
   - The metric I designed for this feature is the ratio of energy of non-partial frequency bins to the energy of partial frequency bins.

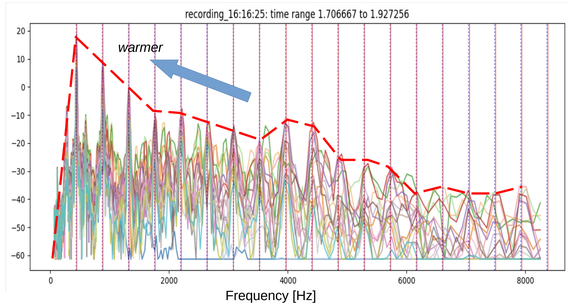


Tone Color

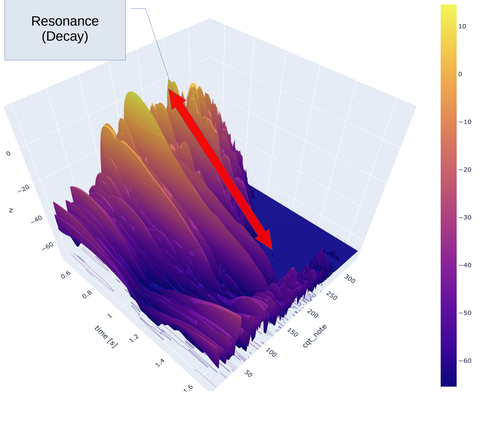


The **tone color** of the note refers to the relative brightness or warmth of the sound during the bowed segment of the Collé, after the initial articulation.
- Greater relative energy in the **higher partials** compared to the lower partial frequencies gives a psychoacoustically ***brighter*** sound. In the extreme, the sound can become unpleasantly metallic.
- Greater relative energy in the **lower partials** gives a psychoacoustically ***warmer*** sound
- Tone color can be controlled by the violinist’s choice of *sound point*: the bow-string contact distance relative to the bridge
- The metric I designed for this feature is a modified spectral centroid, in which the wide-spectrum noise of the articulation “bite” is masked out.
- Because the desired tone color is an aspect of the musician’s creative freedom in how they wish to express any given note, the system provides a metric, but does not judge it as “good” or “bad”

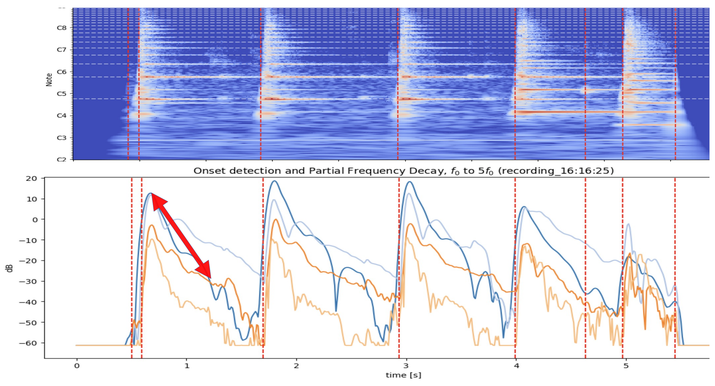
(Photo by the author)



Resonance



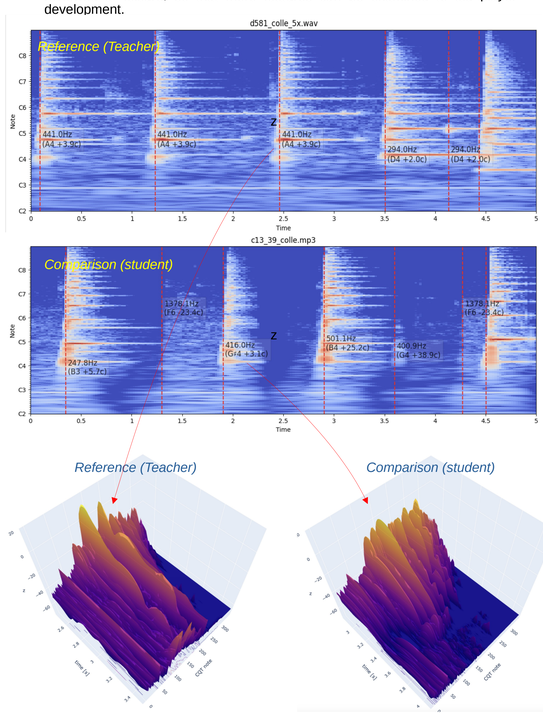
The **resonance** of the collé is the duration of the audible release phase of the waveform envelope after the bow is lifted off the string. In this phase, the sound decays exponentially (or linearly on a dB scale).
- The reference collé tended to have longer resonance durations (often >1s, as these continued to ring even as the next collé stroke was played in the tempo=60 recordings)
- The metric chosen for this feature is the time for the lower partials to decay -30dB from peak



Results

CQT Visualizations of a highly experienced violin teacher (top; left) and student with 9 years of experience (bottom; right) are shown below. The observable detail and metrics enable the following diagnosis to be made:
- The former exhibits stable tempo and pitch within just a handful of cents of deviation from perfect pitch. Articulation is likewise consistent across downbow (odd) and upbow (even) Collé strokes. Tone color is warm, and resonance is full (at t_{avg}=0.63s)
- The latter has a less stable tempo with a significantly delayed upbow (even numbered Collé strokes) with diminished articulation magnitude (8.2dB below the adjacent downbow Collé) and brighter tone. As a consequence of the diminished articulation, the upbow Collé resonance is short. In addition, there are spurious onsets before the upbows indicating extraneous movement or double-touches.

If one is limited to hearing the live sound (and worse yet, if the listener happens to also be the player whose attention is divided to execute the psychomotor sequence to play the Collé in the first place), only the tempo issue might be apparent—and only with a sufficiently trained ear and trained auditory memory. The detailed visualization provides a means of freezing the notes to be reviewed at leisure, so that more features can be examined to aid player development.



Conclusion

Summary
- My system enhances the musician’s ability to perceive the details of their Collé technique by generating visualizations so they can see their sound and by measuring critical features to make clear the specific aspects of the technique that should be adjusted

Future directions
- My system can be augmented to analyze other techniques, as the recording and signal processing pipeline of onset/pitch detection, visualization, and metric feedback can be reused. Similar studies would have to be done to determine the appropriate metrics for the additional techniques.

References

- Fletcher, Neville H., and Thomas D. Rossing. The Physics of Musical Instruments: With 408 Illustrations. Springer-Verlag, 1993.
- McClellan, James H., et al. Signal Processing First. Pearson/Prentice Hall, 2003.
- Pierce, John Robinson. The Science of Musical Sound. Scientific American Library, 1983.
- Resnick, Robert, et al. Physics. J. Wiley & Sons, 2002.
- Sacks, Oliver. Musicophilia: Tales of Music and the Brain. Alfred A. Knopf, 2007.
- Galamian, Ivan. Principles of Violin Playing & Teaching. Shar Products Co., 1999.
- Brown, Judith C. “Calculation of a Constant Q Spectral Transform.” The Journal of the Acoustical Society of America, vol. 89, no. 1, 1991, pp. 425–434., https://doi.org/10.1121/1.400476.
- Yang, Jun, et al. “Spectral Contrast Enhancement: Algorithms and Comparisons.” Speech Communication, vol. 39, no. 1-2, 2003, pp. 33–46., https://doi.org/10.1016/s0167-6393(02)00057-2.
- Rossing, Thomas D. “14.” Springer Handbook of Acoustics, Springer, Dordrecht, 2014.
- Munkong, Rungsun, and Biing-Hwang Juang. “Auditory Perception and Cognition.” IEEE Signal Processing Magazine, vol. 25, no. 3, 2008, pp. 98–117., https://doi.org/10.1109/msp.2008.918418.