MIPCAT - A Music Instrument Performance Capture and Analysis Toolbox

Andre Almeida
UNSW Sydney
a.almeida@unsw.edu.au

Weicong Li
Western Sydney University
weicong.li@westernsydney.edu.au

Emery Schubert
UNSW Sydney
e.schubert@unsw.edu.au

Joe Wolfe
UNSW Sydney
j.wolfe@unsw.edu.au

John Smith
UNSW Sydney
john.smith@unsw.edu.au

ABSTRACT

Playing a musical instrument to convey a convincing and engaging performance requires mastering several musical and technical aspects of playing the instrument. Timing and loudness of notes are recognised as important components of conveying musical expression, but also important are finer aspects such as the timbre of notes, how rapidly a note starts as well as fine variations of loudness and pitch within the note. An expert musician acquires a subtle control (often subconscious) of the gestures needed to produce these sound results, but they are usually difficult to observe or communicate. This presentation introduces the MIPCAT, a hardware and software toolbox that can be used to record simultaneously the sound and the action of a musician: the variables that directly affect the sound such as (in the example of a reed instrument) blowing pressure and bite force, but also body gestures or mouth-mouthpiece geometry, captured via general-purpose cameras. The toolbox also facilitates data processing and analysis in a semi-automatic way. To demonstrate the potential use of the MIPCAT in pedagogy, we show measurements of the gestures of a beginner clarinettist in comparison with those of a panel of expert players.

1. INTRODUCTION

Western music performance traditions involve more than just converting musical notation (the 'score') to a series of note pitches and durations. The score can include what is called expressive indications, telling the musician how loud to play, how loudness should change in a particular stretch of the score, or how the duration of the tone should change relative to the indication given by the note figure in the score.

All this information printed in the musical score still leaves a margin of liberty of individual expressiveness to the musician. In some instruments, this liberty is limited to slight changes in the duration and timing of the note: for example, in baroque keyboard instruments such as the organ or the harpsichord, expressiveness is mostly a fine art of adjusting the durations of a note to the musical context inside a phrase or within the piece, or varying the relative start times of notes in a chord (written as simultaneous). In a piano, the range of available parameters increases, as the loudness can also be changed, or the ringing of the notes with the sustain pedal. In many instruments, such as most of the winds, the range of parameters is greatly increased because the control by the musician for each note is exerted throughout the duration of the note, allowing for changes of the note envelope, multiple aspects of its timbre, and slight modifications of the pitch.

In the clarinet, for example, two important parameters are blowing pressure and reed bite force. Not all possible values can be used for these two parameters: there is a limited range that allows for the production of a periodic tone. Within this range, only a smaller range has "aesthetically suitable" applications and an even smaller range will be used by a particular player, in a particular musical context.

To play a tune requires a player to negotiate a path through a limited volume of the space of control parameters so that the notes sound, the slurs are smooth and the timbre homogeneous. This was the object of a challenge that our team participated in, successfully providing a mapping between player parameters and musical pitches so that a robot musician could play most of the range of the clarinet, for real musical pieces (http://www.phys.unsw.edu.au/jw/clarinetrobot.html).

On the other hand, using a fixed pair of values for a given note results in a rather mechanical performance. Expressivity and "humanness" are achieved by exploring the range of aesthetically sound parameters in a performance. Musicians practice for years in order to acquire an intuitive sense of the right chaining of parameters in a note, in a phrase and an entire musical piece.

Several authors have focused on the slight variations in the characteristics of the sound, in particular tempo and loudness, either due to some implicit rules given the structure of the score, to convey some kind of emotional intention by the performer, or just to personalize the rendition of a musical piece. Timing and loudness aspects have been covered widely in the literature [6]–[8], presumably because those are aspects that are common to many instruments. Other aspects such as intonation and timbre variations have also been studied, although less frequently [9]. Much less focus has been given to how the adjustments to the sound results are achieved by musicians, but see work from the Vienna group [10].

Relatively little is known about the reasons behind the set of parameters chosen by a musician to play a particular note with a particular sound result: Is this an individual choice and can different player parameters be used to achieve similar sound results? Does this choice depend on the musical context or the expressive intention? Can the knowledge of possible parameters be used to teach how to play a musical instrument? Musicians can describe the actions they need to undertake to modify the emotion conveyed by a performance [11], but how do these intentions correlate with their actual, physical actions?

Some authors have incorporated sensors in musical instruments to study musical interpretation in physical terms [12]–[15].

The present work aims to provide a comprehensive and cost-effective way of acquiring and pre-processing the main musical and player parameters involved in the performance of a musical instrument. It is an extension of a reviewed article published recently [16]. In that paper, we foresaw applications in pedagogy. In this paper, we use the toolbox to compare measurements made on a beginner clarinettist with a collection of measurements made on experienced professionals.

The toolbox is here applied to the clarinet, but with some effort, it could be adapted to other wind instruments. For other families of instruments, such as bowed strings, for instance, some of the software tools may be applied and the automated tracking of ArUco tags could also be adapted to the motion of the bow. Different sensors would be used, for example force sensors in the bridge and bow and accelerometers in the latter.

2. THE MIPCAT

2.1 Hardware

The capture system of the Music Instrument Performance Capture and Analysis Toolbox (MIPCAT) consists of several sensors fitted to a clarinet mouthpiece (Figure 1), two microphones capturing the external sound, and three cameras.

The sensors were:

- A tonguing sensor consisting of a small wire glued onto a *Légère* reed. A second wire was connected to the thumb rest on the clarinet so that when the tongue touched the wire, an imperceptible electric current (of a few µA) would flow through the body of the musician;
- An optical sensor measuring the distance between the mouthpiece and the reed, by sensing differences in reflected infrared light on the reed;
- A miniature pressure sensor measuring pressure (DC and AC) inside the mouth of the musician;
- A second miniature sensor measuring the DC and AC pressure inside the mouthpiece;
- A B&K microphone measuring AC pressure fluctuations inside the barrel.

Apart from the B&K microphones and the external audio microphone, all the sensors were plugged into a custom-built conditioning system, whose schematics can be found with the software package. All of the electronics, including proprietary apparatus, the acquisition system and the laptop were powered by a 12V car battery, to ensure the electrical safety of the musician and experimenter.

Of the two external microphones, an audio microphone was placed on a stand about 50 cm from the player. A second microphone was a B&K measurement microphone attached to the lower part of the instrument, below the right hand.

Two cameras captured a front and a right-hand side view of the clarinet player and the instrument (Figure 2). Four ArUco tags (QR-code style markers [17]) were attached to the clarinet. These are easy to detect with image-processing tools. A third miniature camera was attached to the barrel of the instrument and captured a side view of the mouthpiece and the player's mouth (Figure 3). With a coloured tag and a scale glued to the mouthpiece, this allowed the position of the lips along the mouthpiece of the instrument to be measured.

The hardware components are specific to the clarinet, but with small modifications, the sensors could be adapted to other single-reed instruments such as a saxophone. Double-reeds or flutes would require greater modifications, but solutions have been found to measure mouth pressure in flutes [13] or brass instruments [18], [19].

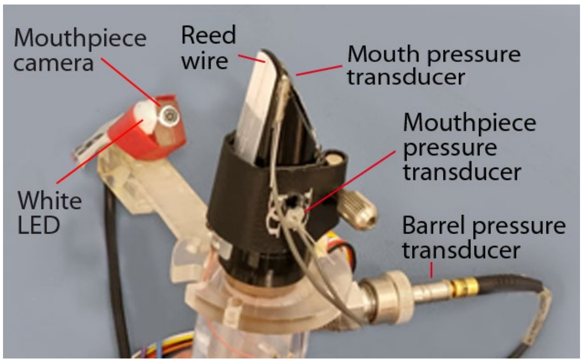

Figure 1. The sensor-fitted mouthpiece used in the MIPCAT (from [16])

Figure 2. One of the authors demonstrating the use of the sensor-fitted clarinet, and the use of ArUco markers for clarinet position tracking (from [16]).

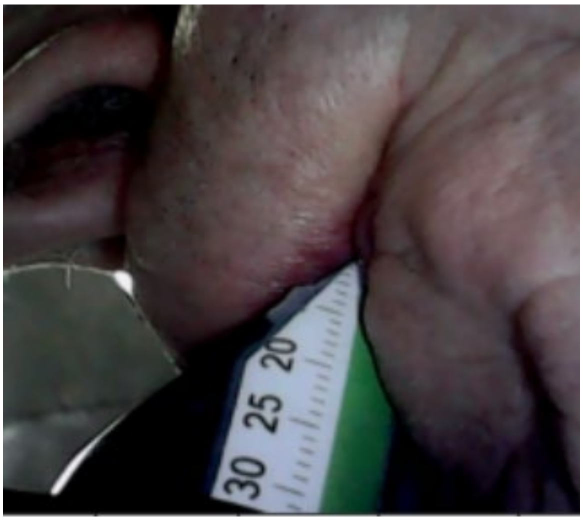

Figure 3. The mouthpiece fitted with a coloured scale used for tracking the lip position on the mouthpiece (from [16]).

2.2 Software

A software package, mostly written in Python, is publicly provided to pre-process and analyse these data and comparable sets. This package is available on GitHub at https://github.com/goiosunsw/mipcat. There are several components to this package:

2.2.1 Video to Audio Alignment
This module calculates the delay between a reference signal, for instance, that of the external microphone, to the audio track associated with a recorded video. The alignment is based on a fingerprinting algorithm [20] that extracts relative peak positions in each of the signal's spectrograms.

2.2.2 Time-series Indicators
This module analyses the recorded data in frames, and calculates some indicators, in particular:

- DC offset: the mean value of the measured signal within that window;
- Amplitude: the RMS amplitude of the oscillation of the measured data;
- Frequency: the fundamental frequency of the oscillation in the measured signal;
- Harmonic components: amplitudes of five harmonics of the fundamental frequency.

The signal input to these analysers is first pre-processed to take into account any calibration needed by the signal. This and other signal properties are defined in a YAML file.

2.2.3 Segmentation and Alignment with the Musical Score
This module takes as input an audio recording of a performance (sometimes repeated several times) and a music score. Time instants in the audio signal are matched to note beginnings and endings.

2.2.4 Mouthpiece Video Processing
The scale glued onto the mouthpiece consists of a human-readable millimetre scale alongside a green strip that is easy to detect and isolate digitally using open-CV. The position of the numbers in the human-readable scale is tracked using a template tracker (based on correlation) for two reasons: firstly, it allows calibrating for physical distances, secondly, it indicates where the green strip should be found, making it easier to discard false detections of the green patch.

2.2.5 Clarinet Position Processing
This module uses open-CV to detect the position of the clarinet markers (ArUCo) on the image. Open-CV provides a tag detector for individual frames. A second layer of detection is added, consisting of a template tracker: a snapshot of the tag found with the standard detector is kept in memory, and a correlation algorithm tracks its position in a new image. This allows tracking of the tags even when they are partly obscured or blurred by motion.

2.2.6 Player Pose
The player is tracked using a deep neural network algorithm provided by Google and called Mediapipe. It can identify the position of key elements of the human body. Important for us are the position of the mouth and the head. This detection is not as accurate as the position of the clarinet based on the ArUco tags but can provide some rough indication, for example of the angle of the head.

2.2.7 Signal Collection and Building of a Database
Large volumes of data are generated by the descriptor extraction. The data set is easier to analyse by aggregating it within individual notes and calculating statistical values such as means, standard deviations, etc.

3. A WORKED EXAMPLE

3.1 Introduction

The primary purpose of the worked example was to use the MIPCAT system to demonstrate how player's physical gestures in a study concerning expressive playing are captured and processed. Detailed results from this study will be published later. Here we give an example of how a set of recordings from expert musicians can be compared to the performance of a beginner, pinpointing when the main differences arise and suggesting actions that are closer to the expert's technique.

The musical excerpt used in this example is an eight-bar section introducing the main theme of the slow movement of Mozart's Clarinet Concerto K. 622. Seven musicians were invited to participate in the study, 6 of them professionals playing in orchestras (used to obtain "reference expert performance"), and one beginner who has played other reed instruments for more than 30 years but never studied or regularly played the clarinet. They were involved in a 3-hour-long session (with a 30-minute break) concerning the expression of emotion through music, and a set of other tasks related to musical performance. As part of this study, they were asked to play the above-mentioned excerpt on the sensor-fitted clarinet (the lab clarinet) and also on their own clarinet.

3.2 The Recording Stage

The musicians sat in a low-reverberation room, as in Figure 2. They had some time to practise on the sensor-fitted instrument, and to select from a set of synthetic reeds according to which felt more comfortable to them. They played the excerpt twice on the sensor-fitted instrument and were then asked to play it again on their own instrument. If needed they could reject a recording and repeat it. An experimenter from the team provided guidance and launched the recording of the sensor data, together with each of the three cameras.

3.3 Pre-processing and Segmentation of the Audio

Once the data have been collected, the first step is to identify the individual notes, matching them to the score. This can be done manually with audio tagging software, but MIPCAT provides a set of tools that can help with this task.

Initially, a series of descriptors are extracted from the audio (Figure 4). Among these are the amplitude and frequency of the sound, but many others are extracted also from the sensor recordings, to be used at a later stage. Extraction of the descriptors can be performed automatically for a single recording or a set of recordings using the script "ts_gen_from_csv".

Amplitude and frequency from a reference channel (usually the internal instrument pressure if available, otherwise from one of the external microphones) are used to detect note transitions, in a first pass, and then to align these note transitions to the score. These two passes can be performed in a single step using the script "note_matcher". Once again, the script can be run individually for one recording, or a set of recordings. The script outputs a TextGrid file per recording, and these files can be used in *Praat* (https://www.fon.hum.uva.nl/praat/) to check and adjust the segmentation.

3.4 Pre-processing of Video Recordings

MIPCAT provides a set of tools to help extract key positions in the video files. Some of these are specific to our project, but the ArUco tagging is quite generic and can be used in many different situations. The script that tracks the ArUco markers is more robust than the simple framewise detection of a tag, using information from previous frames to keep identifying the tag even if it is blurred by motion or partially obscured. Tracking can be performed with the script "aruco_tracker".

Tracking of human position can also be used in many different situations and is done in an automated way by the script "mp_pose_detect" which calls Mediapipe in an automated way and exports the data to be processed later. Finally, the mouthpiece video is specific to clarinets and saxophones. This stage has to be run for each video in two stages: first calling the GUI "mouthpiece_gui" to manually adjust key parameters for the automated processing, and creating a configuration file for each video.

The tracking of the green patch that is partially and variably covered by the musician's lips is then done automatically with "mouthpiece_tracker".

3.5 Alignment of Video and Sensor Signals

The cameras used in this project were not synchronised to the sensor and microphone signals, because cameras that provide a synchronization signal can be expensive. Instead, the audio that is recorded by each camera together with the video is synchronised offline using a fingerprinting algorithm. This algorithm identifies key features in the spectrogram of each audio signal (camera and signal-synchronous) and runs a matching algorithm to identify the delay between the two signals. This is done with the script "align_keypoints.py"

Once the delays are known, the measurements extracted from the videos in the first step can be cut and aligned to the sensors so that they can be seamlessly analysed as the output of any other sensor.

3.6 Collecting Data

The data gathered with this system are quite comprehensive, and can be analysed in different ways. For the purpose of this worked example, we chose to aggregate signal data for each note, calculating average values, trends and variability measurements. For example, for the measured mouth pressure, many different indicators are calculated such as the average mouth pressure during a note, the standard deviation, the trend, the amplitude of the oscillation (since sound also propagates inside the mouth), etc.

These calculations are run for all the notes in all the recordings automatically with the script "build_note_database".

4. RESULTS

With the knowledge of the note boundaries for each performance, it is possible to align different performances on a modified time scale that is measured in musical beats instead of seconds. With this alignment, we can compare the descriptor time series for different players, as seen in Figure 4.

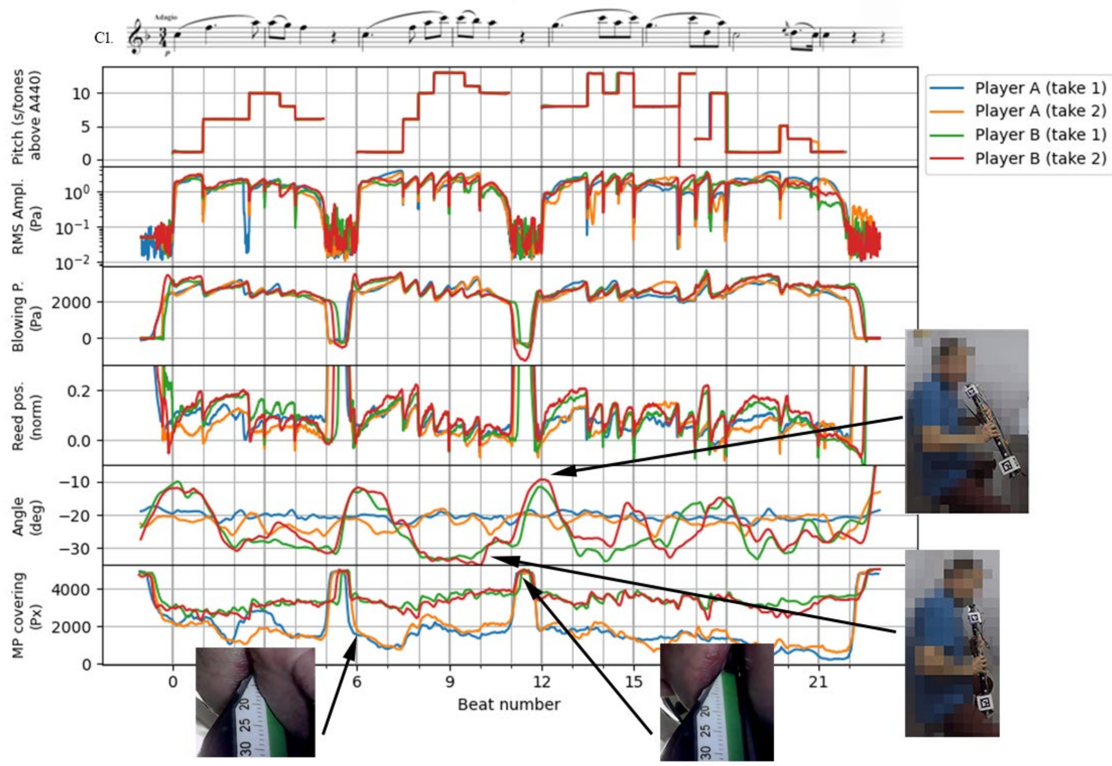

Figure 4. An example capture of descriptor time series corresponding to 4 performances of 2 (expert) players. From top to bottom, the score (transposed for the instrument), the playing frequency, sound RMS amplitude, blowing pressure, normalized distance of the reed to the lay of the mouthpiece, angle of the clarinet and covered area of the mouthpiece (from [16]). The photographs provide two extreme examples of mouthpiece covering (bottom) and clarinet angle (on the right-hand side).

4.1 Note-by-note Analysis

Instead of comparing "instantaneously" how a player is playing their instrument, we can analyse the performances in terms of averages of the descriptors within each note. In this way, it is possible, for example, to compare an individual performance with an "average performance" of a larger set of musicians that participate in a study. (Detailed analysis of a larger set is a project of the present team). In the following plots, we show the "average expert perfor mance" in blue, calculated as the median value of a particular descriptor for each note, for all the 6 expert players, and for both the lab instrument and their own instrument, except when the data involves player parameters such as blowing pressure or reed position. Overlapped is the variability of the descriptor, measured as the inter-quartile range for each note. In red we display the performance of a beginner, to compare it to the "average expert performance".

The first figure (Figure 5) shows the value of the note amplitude in a box plot without whiskers. The horizontal length of each box is the IOI for that note. Because each player can decide to play the excerpt with a different overall amplitude, the absolute amplitude for each note is subtracted from the average amplitude for the entire excerpt by that player. It thus shows the amplitude of the note relative to all other notes. This typically reduces the overall variability of each note's amplitude by 3 dB, i.e. instead of the amplitude of a typical note ranging from 97 to 102 dB (a range of 5 dB), the range is reduced to 2 dB.

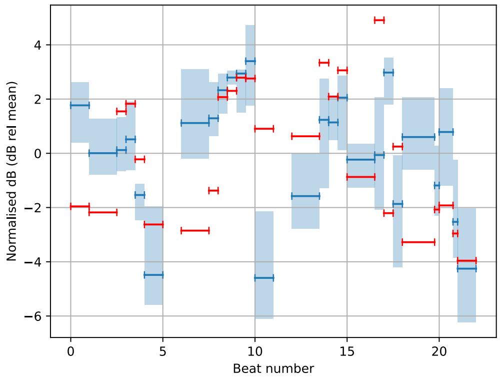

Figure 5. Box plot showing average (horizontal lines) and inter-quartile variability (shaded areas) of the sound amplitude of expert performances (blue), per note of the excerpt, compared to an excerpt from the beginner (red). Amplitude is shown as a difference from the overall excerpt amplitude.

The plot shows, for instance, that the beginner is exaggerating the dynamics at the beginning of the second phrase while not doing a large enough decrescendo towards the ends of the 1st and 2nd phrases.

Similarly, we can see in Figure 6 that the beginner is not coping well enough with the tuning in the higher notes. Notice that the variability in pitch across expert players is small in this plot.

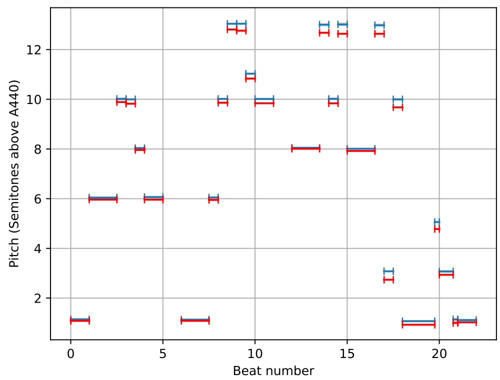

Figure 6. Average pitch of each note of the excerpt for expert musicians (blue) and a beginner (red). Unlike the other figures, the variability (inter-quartile range) is not visible because it has typical magnitude of a few cent (hundredths of a semitone).

The plots in Figures 5 and 6 would be possible to obtain using only an audio recording. The sensors attached to the clarinet however make it possible to look into the actions of the musician that is producing these sound results. Figures 7 to 11 show how it could be possible to pinpoint technical problems from a beginner

Figure 7 shows how the blowing pressure varies along the excerpt, for the "average expert" and for the beginner. For instance, the two high Cs immediately before and on the 9th beat, in the second phrase, were produced using considerably greater blowing pressure from the beginner (red dash) in comparison to the professional mean (blue).

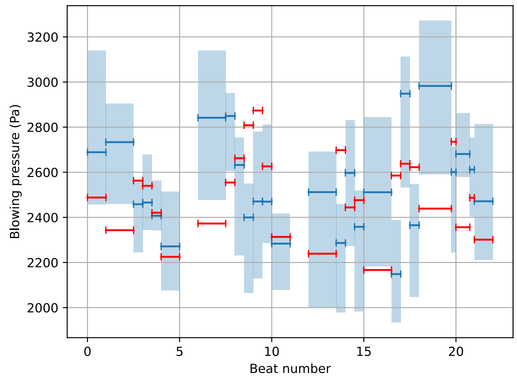

Figure 7. Average blowing pressure (blue) for expert musicians (per note) compared to a beginner's performance (red)

Another important parameter for clarinettists is the bite force applied to the reed. Our sensors do not measure force directly, instead measuring the reflectance of a light shining on the reed. This reflectance is a monotonic function of the distance, roughly linear in the range of interest. However, changing the reed, and to some extent changing the reed position on the lay may change the reflectance. For this reason, the data presented in Figure 8 is a difference in the reed position for the note from the average position for all notes in the recording. In this figure, it is apparent that the beginner changes the reed position more than the expert players.

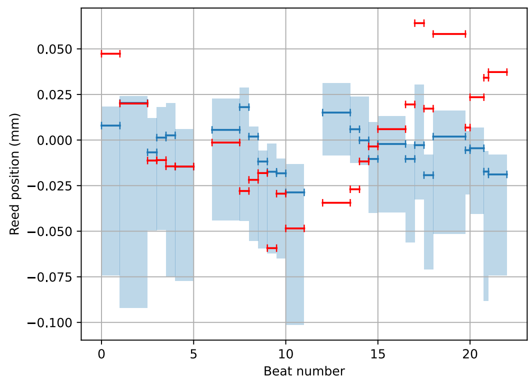

Figure 8. Average reed displacement from excerpt overall mean for expert musicians (blue) compared to a beginner (red).

The mouthpiece camera measured the amount of mouthpiece that is covered by the bite of the musician. It was found that different musicians can use rather different bite positions (or configurations, since the covered part of the mouthpiece measured in the image might not correspond exactly to the bite position), but the position is, on average, only changed by a small amount during playing (Figure 9).

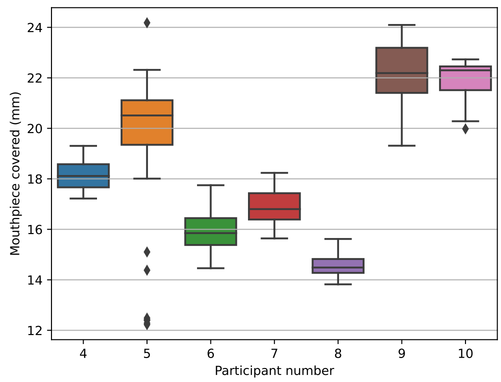

Figure 9. Average mouthpiece length covered as measured from the tip of the mouthpiece, for each of the professional (numbers 4-9) and beginner (number 10) participants.

Still, it is possible to observe a trend during an excerpt whereby the mouth slightly recedes from the mouthpiece in the first phrases (Figure 10). For the beginner the retraction is more pronounced towards the end of the excerpt.

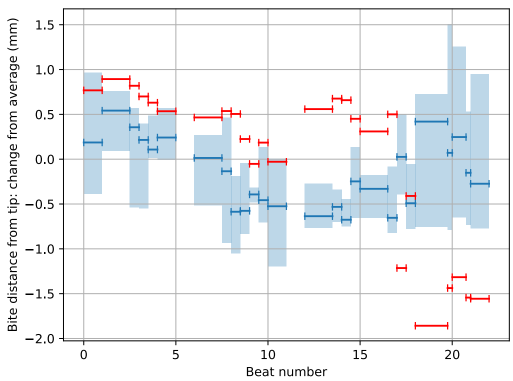

Figure 10. Difference between each note's bite position to the overall average position for each player. Blue: expert musicians, red: beginner.

The views of the player can be used to determine the clarinet's orientation relative to the musician's face. This is also a measure that can vary considerably among players as shown in Figure 11, not least because of physiological differences that may not translate directly into a different action on the instrument. It may also reflect different playing traditions and training.

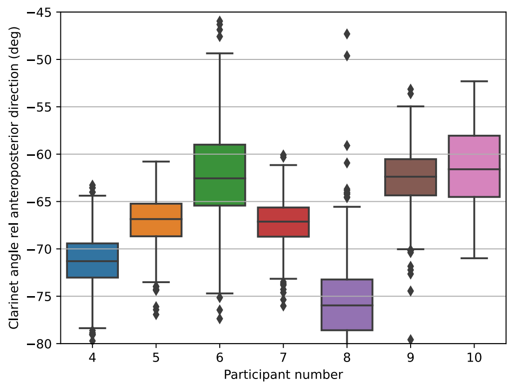

Figure 11. Average clarinet angle relative to forward facing direction of head for each player (4-9: professionals, 10: beginner). Notice how players 6, 8 and 10 move their instrument considerably more during the excerpt.

5. DISCUSSION AND CONCLUSIONS

Measurement systems for musical instruments and automated or semi-automated methods of processing data open new perspectives on how we teach and learn technical aspects of musical performance. Student performances can be compared to expert performances and technical styles can be compared. A range of other applications is foreseen in seeking a deeper understanding of how the musician-instrument combination works in vivo.

In the present study, a pedagogical application was shown, that can describe the range of playing actions of expert performances, against which that the student performances can be compared. This comparison goes beyond the simple auditory, interpretative level, but goes to the physical action level — the very act of playing. With some refinements, this has considerable applications for player training.

The software used for this analysis is available with documentation on https://github.com/goiosunsw/mipcat, and a larger dataset of performances will be made available shortly.

6. ACKNOWLEDGMENTS

We warmly thank the Australian Research Council for support of this project, Yamaha for the clarinet, Légère for the reeds and the participants for the study.

7. REFERENCES

[1] C. Fritz and J. Wolfe, "How do clarinet players adjust the resonances of their vocal tracts for different playing effects?," *J. Acoust. Soc. Am.*, vol. 118, no. 5, pp. 3306–3315, Nov. 2005, doi: 10.1121/1.2041287.

[2] A. Almeida, W. Li, J. R. Smith, and J. Wolfe, "The mechanism producing initial transients on the clarinet," *The Journal of the Acoustical Society of America*, vol. 142, no. 6, pp. 3376–3386, 2017.
[3] A. Friberg, R. Bresin, and J. Sundberg, "Overview of the KTH rule system for musical performance," *Advances in cognitive psychology*, vol. 2, no. 2, p. 145, 2006.
[4] A. Almeida, D. George, J. Smith, and J. Wolfe, "The clarinet: How blowing pressure, lip force, lip position and reed 'hardness' affect pitch, sound level, and spectrum," *Journal of the Acoustical Society of America*, vol. 134, no. 3, 2013, doi: 10.1121/1.4816538.
[5] D. Fabian, R. Timmers, and E. Schubert, *Expressiveness in music performance: Empirical approaches across styles and cultures*. Oxford University Press, USA, 2014.
[6] R. Bresin and G. U. Battel, "Articulation strategies in expressive piano performance," *Journal of New Music Research*, vol. 29, no. 3, pp. 211–224, 2000.
[7] W. Goebl, "Melody lead in piano performance: Expressive device or artifact?," *J. Acoust. Soc. Am.*, vol. 110, no. 1, pp. 563–572, Jul. 2001, doi: 10.1121/1.1376133.
[8] B. H. Repp, "Acoustics, perception, and production of *legato* articulation on a digital piano," *The Journal of the Acoustical Society of America*, vol. 97, no. 6, pp. 3862–3874, Jun. 1995, doi: 10.1121/1.413065.
[9] M. Barthet, R. Kronland-Martinet, and S. Ystad, "Consistency of timbre patterns in expressive music performance," in *Proc. of the 9th Int. Conference on Digital Audio Effects (DAFx-06), Montreal, Canada*, 2006.
[10] M. Pamiès-Vila, "Expressive performance on single-reed woodwind instruments: an experimental characterisation of articulatory actions," Doctoral dissertation, University of Music and Performing Arts, Vienna, 2021.
[11] A. Almeida, W. Li, E. Schubert, J. R. Smith, and J. Wolfe, "Expressive goals for performing musicians: The case of clarinettists," *Musicae Scientiae*, vol. 0, no. 0, 2022.
[12] V. Chatziioannou and A. Hofmann, "Physics-based analysis of articulatory player actions in single-reed woodwind instruments," *Acta Acustica united with Acustica*, vol. 101, no. 2, 2015, doi: 10.3813/AAA.918827.
[13] P. de la Cuadra, B. Fabre, N. Montgermont, and C. Chafe, "Analysis of flute control parameters: A comparison between a novice and an experienced flautist," *Acta Acustica united with Acustica*, vol. 94, no. 5, pp. 740–749, 2008.
[14] A. Hofmann, V. Chatziioannou, M. Weilguni, W. Goebl, and W. Kausel, "Measurement setup for articulatory transient differences in woodwind performance," presented at the Proceedings of Meetings on Acoustics, 2013, vol. 19. doi: 10.1121/1.4799441.
[15] W. Li, A. Almeida, J. Smith, and J. Wolfe, "How clarinettists articulate: The effect of blowing pressure and tonguing on initial and final transients," *Journal of the Acoustical Society of America*, vol. 139, no. 2, 2016, doi: 10.1121/1.4941660.
[16] A. Almeida, W. Li, E. Schubert, J. Smith, and J. Wolfe, "Recording and analysing physical control variables used in clarinet playing: A musical instrument performance capture and analysis toolbox (MIPCAT)," *Frontiers in Signal Processing*, vol. 3, 2023, Accessed: Feb. 19, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/frsip.2023.1089366
[17] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognition*, vol. 47, no. 6, 2014, doi: 10.1016/j.patcog.2014.01.005.
[18] H. Boutin, N. Fletcher, J. Smith, and J. Wolfe, "Relationships between pressure, flow, lip motion, and upstream and downstream impedances for the trombone," *The Journal of the Acoustical Society of America*, vol. 137, no. 3, pp. 1195–1209, 2015.
[19] V. Fréour and G. P. Scavone, "Acoustical interaction between vibrating lips, downstream air column, and upstream airways in trombone performance," *The Journal of the Acoustical Society of America*, vol. 134, no. 5, 2013, doi: 10.1121/1.4823847.
[20] P. Cano, E. Batlle, E. Gómez, L. de C.T.Gomes, and M. Bonnet, "Audio Fingerprinting: Concepts And Applications," in *Computational Intelligence for Modelling and Prediction*, S. K. Halgamuge and L. Wang, Eds. Berlin, Heidelberg: Springer, 2005, pp. 233–245. doi: 10.1007/10966518_17.