How a Clarinettist Conveys Emotion in Music Playing: Measuring Player Gestures and Signal Parameters Using a New Toolbox

Weicong Li weicongli@outlook.com
Andre Almeida a.almeida@unsw.edu.au
Emery Schubert e.schubert@unsw.edu.au
John Smith john.smith@unsw.edu.au
Joe Wolfe j.wolfe@unsw.edu.au
UNSW Sydney

ABSTRACT

Instrumental music has the fascinating capacity to communicate emotions, sometimes with subtlety. Most previous research has focused on comparing features of recordings (e.g. sound level, tempo and timbre) of music conveying different emotions, rather than detailed gestures that musicians use during music playing, which would be useful for music students, teachers and music researchers. This paper reports an experimental study of how an expert clarinet player expresses three different emotions when playing the same pieces of music: *happy*, *sad* and *lacklustre/deadpan*. Parameters showing the musician's continuous control of blowing pressure and reed position were measured, as well as variables mentioned above in the recorded music; all were analysed semi-automatically using a new toolbox developed for this study. The results show how the emotions can be differentiated not only by the musical feature variables, but also the details of how the musician physically controls the instrument to produce them. These results provide an alternative approach for training musicians about expressive playing.

1. INTRODUCTION

Conveying various types of emotion plays an important role in music playing and this often requires that musicians master control over sets of playing techniques to physically manipulate the musical instrument to achieve desired musical goals. Good players can perform the same piece in different ways to transmit distinctly different emotions or expressive goals (EGs). However, the musical parameters used to produce EGs are better understood than the physical gestures used to produce them these musical parameters. Different EGs are conveyed in the presence of different musical parameters such as note length, loudness, pitch and timbre (see, for example, [1, 2]). The musical parameters that convey different EGs in performance have been studied over at least quarter of a century [1-4]. For example, Gabrielsson and Lindström [5] summarised a lexicon of musical elements/parameters that are associated most often with particular EGs.

These musical features are necessarily created by carefully controlled physical gestures or 'playing control parameters' by the musician. Understanding both the musical and gestural/control parameter aspects of performance is currently a gap in research on music expression, and closing this gap would have several applications. Such knowledge would provide insight into the nexus between the physics and the psychology of music performance, and would be useful for music students, teachers and music researchers who seek to understand music performance in terms of the physical control of the instrument, rather than, or in addition to, the more commonly used musical parameters (e.g. 'music systemisers' could benefit from such researcher – Kreutz [6]). Measuring parameters of both aspects simultaneously, and understanding how they are varied to convey different EGs is the aim of this pilot study.

Less attention has been given to the playing control parameters (hereafter called playing parameters) because they are difficult to measure, and measuring physical parameters of the instrument-player interaction usually requires invasion of the instrument and the player by measurement tools, which can interfere with the playing itself. Nevertheless, in the case of the clarinet, a few playing parameters used by players to control the sound have been studied, such as the average or DC air pressure in the player's mouth while blowing the clarinet, the position and vibration of the reed, the lip and tongue action, and some information about the acoustics of the player's vocal tract [7-13]. These studies used either single notes or simple excerpts with the focus of understanding the relation between player's input gestures and output sound, thus little attention was given to the particular EGs (if any) that were used in performance. More recently (and while many laboratory activities were limited by the COVID-19 pandemic), we used a survey to study how experienced clarinettists think they would play in order to distinguish different EGs in terms of both musical and playing parameters when performing the same musical excerpts [14]. Based on similarities, a set of six clusters of EGs were suggested for use in future studies. A few musical and playing parameters were reported by clarinettists as important in achieving specific EGs.

Over the last decade or so, we developed a musical instrument performance capture and analysis toolbox (MIPCAT) to capture and study various musical and playing variables controlled by clarinettists while performing music [15]. The toolbox includes both hardware and software. The hardware consists of various sensors mounted in or on the clarinet, plus microphones and video cameras. These can capture the player's blowing pressure, the sound pressure in the mouth and the instrument, the reed position and vibration, aspects of embouchure including the bite on the reed, some motions of the player's body and the output sound at different positions. The software contains several tools to process and analyse recordings of player performances captured by the hardware semi-automatically. Among other applications, the components of the toolbox enable us to study how clarinettists play the same music to convey different EGs, from a rich set of data that include both musical and playing parameters.

As an exploratory investigation for a larger study, this paper reports a case study of how a clarinettist plays the same musical excerpt with three different EGs and compares the musical parameters produced and some of the playing parameters used in producing them.

2. MATERIALS AND METHODS

In this study, MIPCAT and a modified clarinet (Yamaha YCL250 model with Yamaha 4C mouthpiece) were used for data acquisition and processing: see Figure 1. More details of the setup are described in [16].

One clarinettist having extensive classical and jazz playing experience participated in this study. From written music provided, the participant was asked to play the excerpt 'Happy Birthday to You' in G major (music score shown in Figure 2). The tune was played to convey three different EGs: *happy*, *sad* and *lacklustre/deadpan* (see [14] for detailed discussions of these EGs). Before recording, the participant was allowed to practise on the instrument modified to fit MIPCAT until feeling comfortable with it. The music and instructions were emailed to the participant beforehand so that he or she had time to prepare, if needed. To convey each EG, the participant was instructed not to change the melody but to vary freely other aspects of the performance, including tempo, dynamics, articulation, and timbre, to communicate the intended EG to listeners as convincingly as possible.

The following signals were used in the analysis:

- Mouth pressure: measured by a miniature pressure sensor (8507C-2, Endevco, Irvine, CA) fitted into the corner of the mouthpiece with its sensing membrane exposed to the inside of the player's mouth during playing;
- Reed position: the AC and DC components of the displacement of the reed in a direction at right angles to the instrument axis were measured by a reflective, infrared proximity sensor (QRE1113, ON Semiconductor, Phoenix, AZ) mounted inside the mouthpiece, 5 mm from the mouthpiece tip, directly opposite the reed;
- Radiated sound: measured by a ¾" microphone (RODE NT3, Sydney, Australia) mounted on a stand at the same height as the bell and at a distance of 45 cm. Microphones were also mounted on the clarinet bell and barrel, to separate out the effects of player and instrument motion. The recordings from these last two are not used in the current preliminary investigation. Measurements were conducted in a room treated to have low reverberation. More details about how the signals were acquired and processed were described in [15].

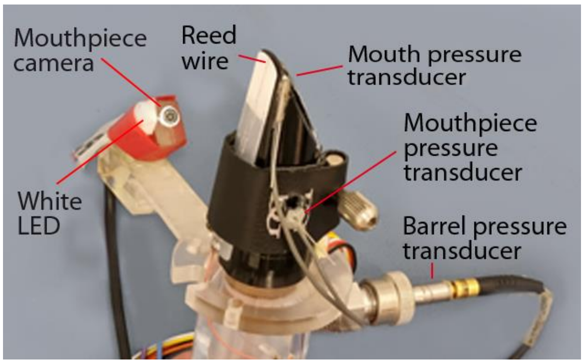

Figure 1. The mouthpiece and some of the sensors used in the MIPCAT. Reproduced from [15].

The software tools of MIPCAT were used for data processing. First, several indicator timeseries such as DC values, amplitudes of oscillation, fundamental frequency, etc. were extracted from the raw data. Then these timeseries were segmented note by note by matching the fundamental frequency measured to that of the music. Then some of the musical and playing parameters were extracted and averaged within each note from the timeseries: RMS sound pressure level, tempo, spectral centroid, blowing pressure and DC reed displacement. A local 'tempo' for each note was calculated by the written duration of the note divided by the Inter-Onset Interval (IOI) in minutes. IOI is the time interval between the onsets of successive notes.

Two takes from the participant gave a total of 50 tokens (notes) for each EG, resulting a total of 150 tokens. In addition to MIPCAT, MATLAB and R were also used for subsequent analysis.

3. RESULTS AND DISCUSSION

3.1 Musical Parameters

In this pilot study, we only explored a few of the musical parameters: sound pressure level, tempo and spectral centroid. These were considered important musical parameters that can characterise the performances and reflect an important subset of the musical parameters used to achieve a particular EG; we discuss them first before discussing control parameters.

Figure 2 shows the average sound level of the radiated sound for each note and the standard deviation of level within that note, for each of the three EGs: *happy*, *sad* and *lacklustre/deadpan*. (The last of these is hereafter abbreviated as *deadpan*.) In general, notes played to convey *happy* have the highest average sound level and largest standard deviation of level within each note. This result is consistent with previous research, e.g. [17]. Those for *deadpan* show medium average sound level and smallest standard deviation, and those for *sad* show lowest average sound level (also consistent with the literature) but

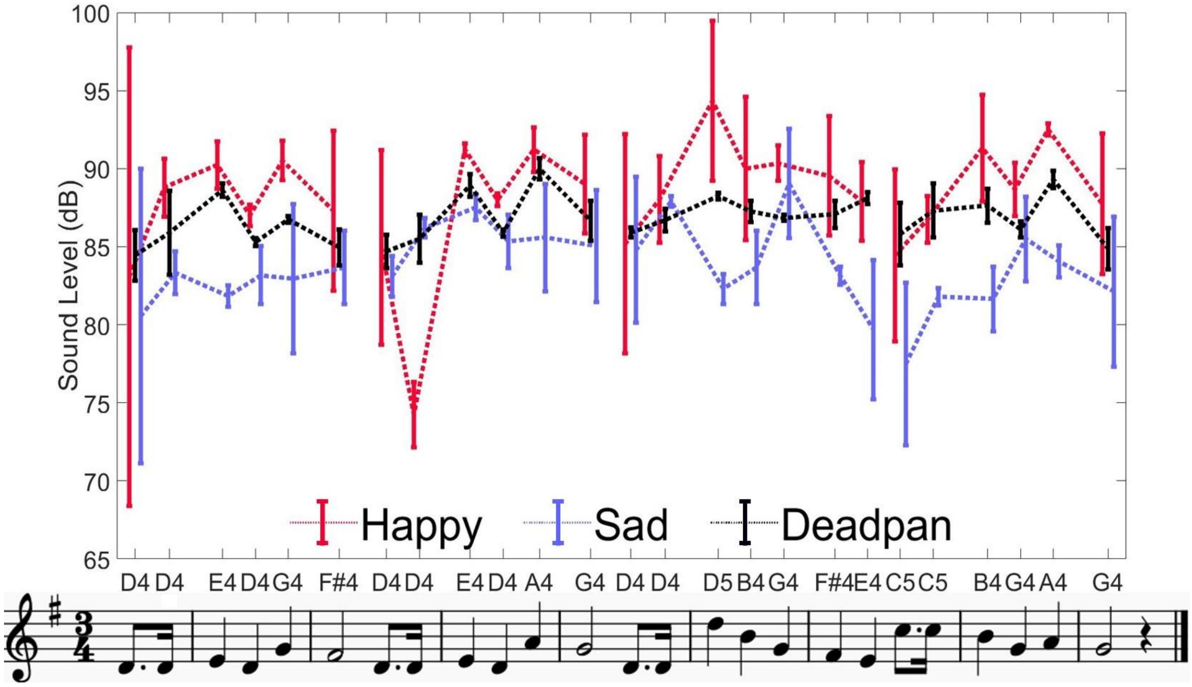

Figure 2. Sound pressure level of the radiated sound for each note and standard deviation of level within that note for the three expressive goals: *happy*, *sad* and *deadpan*. Notes of each phrase are connected with dotted lines. The pitches from the score (Happy Birthday to You) are shown on the x-axis, in order but not proportional to time.

also a large standard deviation within notes. In addition, notes at the start and end of each phrase often show larger variation than other notes in the phrase, e.g. notes D4 and F#4 in the first phrase for *happy*. Another observation is that different EGs show different patterns in the average sound level and standard deviation of the note sequence, e.g. *happy* and *deadpan* have similar patterns for the first, second and last phrases, whereas *sad* has a very different pattern from *happy*: e.g. the variation between notes in the first two phrases are smaller, but much larger in the last two phrases; the peak sound level of the third phrase falls on the third last note (G4) instead of the third note (D5, highest note in that phrase); the average sound level of the last phrase is significantly lower than the other three phrases. These features seem to indicate a particular playing style of the participant when conveying *sad* in the excerpt.

Figure 3 is the boxplot of sound levels of all the notes from the radiated sound for the three EGs. This figure further confirmed the observations from Figure 2: median sound level of *happy* is 88.4 dB, substantially higher than *deadpan* (86.8 dB) and *sad* (84.7 dB). As expected, *happy* and *sad* show more variation in the sound level than *deadpan*.

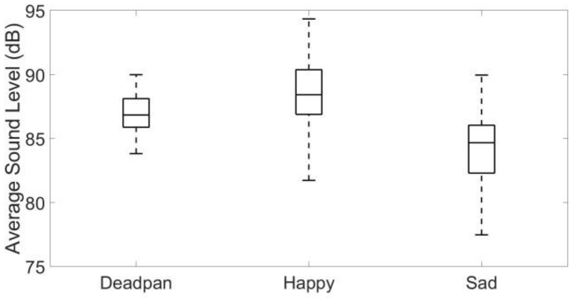

Figure 3. Boxplot showing the note-to-note variations in sound pressure level of the radiated sound for the three expressive goals: *happy*, *sad* and *deadpan*.

Figure 4 is the boxplot of local ‘tempos’ (in the unit of beats per minute) of all the notes, showing the note-to-note variation for the three EGs. *Sad* has the slowest median tempo value (107 bpm) but largest note-to-note variation in the tempo among the three EGs. *Happy* has median tempo (120 bpm), similar to *deadpan* (121 bpm), but much larger variation in local tempo than *deadpan*. The low variation in sound level and tempo for *deadpan* were expected. But it is interesting to note the similarity in tempo between *happy* and *deadpan*.

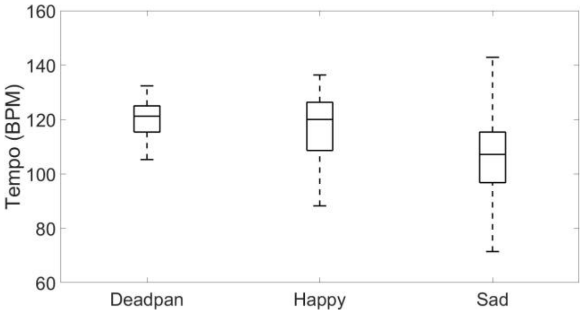

Figure 4. Boxplot showing the note-to-note variations in local tempo for the three expressive goals: *happy*, *sad* and *deadpan*. Local 'tempo' of each note was calculated as the written value of the note in beats divided by the Inter-Onset Interval in minutes.

Figure 5 is the boxplot of spectral centroids of all the notes for the three EGs. The spectral centroid can be pictured as the 'centre of gravity' of the pressure spectrum. It is strongly correlated with perceived brightness of timbre [18]. Among the three EGs, *deadpan* has the lowest median spectral centroid at 1796 Hz and smallest variation in the spectral centroid, followed by *happy* (1918 Hz) and *sad* (1983 Hz). This finding is somewhat consistent with the literature. For example, one study found that spectral centroid for depressive sadness was lower than for happiness [19]. We can reconcile the difference between that study and ours by the stimuli used. The previous study [19] required rating of emotions of single tones played across a range of musical instrument timbres. This meant that those participants had far fewer cues from which to determine differences in emotions (a single pitch, controlled volume etc.), making participants rely more on timbral cues, and hence leading to more distinct results than the present study, where more musical context and flexibility allowed for several other cues to distinguish EGs, and where changes in timbre were 'within-instrument' changes, and therefore generally quite subtle. However, *sad* also has a considerably wider range of spectral centroids employed, suggesting that the player is varying timbre more in comparison to *happy* and *deadpan* to achieve the goal of *sad*. The spectral centroid may be expected to be influenced by blowing pressure and reed position [9]: greater blowing pressure and smaller average reed displacement from the mouthpiece both enhance clipping, which increases high harmonics and thus spectral centroid. Here, the written D4 notes are selected for calculating the correlation with blowing pressure and reed displacement: both correlation coefficients are non-significant (r = 0.05, p = .75 for spectral centroid and blowing pressure; r = 0.14, p = .35 for spectral centroid and blowing pressure). The reason of the non-significant result may be that the player was adjusting both playing parameters while playing.

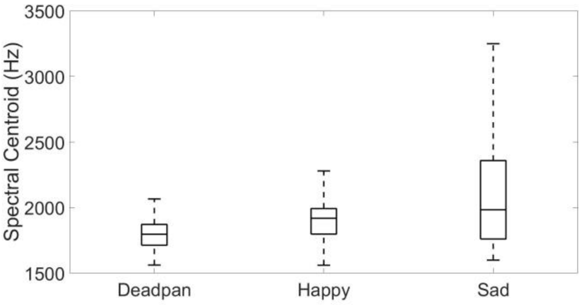

Figure 5. Boxplot of spectral centroids of all the notes for the three expressive goals: *happy*, *sad* and *deadpan*.

3.2 Playing Parameters

This case study only analysed two playing parameters: blowing pressure and DC reed position. Previous studies [9, 11, 12] have shown the influence of blowing pressure, bite force and bite position on fundamental, sound level, and transient behaviour. Here, the DC reed position, which is related to both the player's bite on the reed and the blowing pressure, was used as a simple parameter for initial exploration.

Figure 6 is the boxplot calculated from the blowing pressure (the DC pressure in the mouth) of all the notes for the three EGs. The box shows the median for all notes and the variation is the note-to-note variation; variation within notes is not shown here. *Happy* shows the highest median blowing pressure (4.2 kPa) and largest variation in blowing pressure. The median blowing pressure for *deadpan* and *sad* are 3.6 kPa and 3.4 kPa, respectively, but *sad* shows a larger variation than *deadpan*. This is consistent with the observations from Figure 3, because the sound level of the radiated sound is expected to correlate with blowing pressure over the relatively low values used here [9]. The correlation coefficient is 0.56.

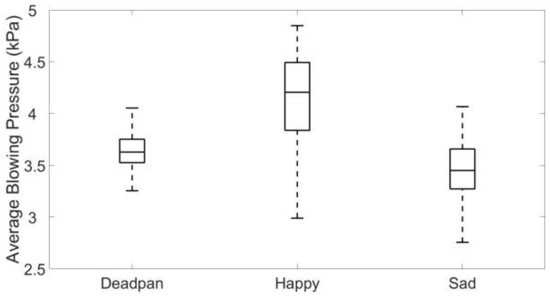

Figure 6. Boxplot of blowing pressure of all the notes for the three expressive goals: *happy*, *sad* and *deadpan*.

Figure 7 is the boxplot of DC reed displacements of all the notes for the three EGs. In Figure 7, zero on the y axis (not shown) corresponds to the reed position where the reed is at rest position without any displacement, and a more negative value corresponds to more reduced opening of the reed-mouthpiece aperture. (The player's bite produces a negative displacement.) *Happy* shows the least reduced aperture (lowest reed displacement, −0.89 mm) and largest variation in reed position among the three EGs, followed by *deadpan* (−0.92 mm) and *sad* (the most reduced opening, with displacement −0.94 mm). On its own, a higher blowing pressure tends to close the reed, giving a more negative reed displacement, so one might naïvely expect a more negative displacement for *happy*, where higher blowing pressure is used. However, the opposite is shown here. The explanation is that, in order to put the reed in an operating position to produce a sound with the similar sound level at lower blowing pressure, more lip force is required [9]. So, lip force needs to be increased further when playing with lower blowing pressure at lower sound level. This indicates that the clarinettist has applied more bite force when using lower blowing pressure to play *sad*.

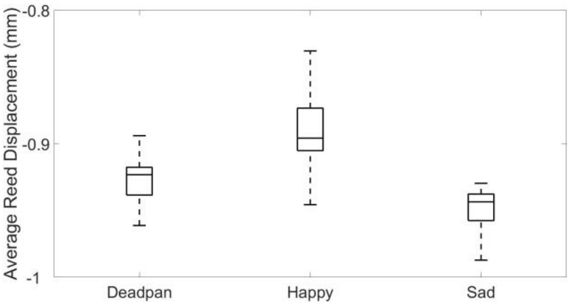

Figure 7. Boxplot of reed displacements of all the notes for the three expressive goals: *happy*, *sad* and *deadpan*. Higher (less negative) position means larger reed-mouthpiece aperture.

3.3 Inferential Statistical Analysis

	ANOVA	*happy* wrt *deadpan*	*sad* wrt *deadpan*	*sad* wrt *happy*
Sound level	F(2, 146) = 25.57 p < .001	↑	↓	↓
Tempo	F(2, 146) = 8.33 p < .001	ns	↓	↓
Spectral centroid	F(2, 146) = 9.67 p < .001	ns	↑	ns
Blowing pressure	F(2, 146) = 60.46 p < .001	↑	↓	↓
Reed position	F(2, 146) = 104.78 p < .001	↑	↓	↓

Table 1. Comparing five parameters of the three EGs using analysis of variance and post-hoc Tukey's honest significance test. 'ns' stands for non-significant result with adjusted p value larger than 0.05; ↑ and ↓ stand for larger and smaller, respectively. 'wrt' means 'with respect to'.

Analysis of variance (ANOVA) and post-hoc Tukey's honest significance test were conducted to check further the significance of differences in the musical and playing parameters per note when comparing the three EGs. As shown in Table 1, most of the items are significantly different (with adjusted p value less than .05), except tempo and spectral centroid between *happy* and *deadpan*, and spectral centroid between *happy* and *sad*. These results statistically confirm the observations discussed earlier.

4. CONCLUSIONS

This preliminary case study involved one clarinettist participant and used the MIPCAT toolbox to investigate experimentally (1) the differences in the musical parameters of three EGs *happy*, *sad* and *deadpan* and (2) how the player conveys different EGs controlling the playing parameters, when performing the same musical excerpt 'Happy Birthday to You'. Overall, notes played in the *happy* EG condition have high sound level and large standard deviation in sound level both within notes and across notes, fast tempo with medium variation, and moderate spectral change. This player used high blowing pressure with greatest variation and bite when interpreting *happy*. For the *deadpan* EG, the player selected medium blowing pressure and bite with little variation, resulting in a performance with medium sound level with minimum variation, fast and stable tempo, and minimum spectral change. For *sad*, low blowing pressure and bite with large variation were used to produce notes showing low sound level but large variation both within notes and across notes, slow tempo with large variation and large spectral variation. The analysis along the note sequence also revealed that *sad* has a very different pattern from *happy* and *deadpan*: the last two phrases have greater variation in the sound level and the peak sound level falling on the third last note (G4) instead of the highest note in the third phrase. Musical and playing features like these could be useful hints for music students to learn playing the music more expressively, especially when a larger sample of players has been measured.

While the present study focussed on just two of the measured playing parameters, MIPCAT offers potential in understanding how playing parameters interact to produce particular musical parameters. This will help to further bridge the gap between what the player is physically doing to the instrument during playing, and the aesthetic, musical product.

This study was limited to one player performing one musical excerpt and only the basic musical and playing parameters were investigated qualitatively. Thus the observations on comparing the three EGs could be unique for this particular player, e.g. similar tempo used for *deadpan* and *happy*. A future study with more players and detailed analysis would be ideal for identifying general regularities.

As a musical instrument performance capture and analysis toolbox, MIPCAT includes both hardware and software. The software part can process and analyse recordings of player performances captured by the hardware semi-automatically, reducing the work involved in the analyses reported here; it makes possible a more ambitious study on a group of clarinettist participants; this is currently being undertaken. Much of the software could, with relatively little modification, be applied to other instruments. For this reason, the software components are publicly available [15].

Acknowledgments

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

5. REFERENCES

[1] A. Gabrielsson and P. N. Juslin, "Emotional expression in music performance: Between the performer's intention and the listener's experience," in Psychology of Music, vol. 24, no. 1, pp. 68-91, 1996.

[2] P. N. Juslin, "Cue utilization in communication of emotion in music performance: Relating performance to perception," in Journal of Experimental Psychology: Human Perception Performance, vol. 26, no. 6, p. 1797, 2000.

[3] P. N. Juslin, "Perceived emotional expression in synthesized performances of a short melody: Capturing the listener's judgment policy," in Musicae Scientiae, vol. 1, no. 2, pp. 225-256, 1997.

[4] P. N. Juslin, "Can results from studies of perceived expression in musical performances be generalized across response formats?" in Psychomusicology: A Journal of Research in Music Cognition, vol. 16, no. 1-2, p. 77, 1997.

[5] A. Gabrielsson and E. Lindström, "The role of structure in the musical expression of emotions," in Handbook of Music and Emotion: Theory, Research, Applications, vol. 367400, 2010, pp. 367-44.

[6] G. Kreutz, E. Schubert, and L. A. Mitchell, "Cognitive styles of music listening," in Music Perception, vol. 26, no. 1, pp. 57-73, 2008.

[7] G. P. Scavone, A. Lefebvre, and A. R. da Silva, "Measurement of vocal-tract influence during saxophone performance," in The Journal of the Acoustical Society of America, vol. 123, no. 4, pp. 2391-2400, 2008.

[8] P. Guillemain, C. Vergez, D. Ferrand, and A. Farcy, "An instrumented saxophone mouthpiece and its use to understand how an experienced musician plays," in Acta Acustica united with Acustica, vol. 96, no. 4, pp. 622-634, 2010.

[9] 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," in The Journal of the Acoustical Society of America, vol. 134, no. 3, pp. 2247-2255, 2013.

[10] V. Chatziioannou and A. Hofmann, "Physics-based analysis of articulatory player actions in single-reed woodwind instruments," in Acta Acustica united with Acustica, vol. 101, no. 2, pp. 292-299, 2015.

[11] W. Li, A. Almeida, J. Smith, and J. Wolfe, "The effect of blowing pressure, lip force and tonguing on transients: A study using a clarinet-playing machine," in The Journal of the Acoustical Society of America, vol. 140, no. 2, pp. 1089-1100, 2016.

[12] W. Li, A. Almeida, J. Smith, and J. Wolfe, "How clarinettists articulate: The effect of blowing pressure and tonguing on initial and final transients," in The Journal of the Acoustical Society of America, vol. 139, no. 2, pp. 825-838, 2016.

[13] M. Pàmies-Vilà, A. Hofmann, and V. Chatziioannou, "Analysis of tonguing and blowing actions during clarinet performance," in Frontiers in Psychology, vol. 9, p. 617, 2018.

[14] A. Almeida, W. Li, E. Schubert, J. Smith, and J. Wolfe, "Expressive goals for performing musicians: The case of clarinetists," in Musicae Scientiae, p. 10298649221122155, 2022.

[15] 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)," in Frontiers in Signal Processing, vol. 3, p. 1, 2023.

[16] A. Almedia, W. Li, E. Schubert, J. Wolfe, and J. Smith, "MIPCAT — A Music Instrument Performance Capture and Analysis Toolbox," in Stockholm Music Acoustics Conference, Stockholm, 2023.

[17] P. N. Juslin, "Emotional communication in music performance: A functionalist perspective and some data," in Music perception, 14(4), 383-418, 1997.

[18] E. Schubert and J. Wolfe, "Does timbral brightness scale with frequency and spectral centroid?" in Acta Acustica united with Acustica, vol. 92, no. 5, pp. 820-825, 2006.

[19] Y. Xu, A. Kelly, and C. Smillie, "Emotional expressions as communicative signals," in Prosody and Iconicity, 33–60, 2013.