Detecting efficiency in trumpet sound production: proposed methodology and pedagogical implications

Alberto Acquilino* **Ninad Puranik*** **Ichiro Fujinaga** **Gary Scavone**

Department of Music Research, Schulich School of Music, McGill University
Centre for Interdisciplinary Research in Music Media and Technology (CIRMMT)
555 Sherbrooke St W., Montreal, Canada H3A 1E3

alberto.acquilino@mail.mcgill.ca ninad.puranik@mail.mcgill.ca

ichiro.fujinaga@mcgill.ca gary.scavone@mcgill.ca

ABSTRACT

In this study, we consider the case of the trumpet to study the role of timbre quality from the perspectives of music pedagogy and music information retrieval. Prominent brass pedagogues have reported that the presence of excessive muscle tension and inefficiency in playing by a musician is reflected in the timbre quality of the sound produced, which is easily distinguished by an experienced ear. A technological tool that provides an automatic feedback on the tone quality can be of immense help for new learners to develop good playing habits during independent practice. To develop such a tool, an extensive dataset consisting of more than 19,000 tones played by 110 trumpet players of different expertise has been collected. We manually labelled a subset of 1,711 notes from this dataset with a grade on a scale of 1 to 4 based on the perceived efficiency of sound production. A classifier model with a mean 10-fold cross validation accuracy above 80% was developed based on the extracted audio features to predict the level of efficiency. Finally, we present an interface for the application of this model in pedagogical contexts, in the framework of commercially available music education systems.

1. INTRODUCTION

The use of technological systems aimed at supporting musical instrument learning is a burgeoning area, especially since the spread of portable devices. Educational technologies are designed to assist music students in various aspects of their study, such as checking accuracy of pitch and rhythm [1], fingering [2], musician's form [3], performance [4], and dynamics control [5, 6].

* Equal contribution

Machine learning has improved the potential for research in music pedagogy through its application to audio signal processing [7]. Systematic statistical analysis of sound features from the audio recordings can be done to assess the music performances automatically [8]. Of particular interest is the study of timbre quality in variable pitch instruments, such as bowed strings, woodwinds, and brasses, as it has an educational value and a connection to the performance quality [9–11].

Considering the trumpet as a case study, the role of timbre quality from both pedagogical and academic research perspectives has garnered considerable interest.

1.1 The musician's perspective of tone quality

Sound production on a trumpet involves a complex coordination and balance between the embouchure, the oral cavity, and the airflow [12]. Numerous literature sources in the field of music performance report how tone quality is closely related to the level of efficiency of the sound production mechanism of the player. It is a widespread belief in trumpet teaching that inefficiencies in playing caused by sub-optimal coordination of the muscular movements of the musician are reflected in the quality of the sound produced.

Arnold Jacobs, one of the foremost brass pedagogues of his time, reports how the presence of excessive muscle tension, which causes rigidity in the musician's body and inefficiency in playing, is reflected in a forced and strained sound [13]. Similarly, James Thompson argues that an incorrect and inefficient sound production mechanism obliges the player to force, leading to a decreased endurance and a strident tone [14]. Analogous observations are mentioned by Campbell et al. who argue that obtaining a good sound depends on the skill of the musician as well as the quality of the instrument, and that timbre plays a significant role in the ability to project the sound [15].

Similar conclusions are also drawn by Kristian Steenstrup, according to whom "it is more efficient, from the point of view of both the lip and respiratory musculature, to produce a beautiful, round tone rich in harmonics than a shrill or dull tone" and points to a lack of thorough research in this area [16]. Steenstrup proposes a parallelism between the sound production mechanism of trumpet players and that of singers, for whom there is more research in the literature. In particular, he hypothesizes that the influence between different types of phonation (i.e., pressed, flow, breath phonation) on the acoustic spectrum - as described by Sundberg and Gauffin [17] - may find a counterpart in the way sound is produced for brass musicians. It should be noted that this idea was proposed as a way to visualize or better understand sound production for trumpet players, without suggesting that the actual acoustic mechanisms involved are similar.

Other sources report that timbre provides most of the information to music teachers on which to base the choice of suggestions offered to their students, consequently influencing the determination of the educational path through targeted exercises [12, 18].

What is reported in the pedagogical field has also a correspondence in the context of orchestra auditions and competitions, where some of the rehearsals hide the performer from the jury's view in order to limit the introduction of bias in the selection of candidates. This suggests that an experienced ear, such as that of the jury members, is able to distinguish the level of efficiency and ease of sound production based solely on auditory information.

1.2 The researcher's perspective of tone quality

In the field of Music Information Retrieval (MIR), researchers are exploring the use of algorithms to identify the quality of trumpet tones from the audio information. The goal is to investigate the possibility of training a model that can assess the timbre quality of a trumpet sound like a qualified teacher. Knight et al. trained classifier models based on 56 mono and multidimensional audio features into two, three and seven classes [19]. Although the dataset used by this study, comprising only 239 notes played by four trumpeters, was limited, the results provided a promising proof of concept for future research in this field.

A few years later, the Music Technology Group at Pompeu Fabra University (MTG-UPF) conducted a research study on similar premises in collaboration with KORG Inc. targeting several musical instruments, including the trumpet [9, 20]. Having created an online platform for collecting, labeling and evaluating audio samples, they proposed a model for the assessment of the sound quality based on five attributes, namely dynamic stability, pitch stability, timbre stability, timbre richness, and attack clarity. Among these attributes, timbre richness - defined as the quality of timbre - is the one that comes closest to that considered by Knight et al. and suggested by brass teachers. The MTG-UPF study has significant limitations due to the lack of diversity in the dataset, consisting of sound samples collected from just two trumpet players. In order to train the model, the two musicians - professional music performance graduates - were required to record each of the attributes correctly and intentionally incorrectly. This could introduce further limitations in that it is not necessarily true that the sound of a professional musician who intentionally plays badly presents the same audio characteristics as the incorrect sound of a novice trumpet player, since the musician should be perfectly capable of dissociating from years of procedural memory training [21].

In this research, we aim to address the above limitations by training a model using an extensive dataset of sounds produced by 110 trumpet players of different expertise, to develop a system that has a better performance in real life applications. We report here our initial developments of this tool, as this is a work in progress. Finally, we present an interface for the application of this model in pedagogical contexts, in the framework of commercially available music education systems.

In Section 2, we describe the characteristics of the recorded dataset. Section 3 discusses the methodology adopted to develop the software and presents the results collected. Section 4 introduces the pedagogical framework of application. Finally, Section 5 discusses an interactive interface that implement the designed algorithm to support effective trumpet learning.

2. DATASET DESCRIPTION

Extensive data collection of trumpet sounds was conducted in music schools and masterclasses in low-noise environments with different acoustic conditions. Two sets of microphones were held 50 cm in front of the trumpet bell about 10 cm off its longitudinal axis for each recording.

The IM69D130 Shield2Go evaluation board manufactured by Infineon Technologies was used, equipped with two Infineon IM69D130 Micro-Electro-Mechanical Systems microphones, and interfaced to a Raspberry Pi Model 3B+ and a Raspberry Pi 4 Model B. Audio data were acquired with a bit depth of 32 and a sampling rate of 48kHz. These hardware components were selected since they use the same technologies as in mobile devices while ensuring a significantly high maximum sound pressure level (i.e., max SPL of 130dB) for recording the trumpet without distortion. This decision was made in order to consequently train a more robust model for use in mobile apps and software directly using the built-in microphone in smartphones, tablets, or PCs.

The complete dataset includes recordings of approximately 19,000 tones collected from 110 different trumpet players. The musicians included students from amateur music schools, students and teachers from conservatories and universities of the arts, music performance graduates, professional orchestral musicians and international soloists. The performers were playing their own trumpet with their own mouthpiece and were recorded using a set of two microphones held side by side. Tuning with respect to a reference pitch was not enforced, as the timbre quality is expected to be independent of reference pitches.

Each musician was asked to play separate long tones along a chromatic scale from E3 to B♭5 three times at different dynamics: once *piano*, once *mezzo forte*, and once *forte* in their order of preference. The duration of the notes, which is required to be at least 1 second in order to have sufficient margin to apply time-variant feature analysis afterwards, ranges from 0.7 to 4 seconds. Despite being asked to play separate tones, several beginners, for whom playing a chromatic scale in front of a recorder constituted a challenging task, played legato notes. Given their evident difficulty in managing the suggested tasks, and that the sounds acquired in that way still contained the timbral characteristics sought, they were not asked to repeat the recording.

To provide clearer guidance on the desired loudness variation (e.g., *piano*, *mezzo forte*, and *forte*), musicians were shown a digital sound level meter displaying the decibel level produced, along with the target reference levels (i.e., 85dB, 105dB, 115dB). This approach aimed to enhance the variability within the dataset, considering the significant impact of loudness level on timbre [22]. Although the performers did not always maintain the exact dB level requested (particularly beginners), the presence of the sound level meter proved beneficial in directing their attention towards differentiating the loudness levels they were playing.

Although musicians were asked to play a chromatic scale along the entire frequency range of the trumpet (i.e., E3 to Bb5), the recordings of trumpet players with less expertise is not complete, missing the notes in the highest register. Indeed, it is common knowledge that playing high notes with the trumpet requires an advanced level of control and coordination between embouchure, oral cavity, and air flow.

The audio dataset described above was independently collected by the first author before the start of his academic program at the host institution. The recording conditions, player's level of expertise and an overall grade on the quality of the sound produced were annotated at the time of recording each track.

The collected dataset encompasses a diverse array of audio samples, providing a suitable environment for training a machine learning model to differentiate the timbre quality level of a trumpeter in real-world scenarios. Section 3 outlines the methodology applied to tackle this challenge, resulting in the development of a robust model that can accurately assess the tone quality of trumpet performances.

3. METHODOLOGY (ML MODEL)

In this section, we present the preliminary work on the analysis of the collected dataset and training a machine learning model to distinguish the efficiency of a trumpet player's sound production from the audio data.

The audio dataset described in the previous section was segmented into single trumpet tones using the pyin vamp plugin by Mauch and Dixon [23]. While the segmentation was accurate in most cases, some issues were noted upon closer inspection. The issues included clipped audios, background talk being segmented as a note, etc. It was thus deemed necessary to manually listen to the notes before training a machine learning model directly using the

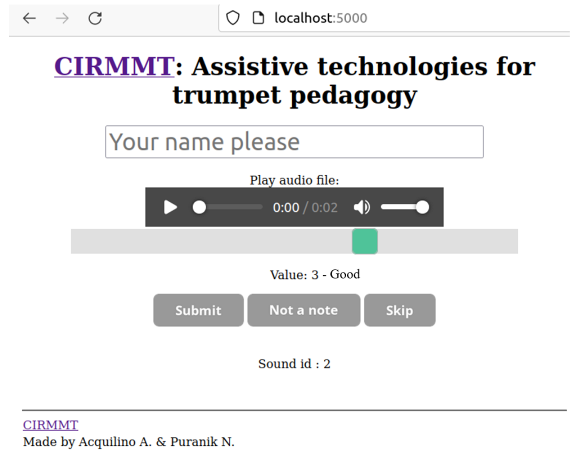

Figure 1. Interface for blind grading the trumpet notes

preliminary grades as the labels. Due to practical difficulty in listening to each of the 19,000+ notes, we decided to base the machine learning model on a sufficiently large and balanced subset of the complete dataset.

In this study, we aimed to develop a 4-level quality classifier for the sounds in the audio dataset, as proposed by Wesolowski [24], aiming to simplify the label assignment process while maintaining sufficient variability. The classifier levels were defined as 1:poor, 2:fair, 3:good, 4:excellent. From the preliminary grades assigned at the time of recording, seventeen players were selected: five of them were assigned to one class and four performers to each of the other classes. Indeed, as beginners were unable to perform the complete chromatic scale, one player more was assigned to the 1:poor category in order to counterbalance the reduced representation of audio samples within that particular class. This selection gave a total of 1,711 single notes which were manually labeled by the first author who is a graduate in trumpet performance with professional experience as a musician and a brass teacher.

To facilitate the labeling process, an online interface was developed for assigning unbiased grades. The single notes of the selected musicians were uploaded to the web page, where the grader could listen repeatedly without knowing the identity of the musician or their class. The grader could then assign a grade from 1 to 4, corresponding to the level of timbre quality, as shown in the Figure 1.

In the case where a sound had been poorly segmented, the grader could indicate that the sound was not a note. The decision to evaluate one tone at a time was made to identify and remove wrongly segmented notes, thus obtaining a clean dataset of 1,481 tones. On analysis of the ratings assigned to individual notes through the web interface and the ratings assigned to the full performance of the musicians at the time of recording, we observed that they had a Pearson correlation coefficient of 87.5% (p-Value$<$0.001). The high correlation may not appear surprising given that the same person (the first author)

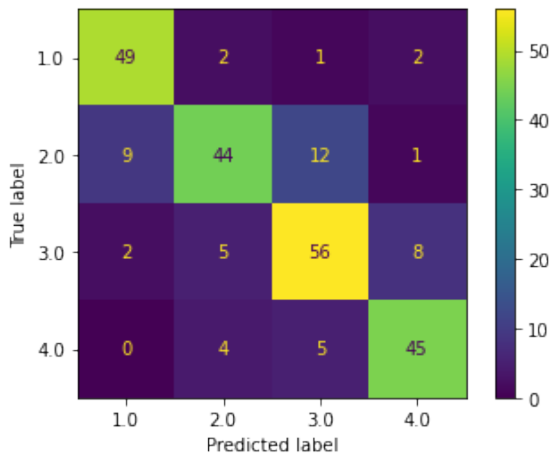

Figure 2. Confusion matrix for a model trained with 75-25 train-test split

made both the ratings. However, it is significant that the first author's perception of the tone quality at the aggregate and the granular levels remained consistent.

The cleaned audio dataset of 1,481 tones was used to fit a Random Forest Classifier model [25], an ensemble learning method that leverages multiple decision trees for enhanced accuracy in classification tasks. As a preprocessing step, the sound samples were first scaled to have a maximum signal amplitude equal to one and white noise at -60dB was added to the audio. The audio features for each tone were then extracted using the essentia Extractor algorithm [26]. To reduce the computational complexity, we used only the statistical aggregates of the audio features such as mean, variance, mean of derivative, etc. As a first step, we included all except the rhythm based features (a total of 1,230 features) to fit a Random Forest Classifier. For a 10-fold cross-validation we achieved a mean accuracy score of 78%.

We used the Random Forest Classifier model-based feature selection to identify the top 256 features for the classification. Using only these features, the 10-fold cross validation mean accuracy score improved slightly to 81.37%.

We fit a model by splitting the dataset into two parts: 75% of the data is used for training (i.e., train set) and the remaining 25% is used for testing (i.e., test set). The confusion matrix, which indicates the performance assessment of the model on the test set, can be seen in Figure 2. It can be observed that the most confusion is between the adjacent classes. Since the audio samples in the adjacent classes are in fact more similar to each other than the other classes, the errors seem to be reasonable. In fact, it is quite possible that for some audio samples there would be such a disagreement in the opinions of two human experts, or the same expert at different times.

To evaluate the out-of-dataset performance of the trained model, the latter was tested using the sound samples of trumpet from the Good-sounds dataset curated by Pompeu Fabra University [9,20], as previously described in Section 1.2. The Good-sounds dataset has 25 notes labelled as 'bad timbre richness'; however, there is no specific label for the 'good timbre richness' notes. Therefore, the model was tested on sounds labelled as 'good', which may include different tonal qualities (e.g., pitch stability, dynamics stability) in addition to 'timbre richness'. All notes labeled as 'bad timbre richness' were given grade 1, while about 70% of the notes labeled as 'good' received a grade 3 or 4 by the model. The results are reported in Table 1.

Good-sounds labels	**Model grades**			
	1	**2**	**3**	**4**
Bad (n=25)	100%	0%	0%	0%
Good (n=190)	30%	0.5%	9.5%	60%

Table 1. Effectiveness of trained RF Classifier Model tested on the Good-sounds database

An analysis of the sounds labeled as 'good' to which a grade equal to 1 was assigned reveals that those notes predominantly had a high vibrato extent, significantly more than was present in the training dataset.

These findings were further substantiated by conducting informal trials, wherein the model was assessed in real time alongside graduate trumpet players, thereby suggesting promising potential for pedagogic applications. Subsequent training with a set of experienced trumpet teachers is in progress.

4. PEDAGOGICAL FRAMEWORK

The field of edTech systems for assessing trumpet tone quality is relatively limited. Furthermore, the few technologies that address this issue may be limited in their pedagogical effectiveness, as discussed by Acquilino and Scavone [27]. One example of this is KORG cortosia[1], which estimates the trumpet sound "goodness" by rating in real time five elements: pitch stability, dynamic stability, timbre stability, timbre richness, and attack clarity. An overall numerical score is provided along with a five-axis visualization corresponding to each of the five elements. The complexity of the designed interface may have affected perceived ease of use and limited its adoption by music students.

Other systems, such as TonalEnergy[2], provide real-time visualization of the spectrum and/or height of partials. However, this kind of feedback is likely too complex to be interpreted by music students and even teachers, given the high number of variables involved in identifying tone quality.

This suggests the need for further development of user-friendly and pedagogically effective edTech solutions for assessing trumpet tone quality. A call for new technologies to improve self-regulated music learning, other than student-teacher interaction, is also supported by a recent thorough study conducted by Waddell and Williamon [28].

[1] www.korg.com/us/products/software/cortosia
[2] www.tonalenergy.com

Where effective musical instrument learning is concerned, the deliberate practice framework introduced by Ericsson et al. [29] assumes a crucial role, given its acknowledged success in research and music performance. Deliberate practice refers to the use of specific and targeted activities, with immediate feedback, that are designed to improve specific aspects of performance. Specifically, Ericsson and Harwell [30] expressed the following four criteria that define purposeful/deliberate practice:

1. *The practice involves individualized training of a trainee by a well-qualified teacher. This teacher can assess which aspects a particular trainee would be able to improve during the time until the next meeting and is able to recommend practice techniques with established effectiveness*.

2. *The teacher must be able to communicate the goal to be achieved by the trainee and that the trainee can internally represent this goal during practice*.

3. *The teacher can describe a practice activity to attain the identified goal for performance and that this activity allows the trainee to get immediate feedback on a given attempt*.

4. *The trainee is able to make repeated revised attempts that gradually approach the desired goal performance*.

Criterion 1 emphasizes the significance of an experienced teacher in facilitating the learning of music students. This suggests the idea of developing technologies that support brass instructors in becoming more effective, extending their pedagogical potential. In addition, Criteria 2, 3 and 4 provide guidelines for the design of such technologies, particularly with regard to the role they can play in assisting the adoption and engagement of deliberate practice. Criterion 2 suggests introducing new ways for music students to understand the concepts of technical skills to be learned with their instruments. This could be achieved through a multi-sensory experience (e.g., combining auditory and visual feedback). Criterion 3 recommends the design of technologies that provide real-time feedback to musicians. Criterion 4 highlights significant pedagogical potential toward systems that can objectively quantify the level of accuracy of sound produced by the performer with reference to a specific technical skill (in this case, timbre quality). In fact, specific synchronous feedback facilitates students' growth in achieving the targets set with the teacher [31]. In this way, music students can objectively track their progress over time and identify strengths and weaknesses for targeted practice.

The assumption is that designing educational systems that follow the criteria above will result in increased perceived usefulness among teachers and music students. The Technology Acceptance Model by Davis explains that these perceptions of usefulness and ease of use create an intention to use the technology, and this intention leads to actual usage [32].

Chapter 5 presents a technological exercise that aims to fulfill the premises discussed in order to stimulate effective study of tone quality by trumpet students.

5. PROPOSED EDUCATIONAL TECHNOLOGY

This section presents an innovative musical exercise aimed at promoting the acquisition of timbre qualities by trumpet students. The exercise is designed to be interactive in real-time, allowing for flexible adjustments by the teacher.

The proposed edTech system prompts the user to enter the metronome value, the lowest and highest note they wish to play between E3 and B♭5. These values are fixed and shown in the top row of the interface in Figure 3. Thus, the exercise starts by presenting a musical score of three four-quarter measures with a random note to be played between the minimum and maximum values selected, as shown in the third row of Figure 3. The first measure is a pause, the second measure shows in gray the selected note that the software plays in time with a trumpet sound of high timbre quality, and the third measure shows the same note to be played by the musician. At the end of the third measure, another score is generated with another randomly selected note and the sequence is repeated for a given number of times. In order to aid proper time keeping, a metronomic "tic" and a synchronized numerical display of the corresponding beat number in the first column of the second row of Figure 3 are included. The sound played by the software during the second measure was previously recorded by a professional trumpet player. It was chosen to use a recording of a fine performer, instead of simply using MIDI sounds, to induce the user to emulate a high-quality timbre by imitation. In order to broaden the range of stimuli for students, multiple high-quality recordings can be provided for each note and randomly selected.

The model presented in Section 3 is implemented within the software by providing near real-time feedback on the level of tone quality played by the user. The sound is analyzed by the algorithm which returns feedback corresponding to the class in which it is associated. The four different emoticons 🤔 😊 😃 🤩 are associated with classes 1, 2, 3, 4 respectively appearing in the second row on the right side of the interface in Figure 3.

This exercise aims to stimulate deliberate practice in music students regarding the acquisition of high-quality timbre, which many brass pedagogues associate with an efficient sound production mechanism. The inclusion of the model within an interactive exercise with gamified experience is intended to foster users' motivation. The possibility of defining the range of notes that can appear in the score ensures that the teacher can gradually approach trumpet students toward the high register. In fact, in the book *Brass Techniques and Pedagogy*, Weidner reports as a common problem among trumpet players the change in tone quality when playing in the high register during lip slurs [33]. The author states that it can be counterproductive to try to play high notes with the trumpet without being able to play with good tone quality lower notes, pointing out as potential undesirable effects the development of bad habits involving excessive

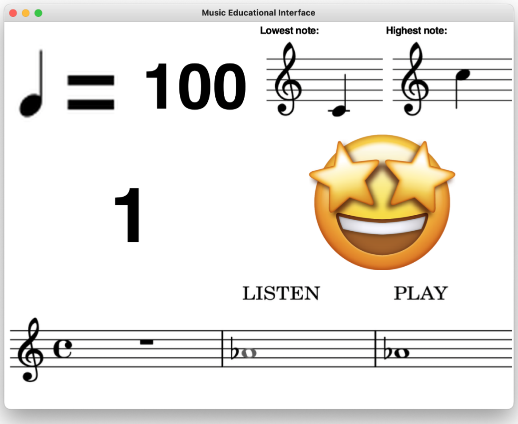

Figure 3. Interface of the proposed edTech system. Top row from left to right shows the input parameters: metronome value, lowest note, and highest note. The middle row shows the current beat value on the left and an emoji feedback on the quality of the timbre produced on the right. At the bottom, a three-measure score is shown cyclically updating with a random note to be played between the highest and lowest values selected. Additional resources about the interface can be found at `https://pninad.github.io/smac2023/`

pressure, lip pinching, and poor air support. The exercise presented thus offers the teacher control over the students' note range development while also providing direct real-time feedback on the quality of tone played. This is intended to make the user focus on the development of this skill by stimulating deliberate practice while avoiding the aforementioned problems.

The developed interface could also provide music instructors with a summary of students' achievements between in-person lessons. In this way, they can observe their learning curve in acquiring sound efficiency, identify learning criticalities, and propose targeted practices to overcome them. The algorithm devised can also be flexibly applied to other interfaces. The mere functionality of emoticons that change according to the level of timbre quality can, for example, be used directly in stand-alone mode, such as chromatic tuners. Or, trumpet teachers can enter the sequence of notes for students to play, instead of relying on a random choice.

6. CONCLUSIONS

This paper presented the topic of timbre quality for the trumpet from both the perspective of music performance teachers and scientists. Through the use of an extensive dataset, the researchers were able to overcome the limitations of previous work and create a model that can distinguish the level of tone quality with a mean 10-fold cross validation accuracy of about 80%.

The authors also proposed an innovative edTech system that utilizes the Technology Acceptance Model framework to encourage deliberate practice in trumpet students, with the goal of improving their tone quality. The proposed interactive exercise can be flexibly adapted by the teacher to meet the individual needs of their students. This approach has the potential to not only enhance trumpet timbre quality, but also to improve teacher-student interaction.

A limitation of the present investigation concerns the use of only a single rater for the training of the model. Such an approach may be susceptible to the influence of rater bias, consequently restricting the generalisability of the results. In an effort to mitigate this inherent limitation, the authors are undertaking a study that incorporates multiple raters in the process of assigning labels to the dataset. This approach is intended to facilitate the training of a model that is less susceptible to the impact of potential individual biases, thereby enhancing the robustness and applicability of the findings.

Future research aims to expand this study exploring the acoustical properties of trumpet sound efficiency by analyzing the main discriminatory audio features identified by the model. In addition, a further investigation is ongoing to evaluate the effectiveness of the proposed system in music pedagogy, focusing on its ability to support trumpet students in learning sound production efficiency.

Overall, this work aims to provide novel solutions in the field of technology-enhanced music learning and provides a valuable resource for trumpet teachers, students, and researchers alike.

Acknowledgments

This work was made possible with the support of a CIRMMT Student Award and a Tomlinson Doctoral Fellowship.

The authors thank the foundational contribution of Mirko d'Andrea and Emanuela Bussino for the audio data collection stages, as well as all the volunteers and colleagues who supported this research.

7. REFERENCES

[1] S. Raptis, A. Chalamandaris, A. Baxevanis, A. Askenfelt, E. Schoonderwaldt, K. F. Hansen, D. Fober, S. Letz, and Y. Orlarey, "Imutus - an effective practicing environment for music tuition," in *Proceedings of 2005 International Computer Music Conference*, Barcelona, Spain, 2005, pp. 383–386.

[2] S. Bagga, B. Maurer, T. Miller, L. Quinlan, L. Silvestri, D. Wells, R. Winqvist, M. Zolotas, and Y. Demiris, "Instrumentor: An interactive robot for musical instrument tutoring," in *20th Annual Conference on Towards Autonomous Robotic Systems*, 2019, pp. 303–315.

[3] Y. Tsuji and N. Atsuhiro, "Development and evaluation of drum learning support system based on rhythm and drumming form," in *Electronics and Communications in Japan (Part III: Fundamental Electronic Science)*, 2006, pp. 11–21.

[4] E. Schoonderwaldt, A. Askenfeld, and K. Hansen, "Design and implementation of automatic evaluation of recorder performance in imutus," in *Proceedings of 2005 International Computer Music Conference*, International Computer Music Association, Barcelona, Spain, 2005, pp. 97–103.

[5] M. Robine, G. K. Percival, and M. Lagrange, "Analysis of saxophone performance for computer-assisted tutoring," in *Proceedings of the 2007 International Computer Music Conference*, 2007, pp. 381–384.

[6] G. K. Percival, Y. Wang, and G. Tzanetakis, "Effective use of multimedia for computer-assisted musical instrument tutoring," in *Proceedings of the International Workshop on Educational Multimedia and Multimedia Education*, 2007, pp. 67–76.

[7] A. Lerch and P. Knees, "Machine learning applied to music/audio signal processing," in *Electronics*, 2021.

[8] P. Seshadri and A. Lerch, "Improving music performance assessment with contrastive learning," in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, 2021, pp. 634–641.

[9] G. Bandiera, O. Romani, H. Tokuda, W. Hariya, K. Oishi, and X. Serra, "Good-sounds.org: a framework to explore goodness in instrumental sounds," in *Conference: 17th International Society for Music Information Retrieval Conference*, New York, 2016.

[10] S. Giraldo, R. Ramirez, G. Waddell, and A. Williamon, "A computational approach for measuring performance quality in violin tones," in *International Symposium in Performance Science*, Reykjavik, 2017, pp. 2231–2240.

[11] S. Giraldo, G. Waddell, I. Nou, A. Ortega, O. Mayor, A. Perez, A. Williamon, and R. Ramirez, "Automatic assessment of tone quality in violin music performance," in *Frontiers in Psychology*, 2019, pp. 1–12.

[12] F. G. Campos, *Trumpet technique*. Oxford University Press, 2005.

[13] A. Jacobs and B. Nelson, *Also Sprach Arnold Jacobs: A Developmental Guide for Brass Wind Musicians*. Polymnia Press, 2006.

[14] J. Thompson, *The buzzing book complete method; trumpet or other brass instruments*. Editions BIM, 2003.

[15] M. Campbell, M. Gilbert, and A. Myers, *The Science of Brass Instruments*. Modern Acoustics and Signal Processing, 2021.

[16] K. Steenstrup, *Teaching Brass*. Det Jyske Musikkonservatorium, 2007.

[17] J. Sundberg and J. Gauffin, "Waveform and spectrum of of the glottal voice source," in *Frontiers of speech communication research*, London: Academic Press, 1978, pp. 301–320.

[18] G. Cassone, *The trumpet book*. Zecchini, 2009.

[19] T. Knight, T. Upham, and I. Fujinaga, "The potential for automatic assessment of trumpet tone quality," in *Proc. of the 12th Int. Society for Music Information Retrieval Conf., ISMIR*, 2011, pp. 573–578.

[20] O. Romani, H. Parra, D. Dabiri, H. Tokuda, W. Hariya, K. Oishi, and X. Serra, "A real-time system for measuring sound goodness in instrumental sounds," in *138th Audio Engineering Society Convention, AES*, Warsaw, Poland, 2015, pp. 1106–1111.

[21] M. Lam, "The physicality of music production: Investigating the roles of mindful practice and kinesthetic learning," in *Music Educators Journal*, 2020, pp. 23–28.

[22] D. Luce and M. J. Clark, "Physical correlates of brass-instrument tones," in *The Journal of the Acoustical Society of America*, 1967, pp. 1232–1243.

[23] M. Mauch and S. Dixon, "Pyin: A fundamental frequency estimator using probabilistic threshold distributions," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 659–663.

[24] B. C. Wesolowski, "Understanding and developing rubrics for music performance assessment," in *Music Educators Journal*, 2012, pp. 36–42.

[25] T. K. Ho, "Random decision forests," in *Proceedings of 3rd international conference on document analysis and recognition*, vol. 1. IEEE, 1995, pp. 278–282.

[26] D. Bogdanov, N. Wack, G. E. E., S. Gulati, P. Herrera, and O. M. et al., "Essentia: an audio analysis library for music information retrieval," in *International Society for Music Information Retrieval Conference*, 2013, pp. 493–498.

[27] A. Acquilino and G. Scavone, "Current state and future directions of technologies for music instrument pedagogy," in *Frontiers in Psychology*, 2022.

[28] G. Waddell and A. Williamon, "Technology use and attitudes in music learning," in *Frontiers in ICT*, 2019.

[29] K. A. Ericsson, R. T. Krampe, and C. Tesch-Römer, "The role of deliberate practice in the acquisition of expert performance," in *Psychological Review*, 1993, pp. 363–406.

[30] K. A. Ericsson and K. W. Harwell, "Deliberate practice and proposed limits on the effects of practice on the acquisition of expert performance: Why the original definition matters and recommendations for future research," in *Frontiers in Psychology*, 2019.

[31] L. Martin, “Using synchronous formative feedback to facilitate student growth,” in *Music Educators Journal*, 2020, pp. 51–57.

[32] F. D. Davis, “Perceived usefulness, perceived ease of use, and user acceptance of information technology,” in *MIS Quarterly*, 1989, pp. 319–340.

[33] B. Weidner, *Brass Techniques and Pedagogy*. Palni Press, 2020.