An Overview of Automatic Piano Performance Assessment within the Music Education Context

Hyon Kim, Pedro Ramoneda, Marius Miron and Xavier Serra
Music Technology Group, Universitat Pompeu Fabra, Spain

Keywords: Piano Performance Assessment, Music Education, Pedagogy, Music Information Retrieval.

Abstract: Piano is one of the most popular instruments among music learners. Technologies to evaluate piano performances have been researched and developed in recent years rapidly, including data driven methods using machine learning. Despite the demand from people and speed of the development, there are still gaps between the methods and the pedagogical setup for real use case scenarios due to lack of accuracy of methods, insufficient amount of training data or the biases in training machine learning models, ignoring actual use case of the technology and such. In this paper, we first propose a feedback approach in piano performance education and review methods for Automated Piano Performance Assessment (APPA). After that, we discuss about gaps between a feedback approach and current methods, emphasizing their music education application. As a future work we propose a potential approach to overcome the gaps.

1 INTRODUCTION

In recent years, many computer based technological and engineering methods to help to assess music performance have been actively researched (Eremenko V, 2020; Lerch and Gururani, 2020). These methods automatically evaluate learners' performance on each or all dimensions of music such as dynamics, expressiveness, rhythm, techniques, timbre, pitch and chords. These dimensions are considered essential characteristics to measure music performance (Eremenko V, 2020). (Lerch and Gururani, 2020) refers to tempo, timing, and dynamics as most salient elements in Music Performance Assessment (MPA). Similarly, piano performance is a composite of multiple dimensions of musical skills such as dynamics or loudness, tempo or rhythm, and techniques such as posture of hands and body and expressiveness.

The MPA methods have the potential to be widely applied to many piano education scenarios. Not only do the methods help students to learn to play the piano, but also support teachers to teach students. However, as characteristics of piano, polyphonic sound, pedalling and the percussive aspect of the instrument make performance analysis challenging. (Müller, 2007; Grachten and Widmer, 2012) focus on assessing particular dimension of piano performance such as rhythm, pitch, dynamics/volume, posture, expressiveness. (Parmar et al., 2021) take performed audio as input and map it to students' skill level directly using an end-to-end machine learning approach.

This paper proposes a feedback approach for piano education and reviews current methods used for Automatic Piano Performance Assessment (APPA). Gaps between the effective feedback approach and methods for APPA are discussed and we propose future research directions.

In Section 2, we propose a framework for generating feedback for students to learn to play the piano. In Section 3, current automatic methods to assess students' piano performance are reviewed. In Section 4, gaps between the current methods and their educational usage are discussed, proposing potential research directions.

2 A FRAMEWORK OF GENERATING FEEDBACK FOR STUDENTS

In the context of education, students should be at the center of the system, aimed at supporting piano learning. In this context, we emphasize the importance of feedback from the system. (Hattie and Timperley, 2007) define a feedback approach as follows. First, feedback requires a common goal between student and teacher. Second, by having a common goal, it

Kim, H., Ramoneda, P., Miron, M. and Serra, X.
An Overview of Automatic Piano Performance Assessment within the Music Education Context.
DOI: 10.5220/0011137600003182
In *Proceedings of the 14th International Conference on Computer Supported Education (CSEDU 2022) - Volume 1*, pages 465-474
ISBN: 978-989-758-562-3; ISSN: 2184-5026

defines a strategy to fill up the gap between current assessment and the goal. Third, an effective feedback is defined by satisfying three points shared by them:

- Where am I going?
- How am I going?
- Where to next?

As an example, the goal may be to pass an examination for a grade, to play a new musical score and perform it in front of people or even to acquire a new specific technique such as trill, pedalling and so on, when applying this model of feedback to piano education. After defining the goal, the teacher has to identify the current situation and measure gaps towards the goal such as gap between currently playable music scores and desired music score to play, finding measures where the student makes mistake often. Therefore, an assessment system able to satisfy the feedback model must consider the initial goal and a chronological progress of student and teacher. The figure 1 shows the model with examples in the case of piano education.

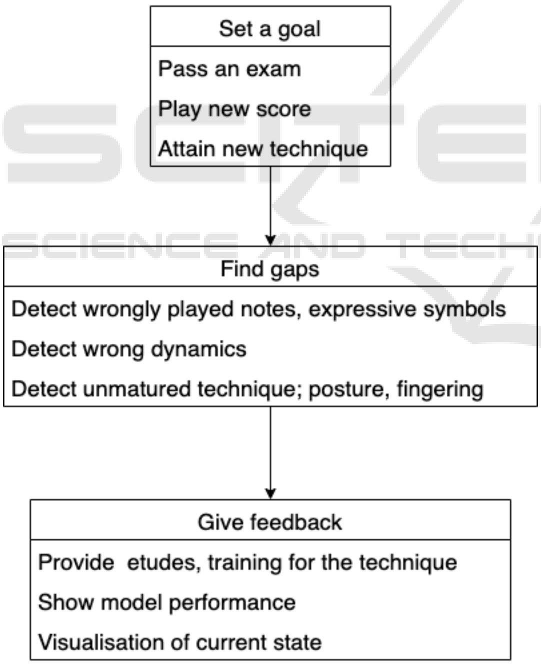

Figure 1: Examples of Feedback Model in Piano Education.

In order to set the feedback model for students' improvement, the APPA must achieve at least the following;

- To be robust and and to detect performance errors with high accuracy, similarly to teachers.
- To offer feedback customized for each student based on a goal and data of chronological progress and feedback.

An automatic piano assessment system with low accuracy to support lessons may dampen piano practice and may increase teachers' work even more. The accuracy of each methods to detect targets should be high enough to maintain the above goals for the system. Giving customized feedback is also a condition of the feedback model and this aspect is achieved by accumulating data such as piano performance, attitude towards practice. A customized feedback to a student based on chronological data of performance may be utilized to give suggestions such as how to practice and recommend appropriate musical works personalised to the skills of each student (Ramoneda et al., 2022). Assessing the motivation of students prevents them from quitting to play the piano and engage more to learn to play the piano and their goals. An APPA system which achieves the above points may also reduce teachers' effort in preparing the class, such as taking and keeping records for each lesson, and pointing out mistakes often seen. Thus, the methods of APPA meet the conditions to form the piano education feedback model.

3 METHODS FOR PIANO PERFORMANCE ASSESSMENT

In this section, research topics within APPA and well-used and new approaches for the research topics are reviewed. The research area covers both methods to assess dynamics and expressiveness, tempo, techniques individually and overall performance assessment using end-to-end machine learning methods.

3.1 Methods Related to Measure Expressiveness in Piano Performance

An interesting research problem is finding potential new features to measure the dimensions (Busse, 2002; Goebl, 2001). The measurement of expressiveness in context of music performance has been researched from different perspectives. One way to put a definition of expressiveness in music performance is a result of manipulation and realization of markings regarding expressiveness such as *p, f, tie, slur, rit.* and rhythm in a music score by a pianist. Therefore, it can be considered as a combination of technique and imagination interpretation of pianists.

Data Analysis for Finding Expressiveness. In order to analyse expressiveness quantitatively, we have seen development on new devices to acquire performance data and format for serialization of the data.

As an example, MIDI data contains various features related to expressiveness; note velocity (data of velocity of pressing a key) in MIDI is considered as loudness of the pressed key. (Busse, 2002) processed the MIDI data as potential features to measure the jazz pianists' expressiveness. The difference between actual performance and music score, such as note placement to a predefined tempo, note duration to a quarter note, note velocity as dynamics and tempo variation are considered as features. These features of expressiveness and a relation between the features are analysed in statistical manner. As a result, statistically significant results are reported when classifying pianists using the set of four features from MIDI data taken from a piano performance.

(Bernays and Traube, 2013) analyses timbral nuances, dry, bright, round, velvety and dark by harvesting data from performance on a special type of piano, Bösendorfer with the CEUS system. Four pianists performed four pieces of solo piano music specifically composed for this research with the five different timbres, three times for each piece. The authors explored 966 features obtained from the device. The number of features is reduced to 13 at the end of analysis by Principal Component Analysis (PCA) and a choice based on authors' empirical decision, in order to describe the five timbral classes. The 13 features chosen are as follows; (right, left, both hand) hammer velocity, Key depression, variations in key attack speed, (right, left, both hand) attack duration, soft pedal depression, sustain pedal use, sustain pedal depression, release duration and right hand chords overlap. Those features illustrates each timbre by getting visualized by a Kiviat chart. For example, timbre of bright showed high intensity, very short attacks, the soft pedal is barely used, the sustain pedal is used sparingly but is strongly depressed when in use. This has a potential usage for grasping mental conception, imitation and careful self-actualization in performance as guided by the musical ear.

In piano performance, regarding the harmonic characteristics, research focused on melody (voicing) line and accompaniment. To account for that, MIDI data has been utilized to analyse the performance. Difference on performance onset times of synchronously written notes in score has been investigated between melody (voicing) line and accompaniment line by MIDI data (Repp, 1996). It has been shown that a clear melody lead, i.e. observed onset on melody line, is detected earlier than the one of the accompaniment. Another paper (Goebl, 2001) also shows that melody line is played louder by measuring hammer velocity on a string. By processing velocity data from MIDI and focusing on differences between hands, the classification task of expert and amateur pianist was conducted (Kim, 2021). This research involved experts and amateur pianists playing Hannon Exercise No. 1 and C-Major scale and observing the difference in data separated by hands. The authors observed that experts tended to have greater difference between hands in dynamics than the amateurs' performance.

Modeling Expressiveness. Apart from finding new features to describe expressiveness, methods for modeling of expressiveness directly have been also researched (Phanichraksaphong and Tsai, 2021; Phanichraksaphong and Tsai, 2021; Kosta, 2016; Grachten and Widmer, 2012). The piano scores contain musical symbols for expression such as *pp, p, f, ff, slur, crescendo, staccato* and more. In addition, models for evaluating expressive performance focusing on staccato and legato have been researched (Phanichraksaphong and Tsai, 2021).

The goal of expressiveness modeling is to classify performance of staccato and legato into three classes; good, normal and bad using machine learning models; Support Vector Machine (SVM), Long and Short Term Memory (LSTM), Convolutional Neural Network(CNN) and Bayesian Network (BN) (Phanichraksaphong and Tsai, 2021). The data is collected from piano performances of children from kinder garden to junior high school. The features chosen for this research are Mel-frequency Cepstral Coefficients (MFCC). Machine learning models are created separately for each expressive marking. In both cases, the CNN based model outperformed the other models; 89% for staccato and 93% for legato. Further evaluation of the model is done with k-fold cross validation. This research showed that it is important to evaluate expressive markings in the music score and performance individually.

In another paper, outliers in dynamics is investigated (Kosta, 2016). The goal of this paper was to find changes in dynamics based on two consecutive dynamics marking; *pp, p, mp, mf, f, ff,* etc. The data set is created from the Mazurka project [1]. Dynamic Time Wrapping (DTW) (Müller, 2007) is employed for audio-to-score alignment. Focusing on the transition of dynamics between two consecutive markings, (Kosta, 2016) found that there is a tendency of dynamics in performance of each pianist based on outliers defined on the dynamics marking pair. This implies that the performers' interpretation as musical choice based on marking of dynamics in a music score can be distinguished and modeled. The research also

[1] http://www.mazurka.org.uk

showed that, using a model built on the same music score and played by different performers, dynamics curve can be predicted. However, if the model is created by the same performers for different score, model prediction does not work well.

(Grachten and Widmer, 2012) modeled dynamics in piano performance. The model takes a sequence of elements having information pitch onset time, duration and other none note element; *p, f, mf, crescendo*, etc. from MusicXML files. The model is a linear basis model and each vector represents note information such as dynamics markings, pitch, grace note, degree of closure and squared distance from the nearest position where closure occurs. The beat tracking part is done by BeatRoot (Dixon, 2001) and the evaluation of model is conducted on the Magaloff's corpus (Sebastian Flossmann and Widmer, 2010) and by computing metrics as Pearson product-moment correlation coefficient and coefficient of determination. The result showed that variance across pieces is too large to identify performer-specific expressive style. A linear basis modeling can be extended in two ways: probabilistic approach, and dictionary learning in sparse coding technique to learn basis function.

3.2 Tempo Estimation

Maintaining a constant tempo, constrained by each style and composer indicators is a skill that piano students progressively acquire. Although tempo estimation in Music Information Retrieval (MIR) focuses on getting the single global tempo, the term in piano performance assessment is linked to the local tempo across the musical piece. Local tempo is the ratio of events within a smaller time window and, i.e., a local deviation from the global tempo.

In the case of expressive piano performances, current beat tracking approaches do not capture local tempo deviations and beat positions, the data is limited, and most of the research has been conducted on the mazurkas dataset [2]. In one of the first contributions, (Grosche and Müller, 2010) assume that some passages are more critical for beat tracking and propose to evaluate them separately in order not to contaminate the tempo induction. In the early times, (Schreiber et al., 2020) propose several metrics to quantify and evaluate local tempo. In addition, (Schreiber et al., 2020) prove that a CNN-based approaches can measure local tempo on expressive piano music, but it is required to carry out domain adaptation in each genre.

[2]http://www.mazurka.org.uk

3.3 Piano Technique Evaluation

The most simple, laconic, and sensible description of the artistic technique was formulated by Alexander Blok "in order to create a work of art", he says, "one must know how to" (Neuhaus, 2008). The piano technique should serve the rest of the dimensions we further introduce in the present section. Therefore, there is a strong correlation between piano technique and dynamics, expressiveness, or tempo. However, in the following paragraphs, we present elements intrinsically focused on piano technique, hands and body posture assessment, fingering assessment, and finally, other elements not extensively researched.

Hand and Body Posture. Young students often miss the opportunity to develop motor (movement) skills, hand strength, or flexibility and develop progressively a proper hand and body position. In this regard, a pianist's posture and movement precision may have a significant influence on the quality of the performance. In the next paragraphs we review the research proposals to assess hand and body posture.

Regarding the body posture assessment, (Payeur et al., 2014) examines the potential of motion capture in the context of piano pedagogy and performance evaluation when somatic training methods are used. The input of the system is images to capture motion during piano performances using Xbox's kinect. The full body's movement is classified into four different classes. Each posture is taken and compared to baseline and checked if the postures are recognizable and differentiated. A drawback is that the paper does not analyze if the Kinect sensor is precise enough to track smaller differences in anatomical positioning, since it is hypothesized that changes to skeletal alignment as a result of somatic training are less exaggerated than the postures used for these initial tests. However, we commend the use of motion capture as it is the simplest and comfortable setup for the performer.

Regarding the hand posture assessment, (Johnson et al., 2020) use a depth map of hands' image to measure the quality of hand posture. The system classifies hand posture for piano performance in three classes: flat hands, low wrists, correct posture. The dataset is derived from the piano performance of 9-12 years old students during performance. The piano performances comprise piano exercises from a book, Dozen a Day and both hands are annotated separately. In its preprocessing step, a Random Decision Forest (RDF) is used to separate right and left hands. After computing depth image features and depth context features, the histogram of oriented gradients and histogram of normal vectors are used as features for SVM classification. From our point of view, although the dataset is small and biased, this article opens new research direction.

Piano Fingering. Piano fingering plays an important role in the realization of the music performance (Neuhaus, 2008) and it is learnt progressively (Ramoneda et al., 2022). Pianists adapt the fingering at each note according to the subsequent fingerings and notes (Nieto, 1988) while preserving the musical content of the piece: articulation, tempo, dynamics, rhythm, style and character (Nieto, 1988). In addition, the proposed piano fingering has to be as comfortable and simple (Brée, 1905) minimising the hand position changes.

Several previous papers tried to model piano fingering: expert systems (Sloboda et al., 1998; Parncutt et al., 1997), local search algorithms (Balliauw et al., 2017) and data-driven methods (Nakamura et al., 2014; Nakamura et al., 2020) Although they can be used to recommend and assess the student fingering election, only public dataset on piano fingering is available for 150 excerpts (Nakamura et al., 2020).

Elements such as piano technique patterns or articulations assessment are also very important technique factors. There is not much research conducted on assessing articulations. E.g. stacatto, legatto or portato, or isolated piano technique patterns. E.g. arpeggios, scales or broken chords. To the best of our knowledge (Akinaga et al., 2006) propose a system for assessing scales and as described in the section 3.1 . (Kim et al., 2021) research propose assessing performance focusing on the relationship between left hand and right hand.

3.4 End-to-End Machine Learning for Automatic Assessment

(Parmar et al., 2021), (Pan et al., 2017) take the audio, video and other features derived from the piano performance and maps these features to scores by utilizing machine leaning techniques. The audio input is usually pre-processed to a better representation and then it is given as input to a machine learning model. Popular representations are: Mel-Frequency Cepstral Coefficients (MFCC), fundametal frequency (*f0*), Constant-Q Transform (CQT), pitch by YIN (DeCheveigne, 2002) or pYIN (Mauch and Dixon, 2014), features generated by Deep Neural Network (DNN). An APPA task is a supervised learning and training data is labeled as performance score, skill level and such assessed by human. Evaluation of the model is usually done by coefficient of determination, F-values. Normally, DTW is employed for a music score and performed data alignment. This alignment is needed for matching the performed data and the score information to assess the performance, given a set of score or evaluation metrics.

As an example of such system is proposed by (Parmar et al., 2021). The method takes audio video together and maps them to skill level (ranging 1 to 10) of a performer using a multi-modal input. The labels to be predicted are related to the performer's skill level, song difficulty level, name of the song, a bounding box around the pianist's hands. The model starts from two different branches as input for audio and video. The input of audio is processed to mel-spectrogram as 2-dimenisional input and corresponding video clips are given as input from another branch. Data is collected by the authors of the study; 61 total piano performances are taken from Youtube and skill level is annotated manually. The video branch is pre-trained using UCF101-Action Recognition dataset (Soomro et al., 2012). The best performing model was the one that combined visual and aural elements and its accuracy to predict a performer's skill level is about 74.6%. This research showed that a DNN multi-modal approach using posture, audio and score related data may be used to assess piano performance

(Pan et al., 2017) propose a method to evaluate student's performance by developing pitch detection. The method takes the audio of the performance as input and processes it to CQT spectrum. A trained DNN acoustic model sorts it to 88 dimension keys of piano. The DTW is used to find mistakes in the performance. The data used for training dataset is the MAPS dataset (Emiya et al., 2010) and the data used for evaluation is their own house collected data set. The research obtain parameters out of the note detection model and conducted linear regression to get a final score set to 0-100. The evaluation metrics are F-value and Mean Absolute Error (MAE) respectively. The main contribution of this research is that the DNN based acoustic model can be used to extract piano key features for piano performance evaluation.

As we can see these examples above, machine learning methods have clearly a huge potential and is feasible to assess the music skills and grade score on the performance.

In terms of an appropriate length of performance for evaluation, (Wapnick et al.,) found that performance of one minute is rated consistent than shorter performance and significant rater activity was not observed longer performance length than the first one minute.

3.5 Piano Performance Transcription

Another method of assessing piano performance is comparing the audio performance and the associated musical score to find differences. On this purpose, music transcriptions assumes deriving symbolic representations from signals and may help the comparison with the score. (Benetos et al., 2012) propose a system that takes the audio performance of the student as input and compares it to MIDI based score information in order to check for mistakes. Pitch estimation is done by Non-Negative Matrix Factorization (NMF) with β -divergence and note tracking is done by Hidden Markov Models (HMMs). Although the assumptions made by the paper are strong, such as detecting onsets of note only, the method involving score information proposes a heuristic and potential way to assess piano performance assessment. Therefore, taking the offset of performed notes into consideration is one of important future work in this context of assessment and pedagogy. Bayesian approaches, classic signal processing methods, Neural Networks (NNs), NMF are common and well researched approaches to tackle to music transcription based on music audio or MIDI input (Benetos et al., 2019) In particular, NMF and NNs actively researched in resent years.

3.6 Pedagogical Set-up with Other Technologies

A pedagogical-driven system for piano performance assessment integrating other technologies including machine learning is proposed (Jianan, 2021), (Masaki et al., 2011), (Hamond, 2019). The pedagogical systems are grouped into two types: systems using teacher supervision, systems without teacher supervision and fully automated systems. A system fully automated and utilizing current modern network technologies such as Internet of Things (IoT) and algorithms to evaluate students' performances automatically is proposed (Jianan, 2021). The system takes an audio of student's performance and uploads it to a cloud system. Then it gives an performance evaluation result and saves user performance history. This proposed application as baseline system may include machine learning methods mentioned in sections above to evaluate students' performance online and fully automated fashion.

Self-evaluation is one of the most feasible methods for student evaluation, as discussed in (Masaki et al., 2011). This research employed the idea from sport education and applied it to piano learning. The method let students watch their performance and make comparison evaluating video feedback by student themselves.

(Hamond, 2019) proposed a pedagogical setup which involves a teacher into the system. The system gives importance to the feedback from the teacher including non-verbal feedback when the teacher is playing, singing, or modelling, the teacher imitating the student's playing, making hand gestures or body movements, such as conducting and tapping the pulse. This is to help student to understand differences between his/her own performance and the ground truth performance. The proposed system takes MIDI data from student's performance and gives visual feedback for every stroke of keys, such as played note, the length of time during which note was pressed and released, and the velocity of a pressed note, the action of the pedals can also be measured. This visual feedback was helpful for the student to understand what actually happened during the performance and also promoted attentive listening. A change of dynamics could also be seen when looking at the solo performance of a student and comparing with the duo performance together with the teacher. This promotes student to be aware of how they are playing at each note and bar.

On mobile and web, there are various kinds of piano performance assessment applications such as SimplyPiano [3], Skoove Piano [4] and flowkey [5]. From their user interface, there are many features to get piano learners engaged to practicing piano such as easy navigation on user-interface under practice purposes, equipped accompaniment to practice and play together, gamification to get attention and continuation of learners and so on. The strength of those applications is definitely providing a good self-taught environment for both children and adults.

4 FUTURE WORK

Gathering all the methods mentioned above would form an integrated system to evaluate and assess student's performance automatically and point out student's mistakes based on a piano score information. However, the accuracy has not reached the teacher's level. Looking at machine learning methods, detailed and meaningful feedback to students and assist teaching are sometimes ignored and missing. As we can see in the Table 1, data used to develop methods are still small, biased, genre specific and some are not

[3]https://join.joytunes.com/
[4]https://www.skoove.com/en
[5]https://www.flowkey.com/en

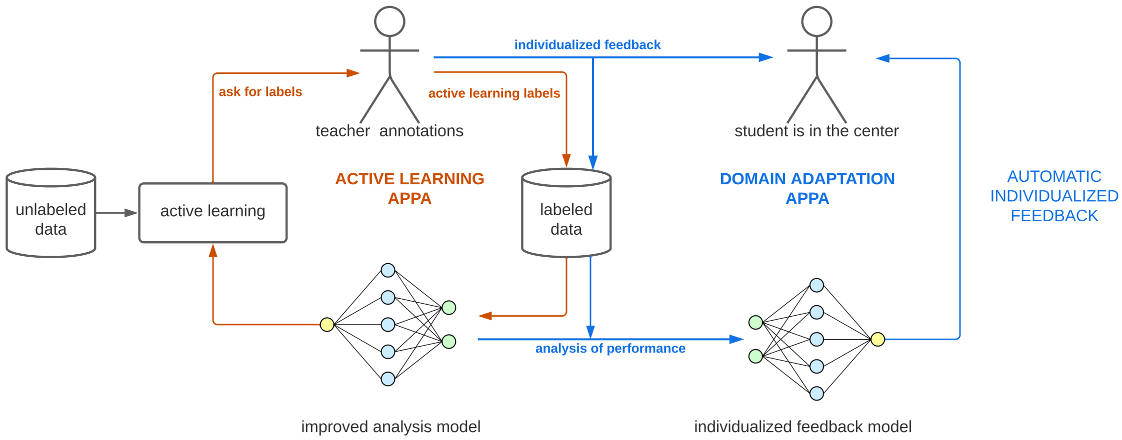

Figure 2: The vision of a potential APPA system with active learning and domain adaptation methodologies. LEFT: The active system scheme. For each iteration, the active learning system chooses unlabeled samples to be assessed by a teacher. Later, the teacher annotates that samples and the model for analyzing the piano performance assessment are improved with the new labeled data. RIGHT: The domain adaptation system scheme. The teacher gives individualized feedback to the student. A model is trained with domain adaptation for each student to give feedback based on the performance analysis model.

Table 1: Referenced Datasets and its Contents.

Reference	Dataset	Availability
(Busse, 2002)	Self collected MIDI-based data: 281 of one-measure performance	N.A.
(Kim, 2021)	Self collected MIDI-based data: 34 experts (Major piano in Univ.) and 34 amateurs (play as hobby), Played Score: Hanon Exercise No. 1 and a C-Major scale	upon request
(Repp, 1996)	Self collected MIDI-based data: 10 piano majored graduate students performed Traumerei by Schumann, La fille aux chevbuex de lin by Debussy, Prelude by Chopin.	N.A.
(Goebl, 2001)	Self collected data via a Bosendorfer SE 290 computer-monitored concert grand piano: 22 skilled pianists performed the Etude op. 10/3 (first 21 measures) and the Ballade op. 38 (initial section, bars 1 to 45, Fig. 3) by Chopin.	http://www.ai.univie.ac.at/~wernerg
(Kosta, 2016)	Mazurkas by CHARM project: An expert pianist performed Mazurkas by Chopin	http://www.mazurka.org.uk/
(Grachten and Widmer, 2012)	Self collected loudness data: Loudness measurements from commercial CD recordings	N.A.
(Grachten and Widmer, 2012)	The Magaloff corpus: Works for solo piano by Chopin	(Sebastian Flossmann and Widmer, 2010)
(Phanichraksaphong and Tsai, 2021)	Self collected audio data: Nine piano students in three groups; kindergarten, elementary school, junior high school. A total of 1080 note datasets were used for training.	upon request
(Parmar et al., 2021)	PIano Skills Assessment (PISA) dataset: A dataset containing visual and auditory information.	https://paperswithcode.com/dataset/multimodal-pisa
(Pan et al., 2017)	YCU-MPPE-II dataset: A piano teacher labeled score each piece as the ground truth label. There are five different polyphonic music.	upon request
(Pan et al., 2017), (Benetos et al., 2012)	MAPS dataset: A dataset which consists of 270 pieces of piano sound and corresponding MIDI annotations.	upon request
(Payeur et al., 2014)	Self collected dataset: A dataset which recorded four beginning piano students while performing standard practice exercises.	N.A.
(Nakamura et al., 2020)	PIG Dataset: PIG Dataset (PIano fingernG) consists of piano pieces by Western classical music composers with fingerings annotated by pianists.	https://beam.kisarazu.ac.jp/~saito/research/PianoFingeringDataset/

openly available. Moreover, the research of APPA system is still far away to the educational purposes to give feedback in their APPA system; defining a goal, find gaps between current student's state and the goal and generate feedback for the student. The state of the art methods are still not chronological and evaluation and feedback is not customized for student. However, using the current methods, appropriate and effective feedback cannot be generated and given to students. One reason causing this situation is the difficulty of data collection. Moreover, data used for in the current state of the art papers tend to be domain (e.g. genre, level of difficulty) specific. Hence the machine learning models tend to be biased. (Johnson et al., 2020) tackled to data set related problems by oversampling technique and adaptive synthetic sampling. These methods are potential to be adapted to the other APPA researches.

A goal of the future systems for APPA should be to acquire data containing the student's progress and the teacher's feedback. To that extent, the data generated during the teaching piano can be used to optimize: (a) the current APPA algorithms' performance, (b) how a teacher gives feedback to students, and (c) the feedback between each teacher and each student, to automate the teacher's efforts to focus in more creative educational tasks instead of repetitive ones. To this end, we want to use active learning and domain adaptation in the following projects to achieve objectives a, b and c. However, this is not a well-researched topic. It is challenging to get student data during a certain time period on regular basis. In this regards, teachers must be involved into the system so that the quality of data and its annotation are guaranteed.

In order to get correct labels and customize the machine learning models for each student, one may rely on active learning (Settles, 2011). Active learning is a methodology to improve a model based on human annotations with as little effort as possible. It is based on labeling samples with higher epistemic uncertainty, i.e., the model knows less about the parameters and can explain worse why a specific model output is realized, further information on the review article (Sinha et al., 2019). In the recent years, the machine learning community has proposed several research papers to carry out active learning, from Bayesian active learning (Gal et al., 2017) to loss-based approaches (Yoo and Kweon, 2019) through adversarial active learning (Sinha et al., 2019). Domain adaptation involves transferring the knowledge from the data distribution of a high-performance model to different but related target data distribution, further reading on (Wang and Deng, 2018). This approach may be used to individualize the APPA system for each student and teacher. Figure 2 shows a potential example of how we may combine these two techniques to create APPA systems grounds on human-centered AI (Xu, 2019)

The form of output from the system for a student centered pedagogical environment may be visually displayed for intuitive feedback, model audio, etc. Recent research analyses the difficulty of piano musical works together with the individualised difficulties of each student (Ramoneda et al., 2022), the dimensions where the student fail the most, it is possible to help teacher to create a very personalised and diverse curriculum for each student. This direction of research on methods of APPA lights up the future for both student and teachers and reach out to more music enthusiasts to enjoy learning piano.

5 CONCLUSION

In this paper, a feedback system and the researches related to APPA including industry applications are reviewed. We discussed the gaps between them and proposed potential methods to bring the APPA system to the next stage under the purpose of students' improvement on their piano performance. Finally, in our future work we aim at fostering collaborations between pedagogy, music and technologies in engineering such as machine learning.

ACKNOWLEDGEMENT

This project is supported by Project Musical AI funded by the Spanish Ministerio de Ciencia, Innovación y Universidades (MCIU) and the Agencia Estatal de Investigación (AEI).

REFERENCES

Akinaga, S., Miura, M., Emura, N., and Yanagida, M. (2006). Toward realizing automatic evaluation of playing scales on the piano. In *Proceedings of International Conference on Music Perception and Cognition*, pages 1843–1847. Citeseer.

Balliauw, M., Herremans, D., Palhazi Cuervo, D., and Sörensen, K. (2017). A variable neighborhood search algorithm to generate piano fingerings for polyphonic sheet music. *International Transactions in Operational Research*, 24(3):509–535.

Benetos, E., Dixon, S., Duan, Z., and Ewert, S. (2019). Automatic music transcription: An overview. *IEEE Signal Processing Magazine*, 36(1):20–30.

Benetos, E., Klapuri, A., and Dixon, S. (2012). Score-informed transcription for automatic piano tutoring. In *2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO)*, pages 2153–2157.

Bernays, M. and Traube, C. (2013). Expressive production of piano timbre: Touch and playing techniques for timbre control in piano performance.

Brée, M. (1905). *The groundwork of the Leschetizky method*. G. Schirmer.

Busse, W. G. (2002). Toward objective measurement and evaluation of jazz piano performance via midi-based groove quantize templates. *Music Perception:An Interdisciplinary Journal*, 19(3):443–461.

DeCheveigne, A., . K. H. (2002). Yin, a fundamental frequency estimator for speech and music. *The Journal of the Acoustical Society of America*, 111(4):1917–1930.

Dixon, S. (2001). Automatic extraction of tempo and beat from expressive performances. *Journal of New Music Research*, 30(1):39–58.

Emiya, V., Bertin, N., David, B., and Badeau, R. (2010). MAPS - A piano database for multipitch estimation and automatic transcription of music. Research report.

Eremenko V, Morsi A, N. J. S. X. (2020). Performance assessment technologies for the support of musical instrument learning.

Gal, Y., Islam, R., and Ghahramani, Z. (2017). Deep bayesian active learning with image data. In *International Conference on Machine Learning*, pages 1183–1192. PMLR.

Goebl, W. (2001). Melody lead in piano performance: expressive device or artifact? *The Journal of the Acoustical Society of America*, 110(1):563–72.

Grachten, M. and Widmer, G. (2012). Linear basis models for prediction and analysis of musical expression. *Journal of New Music Research*, 41(4):311–322.

Grosche, P. and Müller, M. (2010). What Makes Beat Tracking Difficult? A Case Study on Chopin Mazurkas. In *in Proceedings of the 11th International Society for Music Information Retrieval Conference, 2010*, pages 649–654.

Hamond, LF; Welch, G. H. E. (2019). The pedagogical use of visual feedback for enhancing dynamics in higher education piano learning and performance. *OPUS*, 25(3):581–601.

Hattie, J. and Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1):81–112.

Jianan, Y. (2021). Automatic evaluation system for piano performance based on the internet of things technology under the background of artificial intelligence. *Mathematical Problems in Engineering*, 2021:14.

Johnson, D., Damian, D., and Tzanetakis, G. (2020). Detecting hand posture in piano playing using depth data. *Computer Music Journal*, 43(1):59–78.

Kim, S., P. J. M. R. S. N. J. . L. K. (2021). Quantitative analysis of piano performance proficiency focusing on difference between hands. *PloS one*, 16(5).

Kim, S., Park, J. M., Rhyu, S., Nam, J., and Lee, K. (2021). Quantitative analysis of piano performance proficiency focusing on difference between hands. *PloS one*, 16(5):e0250299.

Kosta, K., O. F. B. E. C. (2016). Outliers in performed loudness transitions: An analysis of chopin mazurka recordings. In *International Conference for Music Perception and Cognition (ICMPC)*, pages 601–604, California, USA.

Lerch, A., A. C. P. A. and Gururani, S. (2020). An interdisciplinary review of music performance analysis. *Transactions of the International Society for Music Information Retrieval*, 3(1):221–245.

Masaki, M., Hechler, P., Gadbois, S., and Waddell, G. (2011). Piano performance assessment: video feedback and the quality assessment in music performance inventory (qampi). In *International Symposium on Performance Science 2011*, pages 24–27, Canada.

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

Müller, M. (2007). *Information Retrieval for Music and Motion*, chapter Dynamic Time Warping. Springer Berlin Heidelberg, Berlin, Heidelberg.

Nakamura, E., Ono, N., and Sagayama, S. (2014). Merged-output hmm for piano fingering of both hands. In *Proceedings of the 13th International Society for Music Information Retrieval Conference*, ISMIR 2014, pages 531–536, Taipei.

Nakamura, E., Saito, Y., and Yoshii, K. (2020). Statistical learning and estimation of piano fingering. *Information Sciences*, 517:68–85.

Neuhaus, H. (2008). *The art of piano playing*. Kahn and Averill.

Nieto, A. (1988). *La digitación pianística*. Fundación Banco Exterior.

Pan, J., Li, M., Song, Z., Li, X., Liu, X., Yi, H., and Zhu, M. (2017). An Audio Based Piano Performance Evaluation Method Using Deep Neural Network Based Acoustic Modeling. In *Proc. Interspeech 2017*, pages 3088–3092.

Parmar, P., Reddy, J., and Morris, B. (2021). Piano skills assessment.

Parncutt, R., Sloboda, J. A., Clarke, E. F., Raekallio, M., and Desain, P. (1997). An Ergonomie Model of Keyboard Fingering for Melodic Fragments Sibelius Academy of Music , Helsinski. *Music perception: An Interdisciplinary Journal*, 14(4):341–382.

Payeur, P., Nascimento, G. M. G., Beacon, J., Comeau, G., Cretu, A.-M., D'Aoust, V., and Charpentier, M.-A. (2014). Human gesture quantification: An evaluation tool for somatic training and piano performance. In *2014 IEEE International Symposium on Haptic, Audio and Visual Environments and Games (HAVE) Proceedings*, pages 100–105. IEEE.

Phanichraksaphong, V. and Tsai, W.-H. (2021). Automatic evaluation of piano performances for steam education. *Applied Sciences*, 11(24).

Ramoneda, P., Tamer, N. C., Eremenko, V., Miron, M., and Serra, X. (2022). Score difficulty analysis for piano performance education. In *ICASSP 2022 IEEE Inter-*

national Conference on Acoustics, Speech and Signal Processing (ICASSP).

Repp, B. H. (1996). Patterns of note onset asynchronies in expressive piano performance. *The Journal of the Acoustical Society of America*, 100(3917).

Schreiber, H., Zalkow, F., and Müller, M. (2020). Modeling and estimating local tempo: A case study on chopin's mazurkas. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR), Montreal, Quebec, Canada*.

Sebastian Flossmann, W. G. and Widmer, G. (2010). The magaloff corpus: An empirical error study. In *The 11th International Conference on Music Perception and Cognition (ICMPC11)*.

Settles, B. (2011). From theories to queries: Active learning in practice. In Guyon, I., Cawley, G., Dror, G., Lemaire, V., and Statnikov, A., editors, *Active Learning and Experimental Design workshop In conjunction with AISTATS 2010*, volume 16 of *Proceedings of Machine Learning Research*, pages 1–18, Sardinia, Italy. PMLR.

Sinha, S., Ebrahimi, S., and Darrell, T. (2019). Variational adversarial active learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5972–5981.

Sloboda, J. A., Clarke, E. F., Parncutt, R., and Raekallio, M. (1998). Determinants of finger choice in piano sight-reading. *Journal of experimental psychology: Human perception and performance*, 24(1):185.

Soomro, K., Zamir, A. R., and Shah, M. (2012). Ucf101: A dataset of 101 human actions classes from videos in the wild.

Wang, M. and Deng, W. (2018). Deep visual domain adaptation: A survey. *Neurocomputing*, 312:135–153.

Wapnick, J., Ryan, C., Campbell, L., Deek, P., Lemire, R., and Darrow, A.-A. Effects of excerpt tempo and duration on musicians' ratings of high-level piano performances. 53(2):162–176. Publisher: SAGE Publications Inc.

Xu, W. (2019). Toward human-centered ai: a perspective from human-computer interaction. *interactions*, 26(4):42–46.

Yoo, D. and Kweon, I. S. (2019). Learning loss for active learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 93–102.