A Web-Based Platform to Teach Music Online

Fatemeh Jamshidi

Computer Science and Software Engineering
Auburn University
Auburn, Alabama, USA
Email: fzj0007@auburn.edu

Daniela Marghitu

Computer Science and Software Engineering
Auburn University
Auburn, Alabama, USA
Email: marghda@auburn.edu

***Abstract*—The development of educational technology has encouraged music educators to consider different ways to teach music online. However, this change may need an online platform to support teachers when the size of the class is large. This paper aims to address two research questions. The first question is to determine the learning experiences of students and to understand their needs to support Web-based learning. The second question is to study the key technologies required for a Web-based music teaching system. In this paper, we try to determine the best ways to motivate students to study music and to make music teaching experiences more accessible, engaging, and fun for students.**

Keywords–Web based learning; Music education; Self-directed learning; Self-assessment.

I. Introduction

With advancements in Computer Science (CS) technologies, the music discipline needs to make full use of new innovations to provide high-quality music learning resources to students. Applications of multimedia and Web-based learning technologies in teaching music can help teachers to keep students motivated and engaged. Some researchers have investigated students' Web-based music learning experiences [1]. The results have shown that multimedia teaching has many advantages:

1) A rich audio-visual experience can effectively enhance student's engagement and accelerate music learning through online activities.
2) Friendly interactive environments can increase the enthusiasm of the students in learning and practicing music.
3) The new way of teaching is in accordance with cognitive patterns.

With the use of blended strategies [2], this research tries to enhance student engagement and music learning through online activities in music courses. It also improves the effectiveness and efficiency of music classes by reducing lecture time and increasing the practice time. In the future, the traditional distinction between class time and non-class time will disappear.

The current generation is using technologies that connect single learners and collaborative varieties of learning to socialize, work, and learn [3]. Technology-based platforms to teach music should be able to cover resources for collaborative learning as well as individual capabilities. A Web-based learning environment offers a novel informal platform where individuals with different music backgrounds can learn and practice music. The aim of this research is to see the essential elements of current online learning approaches utilized in academic music courses and to contemplate their application within the development of an Internet pedagogical framework to show music online. Our primary prototype is designed to teach people with little to no music background, the basics of music theory and how to play the piano.

The rest of the paper is organized as following: In Section II, we cover some background about the project. Section III introduces the methodology of the website and covers the design and implementation of the website. Section IV presents the evaluation methods. Section V concludes the project along with suggestions for future enhancements.

II. Related Work

A. Score Following

Score following [4][5] is a technique in music technology to track the performance of the player in a given music score and is one of the most important components in automatic music accompaniment. The two most important components in score following are to express [6]:

1) the similarity between the current observation and the expected observation in each position in the music score that the player is playing.
2) the allowed temporal evolution of the score position.

Score following systems mostly do not recognize visual cues, that human musicians use for coordinating parts of the music without any musician playing [7]. For example, nodding gestures can be used to synchronize the introduction of a song. Some studies use visual information, such as the periodic hand motion of a player [8]. Visual cue is an important component when the audio signal is not enough for tracking the human musicians.

B. Pitch Tracking

A reliable estimate of the pitch of a monophonic sound recording (pitch tracking) is crucial to audio processing with multiple applications in music information retrieval. Pitch tracking is an essential component in music signal processing, where monophonic pitch tracking is used for generating pitch annotations for multi-track datasets.

Estimation of the pitch of a monophonic signal has been a longstanding topic for more than a half-century, and many well-founded methods have been proposed since [9]. Earlier methods mostly utilize a certain candidate-generating function,

 ISBN: 978-1-61208-789-4

using pre and post-processing stages to produce the pitch curve. Those functions include the ”cepstrum [10], the autocorrelation function (ACF) [11], the average magnitude difference function (AMDF) [12], the normalized cross-correlation function (NCCF) as proposed by RAPT [13] and PRAAT [14], and the cumulative mean normalized difference function as proposed by YIN” [9].

A common method in previous approaches is that the derivation of a better pitch-tracking system depends on a robust candidate-generating function and/or sophisticated post-processing steps, i.e. heuristics. Furthermore, none of them are directly learned from data, except for manual hyperparameter tuning, which contrasts with other problems in music information retrieval such as chord ID [15], where data-driven methods have been shown to outperform heuristic approaches.

C. Flowkey

Flowkey [16] is an educational music app that teaches how to play your favorite songs on your digital or acoustic piano.

The app works on multiple devices and comes off useful and practical for beginners to learn piano or even for advanced musicians. It aims to help the beginner players who are not comfortable with reading sheet music notes yet.

The different modes and features of Flowkey are:

- Slow Mode - This feature allows the player to play along with the song at a slow speed to make the user feel comfortable with the virtual sheet music notes. In this section, the video also slows down without disturbing the audio.
- Fast Mode - This mode allows the player to play along with the song in the original tempo for that specific song.
- Loop Mode - the player will be able to choose a portion of the video tutorial and keep it on the loop until he/she gets it right.
- Hand Selection - This feature is beneficial for more complicated songs. It is designed for beginners in case the player initially gets confused by multiple keys being played at a time and would like to master on one hand first and then switch to another hand. Figure 1 shows the design of hand selection in Flowkey.
- Wait Mode - The virtual sheet music notes and the video wait for the player to play the notes after learning from the tutorial. Through the built-in microphone in the app, the wait mode detects the player’s movement without making him/her connect the actual digital or acoustic piano to the app. Therefore, the Flowkey app provides feedback on each notes the player plays.

D. Playground Sessions

Playground Sessions is a Web-based music learning software which helps users to subscribe to music theory lessons and provides a fun and effective experience for people to learn the piano online.

Playground Sessions has several different elements that contribute to the learning experience.

- Interactive Lessons: This section is under the “Bootcamp” tab in the app where excerpts from well-known songs are designated to teach the player specific music concepts related to the song, with written instruction and game-like practice.
- Video Lessons: The video lessons are followed by interactive lessons. They cover more details about the interaction lessons and allow the player to practice what was taught.
- Forums: This is a place for users to share tips, stay up to date on Playground Sessions news, ask questions, and lodge complaints.

Figure 1. Design of Hand Selection in Flowkey.

Unlike Flowkey and Playground Sessions, our Web-based music learning application consists of a curriculum that not only teaches music concepts to students, but also teaches them how to compose to a music piece. The other extra feature our application has compared to the other existing ones is that the player can review the other players’ performances and rate them. The user also may want to join the other users to play a duet, which is the ongoing feature of our application.

III. Methods

We developed a Web-based prototype to teach people with little to no music background the music theory basics and how to play the piano. To address this problem, we developed our prototype in two phases. The first one is for students to learn about the basics of music theory. Secondly, students will be able to practice what they learned by playing their favorite songs on the application with the app’s guidance, which is connected to their own digital or acoustic piano.

A. Phase One: Curriculum and Teaching Strategies

1) Design and development : We designed an online teaching platform, where teachers can have a one-on-one or a group session with their students. The software helps teachers to review the progress of their students throughout the week’s practice. Besides, teachers will be able to upload their own recorded courses for the students. Thirdly, we developed the music theory teaching scriptwriting, a work classification, courseware development, and other links. Figure 2 presents the primary teaching module architecture.

2) Development of the curriculum: In the non-technology-based teaching research, developing the different teaching models had to be according to students’ perceptions and music learning ability. This model could be practical when educators share the same goal, and students have the same understanding level. Under the current teaching models of College Music teaching, a teaching object’s attribute is not

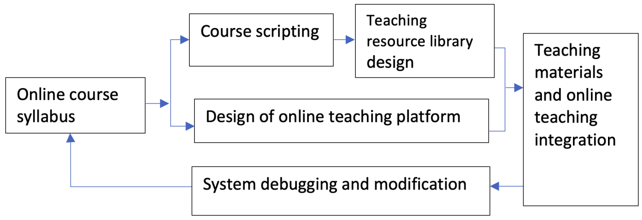

Figure 2. Design and development process of online music theory courses (based on [17]).

consistent. Therefore, to conduct the trusted quality of teaching objectives, it is essential to develop students' multi-attribute professional features and multi-objective future careers.

In the research and practice of "diversity", teaching, content, and target design is based on the students' knowledge base, understanding ability, and so on. Therefore, in this project, we aim to implement an Artificially Intelligent application that can train itself based on the students' different expertise and performance skills. The students' attribute difference is not limited to the learning factor in the trusted teaching mode, but to individual behavior attribute, knowledge attribute, and personal attribute under corresponding training objective.

To determine students' needs to support Web-based learning, we designed a music-technology-based curriculum to evaluate students' interaction with the existing music learning applications and better address the missing components in our prototype.

We conducted this study in two phases, each with separate surveys and strategies. The first phase's purpose was to develop and evaluate an online pilot program's possibilities and requirements integrating computer technologies and music composition concepts for middle-school students. Opportunities to view and critique pilot online instructional units developed for the Knowledge Works Learning Academy (KWLA) were included as a regular part of class activities for two graduate music education courses at Auburn University: CTMU 7520-26 (Curriculum and Teaching in Music Education) and CTMU 7540-46 (Evaluation of Programs in Music Education).

The specific research questions for the study are as follows:

First Survey: Administrators

1) To what extent do comprehensive public high schools in the United States offer technology-based music classes?
2) To what extent does the district's socioeconomic status affect the likelihood of offering a technology-based music class?
3) To what extent does the district's geographic location affect the likelihood of offering a technology-based music class?
4) To what extent do/would school administrators value technology-based music classes?

Second Survey: Teachers

1) What is the curricular nature of these classes?
2) To what extent do these classes address nontraditional music students?
3) What is the professional background of teachers of technology-based music classes?
4) What types of software and hardware are being utilized in technology-based music classes?
5) How long have these classes been offered, and how were they initiated?
6) What level of support do school districts provide for these classes?

After analyzing the first phase, where music graduate students evaluated the curriculum's feasibility and potential student's needs, we made the curriculum changes accordingly. The results of our first phase study will be presented in future work. In phase two, we conduct a pilot study to monitor students' interaction with our music learning Web-based platform. This pilot study is planned for October-November 2020, and the results will be presented in future work.

B. Phase Two: Practicing Strategies

This phase is developed to help the students practice what they learned in the curriculum by playing their favorite songs on the application, connected to their own digital or acoustic piano. In other words, students will be able to play a duet with the application, play famous songs with the help of the application, and connect with other people using the application to play in a team (band).

The following are the key technologies required for a Web-based music teaching system:

1) Data Collection:

- Musicians: 17 music teachers (middle school, high school, or college teachers) are invited from the music education department at Auburn University to perform duet pieces and review the first phase of the application, including the designed curriculum.
- Music pieces: To collect the music pieces, we developed a survey to collect data from music teachers all over the USA. Each musician will perform every detail executed from the surveys for ten times in the different music expression. Therefore, our machine learning model will be trained based on theses music pieces rehearsals.
- Recording settings: Electronic pianos with Musical Instrument Digital Interface (MIDI) output will record the music pieces; therefore, all the parameters (dynamic, starting time, ending time, pedal) of every note can be recorded in real-time [18].
- Recording procedures: Musicians will practice the pieces for 30 minutes together (other than solo practices) and then start recording. Each recording session records approximately ten performances and lasts more than an hour.

2) Data Representation and Models: We are using various function approximations based on [18] to model the differences between one pianist's expression and another's. In this project, we start from music representations and models that only apply to specific music notes and gradually step to a high-dimensional phrase, which implies a whole piece of music.

Our artificial performer will generate (decode) its music expression by communicating with a player according to trained models. Piano notes can be represented by notes, beats, timing, dynamic, and pedal position.

3) Pitch Detection: Our Web-based music learning platform detects players' notes as they play along and provides feedback for the player to guide them toward the right path.

In this research, we used the CREPE pitch detection [9] algorithm to detect the player's notes in real-time. CREPE originally includes a deep convolutional neural network that operates on the time-domain audio signal for a pitch estimation. A diagram of the CREPE architecture is shown in Figure 3.

The CREPE input is a 1024-sample excerpt from the time-domain audio signal, with a 16 kHz sampling rate. There will be five convolutional layers that result in a 2048-dimensional latent representation connected to a 72-dimensional output vector y through sigmoid activation. Each of the 72 nodes in the output layer corresponds to a specific pitch value, defined in cents. Cent describes the relationship between musical intervals to a reference pitch f_{ref} in Hz, expressed as a function of frequency f in Hz:

$$\psi(f) = 1200 \times log_2 \frac{f}{f_{ref}} \tag{1}$$

where $f_{ref} = 5$ Hz throughout the program. The 72 pitch values are noted as $\psi_1, \psi_2, \cdots, \psi_{72}$ and selected so that they cover six octaves with 20-cent intervals between C1 and B6, corresponding to 32.70 Hz and 1975.53 hz. $\hat{\psi}$ is the weighted average of the associated pitches ψ_i based output $\hat{y}$, which provides the frequency estimate in Hz:

$$\hat{\psi} = \frac{\sum_{i=1}^{72} \hat{y}\psi_i}{\sum_{i=1}^{72} \hat{y}}, \hat{f} = f_{ref}.2^{\frac{\hat{\psi}}{1200}} \tag{2}$$

The target outputs we use to train the model are 72-dimensional vectors, where each dimension represents a frequency bin covering 20 cents.

The target is Gaussian-blurred in frequency to reduce the penalty for near-correct predictions, such that the energy surrounding a ground truth frequency decays with a standard deviation of 25 cents [9]:

$$y_i = \exp(-\frac{(\psi_i - \psi_{true})^2}{2.25^2}) \tag{3}$$

The CNN is trained such that the binary cross-entropy between the target vector y and the predicted vector $\hat{y}_i$:

$$\Gamma(y, \hat{y}) = \sum_{i=1}^{72}(-y_i.log\hat{y}_i - (1 - y_i)log(1 - y_i)) \tag{4}$$

Where both y_i and $\hat{y}_i$ are real numbers between 0 and 1. We are using Adam optimizer as our loss function, and the learning rate is 0.05. The best performing model is selected after training until the validation accuracy no longer improves for 12 epochs. One epoch consists of 300 batches of 12 examples randomly chosen from the training set.

A more general and comprehensive representation is designed to improve the model's generality further and predict the expressive timing more than the rhythm context. In particular, features are developed from four aspects of expressive collaborative performance, as shown in Figure 4.

4) Beat Tracking: Timing and dynamics are the two most fundamentals aspects of musical expression. In this project, we are planning to model different musicians' presentation as a co-evolving time series. ”Based on this representation, we use a set of algorithms, including a sophisticated spectral learning method, to discover regularities of expressive musical interaction from rehearsals” [18].

Our Web-based application will be one of the first applications of spectral learning in the field of music. We consider adding some basic improvisation techniques where musicians have the freedom to interpret pitches and rhythms other than expressive timing and dynamics. We aim to implement a model that trains a different set of parameters for each measure and focuses on predicting the number of chords and the number of notes per chord. Given the model prediction, an improvised score is decoded using the nearest-neighbor search, which selects the training example whose parameters are closest to the determination [18]. We expect the model to generate more musical, interactive, and natural collaborative improvisation than a reasonable baseline based on mean estimation.

C. Web-based Platform Development

Our music learning platform is a Java-based Web application developed using HTML5 and CSS3 for the forum and JSP and Servlets as back-end. Initially, all the requirements were collected and analyzed based on Evolutionary Prototyping (EP).

1) Functional Requirements:

- Home Page (index.jsp): This is a Web page where music teachers and students can log in. The machine will recognize if the user is a student or professor based on the MySQL database's username. Students will be able to look through the curriculum and start their online lessons. Music teachers will set up a session, upload a new course, and embed the procedures provided by our application.
- Video Lessons: The video lessons include a score follower, pitch tracking, and beat tracking to provide real-time feedback for students as they play along. The design of the curriculum is provided in Figure 5.
- Forums: This is a place for students to share their performances with others and find other players to play a duet.
- Student Dashboard (dashboard student.jsp): This is a dashboard for students. Students can use their digital or acoustic piano to practice the lessons.

IV. EVALUATION

The evaluation part of the project is in the proposal phase.

A. Evaluation methods

This paper proposes to use both objective and subjective evaluations. The difference between the predicted results and the ground truth performances could be measured in both simulations and real-time accurate assessment arrangements. We propose to let subjects evaluate both the anticipated results and the ground truth performances for subjective evaluation. In particular, topics should be from two groups, which are non-music major and music major.

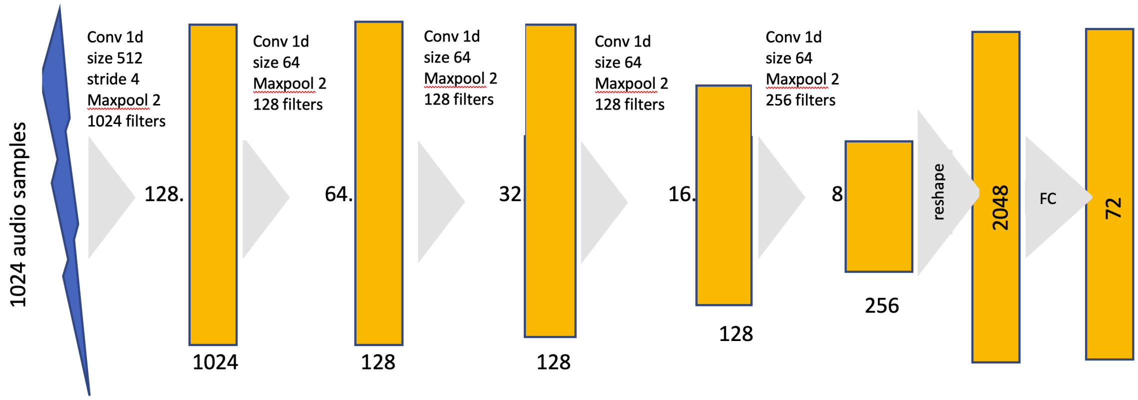

Figure 3. CNN Pitch Detection.

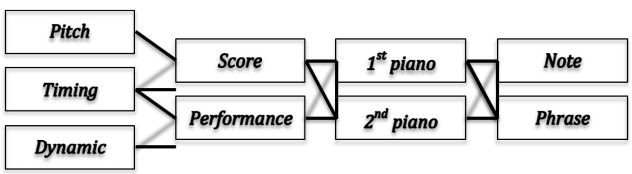

Figure 4. Proposed Architecture.

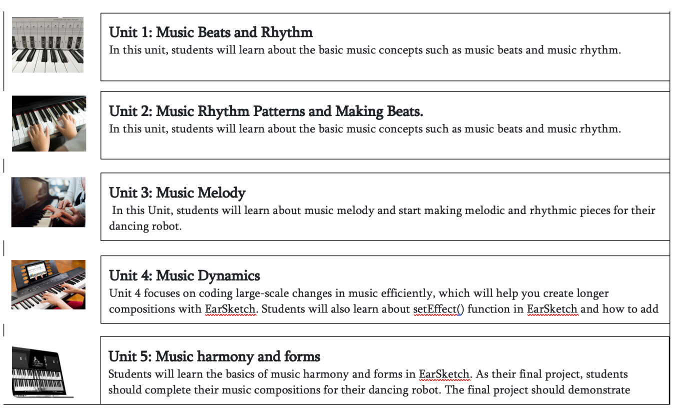

Figure 5. Units Design.

B. Criteria for Successful Completion

In terms of scientific discoveries, successful completion means the introduction questions have been answered and implemented. In terms of system completion, successful completion means the tasks in Section 3 are mainly completed. In particular, pitch detection and instrument recognition are essential successful criteria, and beat tracking is advanced successful criteria. In terms of performance, successful completion means the artificial performer can learn how to sense and coordinate with human performers' music expression from a reasonable amount of rehearsal data. In other words, the artificial performer's synthetic behavior is the same as the ground truth performance and highly rated by subjective evaluations.

V. Conclusion and Future work

We released an initial version of the curriculum in May 2020 for music students at Auburn University. Since then, it has been viewed in academic courses. We have received informal, mostly positive feedback from music teachers (music graduate students) about the platform's user experience and efficiency and the curriculum. We have also received numerous suggestions and feature requests, especially from teachers, which we incorporate into the current version of the curriculum.

We plan to embed our curriculum, teacher training materials, and social media features directly into the interface instead of maintaining them on different websites in the coming year. Another addition to this music learning platform is to make the application more accessible so that students with visual impairment can also learn music concepts and play an instrument.

References

[1] S. W. Conkling, "Envisioning a scholarship of teaching and learning for the music discipline," College Music Symposium, vol. 43, 2003, pp. 55–64. [Online]. Available: http://www.jstor.org/stable/40374470

[2] I. Ruokonen and H. Ruismäki, "E-learning in music: A case study of learning group composing in a blended learning environment," Procedia-Social and Behavioral Sciences, vol. 217, 2016, pp. 109–115.

[3] I. Ruokonen, A. Sepp, A. Ojala, L. Hietanen, V. Tuisku, and H. Ruismäki, "A web-based music learning environment: A case study of users' experiences," The European Journal of Social & Behavioural Sciences, vol. 26, no. 3, 2019, pp. 2983–2993.

[4] R. B. Dannenberg and C. Raphael, "Music score alignment and computer accompaniment," Communications of the ACM, vol. 49, no. 8, 2006, pp. 38–43.

[5] M. Puckette and C. Lippe, "Score following in practice," in Proceedings of the International Computer Music Conference. INTERNATIONAL COMPUTER MUSIC ACCOCIATION, 1992, pp. 182–182.

[6] A. Maezawa and K. Yamamoto, "Muens: A multimodal human-machine music ensemble for live concert performance," in Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, 2017, pp. 4290–4301.

[7] B. Vera, E. Chew, and P. G. Healey, "A study of ensemble synchronisation under restricted line of sight." in ISMIR, 2013, pp. 293–298.

[8] T. Itohara, K. Nakadai, T. Ogata, and H. G. Okuno, "Improvement of audio-visual score following in robot ensemble with human guitarist," in 2012 12th IEEE-RAS International Conference on Humanoid Robots (Humanoids 2012). IEEE, 2012, pp. 574–579.

[9] J. W. Kim, J. Salamon, P. Li, and J. P. Bello, “Crepe: A convolutional representation for pitch estimation,” in 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018, pp. 161–165.

[10] A. M. Noll, “Clipstrum pitch determination,” The journal of the acoustical society of America, vol. 44, no. 6, 1968, pp. 1585–1591.

[11] J. Dubnowski, R. Schafer, and L. Rabiner, “Real-time digital hardware pitch detector,” IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 24, no. 1, 1976, pp. 2–8.

[12] M. Ross, H. Shaffer, A. Cohen, R. Freudberg, and H. Manley, “Average magnitude difference function pitch extractor,” IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 22, no. 5, 1974, pp. 353–362.

[13] D. Talkin and W. B. Kleijn, “A robust algorithm for pitch tracking (rapt),” Speech coding and synthesis, vol. 495, 1995, p. 518.

[14] P. Boersma, “Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound,” in Proceedings of the institute of phonetic sciences, vol. 17, no. 1193. Amsterdam, 1993, pp. 97–110.

[15] E. J. Humphrey and J. P. Bello, “Rethinking automatic chord recognition with convolutional neural networks,” in 2012 11th International Conference on Machine Learning and Applications, vol. 2. IEEE, 2012, pp. 357–362.

[16] “Flowkey - learn piano with the songs you love,” https://www.flowkey.com/en [Last accessed: September 2020].

[17] Y. He, “Research on online teaching of music performance based on diversification and intelligence–take the online music teaching during the covid-19 as an example,” in 2020 International Conference on E-Commerce and Internet Technology (ECIT). IEEE, 2020, pp. 193–196.

[18] G. Xia and R. B. Dannenberg, “Duet interaction: learning musicianship for automatic accompaniment.” in NIME, 2015, pp. 259–264.