MusicCritic: A technological framework to support online music teaching for large audiences

Baris Bozkurt, Sankalp Gulati, Oriol Romani, Xavier Serra
Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain

Abstract

This paper concerns online music education and as contribution, it proposes a new technological framework to support online music performance teaching to reduce loads on teachers for assessing large number of student performances.

The online education field is growing exponentially. One form of online education is the Massive Open Online Courses (MOOCs) where large number of students, on the order of thousands, are enrolled to online courses. Recently, there have been course offerings for teaching music performance through MOOCs which basically rely on peer evaluation for the assessment of student performances and providing feedback.

MOOCs designed for other domains such as computer programming have been successfully using supporting technologies that facilitate assessment and feedback. Here, we argue that supporting technologies dedicated to reducing instructor load in teaching music performance online would pave the way for successful MOOCs in this domain and provide new opportunities for music educators to reach larger audiences.

In this paper, we propose a framework (MusicCritic, https://musiccritic.upf.edu) that can help scale practice-based online music education upto MOOCs level without relying on peer evaluation methods. We discuss two main components of the framework. First, we consider the interfaces for setting up practice exercises, recording student performances, assessing the performances and providing feedback to the students. Second, tools for facilitating assessment are discussed where we demonstrate a semi-automatic assessment system that can learn from assessment of the instructor on a small group of performances and further assess larger sets of performances. We finally present tests performed on real-life data to demonstrate the potential of the approach.

Keywords

music assessment, music technology, online music education, massive open online courses

Introduction

Online education has impacted and changed all fields of education significantly all over the world in the last decade. According to data collected by Class Central (https://www.class-central.com/), the total number of students who signed up for at least one online course has crossed 58 million in 2016. More than 700 universities around the world have launched free online courses. There are numerous well established platforms providing Massive Open Online Courses (MOOCs) to a very large audience of students in various fields. The number of students enrolling to a single course exceeds 100,000 for some very popular courses.

Online education services exist for almost all domains of education, from computer science to health and humanities, as well as for music. In various domains, MOOCs have been very successful in training high number of students in the basic skills of the domain. Computer programming is one such area where MOOCs have become amongst the most prefered resources for getting trained in the basic skills.

Online music education is in a boom period not only with online courses actually delivered by instructors but also with plenty of software tools dedicated to facilitating music practice.

Online degree programs have been created by prestigious institutions since early 2000s and the number of online courses offered by music education institutes increase constantly. For example, Boston University School of Music has been offering an online doctoral program since 2005. The online extension school of Berklee College of Music - Berklee Online (established in 2002), delivers more than 150 online music courses, 9 fully-online Bachelor's Degree programs and 2 Graduate Level programs. Berklee College of Music also offers more than 20 MOOCs, some of which are dedicated to music performance. As the technology provides means to overcome the problem of physical distance, music schools and teachers have now access to a much wider international audience. It has been reported that a large number of students find online resources highly motivating for learning music (Ho, 2007) and positive effects on students' attitudes have been observed in various studies (Bauer et al., 2003, Byrne & MacDonald, 2002). Recent investigations show that online resources used together with face-to-face delivery in a blended learning model leads to significant increase in motivation and progress for music students (Tuisku & Ruokonen, 2017).

Digolo et al., (2011) discuss in detail the opportunities and challenges for online music education where a large portion of the challenges source from technological difficulties faced by students and instructors. In this paper, we specifically focus on such technical difficulties concerning online music courses with very large sized audiences (the MOOCs). We propose a framework, namely MusicCritic (https://musiccritic.upf.edu), which brings in new opportunities for scaling online music education towards MOOCs for teaching basic music performance skills. This framework has already been implemented and integrated in the MOOC platform Kadenze where a Hindustani MOOC is being prepared (using the proposed framework) to start education in early 2018.

In the next section, we present our main framework, the components designed for setting up the music performance exercises for student practice, recording, and the student performance assessment where we also present real-life test results. In the final part, we discuss the results and future work in this direction.

MusicCritic: A framework for supporting MOOCs for music performance

Considering the mode of delivery, there are three types of online courses for learning music performance: video tutorials, real-time video calls with the teachers, and Massive Online Music Courses (MOOCs). The video tutorials (like those used in YouTube) have a large audience, but they lack the interactivity which is essential for music learning. The video calls provide an interactive learning experience similar to the traditional face to face teaching practices, but do not scale (i.e. do not reach large audiences). Finally, the more recent trend of MOOCs, provide video tutorials as well as tools to make the learning experience richer, but still the scalability is very limited due to difficulties in recording and organising student performances, assessing these recordings and providing feedback to students. Below, we present our proposed framework for supporting online music education to increase its scalability to large audiences.

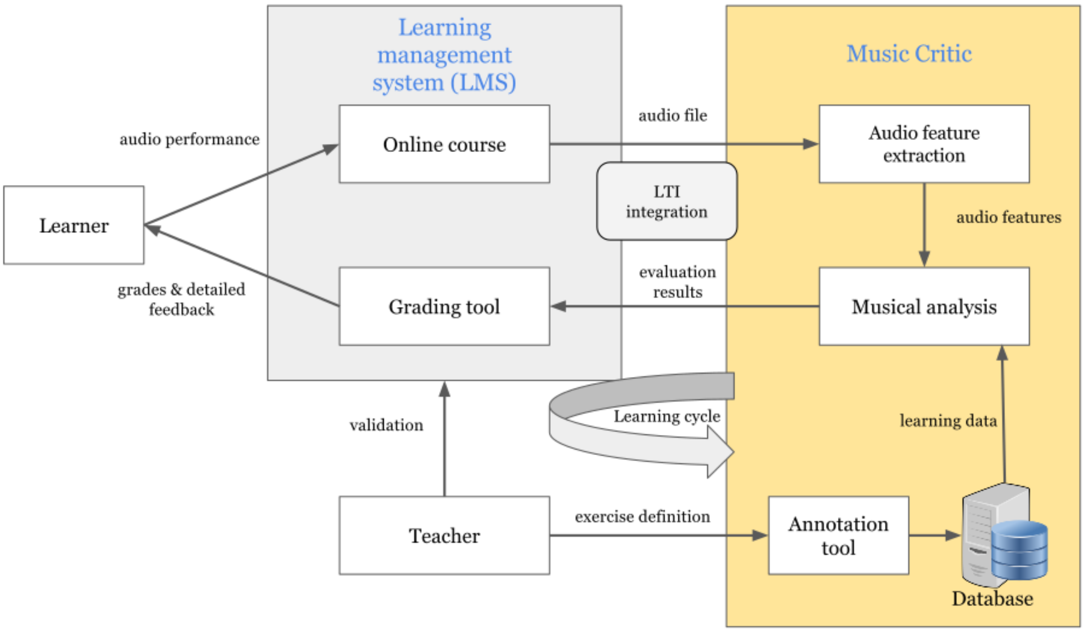

Figure 1: Music education workflow based on the MusicCritic framework

MusicCritic is a framework that can communicate with existing Learning Management Systems (LMS) of the MOOC platforms via the Learning Tool Interoperability standard (LTI) (the basic standard for most of such platforms). The basic workflow is: (1) teacher prepares an exercise using the LMS and MusicCritic; (2) learner uses interfaces for practice and recording and uploads her performance of the exercise to the LMS; (3) LMS sends the audio recording to MusicCritic, where it is analyzed and further presented to the teacher making the assessment easier via several means we discuss below. Here, we limit our discussions to the components aimed at reducing instructor load without addressing the technical implementational details due to space constraints.

The general practice in supplying practice material to students and collecting their recordings in current MOOCs is to provide backing tracks (with or without some reference instructor performance) leaving the task of recording student performances and uploading them completely to the student. As argued by (Hebert, 2007), getting as close as possible to face-to-face delivery interactivity would bring important improvements in online education. One component of MusicCritic is the student practice and recording interface that can be easily tailored to specific exercises by the education content designer (music instructor). These interfaces are now being used in the design of the Hindustani music course by RagaSphere (http://www.ragasphere.com) and Universitat Pompeu Fabra. Our initial tests show that the face-to-face delivery of teacher performance followed by several repetitions of student can be successfully imitated with such interfaces. The first session of this course will be offered during Spring 2018. The demonstrations of the interfaces together with results and observations on user experience will be shared with the audience during the conference.

Assessing large amounts of student recordings and providing feedback

While human beings are very good at taking into account various coexisting dimensions in a student performance to assess quality, this requires a great deal of attention which is difficult to maintain for very long periods of time. The effort required for keeping the attention further increases if the task (musical exercise) also contains repetitions (like repeated phrases or patterns). In such cases, the assessment task is very tiring and as the tiredness builds up, the assessment quality becomes questionable. We tend to refer to such tasks as “mechanistic” which in a way is an expression for stating that it is well suited for a machine then a human being.

The music processing domain can offer support for such repetitive and mechanistic tasks in various ways. The task may be facilitated via dedicated interfaces providing samples in an organized way, providing easy ways of inputting assessment results, storing and accessing them. Visualizations of different musical facets such as melodic curves, measured pitch information, score aligned with performance can be provided, which would help in quickly spotting the errors in a performance. In addition, for relatively simple melody or pattern reproduction exercises, an automatic assessment systems can be deployed. Such a system can be trained on a limited corpus of student performances that are graded by the instructors. When used in combination, such tools have a high potential to reduce the instructor load to a large extend. MusicCritic involves implementation of these different aspects, some of which are discussed below.

Interfaces to facilitate assessment and provide feedback

In Figure 2, we present one of the tools designed to support grading of student performances for the Hindustani course in preparation. The particular exercise in this example involves an ascending-descending pattern aiming at teaching the first four pitches of a scale. The student is expected to record six consecutive repetitions of this pattern (represented with boxes at the bottom). We present here an extended view showing all functionalities which seem complicated at first sight. Yet, all contained functionalities (notation, pitch, waveform views, etc) can be turned on/off and the instructors find it practical upon use of it for a few samples.

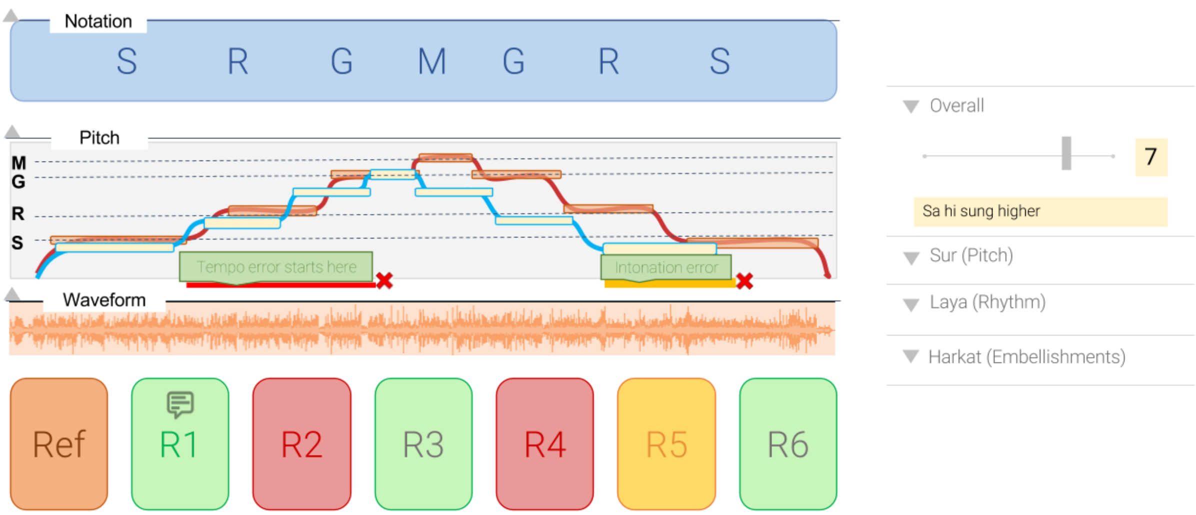

Figure 2. Sample interface to facilitate assessment of a student’s performance and provide feedback for Hindustani music education use case

This interface is designed to be used by the instructor to listen to student recordings, view parametric representations of the performance (for this example the pitch variation), insert comments to specific points to mark problems of the performance and assign grade (using tiers on the right). The same view can be provided to student as the feedback which will help the student figure out what specific problems were involved in each repetition.

Automatic assessment technologies

As mentioned above, another supporting tool reducing instructors load is the use of automatic assessment technologies that can be trained for assessing reproduction of relatively simple melodies and repeated patterns. Since each assessment task has its own particular dimensions, automatic systems using rule-based implementations are of little use. Machine learning based approaches that can learn from graded recordings by human experts are preferable. Various studies in the music processing domain have already reported highly promising results some of which are shortly mentioned below (a larger review is available in (Bozkurt et al., 2017)).

Music performance assessment often involves subjective evaluation of multiple coexisting aspects of the performance. Various studies from the music processing domain have targeted building assessment tools for some of these dimensions such as: vowel quality (Jha & Rao, 2012), strength of singer's formant (Lundy et al., 2000), volume characteristics (Tasi et al., 2015) expression of the voice (Major et al., 2006), vibrato characteristics (Nakano, 2006), rhythm and intonation accuracy (Lin et al., 2014), timbre richness, attack quality and note stabilities (Romaní Picas et al., 2015). Singing assessment has drawn relatively more interest in this domain with various studies reporting highly promising results such as Molina et al. (2013), Schramm et al. (2015) and Tasi et al. (2015).

To demonstrate the potential of automatic assessment tools for this task, we have implemented a benchmark system that uses the common approach in the above mentioned studies: assigning performances grades via mapping note level deviations computed from aligned transcriptions of the performance and the reference. Our benchmark system has been tested on data collected in a real-life scenario: simple melody reproduction task in conservatory entrance exams. These tests target demonstrating the overall potential of these technologies. We are in the process of developing improved models for automatic performance assessment to be integrated into the MusicCritic framework.

Tests on automatic assessment tools

To demonstrate the potential of automatic assessment as defined in state-of-the-art systems, the benchmark system developed for MusicCritic has been tested on a subset of the MAST-melody dataset (Bozkurt et al., 2017) which comprises of 1018 audio recordings of vocal performances by different candidates applied to a conservatory entrance exam in Turkey. These performances are imitations of 40 different melodies played on piano during the examination. The dataset considered here only concerns the assessment of melodic memory phase of the exam. In that phase, each candidate is asked to sing/repeat after a melody has been played two times on the piano. The melodies used in the exams are designed to have a tessitura of 6th interval range, a similar proportion of melodic stepwise motions and leaps, using similar number of quarter, eighth or sixteenth notes in terms of rhythm. Performances are assessed by juries of three conservatory instructors. While juries could assign several levels of grades, only the recordings that were graded as too poor(fail) or high quality(pass) by all jury members (with full agreement) are selected to form the dataset.

The audio recordings in this dataset are extracted from the video recordings of the entrance examinations, which contain reverberations and external noise. Hence the recording quality can be considered to be a representative of a real-world scenario. The maximum length of the recordings is 10.7 seconds. Further details on the preparation of the MAST dataset is available in (Bozkurt et al., 2017).

Considering any potential influence among jury members as they share the same physical space, a subset of this dataset (290 student performances with equal number of pass and fail categories) has been deduced to perform assessment by 6 other individuals (denoted by A^i below). The assessment was carried on a rubric of 4 scales concentrating only the reproduction quality of the melody (discarding factors like vowel quality, timbral aspects, overall tempo): 1) very poor performance, 2) performance with major errors, 3) performance with minor errors, 4) high quality performance. Due to space limitations, we only present a concise analysis of inter-grader variation and the performance of the automatic assessment system to perform grading.

Mean absolute error (MAE) between the ratings given by each annotator and the mean of the ratings given by all other annotators is found as: 0.39, 0.34, 0.35, 0.41, 0.34 and 0.43. The mean of the MAE across all annotators is 0.38 (on a grid of 4 possible grades). This number provides the extent of the inter-annotator agreement and a base to interpret MAE for the benchmark automatic assessment system.

The benchmark system has been trained and tested on this dataset using a standard cross-validation scheme for machine learning systems: a group (%90) of the recordings are used for training the system and then the remaining (%10) of the recordings are used for testing (guaranteeing no overlap exists between training and test samples). The tests are repeated for ten times, each time using another group for testing and finally all results are averaged. The resulting MAE of the automatic system is 0.45. Note that a random baseline (i.e. if all grades are randomly assigned) for this task, results in a MAE of 1.2. Due to space considerations, we are unable to present further details of all the tests carried for the automatic system in this paper. These tests will be presented in the conference.

Discussions and future work

This paper aimed at discussing the opportunities for scaling online music education to MOOC level with the support of technology for reducing instructor load. Our novel framework has already been integrated in the Kadenze platform being used for designing a MOOC for Hindustani music. To demonstrate the potential of our methods for assessment, we have carried tests with real data recorded in conservatory entrance examinations in Turkey. The results are: on a grid of scores 1 to 4, the mean average error (MAE) observed comparing the automatically generated scores with respect to the mean of the grades by 6 human annotators is 0.45 while inter-grader MAE is 0.38. This result (together with various other test results which could not be presented here) demonstrated that, for simple melody or pattern exercises, such systems have the potential to be trained on small subsets assessed by the instructors and can be used to assess large amounts of performances greatly reducing the instructor load. Our actual system (involving new methods for automatic assessment currently under development) will be soon tested in a real-life scenario (the Hindustani MOOC on Kadenze) and test results will be presented in the conference.

Acknowledgements

Omitted for blind review. Acknowledgements will be available upon acceptance.

This research is funded by the European Research Council under the European Union's Seventh Framework Program, as part of the TECSOME project (ERC grant agreement 768530).

References:

1. Bauer, W. I., Reese, S., & McAllister, P. A. (2003). Transforming music teaching via technology: The role of professional development. *Journal of research in Music Education, 51(4)*, 289-301.
2. Bozkurt, B., Baysal, O., & Yuret, D. (2017). A Dataset and Baseline System for Singing Voice Assessment. *Proceedings of the 13th Int. Symposium on Computer Music Multidisciplinary Research*. Porto, Portugal.
3. Byrne, C., & MacDonald, R. A. (2002). The use of information & communication technology (I&CT) in the Scottish Music Curriculum: a focus group investigation of themes and issues. *Music Education Research, 4(2)*, 263-273.
4. Crawford, R. (2013). Evolving technologies require educational policy change: Music education for the 21st century. *Australasian Journal of Educational Technology, 29(5)*, 717–734.
5. Digolo, B., Andang'o, E., & Katuli, J. (2011). E-Learning as a strategy for enhancing access to music education. *International Journal of Business and Social Science, 2(11)*, 135–139.
6. Jha, M. V., & Rao, P. (2012). Assessing vowel quality for singing evaluation. *Proceedings of the National Conference on Communications (NCC) (pp. 1-5).* Kharagpur, India.
7. Ho, W. C. (2007). Students' experiences with and preferences for using information technology in music learning in Shanghai's secondary schools. *British Journal of Educational Technology, 38(4)*, 699–714.

8. Lin, C. H., Lee, Y. S., Chen, M. Y., & Wang, J. C. (2014). Automatic singing evaluating system based on acoustic features and rhythm. *Proceedings of the IEEE International Conference on Orange Technologies, ICOT (pp.165-168)*. Hong Kong, China.
9. Lundy, D. S., Roy, S., Casiano, R. R., Xue, J. W., & Evans, J. (2000). Acoustic analysis of the singing and speaking voice in singing students. *Journal of Voice. 14(4)*, 490-493.
10. Mayor, O., Bonada, J., & Loscos, A. (2006). The singing tutor: Expression categorization and segmentation of the singing voice. *Proceedings of the AES 121st Convention*. San Francisco, USA.
11. Molina, E., Barbancho, I., Gomez, E., Barbancho, A. M., & Tardon, L. J. (2013). Fundamental frequency alignment vs. note-based melodic similarity for singing voice assessment. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (pp. 744–748)*. Vancouver, Canada.
12. Nakano, T., Goto, M., & Hiraga, Y. (2006). An automatic singing skill evaluation method for unknown melodies using pitch interval accuracy and vibrato features. *Proceedings of Interspeech, (pp. 1706-1709)*. Pittsburgh, USA.
13. Romani Picas, O., Parra Rodriguez, H., Dabiri, D., Tokuda H., Hariya, W., Oishi, K., & Serra, X. (2015). A real-time system for measuring sound goodness in instrumental sounds. *Proceedings of the 138th Audio Engineering Society Convention*. Warsaw, Poland.
14. Schramm, R., Nunes, H. D. S., & Jung, C. R. (2015). Automatic solfège assessment. *Proceedings of the 16th International Society for Music Information Retrieval Conference (ISMIR), (pp. 183-189)*. Malaga, Spain.
15. Sherbon, J. W. & Kish, D. L. (2005). Distance Learning and the Music Teacher. *Music Educators Journal, 92(2)*, 36-41.

16. Tsai, W. H., Ma, C. H., & Hsu, Y. P. (2015). Automatic singing performance evaluation using accompanied vocals as reference bases. *Journal of Information Science and Engineering. 31(3)*, 821-838.

17. Tuisku, V., & Ruokonen, I. (2017). Toward a Blended Learning Model of Teaching Guitar as Part of Primary Teacher Training Curriculum. *The European Journal of Social and Behavioural Sciences*, 2520–2537.