



A New Practice Method Based on KNN Model to Improve User Experience for an AR Piano Learning System

Hong Zeng, Xingxi He^(✉), and Honghu Pan

Department of Mechatronics, College of Mechanical Engineering,
Chongqing University, Chongqing 400044, China
xingxi@cqu.edu.cn

Abstract. To improve the novices' short-term piano learning performances, we presented an AR piano learning system with a new practice method proposed. Initially, an Android AR app was developed for the smart glasses Moverio BT-300. The system with the user-friendly graphical augmentations on the real piano and the multi-marker design reduces the cognitive load and supports faster learning of a novice without prior sight-reading skills. In addition, a new practice method with this AR system was studied, which not only considered the practice of a song at slower speed, but also included the repeated practice of some difficult measures selected by a k-Nearest Neighbor classifier. Finally, an evaluation was implemented to compare the new practice method with the normal practice, which treats all the measures in a song equally as important. And we found that this new practice method is more useful for novices to improve their performances during short-term learning.

Keywords: Augmented reality · K-nearest neighbor model · Piano learning

1 Introduction

Augment Reality (AR) [1] has become a hot research topic in recent years and is widely used in medical, education, manufacturing, entertainment, military and other fields [2]. Especially, AR has been proved to be efficient for piano-aided learning as it could set the virtual scene on the screen and interact with the real world [3]. Therefore, AR can help the piano learners, such as novices, reduce the difficulty of learning sheet music notation and its mapping to the piano keys.

There has been quite a few Augment Reality systems or applications designed for novices to skip the sheet music reading and play the piano for either training or entertainment purposes [3, 4]. Huang et al. [5] proposed a piano teaching system, Piano AR, based on a markerless augmented reality method. By detecting and calculating the transformation matrix from keyboard coordinate to camera coordinate, the system can track the real keyboard of piano naturally. Sun et al. [6] developed a portable piano tutoring system, which can give instructions and feedbacks to help novices on fundamental piano skills. Augmented songbook [7], a mobile augmented

reality educational system for young children, which can raise their awareness of musical notation and inspire their interest in music.

Although these AR systems give hints of the notes that need to be played, due to the rapid position changes of virtual keys and the requirement to reach far keys in some pieces of music, the average percentage of correctly pressed notes remains as low as around 50% for a learner's short-term learning performances. Therefore, in order to improve a novice's short-term performance, we designed a new practice method for an AR system with an added practice session in which some selected measures considered "difficult" will be practiced at slower speed. Those measures, i.e. the segments between the bars in a musical notation, labelled "difficult" are where a novice most likely would make mistakes based on a machine learning classifier. In this experiment, more than 50 but less than 1k labelled samples are collected, no text data is involved and therefore either k-Nearest Neighbor (kNN) method or Naïve Bayesian method are suitable. In this paper, the KNN classifier is chosen.

Li et al. [8] improved the kNN algorithm by making use of the mean value's thought of the Baseline algorithm and by adding to the standard deviation of the rating, and presented a music personalized recommendation system, which can recommend music more accurately for users. In this article, each measure is considered as a sample labelled by the difficulty of playing, either "difficult" or "easy". In this two-class problem, the simple non-weight kNN algorithm will be implemented to choose the measures labelled "difficult" for practice session.

In addition, the AR system to display augmentation superimposed on the piano key can also greatly affect the user experience [9], and a well-designed AR displaying method can reduce cognitive load, ease the learning process and foster the user's interest to learn the piano. The paper developed an application to display augmentation with multi-markers for the AR smart glasses to help the novices learn to play a real piano.

2 Methods

In this section, the app design is firstly described. Then the kNN classification algorithm and the implementation of the KNN are introduced in detail.

2.1 The Design of the Application

To improve the user's experience, an android application was developed by using Java and artoolkitX, an open source augmented reality SDK, and installed on the AR smart glasses, Epson Moverio BT300. The Epson Moverio BT300, illustrated in the Fig. 1, weighs 69 g with binocular see-through viewing and runs Android 5.1. The user needs to sit in the front of the real piano and wear the smart glasses, and they can see the video stream with augmentations through the smart glasses.



Fig. 1. The Epson Moverio BT300

In order to improve one of the optical limitations of the smart glasses, i.e. the limited field of view, this application adopts a multi-marker tracking method. We divide the piano keyboard into four zones, each with a marker for tracking and registration, shown in Fig. 2. Zone B and C include one octave of 12 notes or 7 white keys and 5 black keys. Though Zone A or D include a larger range, only one octave in each zone will be used in most musical pieces.



Fig. 2. A piano keyboard divided into four zones and each zone with a marker. Marker A is on the note f, marker B on f1, marker C on f2, and the marker D on f3.

As far as the graphic augmentations, Hackl et al. [9] compared the differences between two ways to display the augmentation, i.e. the instant way and the Beatmania way. The instant way provides no hints for the next note to be played and the display interface is clear and simple. However, when there is no prompt message for the next note, the user might experience some discomfort due to the limited time to react and press the next key and the error rate would rise for the short-term performance. The Beatmania way has some hints for the coming notes to be played which is better than the instant way. However, the Beatmania way might increase the cognitive load or add unnecessary confusion to the user.

The article designed a novel app that draws three kinds of virtual objects in the augmentation layer shown in the Fig. 3. A green solid cube indicates the notes that need to be pressed at the current moment. The user is supposed to press and hold the key until the virtual key disappears. The set period of time of holding a key corresponds to the duration of a note. When the black keys on the piano needs to be pressed, the color of the solid cube will turn red. The second type, a green hollow box or frame whose location means that the notes to be pressed at the next moment. Another virtual object, the arrow is supposed to remind the user of switching the view.

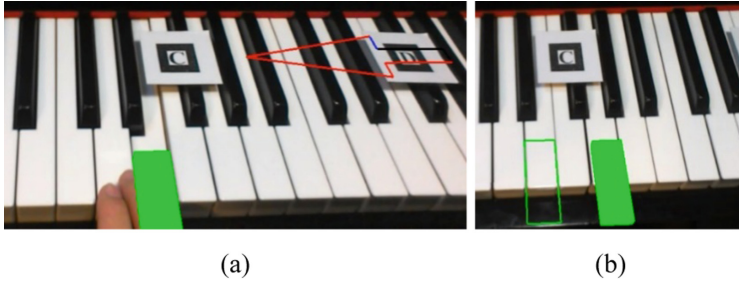


Fig. 3. The video stream with augmentations (Color figure online)

For example, Fig. 3 shows the video stream with augmentations and Fig. 3(a) displays a green solid cube superimposed on the note e2, and a red arrow pointing to the left. This means that the player needs to press the piano key e2 at the current time until the cube disappears or appears in another position. And the virtual red arrow on the marker D indicates that the player should shift the sight to the left, leaving zone D for the key to be pressed is in the zone C. Figure 3(b) shows a cube superimposed on the note g2 and a hollow green box on the note e2 which refers that the player is supposed to press the note g2 until the cube superimposed on the note e2.

2.2 Algorithm of KNN to Select the Measures in a Musical Staff for Practice

In order to select the difficult measures in a song where the novices most likely would make some mistakes, the kNN classification algorithm is used as one of the simplest and the most used machine learning algorithms. By using the kNN classifier, no prior knowledge about or the assumptions on the data distribution is needed. Due to the feature similarity, the majority vote of the class label of the k nearest neighboring samples in a feature space will be assigned to a test sample.

The general steps of the kNN classification include (1) calculate the distance from all the training samples against the test sample in the feature space, (2) sort the training samples by the value of distance in ascending order, (3) select the value of k and (4) the test sample will be assigned to the class most common among its k nearest neighbors.

The proximity or similarity between two samples is represented by calculating their distance in the feature space. Either the Euclidean distance or the Manhattan distance can be used, which are expressed as follows:

$$d_E(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

$$d_M(x, y) = \sum_{i=1}^n |x_i - y_i| \quad (2)$$

where n represents the number of features, d_E refers to the Euclidean distance, d_M is the Manhattan distance, x and y are a test sample and a training sample respectively. The smaller the distance between two samples in the feature space, the more alike their

features and so the more similar the two samples will be. That is why k nearest samples are considered and the right choice of k value can avoid the bias in one direction or another. The different distance metrics using L_k norm ($x, y \in R^d, k \in Z, L_k(x, y) = \sum_{i=1}^d (||x^i - y^i||^k)^{1/k}$) lead to different classification rate in high dimensional space [10]. For high dimensional data, the Manhattan distance metric with $k = 1$, which results in the more meaningful notion of proximity between two samples, is more preferable than the Euclidean distance metric with $k = 2$. In this paper, the number of features discussed is 4 or 5, which is much smaller than the number of samples. Therefore, the data is low dimensional and the Euclidean distance metric can be applied.

2.3 Training Data Acquisition and Features in the Learning Model

To obtain the training data, 5 volunteers with no previous piano experience in learning were invited to play 13 sections from 7 songs which are either elementary songs or the easy version of pop songs. 7 sections are taken from the right-hand melody of a song with a dominant melody and the rest of 6 sections from the left-hand harmony. Those sections have various difficulty levels for novices and their lengths range from 16 bars to 32 bars, resulting in a total of 232 bars or measures. Each bar or measure is considered as one sample in the learning model and the difficult measures are supposed to be selected for the practice purpose. The age of five volunteers are between 21 to 24. In the experiment, each person first spent some time to get used to wearing the smart glasses and practiced a small song to be familiar with the augmentations while playing. Each session ranged from one hour and a half to two hours for a participant to finish playing 13 songs. The volunteers are supposed to play each song only once and their initial performances are recorded and evaluated for labelling. When the player hit the wrong note, the measure where that note is located would be marked. However, the duration of each note were not considered or counted. If the number of people who made mistakes at a certain measure is greater than a threshold (equal to 2 in this article), then this measure is labelled “difficult”. In this way, the paper simplify the problem to a two-category task. More precisely speaking, in the model, the label of the difficult measures are set to 1, the others set to 0, and no text data is involved.

In the learning model, we extracted five features and processed the labels of the samples according to certain rules. Considering that if the distance between two notes is too far, the player has a high probability of making mistakes at this measure, the paper sums up the distances between all the pairs of neighboring notes inside one measure, D_I , as one of the features. Sometimes the distance between the two notes before and after a vertical bar is large and influential to the player. Therefore, the distance between two adjacent measures, D_W , is also considered as a feature. Intuitively, in one measure if there are quite a few notes with short durations such as 1 or 2 described in the app, the player would also easily make mistakes. Therefore, we count the number of notes with the first and the second shortest duration, 1 or 2 in one measure, n_s , as a feature. Similarly the number of notes in a sequence where each note has duration of 1 or 2 consecutively in one bar, n_c , is also chosen as a feature. At last, it

is found that the difficulty levels of the left-hand and right-hand part of a song are different in general, and it will affect the classification results to some extent. So, whether it is the left- or right-hand part of a song is also considered as a feature, h . In summary, the data for each sample contains 5 features and a corresponding label.

2.4 Training Results

The paper divided the 232 samples into training set and development set in the ratio 7:3, and calculated the accuracy rate for the model validation. For the accuracy rate, only the false negative classification is counted as an error, which means the positive measure (1 or labelled “difficult”) is classified as negative (0 or labelled “easy”). However, the false positive classification is not considered as an error. False positive means originally a negative measure (0 or labelled “easy”) is classified as positive (1 or labelled “difficult”). It is believed that the practice of those false positive measures labeled “easy” would not affect the performance of players or at least it would not lead to a decline in performance. The classification accuracy rate for the development set is calculated as follows:

$$\text{accu} = \frac{1}{n} \sum_{i=1}^n (n - k_i), \quad k_i = \begin{cases} 1, & \text{if } y_i = 1, \bar{y}_i = 0 \\ 0, & \text{else.} \end{cases} \quad (3)$$

where n is the number of development set samples, y_i is the original label of the i th test sample, and \bar{y}_i is the classification result.

In the kNN algorithm, choosing the right value of k is important for better accuracy. Higher value of k has lesser chance of error. However, if k is too large, the computing time should be of concern because for each new test sample, all the training data stored has to be used. In addition, the quality of feature extraction will also affect the classification results. Therefore, we analyzed the impact of different features and different k values on model predictions. This model is developed based on Python and implemented in an library (sklearn). Table 1 displays the accuracy rate for the development set with different features and k values. The feature h , the left- or right-hand part of a song, has a greater influence on the accuracy when k value is smaller. When any one of D_I and D_W is missing, different k values have little effect on the prediction results, and the prediction accuracy is relatively low. The effects of features n_s and n_c on the prediction rate are similar because the values of the two features in the same sample are close to each other. Finally four features D_I , D_W , n_c and h are chosen as the features of the samples, the k value is set to 11 in the model and the accuracy rate is 0.8805. The chosen k value also satisfy the following two conditions: (1) an odd number to avoid a tied result of voting; (2) $k = \sqrt{n}$ where n is the total number of training samples and equal to 162 in the model validation.

Table 1. Accuracy rate on development set of different features and k values

Features	k values				
	3	5	7	9	11
	The accuracy rate on development set				
$D_I D_W n_s n_c h$	0.8642	0.8647	0.8667	0.8638	0.8722
$D_I D_W n_s n_c$	0.8448	0.8549	0.8638	0.8664	0.8716
$D_I D_W n_s h$	0.8670	0.8705	0.8740	0.8741	0.8795
$D_I D_W n_c h$	0.8635	0.8699	0.8769	0.8775	0.8805
$D_I n_s n_c h$	0.8651	0.8583	0.8463	0.8555	0.8603
$D_W n_s n_c h$	0.8552	0.8527	0.8511	0.8533	0.8641

3 Evaluation and Experiment

An evaluation experiment was implemented to examine the impact of the new practice method on the learning. We invited 6 volunteers different from those who participated in the training data acquisition and divided them into two groups, the experimental group and the control group, each of which has 3 members. All the volunteers are novices in playing piano and have no knowledge of the five-line staff based sheet music notation. The age of six volunteers are between 21 to 24. The experimental group was asked to use the new practice method for piano practice with the help of the developed AR system. The shortened easy-to-play versions of two songs, a pop song “I do” with 33 measures and “swan lake” with 24 measures, were selected for the evaluation. Only the right-hand melody was played and the difficulty level of the two songs is similar to or a little harder than those used in the training data. After applying the KNN classifier to the 57 new test samples each of which has 4 features, 11 measures from “I do” and 4 measures from “swan lake” were chosen for practice. Those measures labelled “difficult” were repeated multiple times (3 times for “I do” and 6 times for “swan lake”) so that both groups received the equivalent amount of practice. During the practice, the augmentations were displayed more slowly than usual while during the performance evaluation displayed at normal speed. For practice, the control group was supposed to play the original songs twice and the experimental group needed play the original songs once and the chosen measures once. Every participant devoted 40 min on average at their evaluation session.

3.1 Flowchart of Evaluation

The evaluation flowchart is depicted in Fig. 4. Initially for the evaluation, every volunteer was asked to use the practice programmed in the app to be accustomed to wearing the smart glasses and the way of virtual keys display on the piano. The volunteers were supposed to hit a few notes which do not form any certain melody. However, in this evaluation preparation, all the virtual objects in the augmentation design are involved and all the four markers are used. Afterwards, the following steps were repeated for the two test songs, first the “swan lake” and then followed by the “I do”. Every participant from both groups started with playing the song at normal speed

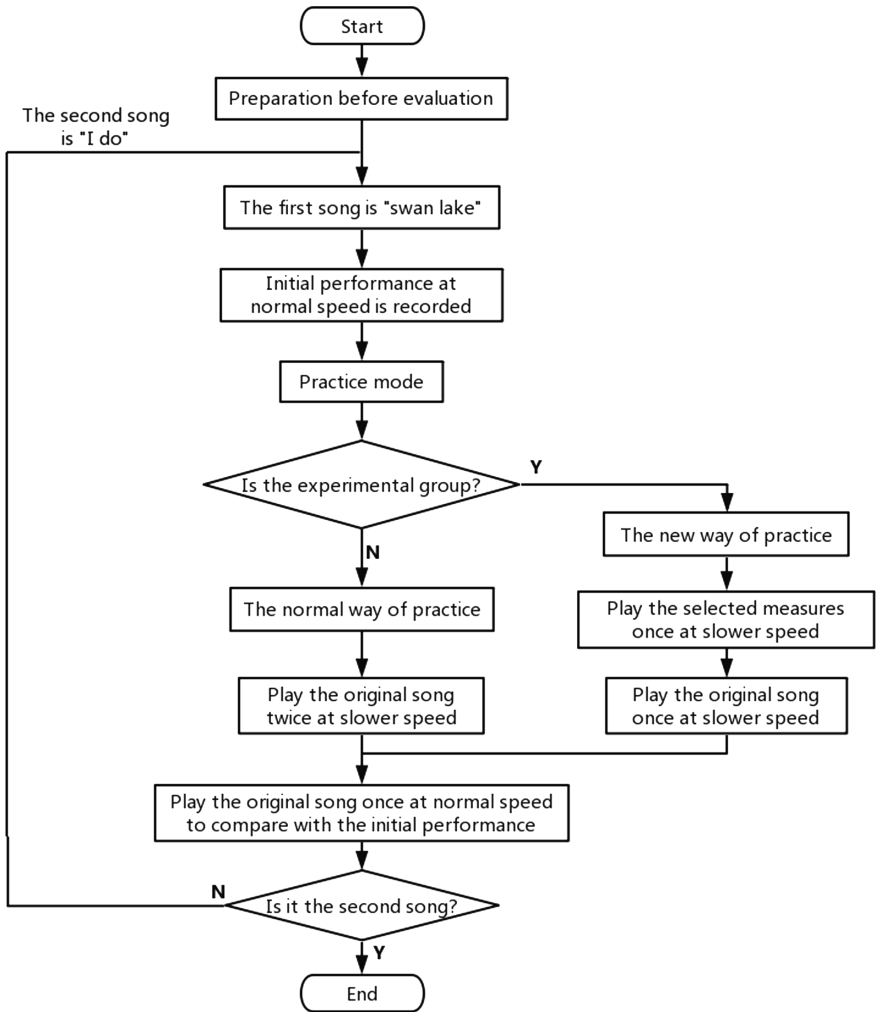


Fig. 4. Experimental flowchart

and this initial performance would be recorded for later analysis. Next, the volunteers entered into the different practice mode. The participants in the control group practiced the original song twice at slower speed while those in the experimental group played the selected measures from KNN once and the original song once also at slower speed. Finally, after the practice, everyone played the original song again at normal speed and the performance was recorded for comparison.

3.2 Results and Discussion

In order to analyze the influence of the new practice method with the AR system on piano learning, the playing accuracy rate was calculated before and after the practice for both the experimental group and control group. The playing accuracy rate is equal to the number of correctly pressed notes divided by the total number of notes in a song. As in the KNN training data acquisition process, the duration of each note was not considered for playing accuracy calculation.

In Fig. 5 the playing accuracy rate and the improved accuracy rate after the practice are illustrated. The x values of 1–3 represent the 3 volunteers in each group (6 persons in total) and the corresponding y values are their playing accuracy of the first song “swan lake”. The x values of 4–6 represent the same 6 persons in the two groups and the corresponding y values are their playing accuracy of the second song “I do”.

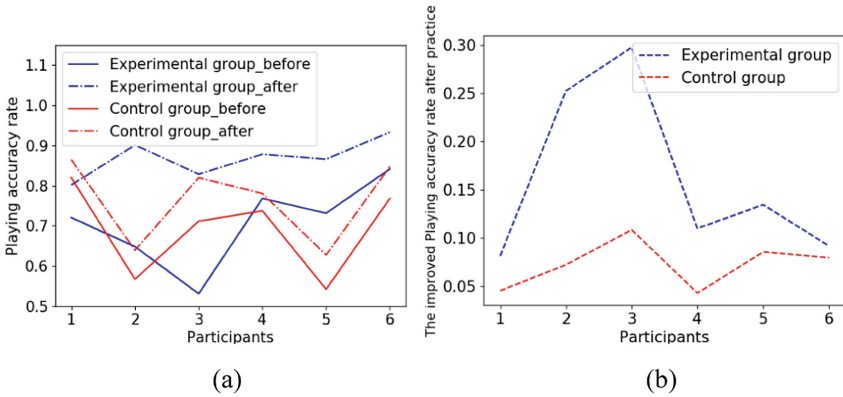


Fig. 5. (a) The playing accuracy rates. “experimental group_before” means the initial performance of the experimental group, “experimental group_after” refers to the performance of the experimental group after the practice session. The x values of 1–3 indicate the 3 volunteers in each group who played the first song of “swan lake”, and the x values of 4–6d represent the same three persons in each group who played the second song of “I do”. (b) The improved playing accuracy rate after practice.

In Fig. 5(a), after the practice session, the participants in both groups showed improvement in playing accuracy rate compared with the initial performance. Another thing to note is that the average values of the initial playing accuracy rate of both groups are close to each other, i.e. 0.6336 for the experimental group and 0.6997 for the control group. Therefore it can be assumed that all the novices in both groups are on the same level of playing the piano at the beginning, which ensures that the experimental results are not disturbed by the previous learning experiences of participants.

Figure 5(b) shows that the improved accuracy rate of the experimental group at each point is greater than that of the control group. The average increase of the experimental group is 0.1610 (SD = 0.08316), and greater than that of the control group 0.07208 (SD = 0.02279). The results indicate that the practice involving the new

method, which chooses some difficult measures based on the KNN classifier and arranges the repeated practice of those measures, is more helpful than the method that treats all the measures in a song equally as important.

For the first song “swan lake”, the average value of the improved accuracy rate for the experimental group is 0.2102 (SD = 0.09314) and for the control group is 0.0751 (SD = 0.02581). For the second song “I do”, the average value of the improved accuracy rate for the experimental group is 0.1118 (SD = 0.01753) and for the control group is 0.0691 (SD = 0.01882). The improvement of playing the first song is greater than that of playing the second one. It is because the second song is harder to play that the improvement is limited after a short-term practice.

In Fig. 5(b), the improved rates of the experimental group at two points are much larger than others. To explain those two points, it is found that in Fig. 5(a), those two participants have comparatively low rates for the initial performance, i.e. 0.6486 and 0.5315, lower than the average value 0.6997. It can be explained that at the beginning, those learners are not familiar with the AR system, and after a period of practice, they become accustomed to the AR display and can perform better.

3.3 Test Data Acquisition and Evaluation

In order to evaluate the generalization ability of the model in the test set, the measures labelled “difficult” in the test samples based on the KNN model are compared with the measures labelled “difficult” by experiment. Five out of six volunteers for evaluation were randomly selected and their initial performance data before the practice was used for labelling. According to the same criteria as applied to the KNN training data labelling, if the number of people who made mistakes at a measure is greater than 2, this measure is labelled “difficult”.

Finally, 57 new test samples in the evaluation were labelled, and Table 2 showed the prediction results. The measure numbers of true positive, true negative, false positive and false negative are listed respectively for both songs to evaluate the performance of the KNN classifier. According to the Eq. (3), the prediction accuracy for the test samples is illustrated by Eq. (4). Only the false negative classification is counted as an error, which means the positive measure (1 or labelled “difficult”) is classified as negative (0 or labelled “easy”), which is the same as before. As is shown in Eq. (4), the accuracy rate for the test set is 85.965%, close to the accuracy rate for the development set. Therefore, the developed KNN classifier based on the training set is verified by the new test samples.

Table 2. The prediction results for the 57 measures

Songs	TP	TN	FP	FN
I do	13	18	0	2
swan lake	11	7	0	6
sum	24	25	0	8

TP = True Positive, TN = True Negative, FP = False Positive, FN = False Negative.

$$accu_{test} = 1 - \frac{TN}{TP + TN + FP + FN} = 1 - \frac{8}{57} = 85.965\% \quad (4)$$

4 Conclusion and Future Work

In this article, to foster the interests of novices to learn the piano and to improve the performance of a short-term learner, we presented a new practice method with an AR system. It turned out that this new practice method is more helpful than the normal practice method for novices to improve their short-term learning performance. The increase in the average correctness rate of the experimental group is 0.1610 (SD = 0.08316), and greater than that of the control group 0.07208 (SD = 0.02279).

Further studies will include more volunteers so that the obtained data is more statistically significant. In the future, our research interest will adopt other more advanced machine learning methods, for example deep neural networks, to improve user experience of the AR aided piano learning system.

Acknowledgments. The authors would like to thank all the volunteers in the study. All of them are students at Chongqing University.

References

1. Azuma, R.T.: A survey of augmented reality. *Presence Teleoperators Virtual Environ.* **6**(4), 355–385 (1997)
2. Billinghurst, M., Clark, A., Lee, G.: A survey of augmented reality. *Found. Trends® Hum.–Comput. Interact.* **8**(2–3), 73–272 (2015)
3. Rogers, K., et al.: PIANO: faster piano learning with interactive projection. In: *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces*, pp. 149–158. ACM, November 2014
4. Das, S., Glickman, S., Hsiao, F.Y., Lee, B.: Music everywhere—augmented reality piano improvisation learning system
5. Huang, F., Zhou, Y., Yu, Y., Wang, Z., Du, S.: Piano AR: a markerless augmented reality based piano teaching system. In: *2011 International Conference on Intelligent Human-Machine Systems and Cybernetics (IHMSC)*, vol. 2, pp. 47–52. IEEE, August 2011
6. Sun, C.H., Chiang, P.Y.: Mr. Piano: a portable piano tutoring system. In: *2018 IEEE XXV International Conference on Electronics, Electrical Engineering and Computing (INTERCON)*, pp. 1–4. IEEE, August 2018
7. Rusiñol, M., Chazalon, J., Diaz-Chito, K.: Augmented songbook: an augmented reality educational application for raising music awareness. *Multimed. Tools Appl.* **77**(11), 13773–13798 (2018)

8. Li, G., Zhang, J.: Music personalized recommendation system based on improved KNN algorithm. In: 2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), pp. 777–781. IEEE, October 2018
9. Hackl, D., Anthes, C.: HoloKeys-an augmented reality application for learning the piano
10. Beyer, K., Goldstein, J., Ramakrishnan, R., Shaft, U.: When is nearest neighbors meaningful? In: ICDT Conference Proceedings (1999)