

SingPAD: A Knowledge Tracing Dataset Based on Music Performance Assessment

Ying Zhang¹, Yan Zhang¹, Wei Xu^{1*}, Zhifeng Wang^{2*}, Jianwen Sun^{2*}

¹Huazhong University of Science and Technology

²Central China Normal University

{zyinghuster, zyan_0917, xuwei}@hust.edu.cn, {zfwang, sunjw}@ccnu.edu.cn

ABSTRACT

Knowledge tracing (KT) aims to model a learner’s knowledge mastery level through his historical exercise records to predict future learning performance. Using this technology, learners can get appropriate customized exercises based on their current knowledge states, and thus the great potential of personalized teaching services such as intelligent tutoring systems and learning material recommendations can be stimulated. Currently, the mainstream datasets in KT include ASSISTments, EdNet, STATICS2011, etc., which are mainly based on objective testing data in the fields of mathematics and language, lacking of datasets on music performance assessment. Therefore, based on the context of performance assessment in music education, specifically sight-singing evaluation, we introduce *SingPAD*¹, the first dataset for performance assessment and the first music dataset in the field of KT, with abundant data collected by a public intelligent sight-singing practice platform, *SingMaster*. Unlike the existing KT datasets, each question in *SingPAD* is defined as a note in a music score, and learners’ music performance can be evaluated objectively and automatically utilizing music information retrieval technology. Several classical knowledge tracing models are tested on *SingPAD*, and the experimental results show that *SingPAD* exhibits good consistency and discriminability with existing datasets. *SingPAD* can be used as a benchmark dataset for applying knowledge tracing models to predict music knowledge mastery levels and promote the development of knowledge tracing research.

Keywords

Knowledge Tracing, Sight-singing Dataset, Music Performance Assessment

1. INTRODUCTION

Knowledge tracing is a challenging research direction in smart education, which aims to trace learners’ knowledge states based on their historical question-answering sequences and predict fu-

ture performance. Leveraging knowledge tracing methods can effectively extract value from the massive data generated by online education platforms and provide feedback on learning progress to learners, helping them to identify weakness and facilitating the customization of personalized learning strategies for different learners. The knowledge tracing task has been extensively studied, and various methods have emerged to address this issue.

There are mainly 3 categories of the existing research based on knowledge tracing methods: Bayesian Knowledge Tracing (BKT) [1], Item Response Theory (IRT) [2], and deep learning-based knowledge tracing. BKT builds its model with the Bayesian probability formula and treats the learner’s answering sequence as a hidden Markov model. It has been widely studied and extended into several derivative models applied in practical teaching environments [3]. IRT predicts learners’ performance by introducing parameters and constructing a logistic regression model. IRT considers that the learner’s response is influenced by both subjective factors, such as learner’s learning ability, and objective factors, such as difficulty and discrimination of questions. Therefore, the model incorporates parameters such as learner’s ability and question difficulty. Deep learning-based knowledge tracing applies deep neural networks to the knowledge tracing task, and was pioneered by Piech et al. [4]. In recent years, there has been an increasing number of studies exploring deep learning-based knowledge tracing models from different perspectives, such as memory structures [5], attention mechanisms [6], graph-based learning [7], and forgetting-aware mechanisms [8]. Alongside these developments, there has been a continuous emergence of knowledge tracing datasets, which mainly focus on subjects such as mathematics and language, including math datasets like ASSISTments and Junyi Academy, as well as datasets records English exercise data such as EdNet [9].

As a supplement to the existing KT research, this article focuses on a specific performance evaluation in music education, namely sight-singing evaluation. By leveraging a large amount of data collected by the public intelligent sight-singing practicing platform called *SingMaster*, we construct the first music KT dataset, *SingPAD*. Sight-singing is a fundamental discipline in music education and a crucial basic skill in developing musical abilities [10], which involves the ability to read and sing a piece of music at first sight. Information such as pitch, duration, and rhythm in musical scores is essential for sight-singing performance [11]. Maintaining accurate pitch is fundamental to ensure the interpretation of a song, while rhythm and phrasing are crucial for smooth and expressive singing voice. Learners’ performance on these sight-singing skills accurately reflects their sight-singing abilities. However, current

¹ Dataset Repository: <https://github.com/itec-hust/singKT-dataset>

*Corresponding author

Y. Zhang, Y. Zhang, W. Xu, Z. Wang, and J. Sun. Singpad: A knowledge tracing dataset based on music performance assessment. In B. Paaßen and C. D. Epp, editors, *Proceedings of the 17th International Conference on Educational Data Mining*, pages 332–340, Atlanta, Georgia, USA, July 2024. International Educational Data Mining Society.

© 2024 Copyright is held by the author(s). This work is distributed under the Creative Commons Attribution NonCommercial NoDerivatives 4.0 International (CC BY-NC-ND 4.0) license. <https://doi.org/10.5281/zenodo.12729828>

sight-singing training primarily relies on one-on-one instruction from teachers, lacking intelligent means of teaching. By using intelligent technology to track the mastery of sight-singing knowledge concepts, feedback and personalized sight-singing resources can be effectively provided to learners, enhancing the intelligence and autonomy of their learning process.

In this article, we construct the first dataset for performance assessment and the first music dataset in the field of KT, *SingPAD*, and attempt to utilize knowledge tracing methods to capture the dynamic changes in learners' sight-singing abilities. Our main contributions are summarized as follows: 1) We collect 19,893 sight-singing records from 1,074 students through an online sight-singing practice platform. 2) We construct a sight-singing knowledge system and extract questions, knowledge concepts, and answer correctness labels from the sight-singing data to model the knowledge tracing task in sight-singing. 3) Several classical knowledge tracing models are tested on *SingPAD* and the experimental results confirm the usability of the dataset in solving sight-singing knowledge tracing problems.

2. RELATED WORK

2.1 Knowledge Tracing

Knowledge tracing works can be traced back to the mid-1970s [12], and the concept was first introduced by Anderson et al. in 1986 [13]. Early works [1][3][14][15] follow Bayesian inference methods and primarily apply a hidden Markov model, using Bayesian rules to update states. Subsequently, with the rise of classical machine learning methods such as logistic regression models, some works [16][17] use the factor analysis method to track students' knowledge states, focusing on learning general parameters for answer prediction from historical data.

Inspired by the breakthrough in deep learning, knowledge tracing models based on deep learning have emerged. DKT [4] pioneers the application of deep neural networks for knowledge tracing. It uses a recurrent neural network (RNN) to encode sequence information obtained from previous interactions, computes the hidden state sequence, and predicts the probability of correctly answering the next question at each time step. Compared to BKT, DKT can capture and utilize deeper levels of student knowledge representations. Many attempts have been made to extend and improve upon DKT to address its limitations. To address the issue of unstable predictions due to the long-term dependencies in DKT, DKT+ [18] enhances the loss function with two additional regularization terms to prevent situations where a learner performs well on knowledge concepts but the prediction of his mastery level decreases. DKT+Forget [8] considers the case of students forgetting knowledge concepts and incorporates more information to further improve performance.

In recent years, researchers have made various explorations in deep knowledge tracing model structure. Inspired by the memory augmented neural network, DKVMN [5] introduces a more powerful memory structure by using key-value memory to represent knowledge states, allowing for greater expressive capacity compared to the hidden variable in DKT. SAKT [6] follows the Transformer architecture and first applies the self-attention mechanism to knowledge tracing to capture the relationships between questions and their relevance to learners' knowledge state. AKT [19] also adopts the encoder structure in Transformer, using two encoders: a question encoder and a knowledge encoder to learn context awareness for both questions and answers. Choi et al. [20] notice that the attention layer in SAKT is too shallow to capture

deep-level information, and they address this issue by using a stacked self-attention layer in the encoder-decoder architecture. Several works [7][21] introduce graph structures to capture various relationship patterns in knowledge tracing. Based on graphs, GKT [7] represents the relationship between knowledge concepts as a graph, with nodes representing knowledge concepts and edges representing dependencies between concepts. When a student answers a question related to certain concepts, the update of the knowledge state involves not only the knowledge concepts of the question but also the information of its neighboring nodes.

2.2 Knowledge Tracing Datasets

There are several commonly used datasets for knowledge tracing tasks:

- ASSISTments is an online teaching assistant platform that provides learners with tools for online learning and evaluation primarily in mathematics. It has a large user base, currently reaching more than 500,000 learners. Derived from ASSISTments platform, the ASSISTments datasets are one of the largest publicly available datasets and are widely used for knowledge tracing tasks. When comparing different datasets in terms of interaction counts and the number of learners, it is evident that the ASSISTments datasets have a larger number of learners, but the average number of interactions per learner is relatively low. The commonly used sub-datasets are ASSIST09 and ASSIST12, which collected data from students using the ASSISTments platform during 2009-2010 and 2012-2013, respectively. ASSIST12 contains important information such as student ID, knowledge concepts ID, correctness of the answers to each question, start time of the question, and end time. The ASSISTments datasets record the precise start and end times of each question, down to the second, which is advantageous for studying the impact of time intervals between questions on students' knowledge memorization.
- EdNet [9], originating from *Senta* (a multi-platform AI tutoring service), is currently the largest publicly available dataset that collects English exercise records from multiple platforms since 2017. It includes data from nearly 780,000 users, with an average of 441.2 interactions per student. EdNet provides rich learning behavior data, such as reading comprehension exercises and whether students have purchased paid courses, allowing researchers to analyze students' characteristics from different perspectives.
- STATICS2011 is derived from the Engineering Statics course at Carnegie Mellon University. It consists of data from 333 students with a total of 1,223 questions and 189,927 student interactions, covering 156 knowledge concepts.
- The Junyi Academy dataset originates from Junyi Academy, an online e-learning platform in Taiwan. Junyi Academy platform covers subjects such as mathematics, chemistry, and English for elementary to high school students, and provides testing scenarios such as teaching videos and in-class exercises. In 2015, the platform released log data specifically for the mathematics subject. This dataset includes 722 questions covering 41 knowledge concepts, resulting in 2,595,292 interactions answered by 247,606 students over two years. Additionally, the knowledge tree constructed from the Junyi Academy dataset provides sequential relationships between exercises as well as similar relationships.
- The KDDcup 2010 dataset comes from the KDDcup 2010 Educational Data Mining challenge [22]. It consists of eighth-grade students' answering records on algebra exercises from 2005 to 2007. The dataset is divided into three subsets: Alge-

bra 2005-2006 (including detailed responses from 574 students across 112 knowledge concepts), Algebra 2006-2007 (including 2,289,726 interactions generated by 1,840 students across 523 knowledge concepts), and Bridge to Algebra (including interactions from 1,146 students across 493 knowledge concepts).

3. DATASET CONSTRUCTION

In this section, the construction methods of *SingPAD* are systematically described, including the construction of the sight-singing knowledge system, data collection, data cleaning and labeling, as well as dataset description and usage.

3.1 The Construction of the Sight-singing Knowledge System

The sight-singing knowledge system primarily references the chapter catalogs of several sight-singing books in the fields of Chinese general and professional music education, as well as the explicit requirements for knowledge points in the Chinese entrance examination syllabus of renowned music institutions (please refer to the dataset repository for the books list). The system was designed in collaboration with two professional and experienced music teachers. On the one hand, this sight-singing knowledge system is comprehensive, incorporating the content from multiple sight-singing textbooks. On the other hand, it takes into account the characteristics of expressive evaluation, focusing on pitch and rhythm as the core components of the knowledge system, which aligns with the *SingMaster* data collection platform.

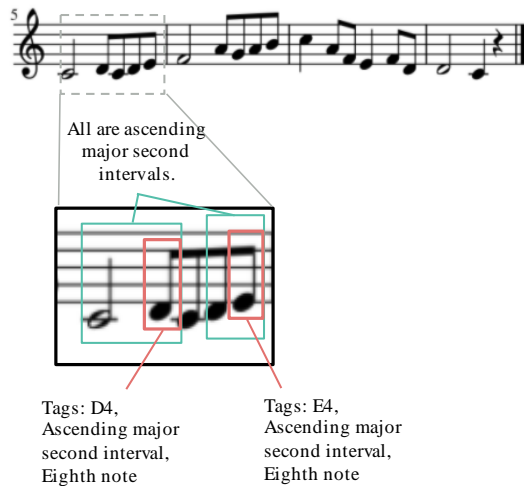


Figure 1. Sample diagram of a sight-singing score.

The staff notation (five-line staff) serves as the primary form of carrying the sight-singing knowledge system. As shown in Figure 1, a sight-singing score typically consists of multiple lines of staff notation, with each note in it having its own pitch and rhythm information. Although the sight-singing performance evaluation actually involves the comprehensive assessment of three abilities: score reading ability, singing ability, and pitch discrimination ability, the evaluation result of the learner’s accuracy on pitch and rhythm for each note is often used as an overall assessment of these three abilities during implementation. This principle guides our construction of the sight-singing knowledge system.

Combining multiple classic sight-singing textbooks, this article defines the core knowledge concepts of the sight-singing knowledge system as shown in Table 1. The knowledge system is developed based on two dimensions: pitch and rhythm. Pitch is

further divided into three parts: single note pitch, accidental pitch (sharps and flats), and intervals. Single note pitch is represented by the letter name of the basic pitch classes (corresponding to the white keys on a piano, CDEFGAB) along with its octave. For example, C4 corresponds to the middle C on a piano. Accidental pitch refers to notes with sharps or flats such as the black keys of a piano, which can be seen as a special case of single note pitch. Intervals, on the other hand, represent the pitch relationship between two notes and are measured in degrees. In this article, the interval of a note is defined as its pitch relationship with the preceding note, categorized as ascending or descending. For example, the two green boxes in Figure 1 indicate ascending major second interval. In addition to pitch, the knowledge system also defines core knowledge concepts in rhythm, including note duration, dotted notes, and the tied note (where two or more notes of the same pitch are connected with a curved line called a tie, with no other notes in between). Overall, the sight-singing knowledge system defines a total of 43 core knowledge concepts.

Table 1. The core knowledge concepts of the sight-singing knowledge system.

Pitch		
Intervals	Single note pitch	Rhythm
Ascending minor second interval	G3	Whole note
Descending minor second interval	A3	Half note
Ascending major second interval	B3	Fourth note
Descending major second interval	C4	Eighth note
Ascending major third interval	D4	16th note
Descending minor third interval	E4	32th note
Ascending major third interval	F4	Dotted half note
Descending major third interval	G4	Dotted fourth note
Ascending perfect fourth interval	A4	Dotted eighth note
Descending perfect fourth interval	B4	Tied note
Ascending augmented fourth and perfect fifth intervals	C5	
Descending augmented fourth and perfect fifth intervals	D5	
Ascending sixth intervals	E5	
Descending sixth intervals	F5	
Ascending seventh intervals	Accidental notation	
Descending seventh intervals		
Ascending perfect octave and ninth intervals		
Descending perfect octave and ninth intervals		

Note: Middle C is referred to as C4.

As observed from Figure 1, a single note can contain multiple core knowledge concepts. For example, note D4 possesses three core knowledge concepts: D4, ascending major second interval, and eighth note. These three core knowledge concepts together form a new knowledge concept. In our dataset, there are 270 such combinations of knowledge concepts.

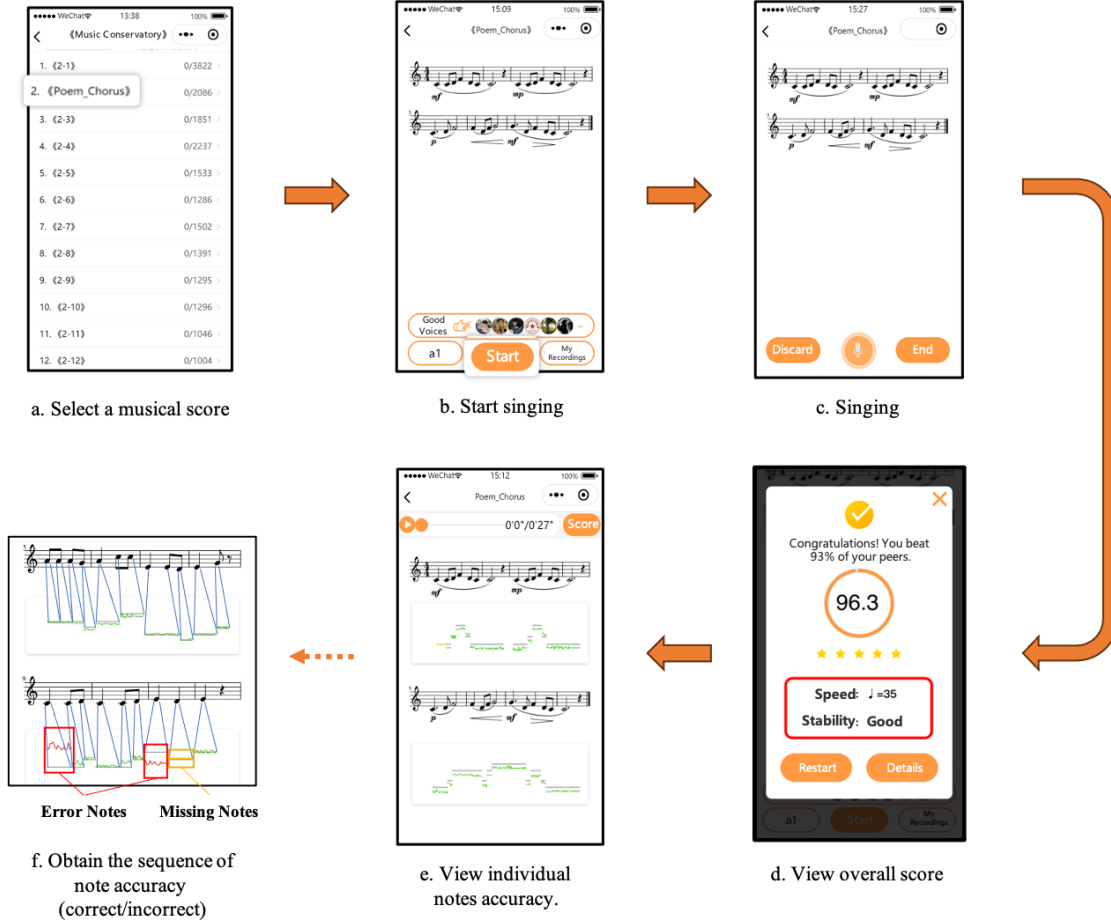


Figure 2. Step-by-step diagram for collecting users' sight-singing information of *SingMaster*.

3.2 Data Collection

This article utilizes a self-developed sight-singing evaluation system, the WeChat mini-program *SingMaster* [23] to collect learners' sight-singing information, which is an online platform that utilizes music evaluation algorithms based on MusicYOLO [24] to provide automated AI evaluation services for sight-singing users. Launched in 2019, *SingMaster* has collected more than 770,000 sight-singing records from nearly 8,800 users in more than four years. It should be noted that the WeChat mini-program has a strict privacy policy and *SingMaster* complies with this policy. Users are informed and should agree to the collection of relevant information by the mini-program before using it. Additionally, we try to protect user privacy in the dataset as much as possible, for example, only randomly generated IDs are used to identify users, and the dataset only contains analyzed sight-singing data instead of the raw audio.

The ability to provide automatic sight-singing performance evaluation relies on a comprehensive technical framework: Firstly, the learner's sight-singing audio is inputted into the sight-singing transcription model MusicYOLO, where target detection and pitch extraction are performed on the musical notes to obtain information such as pitch, onset, and offset for each note. Once the transcription note sequence is obtained, it is aligned with the corresponding musical score fragments. The note alignment algorithm is based on an improved version of the Needleman-Wunsch (NW) algorithm [25]. This algorithm can skip incorrect

matches, addressing issues such as continuous singing errors, overall pitch deviation, missed notes, and structural differences caused by multiple pitches in the record. After the alignment process is completed, the aligned note sequence is evaluated, yielding evaluation results.

SingMaster provides accompanying automatic evaluation services for multiple sight-singing textbooks. Within the same book, different musical pieces have varying levels of difficulty and each book contains a range of 20 to 500 songs. The complexity and practice difficulty of the pieces gradually increase based on the learner's learning patterns. The books range from simple four-beat pieces in A minor and C major to six-beat and nine-beat pieces with more complex rhythmic patterns such as dots and triplets. Due to the diverse categories of books and the wide range of difficulty levels, *SingMaster* caters to the needs of different user groups. It attracts a broad audience, including beginners with weak foundations and advanced learners with considerable music literacy. This wide user base ensures the diversity of *SingPAD*.

SingMaster collects user sight-singing information through the following steps as shown in Figure 2: (a) Select a musical score: The user logs in to the *SingMaster* system and chooses sight-singing books based on his interests. The mini-program displays the available songs within the chosen book, allowing the user to select a specific song. (b) Start singing: The mini-program presents the music score to the user, who then begins to sight-sing the piece. (c) Singing: When the user sings, the mini-program records

audio. (d) View overall score: After the user completes sight-singing, *SingMaster* provides real-time evaluation and feedback on the performance. The user can get sight-singing score based on the evaluation rules of the mini program. (e & f) View individual notes accuracy: The user has the option to replay his record and review the feedback. With the assistance of the algorithm, *SingMaster* offers note-level sight-singing evaluations, with pitches within the reference value thresholds shown in green, beyond the thresholds shown in red, timing missing shown in yellow, and missed notes shown in gray. Learners can utilize the feedback provided by *SingMaster* for AI-assisted sight-singing practice.

3.3 Data Cleaning and Data Annotation

We perform data cleaning on the raw data to make the data more focused, aiming to provide data with greater information density and enable a more even distribution of records. The filtering criteria are as follows:

- Invalid records due to excessive environmental noise. In noisy environments, it is challenging to extract note information from the record.
- Records with scores below 10. Learners who receive extremely low scores have very few correctly sung notes in their performance, which demonstrates their lacking of basic sight-singing ability.
- Deletion of learners with fewer than 5 records. When the learner’s historical interaction is too limited, its coverage of knowledge concepts is insufficient, making it difficult to make meaningful inferences about the learner’s mastery level.
- Limiting the records number for some learners. To ensure data balance and avoid an imbalance caused by excessive records from certain users, a maximum of 30 records is randomly selected for each learner.

Following these criteria, 19892 sight-singing records from 1074 learners are selected.

We annotate score files and sight-singing records to extract three important tags in the knowledge tracing task: questions, knowledge concepts, and learners’ answer correctness. Each note in the score is considered a question, which corresponds to one or more knowledge concepts such as pitch and rhythm. Whether a learner sings the note correctly represents whether he or she answers the question correctly. The knowledge concepts corresponding to each note are extracted by utilizing the `xml.dom.minidom` module in Python to parse the MusicXML file of the score. The MusicXML file contains rich music information, such as note types (whole note, half note, etc.), pitch, intervals, and rhythm. The correct/incorrect labels are extracted from the learners’ sight-singing records collected by *SingMaster*, which provides an audio analysis algorithm that gives the note-level accuracy of each record, from which the learner’s answer correctness on each note can be obtained.

The *SingPAD* dataset covers 2587 questions and 270 knowledge concepts, with a total of 1,080,852 exercises. The following table provides a comparison of *SingPAD* with other datasets:

Table 2. Comparison of SingPAD with other datasets.

	Students	Questions	Knowledge concepts	Exercises	Subject
ASSIST09	4151	19,840	110	325,637	Math

	Students	Questions	Knowledge concepts	Exercises	Subject
ASSIST12	27,485	53,065	265	2,709,436	Math
EdNet	784,309	13,169	293	131,441,538	English
Statics2011	335	80	1,224	361,092	Engineering
Ours	1,074	2,587	270	1,080,852	Music

3.4 Data Description

In the knowledge tracing task, a student’s learning records can be described as a sequence of question-concept-answer triplets, that is, $(q_1, c_1, a_1), (q_2, c_2, a_2), \dots, (q_t, c_t, a_t)$. Here, q_i represents the i -th question, c_i represents the corresponding knowledge concept involved in the exercise, and a_i represents the learner’s response to the question. This data format is commonly used in various deep learning knowledge tracing models like DKT and DKVMN. To construct triplets specific to the music knowledge tracing task, each musical score is considered as a set of questions, and each note of the score is considered as a question. Through the sight-singing knowledge system, the correspondence between the note and the knowledge concept is established. The learner’s performance on the note is treated as the answer to the question.

The *SingPAD* dataset contains the main answering record data table (*RecordDS*) and two supplementary information data tables (*UserDS*, *OpernDS*).

Table 3. Answering record data table *RecordDS*.

Field	Description
<code>user_id</code>	ID of the student.
<code>opern_id</code>	ID of the sheet music.
<code>record_id</code>	ID of the record.
<code>create_time</code>	Time of completion of the exercise.
<code>qa_array</code>	Question-answer array.

Where `record_id` uniquely identifies a learner’s sight-singing record, and `qa_array` is a two-dimensional array that records the id of each question in the musical score as well as the learner’s performance evaluation on questions (correct: 1, false: 0).

Table 4. Student’s information data table *UserDS*.

Field	Description
<code>user_id</code>	ID of the student.
<code>user_avgscore</code>	Student’s average score.
<code>sing_num</code>	Student’s exercise number.

The *UserDS* table records sight-singing information for the 1074 learners contained in the dataset, with `user_id` uniquely identifying a learner.

Table 5. Sheet music information data table *OpernDS*.

Field	Description
<code>opern_id</code>	ID of the sheet music.
<code>book_id</code>	ID of the textbook to which the sheet music belongs.
<code>exe_num</code>	The number of times the sheet music has been practiced.
<code>avg_score</code>	Average score of the sheet music.
<code>qc_array</code>	Note question-concept array.

The *OpernDS* table records the exercise information for the scores in the dataset. `opern_id` uniquely identifies the score, `avg_score` represents the average score of the musical score across all records, and `qc_array` is a two-dimensional array that records the id of each note question in the musical score as well as the id of the

corresponding knowledge concept. The learner’s performance on each question is recorded in the `qa_array` in *RecordDS*.

4. VALIDATION

4.1 Support for Existing Knowledge Tracing Models

The *SingPAD* dataset adopts an equivalent structural definition to the existing datasets. Each score is treated as a set of exercises, with each note being a question within it. A learner’s performance on a musical score is considered an answering process. The learner’s performance evaluation on each note is given by the automatic sight-singing evaluation algorithm, and additionally, the correlation between each note and the knowledge concept is labeled by explicit rules. Therefore on *SingPAD*, it is possible to predict whether a learner can correctly sing a specific note by analyzing his historical exercise records, and track the learner’s mastery level of knowledge concept across multiple practice sessions.

The following mainstream KT models are subjected to validation tests on the *SingPAD* dataset.

- 1) Bayesian Knowledge Tracing (BKT): The BKT model is based on Bayesian probability theory for inference and prediction. It models the learner’s exercise answering sequence as a hidden Markov process and sets the learner’s knowledge state as a binary variable.
- 2) Deep Knowledge Tracing (DKT): DKT is the first model to introduce deep learning into knowledge tracing tasks. It uses a recurrent neural network to dynamically capture the relationships between question-answering interactions. Compared to BKT, DKT can better track changes in knowledge mastery.
- 3) Dynamic Key-Value Memory Network (DKVMN): This model is built upon a memory augmented neural network. It incorporates an external memory module, using a static key matrix to store concepts and a dynamic value matrix to update knowledge states.
- 4) Self-Attentive Knowledge Tracing (SAKT): SAKT is the first to apply an attention mechanism in knowledge tracing models. It exclusively employs the encoder module of Transformer, and allocates different weights to the historical interaction records to enhance the impact of critical information.
- 5) Attentive Knowledge Tracing (AKT): AKT adds three attention heads to integrate contextual information, so as to link a student’s performance on the target question with their historical answering sequences.
- 6) Graph-based Interaction Knowledge Tracing (GIKT): GIKT utilizes a graph neural network to represent the relationships between questions and knowledge concepts. It uses an attention module to calculate weights between the learner’s current state and relevant historical interactions, providing a better representation of the learner’s ability on specific problems.

4.2 Experimental Settings

Five-fold cross-validation is employed to obtain stable experimental results, dividing the dataset equally into five parts, four as the training set and one as the validation set. The cross-validation is repeated five times, once for each sub-dataset, and the average of the results is used as the final result. Commonly used evaluation metrics in KT, namely AUC (Area Under the ROC Curve)

and ACC (Accuracy) are used to evaluate the models. Higher values for both metrics indicate better performance.

4.3 Results

Table 6. Results of mainstream KT models on *SingPAD*.

Model	AUC	ACC
BKT	0.6865	0.7472
DKT	0.8325	0.8086
DKVMN	0.8237	0.8144
SAKT	0.8396	0.8197
AKT	0.8423	0.8157
GIKT	0.8480	0.8253

The table shows the results of the six models, which achieve good performance on our dataset. These results indicate that the models can effectively model the mastery level of sight-singing knowledge based on the historical interaction records, and accurately predict the performance of learners on the next musical fragment. This validates the usability of *SingPAD* dataset and the generalization ability of knowledge tracing algorithms in the field of sight-singing.

Compared to the traditional KT model, BKT, the models based on deep learning method perform better in the sight-singing task. It is because BKT only considers binary knowledge states: “mastered” and “not mastered”, lacking detailed expression. In contrast, deep learning-based knowledge tracing models can better track the evolving knowledge state. In addition, we observe that SAKT and AKT outperform DKT and DKVMN. SAKT and AKT both employ attention mechanisms, indicating the advantages of attention mechanisms in handling sequence tasks like knowledge tracing. They perform well in capturing the correlations between the preceding and succeeding musical exercises and the interaction behaviors. We also observe that GIKT significantly outperforms other baselines in the current dataset. It constructs a bipartite graph structure between musical exercises and knowledge concepts, and utilizes a graph convolutional network to extract higher-order information from the relational graph. This result suggests that modeling KT problems using graph structures can better capture the higher-order relationships between exercises and concepts.

Overall, these findings demonstrate the effectiveness of the selected models in the context of the *SingPAD* dataset, showcasing the potential of knowledge tracing algorithms in the field of sight-singing.

5. FUTURE WORK

Currently, *SingPAD* only includes a small portion of the data from the *SingMaster* platform. In the future, we plan to include more data in this dataset, enhance the modeling of knowledge concepts difficulty and musical score difficulty, and explore the relationships between different knowledge concepts, which will strengthen the support for research in the field of knowledge tracing.

Furthermore, sight-singing exercise evaluation differs from the traditional objective exercises. It combines the characteristics of music discipline and performance evaluation. In music, there is a correlation between consecutive notes. The correctness of the current note’s performance depends not only on its pitch characteristics but also on factors such as note span, bar feature, and the overall difficulty of the musical score. In terms of performance evaluation, sight-singing ability cannot be easily improved in a short period. A learner’s overall performance tends to be stable but can be influenced by incidental factors, leading to local fluctuation.

tuations. These characteristics will be discussed in the following research in sight-singing performance evaluation.

6. CONCLUSION

Traditional sight-singing concept teaching relies on one-on-one guidance between teachers and students. We aim to leverage knowledge tracing technology to capture learners' progress and build individual profiles. This will help learners dynamically adjust their learning paths, assist teachers in addressing specific knowledge gaps, and facilitate personalized exercise recommendations on online education platforms. First, we utilize the advantages of online education platforms to collect rich interaction data from learners, leading to the development of the first KT dataset focusing on performance evaluation. Then, we compile a comprehensive sight-singing knowledge system based on the characteristics of major sight-singing textbooks and the assessment requirements of art colleges. For data annotation, we define extraction rules based on the knowledge concept definition and employ an automated labeling technique to extract labels from musical scores and sight-singing records, reducing reliance on manual annotation. In addition to the fundamental data, we also provide auxiliary datasets encompassing musical score information and learner information, enabling researchers to explore sight-singing data from multiple perspectives.

Finally, we test and validate the dataset on various classic knowledge tracing models. The positive application results demonstrate the dataset's usability and the feasibility of the music knowledge tracing task. We hope that the construction of this dataset will provide valuable references for the development of knowledge tracing in the field of music.

7. ACKNOWLEDGMENTS

This work is supported by the National Natural Science Foundation of China (No. 62277019).

8. REFERENCES

- [1] Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User modeling and user-adapted interaction*, 4, 253-278.
- [2] Deonovic B, Yudelson M, Bolsinova M, et al. Learning meets assessment: On the relation between item response theory and Bayesian knowledge tracing. *Behaviormetrika*, 2018, 45(2): 457-474.
- [3] Pardos, Z. A., & Heffernan, N. T. (2010). Modeling individualization in a bayesian networks implementation of knowledge tracing. In *User Modeling, Adaptation, and Personalization: 18th International Conference, UMAP 2010, Big Island, HI, USA, June 20-24, 2010. Proceedings 18* (pp. 255-266). Springer Berlin Heidelberg.
- [4] Piech, C., Bassen, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L., & Sohl-Dickstein, J. (2015, December). Deep knowledge tracing. In *Proceedings of the 28th International Conference on Neural Information Processing Systems-Volume 1* (pp. 505-513).
- [5] Zhang, J., Shi, X., King, I., & Yeung, D. Y. (2017, April). Dynamic key-value memory networks for knowledge tracing. In *Proceedings of the 26th international conference on World Wide Web* (pp. 765-774).
- [6] Pandey, S., & Karypis, G. (2019). A self-attentive model for knowledge tracing. In *12th International Conference on Educational Data Mining, EDM 2019* (pp. 384-389).
- [7] Nakagawa, H., Iwasawa, Y., & Matsuo, Y. (2019, October). Graph-based knowledge tracing: modeling student proficiency using graph neural network. In *IEEE/WIC/ACM International Conference on Web Intelligence* (pp. 156-163).
- [8] Nagatani, K., Zhang, Q., Sato, M., Chen, Y. Y., Chen, F., & Ohkuma, T. (2019, May). Augmenting knowledge tracing by considering forgetting behavior. In *the world wide web conference* (pp. 3101-3107).
- [9] Choi, Y., Lee, Y., et al. Ednet: A large-scale hierarchical dataset in education. In *Artificial Intelligence in Education: 21st International Conference, AIED 2020, Ifrane, Morocco, July 6-10, 2020, Proceedings, Part II 21* (pp. 69-73).
- [10] Ottman, R. W. (1956). *A statistical investigation of the influence of selected factors on the skill of sight-singing*. University of North Texas.
- [11] Henry, M. L. (2011). The effect of pitch and rhythm difficulty on vocal sight-reading performance. *Journal of Research in Music Education*, 59(1), 72-84.
- [12] Brown, J. S., & Burton, R. R. (1978). Diagnostic models for procedural bugs in basic mathematical skills. *Cognitive science*, 2(2), 155-192.
- [13] Anderson, J. R., Boyle, C. F., Corbett, A. T., & Lewis, M. W. (1990). Cognitive modeling and intelligent tutoring. *Artificial intelligence*, 42(1), 7-49.
- [14] Yudelson, M. V., Koedinger, K. R., & Gordon, G. J. (2013). Individualized bayesian knowledge tracing models. In *Artificial Intelligence in Education: 16th International Conference, AIED 2013, Memphis, TN, USA, July 9-13, 2013. Proceedings 16* (pp. 171-180). Springer Berlin Heidelberg.
- [15] Käser, T., Klingler, S., Schwing, A. G., & Gross, M. (2017). Dynamic Bayesian networks for student modeling. *IEEE Transactions on Learning Technologies*, 10(4), 450-462.
- [16] Ebbinghaus, H. (2013). Memory: A contribution to experimental psychology. *Annals of neurosciences*, 20(4), 155.
- [17] Pavlik, P. I., Cen, H., & Koedinger, K. R. (2009, July). Performance Factors Analysis--A New Alternative to Knowledge Tracing. In *Proceedings of the 2009 conference on Artificial Intelligence in Education* (pp. 531-538).
- [18] Yeung, C. K., & Yeung, D. Y. (2018, June). Addressing two problems in deep knowledge tracing via prediction-consistent regularization. In *Proceedings of the fifth annual ACM conference on learning at scale* (pp. 1-10).
- [19] Ghosh, A., Heffernan, N., & Lan, A. S. (2020, August). Context-aware attentive knowledge tracing. In *Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining* (pp. 2330-2339).
- [20] Choi, Y., Lee, Y., Cho, J., Baek, J., Kim, B., Cha, Y., ... & Heo, J. (2020, August). Towards an appropriate query, key, and value computation for knowledge tracing. In *Proceedings of the seventh ACM conference on learning@ scale* (pp. 341-344).
- [21] Yang, Y., Shen, J., Qu, Y., et al. GIKT: a graph-based interaction model for knowledge tracing. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2020, Ghent, Belgium, September 14-18, 2020, Proceedings, Part I* (pp. 299-315). Springer International Publishing.

- [22] Stamper, J., Niculescu-Mizil, A., Ritter, S., Gordon, G., Koedinger, K., 2010. Challenge data set from KDD cup 2010 educational data mining challenge.
- [23] Xu, W., Tian, B., Luo, L., Yang, W., Wang, X., & Wu, L. (2022, October). SingMaster: A Sight-singing Evaluation System of “Shoot and Sing” Based on Smartphone. In *Proceedings of the 30th ACM International Conference on Multimedia* (pp. 6944-6946).
- [24] Wang, X., Xu, W., Yang, W., & Cheng, W. (2022, May). Musicolo: A sight-singing onset/offset detection framework based on object detection instead of spectrum frames. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 396-400). IEEE.
- [25] Needleman, S. B., & Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of molecular biology*, 48(3), 443-453.

APPENDIX

We here provide some graphs and plots made from the *SingPAD* dataset, as shown in Figure 3-7.

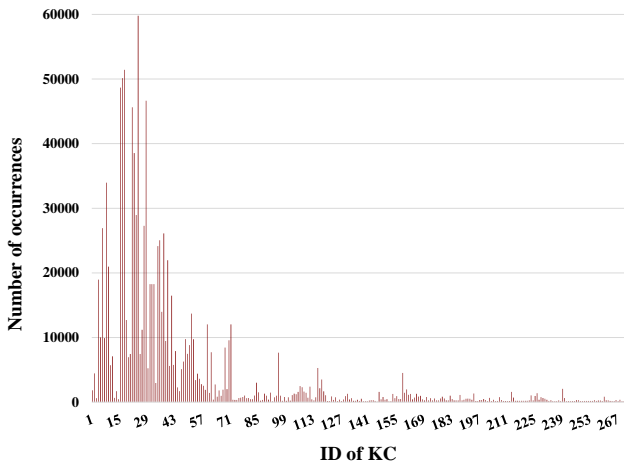


Figure 3. Number of times the 270 KCs occur in sight-singing records (the minimum is 63).

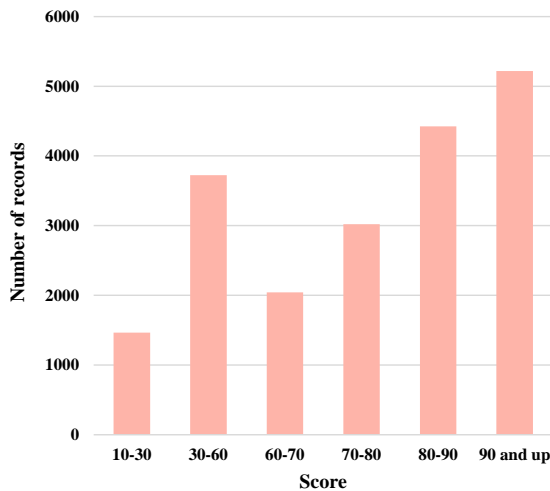


Figure 4. Distribution of scores of sight-singing records.

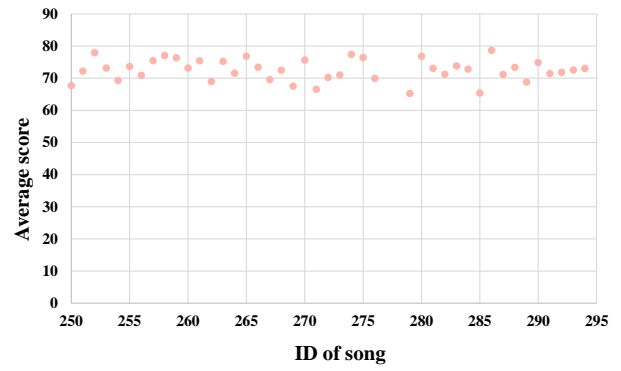


Figure 5. Distribution of the average scores of different songs.

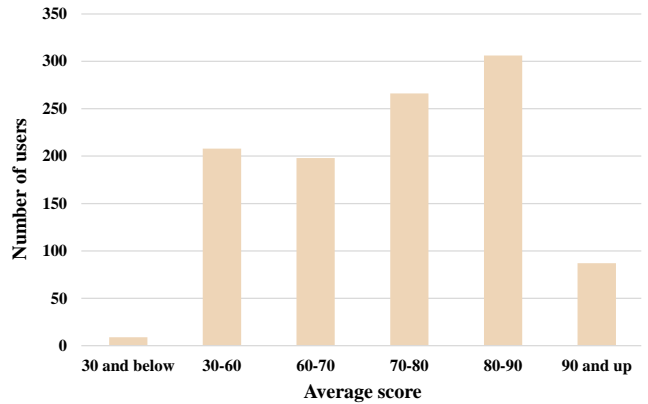


Figure 6. Distribution of the average scores of different users.

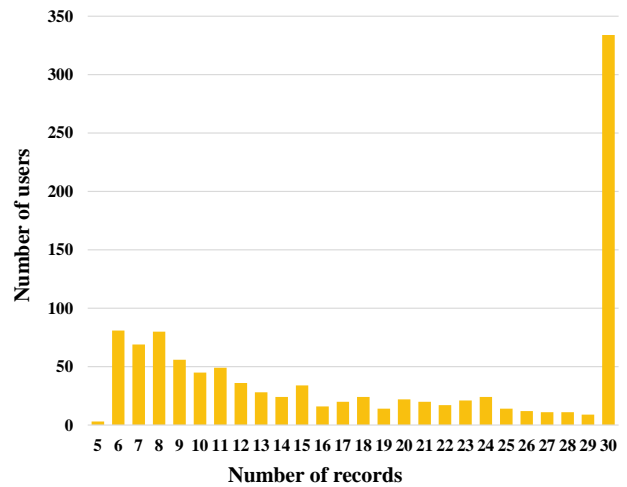


Figure 7. Distribution of the number of records by different users.

We also provide the raw data without data cleaning in the dataset repository. The following graphs describe the raw data.

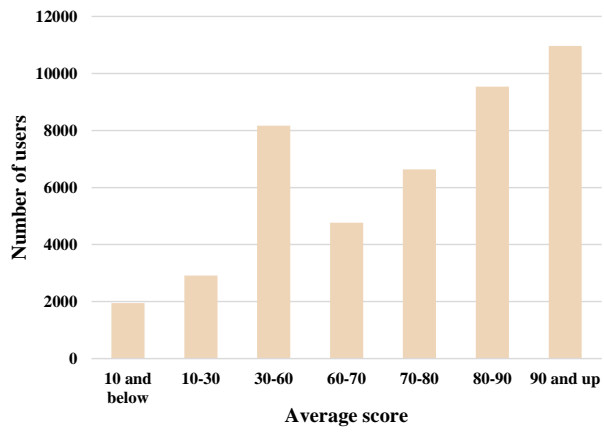


Figure 8. Distribution of the average scores of different users (raw data).

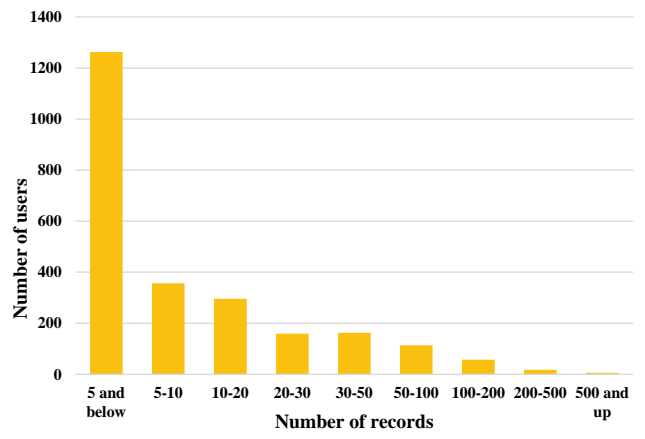


Figure 9. Distribution of the number of records by different users (raw data).