

BookBuddy: Turning Digital Materials Into Interactive Foreign Language Lessons Through a Voice Chatbot

Sherry Ruan¹, Angelica Willis¹, Qianyao Xu², Glenn M. Davis¹,
Liwei Jiang¹, Emma Brunskill¹, James A. Landay¹

¹Stanford University, ²Tsinghua University
{ssruan, arwillis, qianyao, gmdavis, ljianglw, ebrun, landay}@stanford.edu

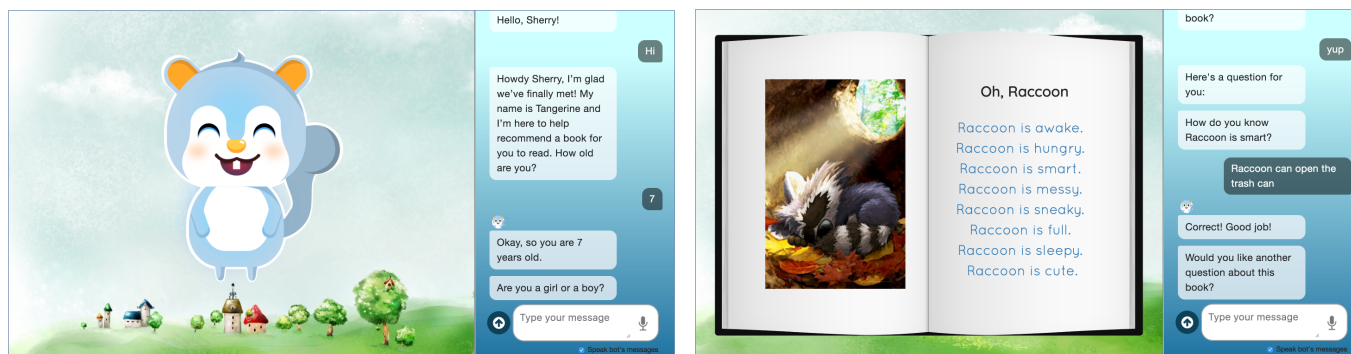


Figure 1: (Left) The chatbot in the BookBuddy system is greeting the child and gathering user information for book recommendation. (Right) The chatbot is quizzing the child on topics related to the book *Oh, Raccoon* shown in the left window.

ABSTRACT

Digitization of education has brought a tremendous amount of online materials that are potentially useful for language learners to practice their reading skills. However, these digital materials rarely help with conversational practice, a key component of foreign language learning. Leveraging recent advances in chatbot technologies, we developed BookBuddy, a scalable virtual reading companion that can turn any reading material into an interactive conversation-based English lesson. We piloted our virtual tutor with five 6-year-old native Chinese-speaking children currently learning English. Preliminary results suggest that children enjoyed speaking English with our virtual tutoring chatbot and were highly engaged during the interaction.

ACM Classification Keywords

J.1 Computer Applications: Education; I.2.7 Artificial Intelligence: Natural Language Processing

Author Keywords

Educational chatbot; foreign language learning; pedagogical agent

INTRODUCTION

Being immersed in a conversational environment is important when learning to speak a new language. In particular, interactive conversation can help children practice their communication skills and promote linguistic and cognitive development [11]. However, for children who are learning a foreign language, opportunities to engage in conversations in foreign languages are usually limited.

Robot peers have been proposed as a solution to engage children [3], help with vocabulary learning [13], and improve storytelling skills [5]. However, pedagogical materials for these physical robots require a non-trivial amount of effort to design and program. Since the Internet has abundant resources for language learning, we instead propose turning existing reading material into interactive lessons through a voice-enabled chatbot. In this work, we present BookBuddy, a virtual reading companion system that can automatically recommend appropriate learning materials from a book database, lead interactive tutoring conversations with children, and assess children's reading comprehension to provide adaptive feedback.

We piloted the BookBuddy system with five 6-year-old native Chinese-speaking children who are currently studying English as a foreign language. We video recorded the entire interaction period and coded the children's emotions. Results show that children were highly engaged while interacting with BookBuddy and enjoyed speaking English with the chatbot in the system. This opens many possible future directions for designing and creating language learning chatbots for children at scale. To our knowledge, BookBuddy is the first scalable

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

L@S '19 June 24–25, 2019, Chicago, IL, USA

© 2019 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-6804-9/19/06.

DOI: <https://doi.org/10.1145/3330430.3333643>

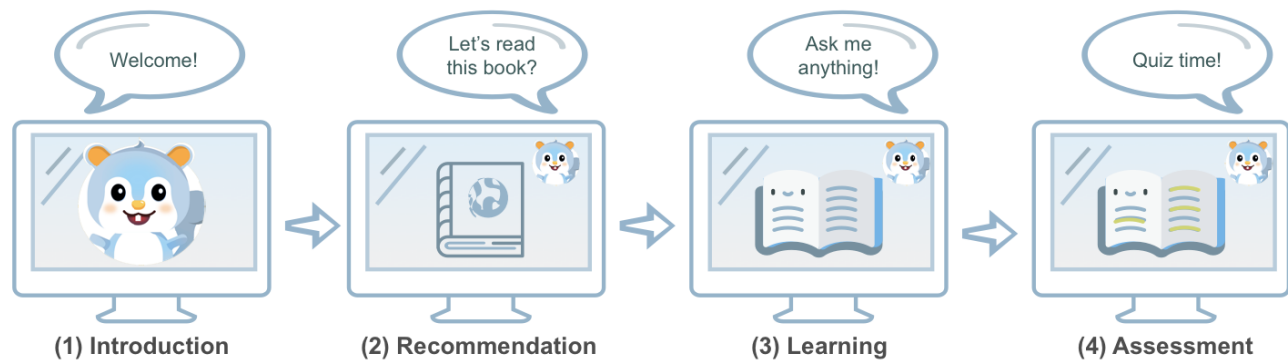


Figure 2: The interaction flow of the BookBuddy system. A child user will (1) be on-boarded by going through an introduction interface; (2) find a suitable book to read in the recommendation stage; (3) study the book under the guidance of the chatbot in the learning stage; and (4) deepen the understanding by answering quiz questions asked by the chatbot in the assessment stage.

human-centered AI system that facilitates interaction with an intelligent agent to develop children’s foreign language skills through reading and speaking.

RELATED WORK

Effective language learning tutoring systems have been developed for improving both first language and foreign language skills.

Project LISTEN’s virtual reading tutor [6] supports primary school students in improving reading skills in their first language, English, by using speech recognition to monitor children as they read a text out loud.

Systems such as Robo-Sensei [7] have implemented NLP algorithms to provide morphosyntactic feedback to adult learners and improve foreign language grammar skills. Physical robots have also been shown to be effective for children’s foreign vocabulary learning [13], and story understanding and creation [5].

However, to the best of our knowledge, there have been no reported scalable virtual tutors that are built on online learning materials and aim to provide conversational practice for children.

SYSTEM

In this section, we first overview the interface design and the interaction flow of the BookBuddy system, and then we describe the architecture underlying the chatbot in BookBuddy and the educational content curated for the system.

BookBuddy Interface

BookBuddy is a web application that can be accessed with a computer. The interface has two windows side by side as shown in Figure 1. The left window is a **reading window** that delivers stage-dependent content including the chatbot’s avatar, the cover of the recommended book, and the book page that the child is currently studying. The right window is a **chat window** that allows the child to have both spoken and typed conversation with the chatbot. The child can browse the conversation history by scrolling the chat window. A

synthesized voice (the child’s voice from Acapela Group [4]) accompanies all text messages sent by the chatbot to the child.

Interaction Flow

The BookBuddy interaction flow has four stages as demonstrated in Figure 2: **introduction**, **recommendation**, **learning**, and **assessment**. The reading window and chat window collaborate throughout these stages to deliver a consistent and engaging learning experience.

The introduction stage familiarizes the child with the interface and introduces the child’s learning companion, a chipmunk avatar named *Tangerine*¹. In the recommendation stage, the system asks for basic information such as the child’s name, gender, and interests. Then the underlying recommendation algorithm (details presented later) finds the most suitable book in our book database. In the learning stage, the chatbot helps the child to practice reading sentences from the book, and explains advanced vocabulary words found in the book. The child can ask the chatbot questions about the book, which are answered through the question-answering model described later in this paper. In the assessment stage, the chatbot asks the child 2-4 reading comprehension questions to help the child review the content learned and deepen understanding. A phrase comparison algorithm similar to [12] is used to provide adaptive assessment feedback.

Conversational Chatbot

The diagram in Figure 3 demonstrates the architecture of the chatbot system. There are three underlying sub-chatbots corresponding to stages described above: **Recommendation Bot** for the recommendation stage, **Book Bot** for the learning stage, and **Assessment Bot** for the assessment stage. The appropriate sub-chatbot is selected by the system to interact with the child according to stage.

Recommendation Bot

Recommendation Bot was designed to greet the child, gather user information, and recommend a level-appropriate book based on this information. To find a book that aligns with the child’s interests, we first convert the child’s interests (such as

¹Tangerines are a popular symbol of good luck in Chinese culture.

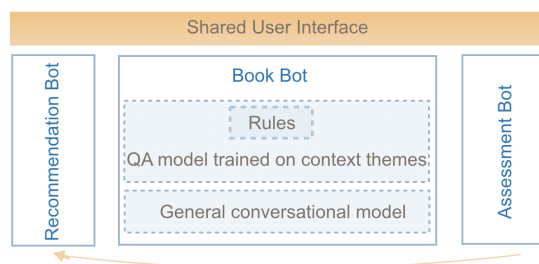


Figure 3: The chatbot system architecture in BookBuddy.

animals and gardens), splitting by each word, to word vectors using GloVe [8]. We then extract topics from each book (such as raccoons and forests) in our database and also convert them to word vectors using GloVe. Finally, we compute the cosine similarity between the child’s interests and each book’s topic labels and select the book that has the highest cosine similarity.

Book Bot

Book Bot was designed to answer any questions that may be raised by the child during the learning stage. Book Bot is an ensemble chatbot system that combines two model architectures based on recent advancements in deep learning neural networks.

For book-related conversation, a generative question-answering model is used to answer context-specific questions. This model was trained on the Stanford Question Answering Dataset (SQuAD) [9] and can answer questions about the characters and plots of the book. Additionally, a rule-based algorithm can identify and answer vocabulary questions (via a dictionary API) and basic arithmetic questions such as “What does sneaky mean?” or “What is 435 minus 381?” Finally, we added a few other rule-based touches like being able to ask the bot to tell a joke or tell the current time.

For general conversation, a sequence-to-sequence machine translation model was trained on movie scripts and other dialog-based text to learn how to respond to popular English conversation expressions. This model was designed to respond to open-ended questions or comments like “How are you?” and “I like reading this!”

A topic-modeling-based question classifier is used to decide which of the two models (the book-related question-answering model or the general conversational model) is adopted to answer the child’s question.

Assessment Bot

Assessment Bot was designed to quiz the child on book content and to provide personalized assessment feedback. The chatbot asks a random short-answer question from the question bank, and the child is prompted to speak or enter a short phrase. We adopted the algorithm described in [12] to compare the similarity between the child’s response and the correct answer. Depending on the evaluation result determined by the algorithm, the chatbot can provide different feedback: either a congratulatory message such as “Good job! You’re correct!” or a cheering-up message such as “You’re close. Try again!”.

Educational Content

There are 20 books currently implemented in the BookBuddy system, all Raz-Kids [10] level A books. These books have about 10 lines per book and 2 to 6 words per line (mean = 3.5). The assessment questions and answers associated with each book were also provided by Raz-Kids. Note that we did not perform any fine tuning of the model after it was trained. This indicates that our chatbot can be easily scaled up to function with a much larger quantity of language learning books.

CASE STUDY

We conducted a pilot study to assess users’ reactions and affective states while using BookBuddy, following work into how emotions affect users of educational technology [1, 2].

Procedure

We recorded five 6-year-old Chinese children, who all spent fewer than seven months abroad, using BookBuddy to read one book in our database. We used screen-capturing technology and positioned webcams to record the students’ faces while using BookBuddy. Following the protocol in the study of the educational software AutoTutor [1], we coded users’ emotions in each 20-second segment in the videos. The coding was done by two independent coders, both native speakers of Chinese born and raised in China, in order to avoid cultural differences in displays of emotions. Each 20-second segment was coded as one of seven possible emotions: *engagement*, *boredom*, *confusion*, *curiosity*, *happiness*, *frustration*, or *neutral*. These labels were taken from D’Mello’s [2] meta-analysis of emotions frequently detected in users of educational technology.

Results and Discussion

Our two coders achieved an interrater reliability (Cohen’s kappa) of $\kappa = .54$. Table 1 shows the proportion of time spent by users in each affective state, averaged across both coders. We can see that engagement was the most frequent emotion, followed by happiness and confusion. Users were engaged more than half of the time while using BookBuddy. In future work we plan to explore in more depth which parts of BookBuddy lead to each affective state.

Users were also interviewed after interacting with BookBuddy, and asked to rate each stage (recommendation, learning, assessment) on a 5-point scale for how enjoyable and how useful it was, with 5 representing the highest levels. Figure 4 (left) shows that users found all three stages enjoyable and useful (all with ratings higher than 4). Users were also asked to rank their preferences for speaking English with parents, teachers, classmates, the BookBuddy chatbot, or themselves. Figure 4 (right) suggests that users most preferred speaking English with our chatbot. They commented that this was because the chatbot was patient, friendly, and non-judgmental.

FUTURE WORK

There are many interesting future directions to explore based on our preliminary results. First, it is important to measure both engagement level and learning benefits for a tutoring system. Although we found children were strongly engaged, we did not evaluate their language skill change through using the system. Second, the sample size was small in our study.

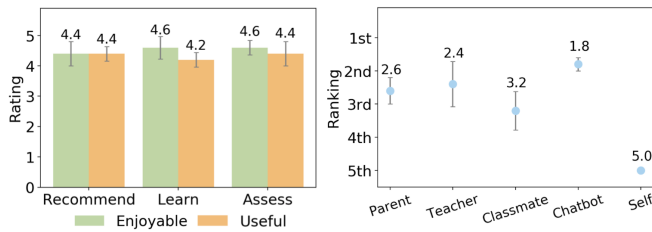


Figure 4: (Left) Users’ ratings of each stage on a 1 to 5 scale. Higher is better. Error bars represent ± 1 standard error. (Right) Users’ rankings of whom they prefer to speak English with. 1st is the most preferred. Error bars represent ± 1 standard error.

Affective state	Proportion of time					
	U1	U2	U3	U4	U5	Avg.
<i>engagement</i>	.35	.74	.50	.71	.58	.57
<i>boredom</i>	.01	.00	.09	.12	.00	.04
<i>confusion</i>	.24	.15	.07	.05	.08	.12
<i>curiosity</i>	.08	.00	.06	.00	.09	.05
<i>happiness</i>	.23	.09	.27	.10	.19	.18
<i>frustration</i>	.00	.00	.01	.00	.00	.00
<i>neutral</i>	.08	.03	.00	.03	.07	.04

Table 1: Proportion of time spent by five users in each affective state. U1–U5 indicate User 1–User 5. The largest proportion for each user is shown in bold.

We plan to run a larger scale study with more users in the future. Last, children only interacted with the system for 14.4 minutes ($SD = .10$) on average. This can be attributed to the short attention span of 6-year-olds. However, as language learning is a long-term process, we intend to run longitudinal studies of children’s usage patterns and language skill changes over a longer period of time.

CONCLUSION

In this work, we present BookBuddy, a scalable foreign language tutoring system that can automatically construct interactive lessons for children based on reading materials. Pilot study results indicate that children were highly engaged while interacting with the system and preferred speaking English with our chatbot over human partners. Our work is a first step toward creating conversation-based intelligent tutoring systems at scale for improving children’s foreign language speaking skills.

REFERENCES

1. Ryan S.J.d. Baker, Sidney K. D’Mello, Ma. Mercedes T. Rodrigo, and Arthur C. Graesser. 2010. Better to be frustrated than bored: The incidence, persistence, and impact of learners’ cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human Computer Studies* 68, 4 (2010), 223–241.
2. Sidney D’Mello. 2013. A selective meta-analysis on the relative incidence of discrete affective states during learning with technology. *Journal of Educational Psychology* 105, 4 (2013), 1082–1099.
3. Goren Gordon, Samuel Spaulding, Jacqueline Kory Westlund, Jin Joo Lee, Luke Plummer, Marayna Martinez, Madhurima Das, and Cynthia Breazeal. 2016. Affective Personalization of a Social Robot Tutor for Children’s Second Language Skills. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence (AAAI’16)*. AAAI Press, 3951–3957.
4. Acapela Group. 2019. Your Voice Matters. (2019). <https://www.acapela-group.com/>
5. Eun-ja Hyun, So-yeon Kim, Siekyung Jang, and Sungju Park. 2008. Comparative study of effects of language instruction program using intelligence robot and multimedia on linguistic ability of young children. In *RO-MAN 2008-The 17th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 187–192.
6. Jack Mostow, Jessica Nelson-Taylor, and Joseph E. Beck. 2013. Computer-guided oral reading versus independent practice: Comparison of sustained silent reading to an automated reading tutor that listens. *Journal of Educational Computing Research* 49, 2 (2013), 249–276.
7. Noriko Nagata. 2009. Robo-Sensei’s NLP-based error detection and feedback generation. *CALICO Journal* 26, 3 (2009), 562–579.
8. Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global Vectors for Word Representation. In *Empirical Methods in Natural Language Processing (EMNLP)*. 1532–1543.
9. Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. 2016. SQuAD: 100,000+ Questions for Machine Comprehension of Text. *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing* (2016).
10. Raz-Kids. 2019. Leveled Books. (2019). <https://www.raz-kids.com/main/RazQuizRoom>
11. Rachel R. Romeo, Julia A. Leonard, Sydney T. Robinson, Martin R. West, Allyson P. Mackey, Meredith L. Rowe, and John D. E. Gabrieli. 2018. Beyond the 30-Million-Word Gap: Children’s Conversational Exposure Is Associated With Language-Related Brain Function. *Psychological Science* 29, 5 (2018), 700–710.
12. Sherry Ruan, Liwei Jiang, Justin Xu, Bryce Joe-Kun Tham, Zhengneng Qiu, Yeshuang Zhu, Elizabeth L Murnane, Emma Brunskill, and James A Landay. 2019. QuizBot: A Dialogue-based Adaptive Learning System for Factual Knowledge. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI ’19)*. ACM, New York, NY, USA.
13. Paul Vogt, Rianne van den Berghe, Mirjam de Haas, Laura Hoffmann, Junko Kanero, Ezgi Mamus, Jean-Marc Montanier, Ora Oudgenoeg-Paz, Daniel Hernandez Garcia, Fotios Papadopoulos, and others. 2019. Second Language Tutoring using Social Robots. A Large-Scale Study. In *IEEE/ACM Int. Conf. on Human-Robot Interaction (HRI 2019)*.