Game-based spoken dialog language learning applications for young students

Keelan Evanini[†], Veronika Timpe-Laughlin[†], Eugene Tsuprun[†], Ian Blood[†], Jeremy Lee[†], James Bruno[†], Vikram Ramanarayanan[‡], Patrick Lange[‡], David Suendermann-Oeft[‡]

Educational Testing Service R&D
[†]660 Rosedale Rd., Princeton, NJ, USA
[‡]90 New Montgomery Street, Suite 1500, San Francisco, CA, USA

`kevanini@ets.org`

Abstract

This demo presents four different spoken dialog applications that were developed to provide young learners of English an opportunity to practice speaking and to receive feedback on particular aspects of their speaking proficiency. The speaking tasks were designed as game-based interactions in order to engage young students, and they provide feedback about grammar (*yes/no* question formation and simple past tense verb formation) and vocabulary. A pilot study with primary school students in Germany demonstrated the usefulness of these applications.
Index Terms: SDS applications, language learning, grammar feedback

1. Introduction

Due to the increasing use of English as a global *lingua franca* in academia and industry, it is now common in many countries for students to start learning English in primary school. However, resource limitations may lead to a lack of opportunities for young learners to practice speaking with and receive feedback from teachers. Automated spoken dialog applications have the potential to fill this gap and enable young students to practice speaking when a teacher is not able to provide one-on-one instruction. With this goal in mind, we developed four interactive, game-based spoken dialog applications targeted towards young learners of English. This paper summarizes these tasks and describes some lessons that we learned while deploying them in a pilot study.

2. Task Descriptions

The language learning applications were all developed using the open-source, cloud-based HALEF spoken dialog system framework [1]. The applications are accessed via a web browser and streaming audio is processed in real time on a server using voice activity detection to determine the end of a student's response. The speaking tasks were all designed to be as interactive, engaging, and gamified as possible in order to appeal to young learners of English.

2.1. Guessing Game

This task was designed to enable students to practice forming *yes/no* questions in a gamified environment. The student sees an image containing eight animated characters on the computer screen as shown in Figure 1 and is then presented with the following prompt to start the conversation:

> Let's play a game. I am one of these people. Can you guess who I am? Look at the pictures and ask *yes/no* questions to find out which person I

am. For example, you can ask "Do you have red hair?" or "Are you wearing a green t-shirt?" Okay lets get started.
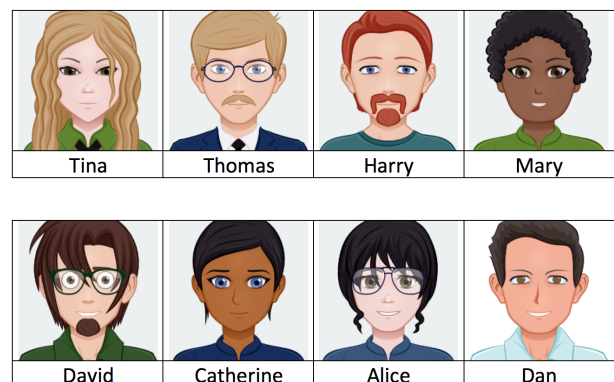

Figure 1: *Image of eight animated characters presented to language learners in the Guessing Game activity*

The system processes each *yes/no* question provided by the student to determine whether the answer to the question is true or false based on the character that had been selected randomly by the system at the beginning of the conversation. The system then provides an appropriate answer to the learner's question along with feedback about appropriate *yes/no* question formation in case the student's question was formed incorrectly, and the conversation continues until the student correctly guesses the name of the character. Further details about the Guessing Game task are presented in [2].

2.2. I Spy

This task provides students the opportunity to practice producing vocabulary words from a particular semantic domain while playing the children's game I Spy. Versions of the task were developed for two different semantic domains that are traditionally emphasized for young learners of English: school supplies and fruits & vegetables. The student sees an image containing a number of items from the semantic domain (Figure 2 presents the image for the school supplies version) and plays an interactive game of I Spy in which dialog system gives clues about a particular item in the image and the student tries to name the item targeted by the system. For example, the system could present a prompt such as "I spy with my little eye something that is brown and starts with the letter R." If the student responds with the target vocabulary item (*ruler*), then the system offers praise and moves on to another item in the picture. If the

student responds with an incorrect vocabulary item, the system provides another hint; for example, if the student incorrectly guessed *book* instead of *ruler*, the system would say "That's brown, but it doesn't start with the letter R. I spy something brown that starts with the letter R and is very long."

Figure 2: *Image of school supplies presented to language learners in the I Spy activity*

2.3. Story Telling

This task was designed to provide students with an opportunity to practice English past tense verb formation in the context of a guided story telling activity. The student participates in a guided conversation with an avatar. The initial system prompt is as follows:

> Hi, I'm Carla. I'm going to help you with your English. Something strange happened to Robert yesterday. Look at the first picture. What did Robert do first?

The student is presented with an image of an action in the story along with keywords about the main action in the image and is expected to produce a simple past sentence describing the image with the keywords. If the student produces a grammatically correct past tense sentence, the system moves on to the next scene in the story; if not, the system provides feedback and reminds the student to use a past tense verb in their response. For example, Figure 3 presents an image from the middle of the story where the targeted response is "Robert saw something amazing." If the student provides two incorrect responses (for example, the use of *seed* instead of *saw* was a relatively frequent mistake made by the students in our pilot study), the system provides the targeted answer and moves on to the next scene.

3. Lessons Learned

While developing the prototypes, we first deployed them with hundreds of users in a crowdsourcing environment (Amazon Mechanical Turk) to collect responses for training the language models, increasing the coverage of the conversational branches, and making the applications more robust. After this iterative development process, we conducted a pilot study with 27 young German EFL learners between the ages of 9 and 11. Each of the students interacted with all four of the language learning applications and completed surveys about their perceptions of the system's performance and the conversational tasks. In general, the students found the tasks to be very engaging and rated them positively, despite the presence of some system errors. We

Figure 3: *Image of scene and associated keywords in the Story Telling activity*

found that it helped to create a relaxed, low-anxiety environment for the students to tell them that they were not being tested, but, rather, that they were helping to teach the computer how to listen and respond to a human interlocutor in order to ultimately build a system that would allow children around the world to practice speaking English by using a computer. Since it was not always completely clear how to navigate the user interface to access the speaking applications, it was necessary for one of the developers to monitor each student's interactions with the spoken dialog applications in case any questions arose. In the future, we will work on developing a more user-friendly interface that can be navigated easily by young children in order to enable the applications to be used at scale with larger numbers of young English learners with minimal supervision.

4. Acknowledgements

5. References

[1] V. Ramanarayanan, D. Suendermann-Oeft, P. Lange, R. Mundkowsky, A. V. Ivanov, Z. Yu, Y. Qian, and K. Evanini, "Assembling the jigsaw: How multiple open standards are synergistically combined in the HALEF multimodal dialog system," in *Multimodal Interaction with W3C Standards: Towards Natural User Interfaces to Everything*, D. A. Dahl, Ed. Springer, 2017, pp. 295–310.

[2] V. Timpe-Laughlin, J. Lee, K. Evanini, J. Bruno, and I. Blood, "*Can you guess who I am?*: An interactive task for young learners to practice *yes/no* question formation in english," in *Proceedings of the 6th Workshop on Child Computer Interaction (WOCCI 2017)*, 2017, pp. 62–67.