applied sciences

MDPI

Article

Designing a Leveled Conversational Teachable Agent for English Language Learners

Kyung-A Lee [1,†] and Soon-Bum Lim [2,*,†]

[1] Department of IT Engineering, Sookmyung Women's University, Seoul 04310, Republic of Korea; Knoxflora88@gmail.com
[2] Department of IT Engineering, ICT Convergence Research Institute, Sookmyung Women's University, Seoul 04310, Republic of Korea
* Correspondence: sblim@sookmyung.ac.kr; Tel.:+82-2-2077-7278
† These authors contributed equally to this work.

Abstract: Given the considerable importance of the English language as a common method of global communication, supporting and engaging learning environments for studying English as a second language are essential. In this study, we propose an interactive agent that considers differences in learners' skill levels to help motivate learners and support independent study. The proposed agent is an AI chatbot that can vary the complexity of its natural language interactions as learners' skills improve and engage in unstructured conversations with users. We first collected a dataset of English conversations among children and then acquired data on learners' utterances of words and sentences using an interactive interface for skill evaluation. Based on these datasets, we generated a set of personalized conversational agents by using algorithms to evaluate speakers' pronunciation and sentences. Finally, an expert in English language learning evaluated the interactive system to ensure that it could support learners with different levels of proficiency and that the application may be expected to help motivate students to learn. The results showed that the interactive teachable agent could help motivate students who may not otherwise be interested in learning by using a more personalized approach tailored to their skill level.

Keywords: artificial intelligence chatbot; virtual agents; pedagogical agents; teachable agents; educational software; user studies; speaking skill

check for updates

Citation: Lee, K.-A.; Lim, S.-B. Designing a Leveled Conversational Teachable Agent for English Language Learners. *Appl. Sci.* **2023**, *13*, 6541. https://doi.org/10.3390/app13116541

Academic Editor: Douglas O'Shaughnessy

Received: 11 April 2023
Revised: 22 May 2023
Accepted: 25 May 2023
Published: 27 May 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Digital transformation has rapidly occurred since the onset of the COVID-19 pandemic. In the field of education, researchers are actively combining IT and pedagogical theories to provide effective learning environments through customized learning programs based on artificial intelligence (AI). In a second language environment, English language learners have little or no opportunity to use English in daily communication and feel psychological pressure to speak English. Therefore, various English learning methods are being developed to overcome non-native environmental factors and enhance learning effectiveness. The first method involved speaking practices using voice chatbots. A voice chatbot is a learning tool that utilizes artificial intelligence to improve English-speaking skills by conversing with users. The use of AI voice chatbots in English-speaking practice activities has been shown to facilitate individualized learning, increase learners' motivation to speak, provide opportunities for communication, and ultimately improve their English-speaking skills. Recent studies have shown that voice chatbots can help learners check and improve their pronunciation and intonation as they interact with them, thereby promoting learner autonomy. They can also analyze learners' conversations to determine their learning level and provide personalized learning, allowing them to learn efficiently at their own pace. Voice chatbots provide immediate feedback when learners make linguistic errors, such as pronunciation or intonation. This allows learners to recognize and improve

their errors immediately, making their learning more efficient [1–5]. These features help increase interest and engagement in learning and are likely to be the most effective in learning to speak, as they provide multiple opportunities for interaction and immediate feedback. However, most learning voice chatbots on the market can only play native speaker pronunciations, mimic them, or talk about a set topic. In addition, the answers are often predetermined. These limited learning capabilities have been shown to significantly decrease student engagement and usefulness as grade levels increase [6]. If a learning chatbot is developed with a simple response style, it will be difficult to have continuous dialogue and it cannot provide a dialogue that matches the student's level, which often leads to learner boredom [7,8]. If the content is repetitive and similar, learners may not receive adequate feedback on their progress or achievements and may lose their motivation to learn.

In conclusion, If the content is repetitive and similar, learners may not receive adequate feedback on their progress or achievements and may lose their motivation to learn. Motivation stems from the learner's personal reasons for wanting to learn English, such as the need for and interest in becoming proficient in it.

Another effective way to motivate English language learners is through the learning-by-teaching method, which involves learners taking an active role in teaching a computerized agent, called a teachable agent [9]. This approach leads to deeper learning and mastery of the content, as learners analyze and refine their understanding from different perspectives. One example is Betty's Brain, which uses a learning-by-teaching paradigm to help students learn scientific concepts by testing their own knowledge by teaching a fictional student named Betty [10–13]. However, the concept map approach used in teaching teachable agents is limited to teaching concept-related material or causal effects in specific areas, such as science curricula topics. Since English language learning is based on tense, memory, and context, rather than concepts or causal relationships, it cannot be effectively taught using concept maps.

Learning in levels (leveled learning) is a more effective way to learn English because it provides personalized instructions based on a learner's ability level. With leveled learning, each learner can learn at their own pace, using materials that are appropriate for their skill level. This personalized approach to learning has been shown to increase the learners' interest, confidence, and effectiveness [14,15]. English language learning by level provides a personalized approach that takes into account the differences in speaking abilities among individuals, which can help learners become more interested and confident in learning English. As such, there is a growing need to develop chatbots that can be applied to English learning by level. However, many voice chatbots have limited coverage of topics and vocabulary and do not provide speaking practices tailored to the user's level. In particular, if the user is an elementary school student who is not proficient in English, it may be impossible to have a dialogue using a chatbot system.

Therefore, in this study, we developed a chatbot designed to help elementary students learn English according to their skill level, to help motivate them to learn and support continuing self-directed study. To this end, we propose an interactive agent with a teaching method that reflects learners' different levels of speaking abilities.

The contributions of this study are summarized as follows.

1. We developed a teachable agent that allows learners to interact with an AI conversational partner instead of performing repetitive learning alone.
2. The agent was developed to reflect students' varying levels of proficiency in English. The proficiency levels of dialogue data were calculated through the ARI index, and the dialogue dataset level was built by dividing it into 1–4 levels.
3. The proposed teachable agent generates and communicates level-specific sentences to enable leveled dialogue through pronunciation and sentence-level evaluation of utterance data acquired from individual learners.

Our results confirm the effectiveness of the proposed system, and our findings demonstrate that our approach can help learners increase their level of proficiency in English.

2. Designing a Leveled Interactive Teachable Agent

2.1. Leveled Interactive Teachable Agent: Concept and Modeling

The leveled interactive teachable agent proposed in this study is based on a speech-based interface. Typically, speech-based English language learning systems require learners to listen to and follow the words and sentences spoken by the agent. However, in our study, we designed a model where learners teach English sentences to the agents. The agent adapts according to the learner's level, unlike existing interactive agents that initiate the teaching and learning of English sentences (see Figure 1), we set up an agent that acts as a professor to teach the correct pronunciation of sentences or words that learners may not know. This is particularly important for young or inexperienced English language learners who have limited vocabulary and struggle to form sentences and, therefore, require additional support to learn English effectively.

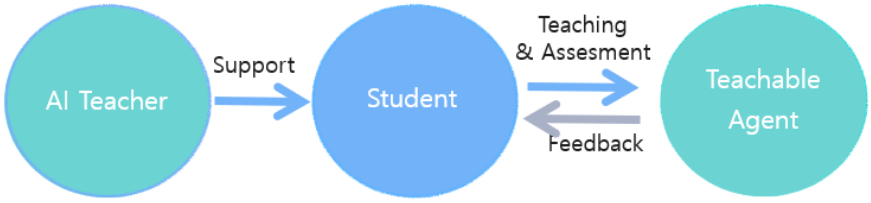


Figure 1. Interactive teachable agent.

In our proposed method, the learner first speaks some sample sentences into the microphone, and the voice data of the sentences are converted into text through the Google speech-to-text API [16] and sent to a server for processing. The server analyzes the voice data to evaluate the learner's pronunciation and intonation. To do so, it compares the input with the speech data from more than 1000 native speakers for the sample sentence, and the speech data are converted into mel-frequency cepstral coefficients (MFCCs) [17] and the similarities with the existing native speaker speech data are calculated. The learner's skill level is determined according to the results of the similarity analysis, and the teachable agent is then trained to reflect the learner's skill level as the learner improves.

In addition, learners can view the waveforms generated by a fast Fourier transform (FFT) algorithm [18] for the sentences they speak, and the waveforms of existing native speakers' speech data are also presented to help learners recognize whether their pronunciation is correct. In addition, errors in the sentence recognition function of the speech recognition API are corrected using a dictionary of children's English mispronunciations constructed for this purpose.

The learner's utterance score and level are applied to the teachable agent, which is implemented as a character. The human study then begins basic learning with leveled sentences provided by the AI teaching model. The learner provides several spoken sentences to the computational agent. After repeatedly processing these sentences, the teachable agent constructs a knowledge database for that user. When the agent has collected sufficient data, it can create new dialogues with the learner and use a leveled dialogue generation system, which allows the agent to generate leveled dialogues similar to the dialogues it has learned and the sentences it has accumulated. To perform this function, we first collected data to train a machine learning model, which included dialogues of various levels of linguistic complexity.

2.2. Collecting and Processing Conversation Dataset for Leveled Dialogue Generation Systems

To build a leveled dialogue generation system, we fine-tuned a BERT natural language generation model. The structure of the BERT language model is shown in Figure 2.

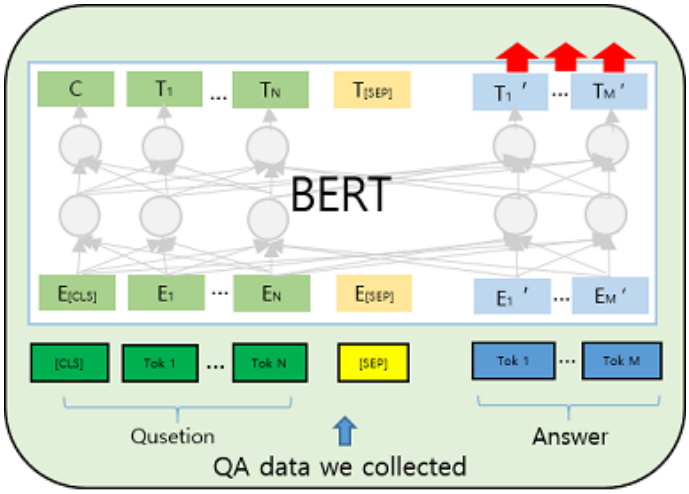


Figure 2. The structure of the Bert language model.

To train the model and generate level-specific agents, we constructed a dataset of dialogues specific to English education. This dataset included a variety of different situations and was trained using a dialogue model with many parameters to learn conversations based on specific situations. To implement a natural English conversation model for toddlers, we collected datasets of conversations in natural English spoken by toddlers in the target domain. We selected a set of movies, including children's movies, collected different dialogue elements, and extracted them into text files as movie subtitle scripts. Inappropriate dialogues were deleted when processing the data, and we ultimately obtained about 130,000 dialogues.

In addition, we utilized question-and-answer datasets, such as question rewriting in conversational context and commonsense QA. The data were divided into two groups, with QAs and QAs in the format of continuous conversations, and we were able to obtain approximately 210,000 data points by integrating the two types of data.

To fine-tune the language model, we performed data preprocessing to convert the data into a machine-readable format. We extracted features from the preprocessed data, determined the level of each conversation, and added labels to the data. The detailed composition of the collected dataset is shown in Table 1.

Table 1. Data collected.

Data	Commansense	ai2_arc	Squad	Qress	Movie	Coqa
Type	1:1 QA short	1:1 QA short	1:1 QA short	1:1 QA short	1:1 QA short	N:N QA conversation
Format	xlsx	xlsx	xlsx	xlsx	xlsx	json
Size(MB)	0.3	0.1	1.8	0.8	7.3	19.5
Volumn	6959	1307	36,355	15,125	136,803	191,607

Because the film subtitle file data were unprocessed, we preprocessed the data for training. First, we excluded data with only blanks and processed them as sentences; we then separated them into numerical data and used heuristic criteria to filter out identifiable patterns of questions and answers. In general, in most dialogues, a sentence that follows another sentence with a question mark can be considered an answer. Based on this, we identified the following types of question-and-answer patterns.

- Type 1: Question mark 1, question mark 2, normal sentence;
- Type 2: Question (Q): question mark 1 answer (A): question mark 2;
- Type 3: Question (Q): question mark 2 answer (A): general sentence.

Next, we eliminated onomatopoeia, profanity, "," ♪, QA, redundancy, and so forth, from the sentences, and collected sentences that met the requirements. The collected sentences were processed individually with the object name recognition and intent classification models, and all question-and-answer sentences were classified into a limited number of topics. To organize sentences by level, an automated readability index (ARI) score level was set for the generated sentences. The calculation formula for setting the level followed the formula in Equation (1) [19].

$$4.71\left(\frac{characters}{words}\right) + 0.5\left(\frac{words}{sentences}\right) - 21.43 \quad (1)$$

We processed the collected datasets by converting them into questions and answers and assigning automated readability index (ARI) scores to each sentence for 1:1 Q & A. We then collected data that fell within the required ARI score range (see Figure 3). For the interactive Q & A, we divided the data into dialogue situations and sub-dialogues within a given dialogue situation. The level evaluation was based on ARI's four levels, from 1 to 4, and we measured the ARI scores of the detailed conversations. Data were collected if the average score fell within the specified range.

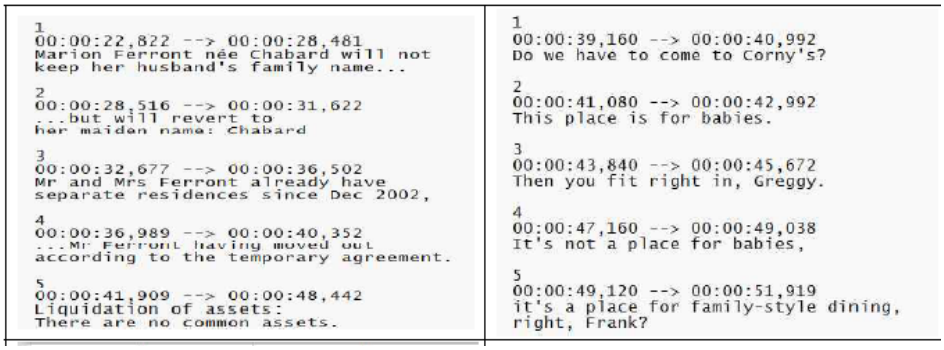


Figure 3. Processing the collected data.

Finally, the collected and refined data were used as input to the BERT model, including the level of the sentence, to generate an answer of an appropriate level to the learner's question [20].

2.3. Evaluating the Learner's Sentence Level

In the proposed approach, to evaluate a learner's level of proficiency in speaking English, we extract sentences spoken by each learner to the agent as the initial training data for each user, and then evaluate their level of complexity. The sentences spoken by the learning to train the agent were transcribed by typing each utterance as text and using the word tokenization function provided by the Natural Language Toolkit to divide the raw sentences in the dataset into separate word tokens. Then, we performed text normalization to convert individual tokens into lowercase letters for consistency during training and prediction. Because learners are not yet fluent in English and are still learning, their utterances may not be accurate. Hence, errors can occur in word recognition owing to the repetition of homonyms or mispronunciations. In the classification technique, commonly used words that did not contribute significantly to the context and semantics of the text, parts of speech that were not fluent, meaningless interjections, and homophones were removed from the extracted tokens. After tokenization, the parts of speech for each lexeme were tagged using a database (see Figure 4).

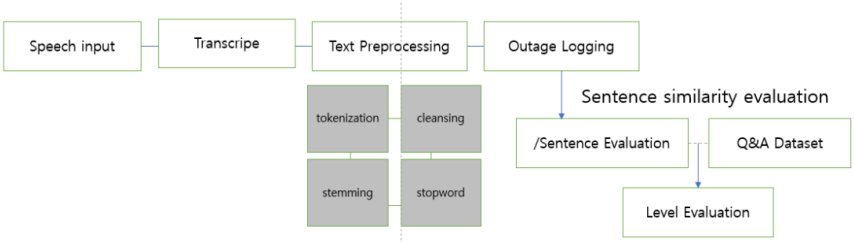


Figure 4. English level assessment process.

There are different ways to evaluate learners' English-speaking levels. In this study, we focused on pronunciation accuracy and sentence organization. Pronunciation accuracy was assessed by calculating percentage scores based on similarity and confidence values generated by automatic speech recognition, which were then averaged. The assessment was centered on the keywords in the main expression, and the percentage score was based on how well the learner included these keywords in their sentences (see Figure 5).

```
title: "What I Want to Be Is..."
uid: "xnxhd82y60UNFmEU5EZhei0xJYi2"
uscore_1: 99
uscore_2: 92
uscore_3: 94
uspeech_1: "{ "text": "I want to be an astronaut", "score": 0.987629 }"
uspeech_2: "{ "text": "I want to be an actress", "score": 0.9225468 }"
uspeech_3: "{ "text": "I want to be all that", "score": 0.9408572 }"
```

Figure 5. English level assessment process.

The learner's dialogue data were extracted in a similar manner for sentence-level evaluation of the collected data and were graded by assessing the ARI (see Figure 6).

	Question	Answer	Q_ARI_Score	A_ARI_Score
0	Where are we going?	I won't stop. Run, run as fast as you can.	1	1.2
1	Don't you see the water is getting even deeper?	You can carry me,	3.5	1
2	Oh, what have you brought us?	I think that's obvious, dear	0.4	3.7
3	Why, what's wrong with my name?	Oh, there's nothing wrong with your nan	2	5.6
4	Will you recite the alphabet for us today?	And he would name all the letters from	3.2	0
5	What about Adrian?	It's a lovely name, very dignified.	5.2	5.1
6	Woolly, Wolsten, Sten or Croft. Which one do you like be	I like Croft,	2.3	1
7	Where to next?	Those nuts smell so good!	1	0.9
8	But where were the mithai boxes?	Upstairs, in a corner, were a pile of cloth	2.8	1.9
9	Should I find that comforting?	Of course	5.6	1
10	Why do I have to wear this?	Because the queen has come of age.	1	0.9
11	Are you hurt?	Uh... No, no, I'm okay.	1	1
12	Are you sure?	Yeah. I just wasn't looking where I was g	1	2.9

Figure 6. Data labeled by the ARI score.

2.4. Leveled Agent Dialogue Generation

Typical chatbot systems incorporate a dialogue management module to semantically process the users' natural language utterances and generate responses to facilitate human–computer interactions. If a chatbot fails to "understand" the user's intention to communicate, it resolves the communication breakdown by engaging in appropriate semantic negotiation. In instances where a chatbot encounters a communication breakdown crisis due to an interrupted utterance on the part of the interlocutor, it leverages acquired

knowledge or an external database to solve the issue. The dialogue management model used in the proposed approach does not learn a scripted set of utterances in advance or generate dialogue according to a fixed logic. Rather, the agent generates dialog and provides a response based on its evaluation of the learner's level of proficiency and learned data for each level (see Figure 7).

Level 1	Level2	Level 3	Level 4
Today is monday.	Today is monday. Did you have a nice weekend?	It is Monday today. How was your weekend?	It's beautiful Monday. How did you spend your weekend?
Do you like school?	Do you like going to school?	Do you enjoy going to school? Is it fun?	I wonder if you like going to school. Can you tell me why?
Do you brush your teeth?	Do you brush your teeth after meals?	Do you brush your teeth before going to bed?	How many times a day do you brush yo ur teeth?
What do you do at school?	How is your school life?	Tell me about what you do at school, ple ase.	Can you tell me about your school life? Because it seems fun.
Are you a good singer?	Are you so good at singing?	Do you sing better than your best friend?	Do you sing well? Because I think your voice sounds so good.

Figure 7. Leveled conversation dataset.

During the dialogue generation process, the agent transcribes and inputs the learners' utterances and tags them with relevant content terms. Recognized sentences are evaluated in terms of their level by analyzing the utterances and sentences. A sentence is generated by matching it with an agent of the same level and selecting the token with the highest probability pack by performing a greedy search on the generation model trained with the previously collected leveled dialogue dataset. The generated sentences are delivered to the learner as voice dialogue.

Based on what it has learned, the agent provides interactive feedback by responding to the learner's input sentences in the QA dataset. It also compares the words in the question data to a separate database and probabilistically selects the most appropriate response sentence to answer the question.

The agent recognizes the grammatical structure, extracts the collected words and keywords from the recognized sentences, separates the learner's sentences into words, compares them with the stored data words in the database, and verbalizes the response with the highest probability. This process enables the learner to engage in a leveled conversation with the agent (see Figure 8). The conversational features were designed to be context-dependent and topic-independent to allow the system to observe real-time conversations with the agent.

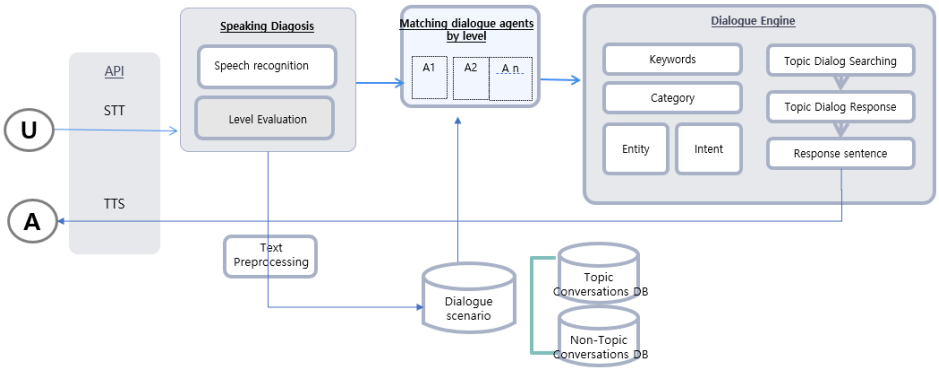


Figure 8. Leveled agent conversation generation system.

The following example illustrates how dialogue is created, according to the user's level of proficiency, by selecting an appropriate agent.

We also assessed the learner's input sentences to generate incorrect responses from the agent that would highlight the learner's weaknesses. At first, the learner was given the chance to self-correct the error and was encouraged to utter the correct sentence. If the error was not corrected, the learner was explicitly presented with a comparison between the sentence containing the error and the correct sentence to encourage the learner to express the correct sentence. Figure 9 shows an example of dialogue creation at the same level as a user and an agent.

	User & Agent Level 3
User	What are you going to do this summer?
Agent	I'm going camping with my family in June.
User	I'm going back to Korea for one month.
Agent	You must be really excited to see your friends!
User	Yes, I can hardly wait!
Agent	Well, I hope you have a good time.

Figure 9. Example of creating dialogue on the same level as a learner and an agent.

3. Implementation Results

3.1. Example of Implementation Results

In this study, we implemented a teachable voice chatbot that could practice speaking at a level. The main functions of the developed system include evaluating and diagnosing the learner's utterances and sentences, a teaching system to train the agent, and a free-talking system designed to allow free conversation with the trained agent at the learner's speaking level. The development environment of this service was developed using the Unity platform to support the execution of the app on Android devices through the Android platform. Because it should be able to recognize and output speech, we implemented a conversation exchange function by allowing the agent to answer using the Google Speech API (see Figure 10).

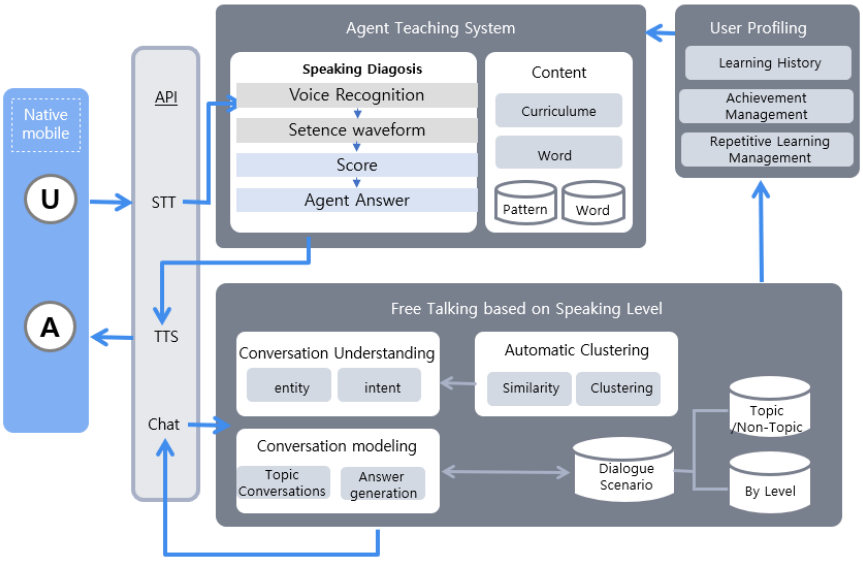
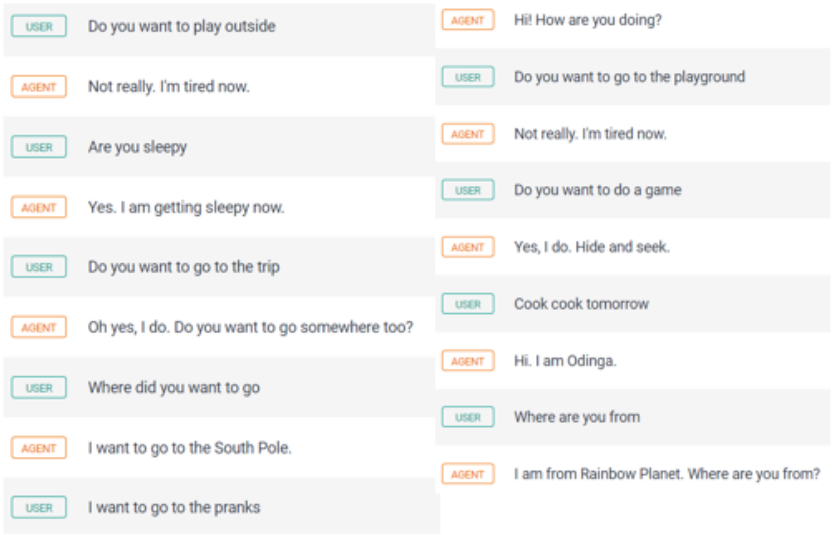


Figure 10. Teachable agent system.

Figure 11 shows an actual conversation between the learner and the agent. The implementation results for Level 1 show that the agent "understood" and responded to the learner's questions when the learner used simple sentences to say hello or spoke about their mood. The implementation results for Level 2 showed that the conversations remained at a level at which the agent could understand and engage in everyday conversations. In particular, unlike Level 1, the agent did not end with an answer to the learner's question, but instead asked the learner a related question again.



(a) Level 1 Sample Conversations



(b) Level 2 Sample Conversations

Figure 11. Teachable agent sample conversation.

3.2. Evaluation of Teachable Agent Levels

We implemented and evaluated the effectiveness of the proposed leveled interactive teachable agent in enabling leveled dialogue between learners and agents. This evaluation consisted of the following three aspects.

First, we evaluated whether the agent could have a leveled dialogue with the learner based on the learner's proficiency level. For this purpose, we evaluated whether the agent's answers to user questions at each level were at the same level, using the level conversability scale and level satisfaction scale in Table 2.

Second, we evaluated the appropriateness of the flow of the conversation, which included topics such as self-introductions, weather, soccer, weekend plans, and birthday parties. We evaluated whether the agent answered the learners' questions and continued the conversation. The appropriateness of the conversational flow metric was evaluated using a 5-point Likert scale to determine whether the flow was relevant to the topic, as listed in Table 2.

Third, we evaluated the agent's effectiveness in helping the learners improve their English skills. Based on an expert's experience, we assessed whether the agent's teaching and leveled dialogue features helped learners improve their English skills. Learning effectiveness was evaluated on a 5-point Likert scale to determine whether the agent helped learners improve their English-speaking skills, as listed in Table 2.

Table 2. Teachers' evaluation metrics.

Metrics	Questions	Score
Conversation Functionality by Level	How appropriate is the level of questions and answers in the conversation?	Level 1–4
	How well does the agent answer questions at each level?	Level 1–4
Level Satisfaction	How satisfied are you with the level of dialog functionality?	5-point Likert scale
Conversation Flow Appropriateness	Does the conversation with the agent stay on topic?	5-point Likert scale
Learning Effectiveness	Does it help the learner improve their English speaking skills?	5-point Likert scale

We conducted an evaluation, from 23 to 24 February 2023, to verify the effectiveness of the interactive teachable agent by level. The evaluation involved five English experts who, on average, had five years of teaching experience. We first explained and demonstrated the agent to the users, and then presented the experts with a set of English assessment questions for each level. The evaluation questions were organized according to four levels, with two dialogues for each level, for a total of eight questions. After viewing the agent's answers, the experts rated the level of each answer. The language proficiency level was classified from 1 to 4, as described in the list below.

- Level 1—Conversation Assessment Questions(see Table 3).

This level comprises basic English expressions and words for simple greenery, introductions, numbers, colors, dates, times, and so forth, at an understandable level.

- Level 2—Conversation Assessment Questions(see Table 4).

These questions were designed to focus on everyday conversations about family, hobbies, food, shopping, travel, and so forth.

- Level 3—Conversation Assessment Questions(see Table 5).

These questions were designed to focus on conversations about common topics in everyday life, such as the weather, environment, health, and culture.

- Level 4—Conversation Assessment Questions(see Table 6).

This level represents the ability to communicate on various topics and appropriately understand and use grammatical details, such as modal verbs, nouns, and pronouns.

- Level 2—Conversation Assessment Questions.

The questions were designed to help one understand and engage in simple, everyday conversations about family, hobbies, food, shopping, travel, etc.

- Level 3—Conversation Assessment Questions.

The questions were designed to help one understand and engage in conversations about topics common in everyday life, such as weather, environment, health, and culture.

- Level 4—Conversation Assessment Questions.

One can communicate on various topics and appropriately understand and use grammatical details, such as modal verbs, nouns, and pronouns.

Table 3. Level 1—Conversation Assessment Questions.

Question	Agent Expected Answer
Hi, what's your name?	My name is Odinga. What's your name?
My name is David. Nice to meet you!	Nice to meet you too, David!
How old are you, Odinga?	I'm 8 years old. How old are you?
I'm 9 years old.	Cool!
Hey! How are you doing?	I'm doing well, thank you. How about you?
I'm good too, thanks. What day is it today?	Today is Monday.
How's the weather today?	It's sunny.
Do you like sunny days?	Yes, I like sunny days. How about you?

Table 4. Level 2—Conversation Assessment Question Examples.

Question	Agent Expected Answer
Hey, Odinga! Do you like soccer?	Yes, I love soccer!
That's awesome.	
Do you have a favorite team?	My favorite team is Manchester United What about you?
My favorite team is Manchester United too.	
Do you have a favorite player?	My favorite player is Son Hong Min.
Today is Monday. Did you have a nice weekend?	Yes, I did.
What did you do yesterday?	I went hiking with my friends.
That sounds amazing!	
Did you take any pictures?	Yes, I took a lot of pictures with my friends.
Can you show me the pictures?	Sure, I can show you later.

Table 5. Level 3—Conversation Assessment Question Examples.

Question	Agent Expected Answer
Hey, Odinga! What do you like to do in your free time?	I love to read books and watch movies. How about you?
I like playing video games and hanging out with friends.	
What kind of books and movies do you like?	I like to read all kinds of books For movies, I like comedies
Do you have any favorite authors or movies?	Yes, my favorite author is Jane Austen.
Have you ever played soccer before?	Yes, I have played soccer a few times with my friends.
Are you good at soccer?	I'm not very good at it.
Don't worry, it takes practice to get better.	I hope so, too.
If you ever want to play together, I can give you some tips	That sounds like a good idea.
Maybe we can play a friendly game sometime.	Cool.

Table 6. Level 4—Conversation Assessment Question Examples.

Question	Agent Expected Answer
I was actually calling to invite you to my birthday party next week.	Oh, that sounds like fun! When is it?
It's next Saturday at 2 pm. I'm having it at my house.	Okay, great! I would love to come. Do you want me to bring anything?
No need to bring anything, just yourself! But if you want to bring a small gift, that's totally up to you.	Sure, I'll bring a gift for you. What kind of things do you like?
Hmm, I like books, movies, and board games. But really, anything you bring will be appreciated.	Okay, I'll find something special for you. I'm really excited to celebrate your birthday with you!
Do you like playing soccer?	Yes, I love playing soccer. How about you?
I like soccer too, but I also like playing with dolls.	That's great. Do you have a favorite doll?
Yes, her name is Daisy. She's my best friend.	Aw, that's sweet. I have a favorite toy too. It's a toy car that I love to play with.
That sounds fun! Do you want to come over to my house and play with Daisy sometime?	Sure, that would be awesome! Maybe we can also play some soccer together
That sounds like a great idea!	Cool.

3.3. Evaluation Results

We evaluated the functionality and user satisfaction of a conversational agent that played a teaching role. The agent's functionality was evaluated by analyzing its answers at the learner's conversation level. Satisfaction was assessed based on the appropriateness of the flow of conversation, the level of the agent's answers, and the effectiveness of the learning agent when used by the students in a classroom setting. Specifically, we evaluated the following.

First, Figure 12 presents the experts' evaluations of the agent's responses to the learner's level of dialogue, which shows that for level 1 questions, the agent's answer was measured as level 1 (SD 0), whereas for level 2 questions, it was recorded as level 2.2 (SD 0.4), and for level 3, it was recorded as level 3.1 (SD 0.3); for level 4, the agent's answers were evaluated as level 4.4 (SD 0.5). These responses show that the leveled agent system responded at the same level as the user's question, but as the conversation level increased, the agent's responses exhibited a somewhat higher level of complexity than those of the users.

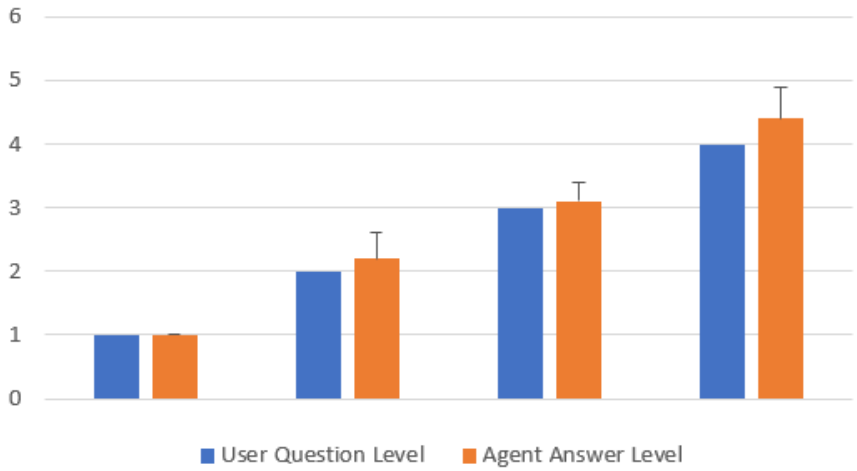


Figure 12. Differences in levels between the learner's questions and the agent's answers.

Second, we present the results of evaluating the adequacy of the dialogue flow of the agent system. As shown in Figure 13, for level 1, the system maintained the dialogue flow with a mean of 5 (SD 0.0), for level 2, it maintained the dialogue flow with a mean of 4.9 (SD 0.3), for level 3, it maintained the dialogue flow with a mean of 4.8 (SD 0.4), and for level 4, it maintained the dialogue flow with a mean of 4.3 (SD 0.5). The lowest mean score was for level 4, where the agent's answer level was higher than that of the users' questions.

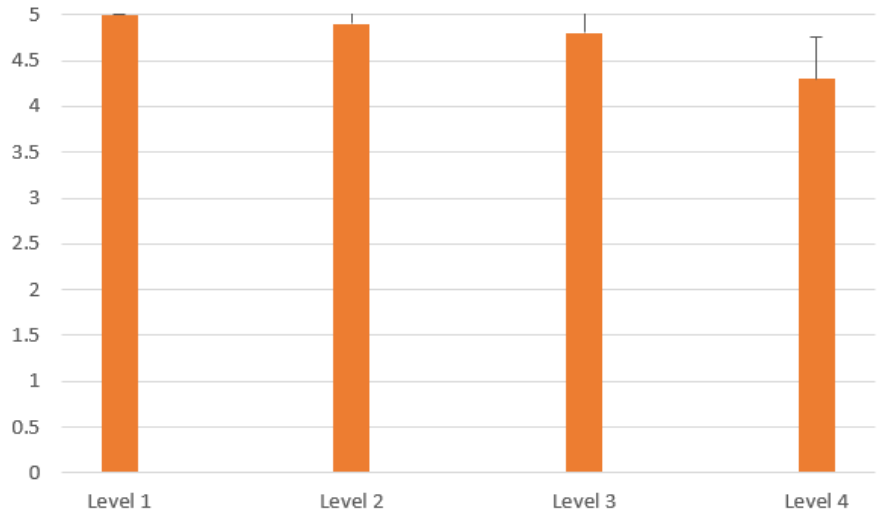


Figure 13. Adequacy of conversation flow.

Third, the results for learning effectiveness assessed whether the conversational agent system assisted learners in enhancing their English proficiency. The results of the evaluation are shown in Figure 14, with a mean of 4.9 (SD 0.3) for levels 1 and 2, a mean of 4.8 (SD 0.4) for level 3, and a mean of 4.4 (SD 0.5) for level 4. This result also shows that the lowest mean score is for level 4, which was higher than the level of the users' questions.

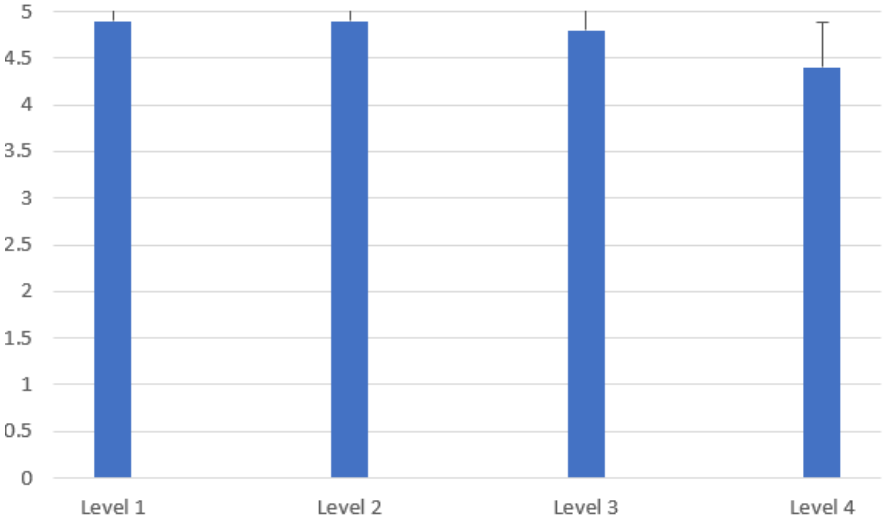


Figure 14. Learning effectiveness.

The agent performed well overall because the average score was above 4.3 at all levels. The differences between the average scores at each level were not significant, so we can say that the agent system performed consistently regardless of the level. The overall evaluation results suggest that the agent was more effective at the lower levels.

3.4. Discussion

By applying a leveled interactive teachable agent, we investigated the applicability of personalized conversational agents to English language learning and studied the implementation of voice chatbots at different levels. We designed an agent based on grammar, memory, and context rather than concepts and causality, and implemented the agent to respond to the learner's level. The leveled interactive teachable agent is a method in which the agent learns the data taught by the learner and responds to each situation based on the learned data by level.

The teachable agent voice chatbot improves the agent's speaking skills, such as asking and answering questions as the learner's skills improve. To create leveled conversation data, we collected a dataset of children's English conversations and set the conversation level using the existing ARI index model. For learner-level evaluation, we acquired learner-utterance data for words or sentences through an interactive interface and generated learner-specific agents through pronunciation and sentence evaluation algorithms for the acquired learner-utterance data. Our results confirmed that teaching and learning with a leveled agent can motivate beginner-level learners to use the system continuously, with a positive impact on their learning progress. In the future, the effectiveness of the proposed system should be evaluated in comparison with AI voice chatbots used on existing English education sites. Issues related to the difficulty of conversational dialogues of level-specific teachable agents should be improved in future studies.

4. Conclusions

In a globalized world where English is considered a crucial language, Koreans encounter difficulties in speaking it as their first language. Therefore, changes in learning methods are necessary, and self-directed learning through motivation and leveled learning is crucial. This study verifies that an interactive chatbot is capable of learning to interact with users according to different complexity levels. It demonstrates that this approach is motivating and sustainable for teaching purposes. We evaluated the usability of the interactive interface by involving a group of teachers and students. English experts confirmed that this method could potentially motivate beginner learners to learn English for the first time and that it can be used continuously. However, the challenges related to the conversational proficiency level of the teachable agent should be addressed in future studies.

Although the proposed teachable agent has some limitations in terms of creating conversations for different levels and reflecting learners' individual characteristics, we expect the system to contribute to improving English proficiency by motivating second language learners and enabling English conversations by level. The leveled conversational teachable agent can be applied to students in any L2 environment, not just in Korea, as long as language-specific data of children's mispronunciations are collected. In the future, the effectiveness of the proposed system should be evaluated and compared with AI voice chatbots used on existing English education sites.

Author Contributions: All authors contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ICAN (ICT Challenge and Advanced Network of HRD) program (IITP-2022-RS-2022-00156299) supervised by the IITP (Institute of Information Communications Technology Planning Evaluation).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Shin, D.-H.; Kim, S.-K.; Yang, K.-M. A Study on the Effect of Conversational Agent-based English Learning System on Speaking Proficiency. *J. Korea Soc. Comput. Inf.* **2018**, *23*, 39–47.
2. Kim, S.-J.; Lee, S.-J.; Yang, K.-M. A Study on the Effect of English Learning Conversation System Using Speech Recognition Technology. *Int. J. Adv. Sci. Technol.* **2019**, *28*, 813–821.
3. Liu, Q.; Matsui, T.; Kaneko, K. Integrating Dialogic Feedback into an Intelligent Tutoring System for English Speaking Practice. *Int. J. Distance Educ. Technol.* **2020**, *28*, 1–17.
4. Sung, M.C. Artificial intelligence chatbot developed by elementary school English pre-service teachers. *Lang. Study* **2020**, *56*, 97–115.
5. Hwang, Y.; Lee, H. Artificial Intelligence and its potential in English education: Deployment and creation of an AI chatbot by pre-service EFL teachers. *Multimed. Assist. Lang. Learn.* **2020**, *24*, 104–133.
6. Seong, S.; Lee, S. Analyzing learners' and teachers' perceptions of AI PengTalk for English learning and the suggestions for its use. *J. Learn. Centered Curric. Instr.* **2021**, *21*, 915–935. [CrossRef]
7. Lee, H.J. A Systematic Review of Artificial Intelligence Use in English Learning: Focus on Higher Education. *J. Humanit. Soc. Sci.* **2020**, *11*, 2027–2042. [CrossRef]
8. Kim, H.Y.; Shin, D.K.; Yang, H.J.; Jang, H.L. A study of AI chatbot as an assistant tool for school English curriculum. *Korean Assoc. Learn. Centered Curric. Instr.* **2019**, *19*, 89–110. [CrossRef]
9. Okita, S.Y.; Schwartz, D.L. Learning by Teaching Human Pupils and Teachable Agents: The Importance of Recursive Feedback. *J. Learn. Sci.* **2013**, *22*, 375–412. [CrossRef]
10. Biswas, G.; Segedy, J.R.; Bunchongchit, K. From Design to Implementation to Practice a Learning by Teaching System: Betty's Brain. *Int. J. Artif. Intell. Educ.* **2015**, *26*, 350–364. [CrossRef]
11. Biswas, G.; Leelawong, K.; Schwartz, D.; Vye, N.; The Teachable Agents Group at Vanderbilt . Learning by teaching: A new agent paradigm for educational software. *Appl. Artif. Intell. Int. J.* **2005**, *19*, 363–392. [CrossRef]
12. Leelawong, K.; Biswas, G. Designing Learning by Teaching Agents: The Betty's Brain System. *Int. J. Artif. Intell. Educ.* **2008**, *18*, 181–208.
13. Jin, S. The Effects of Self-directed Learning through an AI Speaking Application in an EFL Speaking Class on Learners' Self-directed Learning Ability and Speaking Skills. *Stud. Engl. Educ.* **2022**, *27*, 381–414. [CrossRef]
14. Krashen, S.D. *Explorations in Language Acquisition and Use*; Heinemann: Portsmouth, UK, 2022; pp. 10–109.
15. Lee, I. English Speaking Practice System for Non-Native Speakers Based on Reading Comprehension Checks. Master's Thesis, Korea University, Seoul, Republic of Korea, 2022.
16. Matarneh, R.; Maksymova, S.; Lyashenko, V.; Belova, N. Speech recognition systems: A Comparative Review. *J. Comput. Eng.* **2017**, *19*, 71–79.
17. Muda, L.; Begam, M.; Elamvazuthi, I. Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques. *arXiv* **2020**, arXiv:1003.4083.
18. Cooley, J.W.; Lewis, P.A.; Welch, P.D. The fast Fourier transform and its applications. *IEEE Trans. Educ.* **1969**, *12*, 27–34. [CrossRef]
19. Smith, E.A.; Senter, R.J. *Automated Readability Index: Aerospace Medical Research Laboratories*; University of Cincinnati: Cincinnati, OH, USA, 1967.
20. Puri, R.; Spring, R.; Patwary, M.; Shoeybi, M.; Catanzaro, B. Training question answering models from synthetic data. *arXiv* **2020**, arXiv:2002.09599.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.