Cognitive Effects of Robot-Assisted Language Learning on Oral Skills

Sungjin Lee
Pohang University of Science and Technology

Hyungjong Noh
Pohang University of Science and Technology

Kyusong Lee
Carnegie Mellon University

Gary Geunbae Lee
Pohang University of Science and Technology

Cognitive Effects of Robot-Assisted Language Learning on Oral Skills

Sungjin Lee, Hyungjong Noh, Jonghoon Lee, Kyusong Lee, Gary Geunbae Lee

Department of Computer Science and Engineering,
Pohang University of Science and Technology (POSTECH), South Korea
{junion, nohhj, jh21983, kyusonglee, gblee}@postech.ac.kr

Abstract

This study introduces the educational assistant robots that we developed for foreign language learning and explores the effectiveness of robot-assisted language learning (RALL). To achieve this purpose, a course was designed in which students have meaningful interactions with intelligent robots in an immersive environment. A total of 24 elementary students, ranging in age from 9 to 13, were enrolled in English lessons. A pre-test/post-test design was used to investigate the cognitive effects of the RALL approach on the students' oral skills. No significant difference in the listening skill was found, but the speaking skills improved, with a large effect size at the significance level of 0.01..

Index Terms: robot-assisted language learning, RALL, r-Learning, computer-assisted language learning, CALL, foreign language learning, second language acquisition

1. Introduction

There has been tremendous worldwide growth in using computer-based methods for learning different language skills and components. They assist students' linguistic development because they provide more opportunities for language learning than human-based instruction. One of the ultimate goals of computer-assisted language learning (CALL) is to provide learners with a good environment that facilitates acquiring communicative competence in the L2. Since the advent of Second Language Acquisition (SLA), a number of crucial factors have been revealed for improving students' productive conversational skills: 1) comprehensible input, 2) comprehensible output, 3) corrective feedback, and 4) motivation and attitude [1], [2], [3], [4].

In relation to oral understanding, accumulated work on the process of listening suggests that comprehension can only occur when the listener places what she or he hears in context, i.e. the knowledge of who the participants are, the setting, the topic and even the purpose. What is really retained after understanding is not the literal meaning but some mental representation mainly provided by contextual information. Hence it has become quite clear that in giving students comprehension activities out of context we set them a difficult task.

While comprehensible input is invaluable to the acquisition process, it is not sufficient for students to fully develop their L2 proficiency. The output hypothesis claims that production makes the learner move from 'semantic processing' prevalent in comprehension to more 'syntactic processing' that is necessary for improving accuracy in their interlanguage [2]. Specifically, producing output is one way of testing one's hypotheses about the L2. Learners can judge the comprehensibility and linguistic well-formedness of their interlanguage utterances against feedback obtained from their interlocutors, leading them to recognize what they do not know, or know only partially. The recognition of problems may then prompt the learners to attend to the relevant information in the input. Additionally, output processes enable learners not only to reveal their hypotheses, but also to reflect on them using language. Reflection on language may deepen the learners' awareness of forms, rules, and form-function relationships if the context of production is communicative in nature.

On the other hand, it has been argued that corrective feedback plays a beneficial role in facilitating the acquisition of certain L2 forms which may be difficult to learn through input alone, including forms that are rare, are low in perceptual salience, are semantically redundant, do not typically lead to communication breakdown, or lack a clear form-meaning relationship. If there is no concern for feedback in terms of linguistic correctness, meaning-based activities per se may accelerate language progress but in the long term lead to ''fluent-but fossilised students'' [5] .

Motivation and attitude is another crucial factor in L2 achievement. For this reason it is important to identify both the type and combination of motivation that assists in the successful acquisition of a foreign language. In order to make the language learning process a more motivating experience instructors need to put a great deal of thought into developing programs which maintain student interest and have obtainable short term goals. The use of an interesting computer-based method can help to increase the motivation level of students. computer-based learning has an advantage over human-based learning in that it seems to be a more relaxed atmosphere for language learning.

There have been few serious attempts to provide students with natural contexts that embody most of the aforementioned attributes. Therefore, we have provided an opportunity to learn English in an immersive environment in which learners experience free conversations of everyday life in real situations with intelligent robots. They can perceive the utterances of learners, especially Korean learners of English, and can provide corrective feedback to erroneous utterances. Recent development of robot-related technologies has drawn attention to the utilization of robots in real life, and increased interest in robots can give students integrative motivation to have a successful conversation with a robot. A major purpose of this investigation is to estimate the magnitude of the contributions that robot-assisted language learning (RALL) makes to achievement of oral skills in the foreign language.

The remainder of this paper is structured as follows. Section 2 briefly describes related studies. Section 3 introduces the technologies for Human Robot Interaction (HRI). Section 4 presents a detailed description of the experimental design. The results and discussion follows in Section 5. Finally, Section 6 gives our conclusion.

2. Related Work

In Japan, the educational use of robots has been studied mostly with Robovie in elementary schools focusing on English language learning. Robovie has behavior episodes with some English dialogues based on 800 rules in memory. To identify the effects of a robot in English language learning, Kanda et al. placed a robot in the first grade and sixth grade classrooms of an elementary school for two weeks, and compared the frequency of students' interaction with the English test score. Although the two-week robot-aided learning did not provide any significant effect on the students' test score in English speaking and listening, the students who showed a lot of interest at the starting phase had a significantly elevated English score. This implies that robot-aided English learning can be effective for students' motivation [6].

IROBI was recently introduced by Yujin Robotics in Korea. IROBI is both an educational and home robot, containing many features. IROBI was used in [7] to compare the effects of non-computer-based media and web-based instruction with the effects of robot-assisted learning for children. Robot-assisted learning is thought to improve children's concentration, interest, and academic achievement. It is also thought to be more user-friendly than other types of instructional media.

Studies on robot-aided education are still relatively new and most are in the early stages in a starting phase. Therefore, attempts need to be made to use robots for educational purposes and to investigate the effects of their use in this field. To the best of our knowledge, there have not been approaches combining authentic situations in the real worlds and real robots, which can provide a more realistic and active context than other approaches. The following section gives an account of the Human Robot Interaction (HRI) technologies used in the project.

3. Human Robot Interaction (HRI) Technology

KIST's Center for Intelligent Robotics (CIR) and POSTECH are cooperating in developing robots as educational assistants, called Mero and Engkey. These robots were designed with expressive faces, and have typical face recognition and speech functions allowing them to communicate. Mero is a head-only robot. The penguin-like robot Engkey is 80cm tall and weighs 90kg, and is equipped with stereo vision.

3.1. Automatic speech recognition

Speech recognition is performed by the DARE recognizer [8], a speaker independent real-time speech recognizer. Since data is costly for a fully trained acoustic model for a specific accent, we have used a small amount of transcribed Korean children's speech (17 hours) to adapt acoustic models that were originally trained on the Wall Street Journal corpus. The occurrence of pronunciation variants is detected with a speech recognizer in forced-alignment using a lexicon expanded according to the all possible substitutions between confusable phonemes.

3.2. Language understanding

Since language learners commit numerous and diverse errors, a system should be able to understand language learners' utterances in spite of these obstacles. To accomplish this purpose, rule-based systems usually anticipate error types and hand-craft a large number of error rules but this approach makes these methods sensitive to unexpected errors and diverse error combinations. Therefore we statistically infer the actual learners' intention by taking not only the utterance itself but also the dialog context into consideration as human tutors

Figure 1: *A portion of the dialog task tree for the Shopping system*

do. The intention recognizer is a hybrid model of the dialog state model and the utterance model. When the learner utters, the utterance model elicits n-best hypotheses of the learner's intention, and then they are re-ranked by the results of the dialog state model. The detailed algorithm is described in [9].

3.3. Dialog management

The dialog manager generates system responses according to the learner's intention and generates corrective feedback if needed. Dialog management is performed by RavenClaw [10]. The RavenClaw dialog management framework enforces a clear separation between the domain-specific and the domain-independent aspects of the dialog control logic. System developers focus exclusively on specifying the domain-specific aspects which are captured by the dialog task specification, essentially a hierarchical-plan for the interaction. Figure 1 depicts the top portion of the dialog task specification for the Shopping domain. In parallel, the RavenClaw dialog engine automatically ensures a rich set of conversational behaviors, such as error-handling, timing and turn-taking, etc.

When it is desirable to offer corrective feedback, Engkey provides fluent utterances which realize the learner's intention. Corrective feedback generation takes two steps: 1) Example Search: the dialog manager retrieves example expressions by querying Example Expression Database (EED) using the learner's intention as the search key. 2) Example Selection: the dialog manager selects the best example which maximizes the similarity to the learner's utterance based on lexico-semantic pattern matching. If the example expression is not equal to the learner's utterance, the dialog manager suggests the example as recast feedback and conducts a clarification request to induce learners to modify their utterance. Sometimes, students have no idea about what to say and they cannot continue the dialog. In such a case, timeout occurs and the utterance model does not generate hypotheses. Hence, the dialog manager searches EED with only the result of the dialog state model and suggests the retrieved expression so that students can use it to continue a conversation [9].

3.4. Emotional expression

The human perception of robot's emotional expressions plays a crucial role in human robot interaction. Mero and Engkey were designed with expressive faces that can represent different emotions: like, dislike, neutrality, hope, fear, joy, distress, surprise, embarrassment, pride, shame, and sadness (Figure 2). By virtue of its movable body, Engkey can also make diverse gestures by conducting a series of facial and body motions in accordance with the meaning of a verbal response: winking, yawning, cheering, sulking, etc.

Figure 2: *Facial expressions for some emotions*

4. Experimental Design

We performed a field study at a Korean elementary school. The following subsections describe the method of the study in more detail.

4.1. Setting and participants

A total of 24 elementary students were enrolled in English lessons two days a week for a total of about two hours per day and had chant and dance time on Wednesdays for eight weeks during the winter vacation. However, three students did not complete the study, resulting in a total of 21 students who completed the study. The students in this study were recruited by the teachers of the school and divided into beginner-level and intermediate-level groups, according to the pre-test scores. They ranged from second to sixth grade; in general, there are six grades in a Korean elementary school. All of them were South Korean, spoke Korean as their first language and were learners of English as a foreign language. None of the participants had stayed in an English-speaking country, such as the United States and United Kingdom, for more than three months, which may indicate that this group had limited English proficiency. Fig. 4 shows the layout of the classroom: 1) PC room where students took lessons by watching digital contents, 2) Pronunciation training room where the Mero robot performed automatic scoring of pronunciation quality for students' speech and provided feedback, 3) Fruit and vegetable store, and 4) Stationery store where the Engkey robots acted as sales clerks and the students as customers.

4.2. Material and treatment

The researcher produced training materials including a total of 68 lessons, with 17 lessons for each combination of the level, beginner and intermediate, and the theme, fruit and vegetable store and stationery store. Among other things, the course involves small talks, homework checking, purchases, exchanges, refunds, etc. (as in the dialog tree in Fig. 1). Participants in this course should become thoroughly trained in various shopping situations. With this aim in mind, when dealing with task assignment, the instructors proceeded in subtle gradations, moving from the simple to the complex. Throughout the course of the study, each student was asked to enter the four rooms in the order of PC room, Pronunciation training room, Fruit and vegetable store, and Stationery store so that students were gradually exposed to more active oral linguistic activities.

4.3. Data collection and analysis

In order to measure the cognitive effects of the RALL approach, i.e., improvement of listening and speaking skills, all students took a pre-test at the beginning of the study and a post-test at the end. For the listening skill test, 15 multiple-choice questions were used which were developed by experts in evaluation of educational programs. The items in the test were mainly selected from the content taught during the course. The test was used as the assessment tool in both the pre-test and the post-test phases of the study. The internal consistency estimates of reliability, Cronbach's alpha [11], were computed for the listening test. The value of Cronbach's alpha for the pre-test was .87 and the value for the post-test was .66, each indicating satisfactory reliability. The speaking skill test consisted of 10 1-on-1 interview items. The topics of the interviews were selected from the content taught. The evaluation rubric measured speaking proficiency on a five point scale in four categories: pronunciation, vocabulary, grammar, and communicative ability. The value of Cronbach's alpha for the pre-test was .93 and the value for the post-test was .99, each indicating satisfactory reliability. A paired t-test was performed using the mean scores and standard deviations to determine if any significant differences occurred.

Figure 3: *Students interacting with Mero and Engkey*

5. Results and Discussion

The students' achievement of the beginner group on pre- and post-test is presented in Table 4. According to the findings in the table below, there were large improvements of the speaking skills in the participants' achievement on the post-test. The score in the post-test is significantly better than that of the pre-test. The effect sizes, which were calculated following the formula proposed in [12], range over 0.82-0.90, showing large effects. The listening skill, however, showed no significant difference. Significant differences of the speaking skill were also found in the result of the intermediate group and the effect sizes are also large, whereas the listening skill showed a significantly negative effect (Table 2). The combined results of both groups showed no significant differences in the listening skill (Table 3). This This finding can be explained by a number of factors such as the unsatisfactory quality of the text-to-speech component and hindrance of robots' various sound effects. The large improvement of speaking skill in the overall results agrees with the findings of previous studies in general. Specifically, the gain in the vocabulary area indicates that the authentic context facilitated form-meaning mapping and the vocabulary acquisition process. The improved accuracy of pronunciation and grammar supports the output hypothesis and the effects of corrective feedback. Learners had feedback at any related point which made them reflect on their erroneous utterances. The increase of communicative ability shows that learners were getting accustomed to speaking English. It also can be attributed to the fact that when using robot-assisted learning the student gained confidence in a relaxed atmosphere. A lack of confidence and a feeling of discomfort were more related to students' participation in face-to-face traditional discussions, and less to participation in computer-based learning.

Table 1. *Cognitive effects on oral skills for the beginner group*

Category		N	Pre-test		Post-test		Mean difference	t	df	Effect size
			Mean	SD ⸸	Mean	SD ⸸				
Listening		10	8.60	2.84	9.50	1.90	0.90	1.03	9	0.32
Speaking	Pronunciation	10	26.90	8.39	41.80	1.81	14.90	6.10*	9	0.90
	Vocabulary	10	27.50	8.02	38.10	3.51	10.60	4.85*	9	0.85
	Grammar	10	27.20	7.45	37.30	3.37	10.10	5.74*	9	0.89
	Communicative ability	10	30.60	12.00	45.60	2.91	15.00	4.37*	9	0.82
	Total	10	112.20	35.23	162.80	10.81	50.60	5.34*	9	0.87

Table 2. *Cognitive effects on oral skills for the intermediate group*

Category		N	Pre-test		Post-test		Mean difference	t	df	Effect size
			Mean	SD ⸸	Mean	SD ⸸				
Listening		11	13.09	1.64	11.73	1.19	-1.36	-3.32*	10	0.72
Speaking	Pronunciation	11	36.91	6.43	49.09	2.43	12.18	7.72*	10	0.93
	Vocabulary	11	36.00	6.24	46.27	3.23	10.27	6.41*	10	0.90
	Grammar	11	35.64	6.28	43.64	2.84	8.00	4.94*	10	0.84
	Communicative ability	11	36.27	6.83	49.18	2.09	12.91	7.53*	10	0.92
	Total	11	144.82	25.60	188.18	10.19	43.36	6.82*	10	0.91

Table 3. *Cognitive effects on oral skills for overall students*

Category		N	Pre-test		Post-test		Mean difference	t	df	Effect size
			Mean	SD ⸸	Mean	SD ⸸				
Listening		21	10.95	3.2	10.67	1.91	-0.29	-0.55	20	0.12
Speaking	Pronunciation	21	32.14	8.86	45.62	4.28	13.48	9.48*	20	0.90
	Vocabulary	21	32.95	8.21	42.38	5.31	10.43	8.00*	20	0.87
	Grammar	21	31.62	7.96	40.62	4.43	9.00	7.59*	20	0.86
	Communicative ability	21	33.57	9.83	47.48	3.06	13.91	7.60*	20	0.86
	Total	21	123.13	34.13	176.1	16.53	46.81	8.48*	20	0.88

* $p < .01$, SD ⸸ = Standard Deviation

6. Conclusion

In this study, we described the rationale of RALL from a theoretical view of language learning and briefly introduced the HRI technologies we used to implement the educational assistant robots. To investigate the cognitive effects of robot-assisted learning, a course was designed in which students had meaningful interactions with intelligent robots in an immersive environment. A pre-test/post-test design was used to investigate the effects on the students' oral skills. The results showed that RALL can be an enjoyable and fruitful activity for students. Although the results of this study bring us a step closer to understanding robot-assisted language learning, the results are only valid for Korean elementary students. More studies are needed to consolidate/refute the findings of this study over longer periods of time using different activities with samples of learners of different ages, nationalities, and linguistic abilities.

7. Acknowledgements

We would like to thank Dr. Jaechun Ban for his expertise and efforts on evaluation of educational programs. This research (paper) was performed for the Intelligent Robotics Development Program, one of the 21st Century Frontier R&D Programs funded by the Ministry of Knowledge Economy (MKE).

8. References

[1] S.D. Krashen, "The input hypothesis: Issues and implications," Longman Publishing Group, 1985.
[2] M. Swain, "Communicative competence: Some roles of comprehensible input and comprehensible output in its development," pp. 235-253.
[3] M.H. Long, "Focus on form in task-based language teaching," Language policy and pedagogy, 2000, pp. 179–192.
[4] A.M. Masgoret and R.C. Gardner, "Attitudes, Motivation, and Second Language Learning," Language Learning, vol. 53, 2003, pp. 167–210.
[5] H. Johnson, "Defossilizing," ELT Journal, vol. 46, 1992, p. 180.
[6] T. Kanda, T. Hirano, D. Eaton, and H. Ishiguro, "Interactive robots as social partners and peer tutors for children: A field trial," Human-Computer Interaction, vol. 19, 2004, pp. 61–84.
[7] J. Han, M. Jo, V. Jones, and J.H. Jo, "Comparative Study on the Educational Use of Home Robots for Children," 2008.
[8] D. Ahn and M. Chung, "One-Pass Semi-Dynamic Network Decoding Using a Subnetwork Caching Model for Large Vocabulary Continuous Speech Recongnition," IEICE TRANSACTIONS on Information and Systems, vol. E87-D, 2004, pp. 1164-1174.
[9] S. Lee, C. Lee, J. Lee, H. Noh, and G.G. Lee, "Intention-based Corrective Feedback Generation using Context-aware Model," Proceedings of CSEDU, 2010.
[10] D. Bohus and A.I. Rudnicky, "The RavenClaw dialog management framework: Architecture and systems," Computer Speech & Language, vol. 23, Jul. 2009, pp. 332-361.
[11] L.J. Cronbach, "Coefficient alpha and the internal structure of tests," Psychometrika, vol. 16, 1951, pp. 297–334.
[12] R.L. Rosnow and R. Rosenthal, "Beginning Behavioral Research: A Conceptual Primer," Prentice Hall, 2007.