Are commercial 'personal robots' ready for language learning? Focus on second language speech

Souheila Moussalli[1] and Walcir Cardoso[2]

Abstract. Today's language classrooms are challenged with limited classroom time and lack of input, and output practice in a stress-free environment (Hsu, 2015). The use of commercial, readily available tools such as Personal Robots (PRs; e.g. Amazon's Echo, Jibo) might promote language learning by freeing up class time, allowing for a more focused personalized instruction, and giving learners more opportunities for input exposure and output practice in a stress-free environment. PRs are pedagogically valuable because of their built-in Automatic Speech Recognition (ASR) software. This feasibility study investigates the pedagogical use of PRs as tools to extend the language classroom by combining it with traditional in-class, teacher-facilitated interactive practices. We evaluated a commercial PRs' ability to comprehend L2 speech, to provide 'easy-to-understand' feedback, and to deliver accurate results for a set of pre-established questions. Using a survey and interviews, the results highlight the benefits of personalized, computer-mediated instruction as an approach to extend the reach of the classroom. As such, our study contributes to this under-studied area of Computer-Assisted Language Learning (CALL): the pedagogical use of personal robots in L2 education.

Keywords: ASR, personal assistants, personal robots, L2 learning, L2 speech.

1.   Introduction

The pedagogical use of technologies can be effective in language education as it has the potential to enhance the input (in both quality and quantity), provide authentic means of communication, and provide learners with personalized and therefore more useful feedback. Examples of these technologies include Text to Speech Synthesizers (TTSs) and ASR, the two tools utilized in the technology employed in PRs.

1. Concordia University, Montréal, Canada; so_mouss@education.concordia.ca
2. Concordia University, Montréal, Canada; walcir.cardoso@concordia.ca

TTS is a form of speech synthesis that converts text into oral output. It is available and embedded in mobile applications for computers and mobile devices such as smartphones and tablets. ASR, on the other hand, is a technology that, simply put, allows users to speak rather than type in information. As such, ASR transcribes speech and provides information based on spoken questions such as "How is the weather today in Cyprus?". TTS and ASR technologies are seen in products such as GPS and voice-controlled personal assistants such as Siri and Cortana. Their goal is to recognize intelligible speech with accuracy and efficiency, independent of the speaker's accent, background noise, and other variables.

Few research that has examined the pedagogical use of TTS and ASR have found positive outcomes in language performance and oral self-assessment (Dalby & Kewley-Port, 1999; Derwing, Munro, & Carbonaro, 2000). A study by Liakin, Cardoso, and Liakina (2015) showed that the use of a more recent commercial ASR application had a positive effect on the acquisition of the French vowel /y/. Research in mobile technology has shown overall that language students enjoy using their mobile devices to learn vocabulary (e.g. Thornton & Houser, 2005) and pronunciation (Liakin et al., 2015), and it is beneficial to visual and kinesthetic English as a foreign language learners (Hsu, 2015) because it allows them to improve their L2 pronunciation through self-regulated learning.

Interestingly, we are not aware of any study that investigates the pedagogical potential of PRs such as Amazon Echo and Jibo in L2 education. To address this gap, this study investigates L2 learners' perceptions of the personal robot Echo as a pedagogical tool by addressing the following research question: What are learners' perceptions of using a personal robot (Echo) as a pedagogical tool? In the analysis, we considered a number of variables that could affect learners' perceptions: ease of use, options for self-regulation, motivation and, more importantly, opportunities for input and output practice.

2.   Method

2.1.   Participants and design

Four female intermediate-level English learners (equivalent to B1-B2 levels in the Common European Framework of Reference for languages) interacted with Echo in order to assess its pedagogical usefulness in terms of ease of use, options for self-regulation, opportunities for input/output practice, and their motivation to

use it. Their interactions, which lasted 30 minutes on average, included asking Echo a pre-established set of questions (to find out the tool's ability to understand different phonemes and stress patterns) as well as learner-generated questions. Later, participants completed a survey about their perceptions and attitudes towards Echo, followed by a semi-structured interview.

Due to its popularity and low cost, we adopted Amazon Echo (Figure 1), a 23.5 cm cylinder speaker that provides oral answers to questions asked and connects to its associated app, the cloud-based voice service Alexa. The Alexa app is the actual voice recognizer that functions as a companion to Echo for setup, remote control, and a set of enhanced features.

Figure 1.   A personal robot: Amazon Echo

2.2.   Procedure

Participants initially filled out a background questionnaire to report on their language learning experience and personal information. They then interacted with Echo for approximately 30 minutes using a set of questions and commands (n=26, related to general information and games) and learner-generated questions. After their interaction, participants were given a 17-item survey using a 5-point Likert scale (1=strongly disagree and 5=strongly agree) to quantify their responses to a number of statements about their perceptions (e.g. 'Echo is able to understand me'). After the survey, a semi-structured interview was conducted with each participant.

3.   Analysis and results

For the survey, means and standard deviations were calculated for each item, as illustrated in abridged format in Table 1.

Table 1. Survey results: learners' perceptions

Statements	MEAN	SD
I felt more comfortable speaking English using Echo than I would in other types of classroom activities	3.75	0.96
I felt more comfortable speaking English while using Echo than I would in front of the teacher	3.75	0.96
I would like to use Echo to learn other languages	4.25	1.50
Echo is a great tool to learn a language	4.00	1.15
Echo is able to understand me	3.25	0.50
Echo's voice is easy to understand	4.25	0.96
I enjoyed using Echo in this project	4.75	0.50

As indicated in Table 1, participants enjoyed their experience using Echo as a pedagogical tool in all of the items included. To summarize the key points, they felt comfortable speaking in this type of computer-mediated communication (3.75), and consider it a great tool to learn languages (4). In addition, they found that Echo was able to comprehend their requests (3.25), was intelligible (4.25), and that their experience was overall enjoyable (4.75).

The transcriptions of the interviews were analyzed and categorized as relating to one of the four topics of the research question: ease of use, autonomy, opportunity for input and output practice, and motivation. We found that participants considered Echo user-friendly, enjoyable, helpful for language learning and fun; as originally written: "It's the first time I talk with the machine, so I found it… c'est amusant [it's fun]"; and "I was very motivated, I want to explore". Participants also found that Echo was helpful for learning pronunciation and vocabulary: "I can hear her to approve my English, so I can hear the way she talk and I learn from her the pronunciation"; "It increase the understanding of pronunciations and some vocabulary". Participants also expressed that they received good implicit feedback that encouraged repetition: "I think it's more encouraging, if you have to repeat, it's like she don't understand you, you can be better the next time". Our findings also revealed Echo as a helpful teacher-facilitated tool to extend the reach of classroom: "I found it very helpful maybe in classrooms, to help teachers maybe". Nevertheless, some participants questioned the use of machine-based interactions, and reported problems with the speech recognizer that sometimes could not understand their requests.

4.   Discussion and conclusions

This feasibility study investigated the pedagogical use of a PR (Echo) for L2 education and its potential to provide speaking opportunities outside the classroom

and consequently improve one's pronunciation skills. Our results corroborate previous research in terms of ease of use and assisting in pronunciation practice (Liakin et al., 2015), providing opportunities for input exposure and output practice (e.g. Derwing et al., 2000), and motivating students to learn on their own. As such, these results can serve as a starting point to a better understanding of this type of learning environment, one that is able to cater to different learning styles and that can extend the reach of the classroom and thus promote self-regulated learning. We hope that our study will pave the way for this under-studied area of CALL: the use of commercial, readily available personal robots as tools for L2 education.

References

Dalby, J., & Kewley-Port, D. (1999). Explicit pronunciation training using automatic speech recognition technology. *CALICO, 16*(3), 425-445.

Derwing, T. M., Munro, M. J., & Carbonaro, M. (2000). Does popular speech recognition software work with ESL speech? *TESOL Quarterly, 34*(3), 592-603. https://doi.org/10.2307/3587748

Hsu, L. (2015). An empirical examination of EFL learners' perceptual learning styles and acceptance: SR-based computer-assisted pronunciation training. *CALL, 28*, 1-20.

Liakin, D., Cardoso, W., & Liakina, N. (2015). The acquisition of French /y/ in a mobile-assisted learning environment. *CALICO, 32*(1), 1-25.

Thornton, P., & Houser, C. (2005). Using mobile phones in English education in Japan. *Journal of Computer Assisted Learning, 21,* 217-228. https://doi.org/10.1111/j.1365-2729.2005.00129.x

Research-publishing.net

CALL communities and culture – short papers from EUROCALL 2016
Edited by Salomi Papadima-Sophocleous, Linda Bradley, and Sylvie Thouësny