

# Can you understand me? Speaking robots and accented speech

Souheila Moussalli<sup>1</sup> and Walcir Cardoso<sup>2</sup>

**Abstract.** The results of our previous research on the pedagogical use of Speaking Robots (SRs) revealed positive effects on motivating students to practice their oral skills in a stress-free environment. However, our findings indicated that the SR was sometimes unable to understand students' foreign accented speech. In this paper, we report the results of a study that investigated the ability of an SR to recognize and process non-native English speech from different levels of accentedness. The analysis is based on how the SR handled the participants' speech in terms of accuracy, the number and types of communication breakdowns observed, and how the participants behaved to solve the interaction problems that they experienced with the SR. Based on the study's surveys, interviews, and observations of users' interactions with the device, the results emphasize SRs' potential to recognize different types of accented L2 speech and their use as pedagogical tools.

**Keywords:** personal assistants, speaking robots, L2 accented speech.

## 1. Introduction

In the past few years, our reliance on voice commands in our daily interactions (e.g. voice-activated searches on smartphones) has increased dramatically. Despite this trend, the recognition accuracy of accented speech remains problematic for certain accents (Moussalli & Cardoso, 2016). Moussalli and Cardoso's (2016) study investigated learners' perceptions of the pedagogical use of a speaking robot (Amazon Echo and its associated app, Alexa) as cylinder speaker that provides oral answers to any questions asked. The results showed that the SR can extend the reach of the classroom, promote self-learning, and motivate oral practice in a stress-free environment. The results also showed that Echo offered helpful

---

1. Concordia University, Montreal, Canada; souheilamoussally@hotmail.com

2. Concordia University, Montreal, Canada; walcir.cardoso@concordia.ca

**How to cite this article:** Moussalli, S., & Cardoso, W. (2017). Can you understand me? Speaking robots and accented speech. In K. Borthwick, L. Bradley & S. Thouéšny (Eds), *CALL in a climate of change: adapting to turbulent global conditions – short papers from EUROCALL 2017* (pp. 217-221). Research-publishing.net. <https://doi.org/10.14705/rpnet.2017.eurocall2017.716>

negative feedback and, more importantly, its use was perceived as an effective and efficient L2 learning tool. However, the results suggested that heavily-accented beginner learners experienced difficulties understanding and being understood by Echo, as has been observed in studies involving speech recognition (Coniam, 1999; Derwing, Munro, & Carbonaro, 2000).

Interestingly, human-to-human interactions involving speakers with accented L2 speech reveals that, in this context, effective and efficient communication is possible and is not always hampered by accented speech (Derwing & Munro, 2009). Following Derwing and Munro (2009), we define accented speech as “the way in which speech differs from the local variety of [that speech] and the impact of that difference on speakers and listeners” (p. 476). The concept of accentedness includes two sub-components: intelligibility (how much a listener can understand an utterance) and comprehensibility (the listener’s perception of the degree of difficulty in understanding the interlocutor). To address this discrepancy between human-human and human-SR interactions with L2-accented speech and to address one of the limitations of SR use reported above, this study aims to answer the following research questions:

- How much can Echo understand the L2-accented speech of English learners?
- How do Echo and raters (English as a second language teachers) compare in their ability to understand L2-accented speech?
- When Echo-human communication fails, what strategies do learners use to resolve it (types and numbers of Communication Breakdowns (CBs) and resolutions)?

## **2. Method**

### **2.1. Participants and design**

Eleven L2-accented participants (five males, six females; ages: 19-30) from seven different language backgrounds (French, Cantonese, Mandarin, Arabic, Hindi, Tulu, and Marathi-Gujarati) and proficiency levels (low-intermediate to advanced) interacted with Echo for approximately 30 minutes by asking the SR a pre-established set of requests and other personal questions (total=30). They were also asked to fill out a language background questionnaire, and two surveys

---

using a five-point Likert scale (1=strongly disagree and 5=strongly agree). The first survey consisted of 17 items to quantify their responses regarding several statements about their perceptions regarding their experience using the SR (e.g. ‘Echo is able to understand me’). The second included two items for rating Echo’s speech globally (to test comprehensibility), and one item that asked participants to transcribe what they heard after asking Echo a question (to test intelligibility). After the surveys, participants engaged in semi-structured interviews where they articulated their experience with the SR.

The judges and transcribers were two native English speakers who were asked to rate 15 randomly selected speech samples that represented different types of interactions from the participants using a five-point Likert scale on accentedness and intelligibility. They were also asked to transcribe participants’ speech to determine their intelligibility.

### 3. Analysis and results

Means and standard deviations were calculated for each survey item. As illustrated in [Table 1](#), contrary to our previous study, participants found that Echo is able to understand them relatively easily (3.55/5) and they can understand it without difficulties (4.18/5). Overall, the results also revealed that participants felt comfortable interacting with Echo (3.36/5), would consider it to learn other languages (4/5), and enjoyed using it (4.45/5).

Table 1. Mean and standard deviation: survey statements

Statements	Mean	SD
Echo can understand me.	3.55	0.934
I can understand Echo.	4.18	0.751
I felt more comfortable speaking English using Echo than I would in other types of classroom activities (e.g. role-playing, group work).	3.36	1.629
I would like to use Echo to learn other languages.	4.00	0.894
Overall, I enjoyed using Echo in this project.	4.45	0.934

CBs were assessed by two native English-speaking judges (inter-rater reliability: Accentedness: ICC=0.588; Comprehensibility: ICC=0.576; Intelligibility – via transcriptions: 84.6% – Cohen’s kappa  $\kappa$ =0.567, suggesting a moderate level of reliability). Of the 1000 interactions between Echo and the participants, the number of CBs was 177, which were mainly caused by pronunciation issues (94/177= 53.11%; indicated by \*), as summarized in [Table 2](#).

Table 2. Types of communication breakdown

Type	Example	Total/177
*Pronunciation error: segments	How many cups in a liter ([lajtər])?	40
*Hesitations	um... could you... help me with pronouncing b.i.t...s?	37
Incorrect sentence structure	From Montreal and Quebec, what is the distance between?	33
Atypical demand	Can you shout for me?!	28
Phrases not requiring a response	Wow, that's great!	11
Complex questions	I'm thinking what to have for lunch; suggest something which is Mexican cuisine.	11
*Extremely fast speech	N/A	10
*Extremely slow speech	N/A	7

The results of the CB analysis are provided in Table 3. As CBs occurred, participants behaved differently from each other in terms of resolving the interaction problems with Echo, which was indicated via a follow-up question, silence, or an incorrect response. Participants tended to repeat their questions, abandon them altogether, or re-phrase them, as the following exchange illustrates:

Participant:        Alexa, where is located Niagara Falls?

Echo:                I can't find the answer to the question I heard.

Participant:        Alexa, where is Niagara Falls located?

Echo:                Niagara Falls, New York, is a waterfall in ...

Table 3. Communication breakdowns and resolutions

Type of Behavior	Mean	Standard deviation
Repetition	7	5.514
Rephrasing	3.45	3.984
Abandonment	4.91	4.592

Finally, an analysis of the transcribed interviews indicated that participants found Echo convenient to use and it provided speaking and listening language practice: “I think it’s great tool because there are [...] so many nationalities not fluent in English, they could just sit and practice”; “it’s difficult to have a conversation with a person if your English is weak, you [...] wouldn’t feel comfortable. But if you

talk with Echo, you can always practice at your own pace”. The participants also commented that the SR accommodated and helped them understand where and why their communication failed: “when I was asking the question what she was thinking about, the first time, she didn’t understand, [...] I think I said everything, but maybe she didn’t hear something”. However, the results also revealed that participants wanted specific feedback “as I was asking my questions she didn’t get what I wanted to say, but I didn’t know what she didn’t understand”.

#### 4. Discussion and conclusions

This study investigated an SR’s ability to understand oral English as spoken by accented L2 learners and also be understood by the same speakers, without incurring human-SR communication breakdowns. Our findings indicate that, contrary to what was reported in a previous study on learners’ perceptions of the pedagogical use of SRs (Moussalli & Cardoso, 2016), Echo’s ability to understand and be understood by L2 learners and vice-versa is relatively unproblematic from both quantitative and qualitative standpoints. Future studies could look at learning gains at the segmental and prosodic levels, as well as the effects of SRs on fluency development. Despite the number of obvious limitations (small sample size, limited time-on-task), we conclude that SRs are ready to be considered for English L2 instruction due to their pedagogical potential, particularly their ability to motivate students to practice their aural listening and speaking skills (including pronunciation) in a stress-free environment.

#### References

- Coniam, D. (1999). Voice recognition software accuracy with second language speakers of English, *System*, 27(1), 49-64. [https://doi.org/10.1016/S0346-251X\(98\)00049-9](https://doi.org/10.1016/S0346-251X(98)00049-9)
- Derwing, T. M., & Munro, M. J. (2009). Putting accent in its place: rethinking obstacles to communication. *Language Teaching*, 42(4), 276-490. <https://doi.org/10.1017/S026144480800551X>
- 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>
- Moussalli, S., & Cardoso, W. (2016). Are commercial ‘personal robots’ ready for language learning? Focus on second language speech. In S. Papadima-Sophocleous, L. Bradley & S. Thouëсны (Eds), *CALL communities and culture – short papers from EUROCALL 2016* (pp. 325–329). Research-publishing.net. <https://doi.org/10.14705/rpnet.2016.eurocall2016.583>

Published by Research-publishing.net, not-for-profit association  
Contact: [info@research-publishing.net](mailto:info@research-publishing.net)

© 2017 by Editors (collective work)  
© 2017 by Authors (individual work)

**CALL in a climate of change: adapting to turbulent global conditions – short papers from EUROCALL 2017**  
Edited by Kate Borthwick, Linda Bradley, and Sylvie Thoušny

**Rights:** This volume is published under the Attribution-NonCommercial-NoDerivatives International (CC BY-NC-ND) licence; individual articles may have a different licence. Under the CC BY-NC-ND licence, the volume is freely available online (<https://doi.org/10.14705/rpnet.2017.eurocall2017.9782490057047>) for anybody to read, download, copy, and redistribute provided that the author(s), editorial team, and publisher are properly cited. Commercial use and derivative works are, however, not permitted.

**Disclaimer:** Research-publishing.net does not take any responsibility for the content of the pages written by the authors of this book. The authors have recognised that the work described was not published before, or that it was not under consideration for publication elsewhere. While the information in this book are believed to be true and accurate on the date of its going to press, neither the editorial team, nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, expressed or implied, with respect to the material contained herein. While Research-publishing.net is committed to publishing works of integrity, the words are the authors' alone.

**Trademark notice:** product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

**Copyrighted material:** every effort has been made by the editorial team to trace copyright holders and to obtain their permission for the use of copyrighted material in this book. In the event of errors or omissions, please notify the publisher of any corrections that will need to be incorporated in future editions of this book.

Typeset by Research-publishing.net

Cover design based on © Josef Brett's, Multimedia Developer, Digital Learning, <http://www.eurocall2017.uk/>, reproduced with kind permissions from the copyright holder.

Cover layout by © Raphaël Savina ([raphael@savina.net](mailto:raphael@savina.net))  
Photo "frog" on cover by © Raphaël Savina ([raphael@savina.net](mailto:raphael@savina.net))

Fonts used are licensed under a SIL Open Font License

ISBN13: 978-2-490057-04-7 (Ebook, PDF, colour)

ISBN13: 978-2-490057-05-4 (Ebook, EPUB, colour)

ISBN13: 978-2-490057-03-0 (Paperback - Print on demand, black and white)

Print on demand technology is a high-quality, innovative and ecological printing method; with which the book is never 'out of stock' or 'out of print'.

British Library Cataloguing-in-Publication Data.  
A cataloguing record for this book is available from the British Library.

**Legal deposit:** Bibliothèque Nationale de France - Dépôt légal: décembre 2017.