2 A study on technology-based speech assistants

Serpil Meri-Yilan[1]

Abstract

This research aimed to look at students' perspectives on learning language through two technology-based speech recognition programmes, ImmerseMe and ELSA (English Language Speech Assistant). Data were collected from qualitative research instruments in April 2018. Five university-level students performed activities to improve their English and other languages in ImmerseMe for 30 minutes twice in two weeks, whereas they did activities to build up their English in ELSA once. The researcher observed them, and then interviewed them asking questions about their learning via these programmes. The findings showed that students had contrasting views on the programmes drawing attention to the programmes' benefits and potential improvements. This study demonstrated that Speech Recognition Technology (SRT) improved their speaking and listening skills. It makes recommendations for students, teachers, institutions, and designers to consider the effectiveness of SRT in language learning environments. It indicates the need to design a learning environment with a well-equipped programme.

Keywords: technology-based language learning, speech recognition system, ImmerseMe, ELSA.

1. Agri Ibrahim Cecen University, Agri, Turkey; serpilmeri@gmail.com; https://orcid.org/0000-0003-1132-568X

How to cite this chapter: Meri-Yilan, S. (2019). A study on technology-based speech assistants. In A. Plutino, K. Borthwick & E. Corradini (Eds), *New educational landscapes: innovative perspectives in language learning and technology* (pp. 11-17). Research-publishing.net. https://doi.org/10.14705/rpnet.2019.36.950

1. Introduction

The advance in technology-based speech assistants has drawn attention to the use of commercial products, such as "Apple's Siri, Amazon's Alexa, Microsoft's Cortana, and Google's Assistant" (Hoy, 2018, p. 81), to complete tasks automatically (Johnson, 2013). These technology-based speech assistants help users through Automatic Speech Recognition (ASR) systems such as Speech-To-Text (STT), or Text-To-Speech (TTS) (Liakin, Cardoso, & Liakina, 2017).

Research in English as a foreign language has indicated the concerns of Non-Native Speakers (NNSs) to speak and listen to Native Speakers (NSs) of English (Shadiev, Hwang, Huang, & Liu, 2016). Although it is debatable whether ASR gives a "sufficiently correct" utterance or feedback (Rodman, 1999, p. 273), ASR helps NNSs first be understandable and have native-like speech in a long term (Bajorek, 2017). Recent studies have shown:

- NNSs' interaction with ASR and immediate feedback enhances speaking skills and positive views (Ahn & Lee, 2016);
- STT guides NNSs to apply different languages strategies (Shadiev et al., 2016);
- feedback, especially from ASR, is beneficial in improving pronunciation (Liakin et al., 2015, 2017);
- feedback provided by software is not enough for L2 pronunciation development (Bajorek, 2017);
- ELSA and Google Docs Voice Typing are a good opportunity for learners to hear their voice and correct their pronunciation (Bajorek, 2018a); and
- SRT embedded into lessons in ImmerseMe comforts speaking anxiety as NNSs practise language with NSs (Bajorek, 2018b).

Although these studies have suggested the implementation and design of learning programmes with ASR, there is still a research gap in how technology-based speech assistants support NNSs' speaking and listening skills. Considering the research gap in SRT, this study aimed to explore NNSs' perceptions of learning and developing to speak through embedded SRT programmes such as ImmerseMe and ELSA.

2. Method

2.1. Participants

This study involved five Turkish participants, three females and two males, aged between 19 and 22, studying in the preparatory class in the Department of Interpretation and Translation at Agri Ibrahim Cecen University, Turkey. Their English level was intermediate. All of them were unfamiliar with SRT systems embedded in learning programmes.

2.2. Speech recognition language learning programmes

This study applied two programmes, ImmerseMe and ELSA. ImmerseMe is a virtual reality-based language learning programme which has over 500 scenarios in nine different languages and makes a user speak in the dialogue perfectly to progress further in scenarios, which is feedback (ImmerseMe, 2018). In ImmerseMe, users travel through a 3D environment using the target language. However, ELSA is a technology-based speech assistant which focusses on and gives assessment and feedback on users' pronunciation and intonation (ELSA, 2018). When they succeed in speaking, the programme writes 'excellent'. In the contrary case, it provides feedback on the errors they make by giving suggestions on what to consider and examples of similar sounds of different words and showing their speech and the correct sound in the phonemic transcription and audio form. This study drew on the two programmes' features and their potential effects on speaking skills to explore students' perspectives of learning and improving speaking via these programmes.

2.3. Data collection and analysis

Data were gathered from observations and follow-up semi-structured interviews in April 2018. During the observations, each participant performed English activities and one of the other languages for 30 minutes in ImmerseMe twice in two weeks, whereas each of them did English activities for ten minutes in ELSA once. The researcher did not interrupt them but observed their performance. After observation, they were interviewed to validate observation data (Charters, 2003) by responding to the question of how they thought about their learning.

Data sets were analysed in NVivo, coding the transcripts of participants' performances and perceptions of their learning in each programme according to the following categories: benefits, drawbacks, similarities, and differences.

3. Results and discussion

Data from observations and interviews demonstrated that participants had positive views on SRT in ImmerseMe and ELSA. They believed that these programmes improved their speaking, as consistent with the studies by Bajorek (2018a, 2018b). However, this study compared the benefits and drawbacks of SRT provided by these programmes from the perspective of participants (see Table 1).

Table 1 shows that there is still a need to improve the programmes for the enhancement of speaking and listening skills. Along with the effect of these programmes on listening and speaking skills, participants thought that SRT in both programmes increased motivation and confidence.

This study showed that the more they used ImmerseMe, the more they felt comfortable in speaking and had fun with the activities and focussed on not only improving speaking skills but also travelling in an immersive 3D environment. However, in ELSA, they just focussed on their pronunciation and correct use of stress.

Table 1. Benefits and drawbacks of ImmerseMe and ELSA stated in this study

	ImmerseMe	ELSA
Benefits	• Pronunciation improvement • Communication and interaction with NSs in a country where language is spoken in a 3D environment • Listening and speaking practice • Activities in different languages • Immediate feedback • Learning strategies development • Repeating NSs' speech • 360 degree videos	• STT system • Immediate written feedback on their speech and individual sounds • Listening and speaking practice • Pronunciation dictionary • Words in an example sentence and the international phonetic alphabet • Assessment (NS pronunciation score, needed work, proficiency level, conversation score) • Multiple activities • Seven day free trial
Drawbacks	• Just desktop-based programme • Weakness in recognising voices (i.e. soft voice, or a change because of sickness) • No phonetic and phonemic transcription of words • No STT system • The need for more scaffolding and feedback • No dictionary • Just British accent • No free trial activities	• Just mobile-based programme • Just American accent • No videos • No feedback about the assessment scores

4. Conclusions

This study concludes that SRT provides NNSs with listening, speaking, and pronunciation development. SRT increases NNSs' motivation and confidence.

The study suggests that language learning programmes with SRT should be designed with adequate scaffolding and feedback, STT and TTS technology, free and easy use, and phonetic and phonemic transcriptions of sounds. Learning programmes should be considered with different accents and multiple activities with different languages. This study recommends NNSs to empower their pronunciation with learning programmes; teachers to bring programmes into learning environments; institutions to adapt technology-based learning environments into their classrooms; and designers to reconsider the suggested benefits and drawbacks of creating an ideal learning programme.

Acknowledgements

I would like to thank ImmerseMe for allowing me to conduct this research.

References

Ahn, T. Y., & Lee, S. M. (2016). User experience of a mobile speaking application with automatic speech recognition for EFL learning. *British Journal of Educational Technology*, *47*(4), 778-786. https://doi.org/10.1111/bjet.12354

Bajorek, J. P. (2017). L2 Pronunciation tools: the unrealized potential of prominent computer-assisted language learning software. *Issues and Trends in Educational Technology, 5*(2), 60-87.

Bajorek, J. P. (2018a). Speech technology for language learning: research & today's tools. *OLLReN*. http://ollren.org/ld.php?content_id=44618256

Bajorek, J. P. (2018b). *L2 virtual reality-based software, ImmerseMe: speaking, interactivity, and time*. Presented at Computer Assisted Lang. Instruction Consortium (CALICO), University of Illinois, Urbana-Champaign. https://www.academia.edu/36768999/Virtual_Reality_Technology_ImmerseMe_Software_Speaking_Interactivity_and_Time._Bajorek_2018_CALICO_2018

Charters, E. (2003). The use of think-aloud methods in qualitative research an introduction to think-aloud methods. *Brock Education Journal, 12*(2). https://doi.org/10.26522/brocked.v12i2.38

ELSA. (2018). *ELSA-Speak English fluently, easily, confidently*. https://www.elsaspeak.com/home

Hoy, M. B. (2018). Alexa, Siri, Cortana, and more: an introduction to voice assistants. *Medical reference services quarterly*, *37*(1), 81-88. https://doi.org/10.1080/02763869.2018.1404391

ImmerseMe. (2018). *ImmerseMe: VR-based language learning*. https://immerseme.co/#home.

Johnson, S. A. P. (2013, June 14th). Language-learning software review: Babbel and Duolingo. *The Economist.* http://www.economist.com/blogs/johnson/2013/06/language-learning-software

Liakin, D., Cardoso, W., & Liakina, N. (2015). Learning L2 pronunciation with a mobile speech recognizer: French /y/. *CALICO Journal, 32*(1), 1-25. https://doi.org/10.1558/cj.v32i1.25962

Liakin, D., Cardoso, W., & Liakina, N. (2017). The pedagogical use of mobile speech synthesis (TTS): focus on French liaison. *Computer Assisted Language Learning, 30*(3-4), 325-342. https://doi.org/10.1080/09588221.2017.1312463

Rodman, R. (1999). *Computer speech technology*. Artech House.

Shadiev, R., Hwang, W. Y., Huang, Y. M., & Liu, C. J. (2016). Investigating applications of speech-to-text recognition technology for a face-to-face seminar to assist learning of non-native English-speaking participants. *Technology, Pedagogy and Education*, *25*(1), 119-134. https://doi.org/10.1080/1475939X.2014.988744

Research-publishing.net

Published by Research-publishing.net, a not-for-profit association
Voillans, France, info@research-publishing.net

New educational landscapes: innovative perspectives in language learning and technology
Edited by Alessia Plutino, Kate Borthwick, and Erika Corradini

Publication date: 2019/07/10

Typeset by Research-publishing.net
Cover design by © Raphaël Savina (raphael@savina.net)

ISBN13: 978-2-490057-48-1 (Ebook, PDF, colour)
ISBN13: 978-2-490057-49-8 (Ebook, EPUB, colour)
ISBN13: 978-2-490057-47-4 (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, UK: British Library.
Legal deposit, France: Bibliothèque Nationale de France - Dépôt légal: juillet 2019.