A study on technology-based speech assistants

Serpil Meri-Yilan

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 focuses 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

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

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**


