Advances in Artificial Intelligence and Machine Learning; Research 4 (2) 2387-2394 Received 29-04-2024; Accepted 22-06-2024; Published 29-06-2024

On the Adequacy of Elsa Speak in Formal Education: A Survey of Teacher-Users

R. Yamamoto Ravenor yamamoto.ravenor@twmu.ac.jp
Tokyo Women's Medical University
School of Medicine
8-1 Kawada, Shinjuku, Tokyo
Japan

Corresponding Author: R. Yamamoto Ravenor

Abstract

There are hundreds of education-related studies attesting to the efficacy of "Elsa Speak" AI platform for learning English pronunciation. While the literature often reveals itself in a preoccupation with formal education, it is limited to examining Elsa's effects on learners, hardly establishing any relation between this app and organised teaching. This paper presents the results of an investigation into Elsa's impact on Japanese junior high school teachers of English who integrated it into their classroom. Although no teacher interviewed denied Elsa's learning benefits, they essentially consider its unrestricted use disruptive during class. Most teachers also stated that Elsa is neither necessary nor sufficient a repetition or correction method, and that aligning students to a single English pronunciation is not an educational goal they pursue. The study concludes that, unless pronunciation is a criterion for grading, and Elsa can support objective student evaluation, however helpful a learning tool, its adequacy in schools is yet to be demonstrated.

Keywords: Elsa Speak, Automatic speech recognition, AI in education, Mobile-assisted language learning

1. INTRODUCTION

While artificial intelligence (AI) cannot yet help humans learn to speak perspicuously, advanced natural language processing (NLP) models can efficiently analyze vast amounts of speech data, enabling the detection of a wider range of linguistic features and errors than humans [1]. A variety of automatic speech recognition (ASR) and other solutions contribute to comprehensive speech assessment and facilitate improvement, many of which are designed to help learners with English pronunciation [2].

Before the advent of AI-based English pronunciation coaches, English learners used more limited forms of technology to complement or substitute teachers, such as audio recordings, language learning software, pronunciation guides etc. These are static, "one way" speaking tools that lack adaptability and feedback. At first blush, the AI-based method seems to have a lot in common with

Citation: R. Yamamoto Ravenor. On the Adequacy of Elsa Speak in Formal Education: A Survey of Teacher-Users. Advances in Artificial Intelligence and Machine Learning. 2024;4(2):138.

learning English pronunciation from a human teacher: it is an interactive, flexible, feedback-driven process that allows for personalized improvement and dynamic adjustments. Elsa Speak, or English Language Speech Assistant (ELSA), is one such AI, designed to help English language learners with pronunciation and prosody. It is said to be in the world's top 5 English learning apps, exceeding 10 million users from over 100 countries [3]. In Elsa Corporation's words, "27 hours of studying with ELSA is equivalent to an ESL speaking course at an American university" [4]. The metric is reportedly "based on learners who use Elsa". This statement is worth a moment's consideration. First, Elsa's users are, obviously, "learners", "students", or both; and second, the comparison with "an American university" implies a benchmark with formal education.

Some educational institutions around the world integrated Elsa into their English classes. As far as known, at the time of writing, 2024, in Japan, only the public junior high schools in Kyotango city, Kyoto, and in several private junior high schools elsewhere, provide their students and English teachers with the Elsa app. Reports issued by schools and Kyotango City Education Board, as well as various studies seem to indicate that Elsa's effect on students justifies the implementation.

The question before us is: What is Elsa's impact on the teachers who adopted it in formal educational settings? This study presents, in a straight-forward manner, the findings from interviews with English language teachers after roughly a year of user experience. This approach makes the results readily accessible to a wide audience, including education stakeholders, parents and students, who may not have a background in statistics or sociological research methods, but who are nevertheless equally invested.

In terms of Elsa's adequacy in formal education, the 17[1] respondents can hardly be categorized into more than one distinct group. All teachers interviewed are Japanese and have pedagogical training and, on average, approximately 15 years of teaching experience. The median respondent reported the ideal teaching time spent on pronunciation is 30%, only 3 teachers said Elsa is sufficient for correcting student pronunciation (implying that the teacher would no longer need to do it), the others acknowledging the limitations inherent in indirect instruction for pronunciation. With the exception of one, all teachers think that they should accommodate a variety of pronunciation models, rather than a single one. 7 teachers, do not use Elsa, and of those who do, only 3 use it in class. 71% of interviewees deny the need for the app, and all of them considered it a hindrance if used freely during class. Roughly half of them use time limits for their students to use Elsa, varying from 5 to 15 minutes, stating that otherwise they would have to ignore the lesson plan.

2. LITERATURE SURVEY

This literature review focuses strictly on information relevant to the impact of Elsa on schoolteacher-users. It is outside the scope of this paper to describe the technology in detail, provide evaluations, or compare it with similar mobile-assisted language learning (MALL) tools. Elsa is an AI platform focused on improving English pronunciation through feedback and personalized learning. It utilizes speech recognition technology to analyze a user's pronunciation and provide real-time feedback on sounds and intonation. [5], provides a comprehensive list of strengths and weaknesses of relevance for the purpose of this study (TABLE 1).

[1] Based on available information, these 17 teachers represent the majority of Elsa teacher-users in Japanese formal education.

Table 1: Elsa's strengths and weaknesses [5]

STRENGTHS	*WEAKNESSES*
Suitable for mobile devices	Not suitable for large screen demonstration.
Progress tracking	No waveform.[2]
Tutorial content	No TTS.[3]
Test and activities	No pronunciation models.[4]
Pronunciation evaluation	American English only."
Listening and speaking content	

In studying Elsa's impact on teachers, the fact that it is not easily displayed for group instruction suggests that the application is not adequately designed as a teaching tool for formal education, where the teacher is likely to use large monitors. The lack of text-to-speech capability to allow learning material input indicates that Elsa is also not designed as a learning tool for organized teaching. According to [6], Elsa is the only app that provides prosody feedback. This is worth noting because [7], prioritize word stress as a key element in their prosody-based descriptors for evaluating oral production in English.

Assuming that standardized American English pronunciation, which Elsa uses, is an educational objective systematically pursued in an English language curriculum, a fair evaluation would likely consist of both pronunciation and prosody assessments. Elsa, as part of a teacher's requisite, would evaluate word stress, perhaps pitch movement, rhythm, timing, and intonation patterns too. However, the lack of a waveform may limit the depth and precision of feedback, relying solely on detection, an automated processes to *indirectly* identify errors, rather than a more direct, more comprehensive evaluation and nuanced feedback that a human teacher could provide.

Finally, the absence of graphical depiction of how sounds are produced or articulated shows that the correction method done by Elsa primarily utilizes *implicit* rather than explicit correction in its approach to pronunciation improvement, leveraging the *subconscious* (rather than conscious) processes involved in language acquisition. The approach is deemed crucial within authentic contexts and situated learning environments, aligning with the pervasive use of mobile technology in learners' lives [8], but no study was found to suggest the suitability of MALL-based subconscious learning in situations created strictly for educational purposes.

An aspect relevant for this study, not yet found in the literature is that implicit and indirect methods inherent in most, if not all, such machines may yield unequal results across learners, all other variables being equal, contingent on specific factors such as the influence of the McGurk effect, a perceptual phenomenon where individuals' interpretation of speech sounds is influenced by the simultaneous visual presentation of incongruent mouth movements, leading to a fusion of auditory and visual cues, potentially altering the perceived phonetic sounds [9]. For instance, learners whose native languages exhibit a weaker McGurk effect, as observed in Japanese [10, 11], may face greater challenges in mastering English pronunciation and prosody through indirect methods, such as Elsa's, even with the auditory and visual support it presently lacks. While a substantial body of research investigated the effects of Elsa on students, existing literature lacks studies examining the impact of Elsa on teachers.

3. SURVEY INSTRUMENT AND ADMINISTRATION

The survey instrument used for the study consists of 11 main questions and 9 additional sub-questions. Only 6 questions do not have answer choices. The first question is demographic, inquiring about the years of teaching English experience. 4 questions are ends-related questions, 9 are means-related, and the remaining are quantitative. The initial interview consisted of 29 questions and sub-questions, some of which were gradually eliminated. Several teachers did endeavor to respond to all questions[5]. Interview permission was obtained from Kyotango City Education Board management, and, with the cooperation of each school, interviews were scheduled with the teachers. It is understood that only one teacher-user was not interviewed. No other public Japanese schools are known to have adopted Elsa. Interviews were also conducted with all English language teachers at one of the few Japanese private junior high schools using Elsa, Shibaura Institute of Technology Kashiwa Junior and Senior High School. Interview questions, written both in English and Japanese, were sent well in advance to the teachers, and the responses were obtained in person. Every effort was made to explain the research aims, highlight the importance of sharing insights from such rare experience, and elicit thoroughly thought responses.

4. ANALYSIS OF THE DATA

Although many questions had answer choices to aid a smooth response, some teachers offered responses other than the available options. All teacher answers are considered valuable and are reported in the 20 question sections below, in a simple and transparent manner allowing education stakeholders to easily interpret research findings.

Q1. How many years of English teaching experience do you have?	**Q2.** Do you think pronunciation and prosody are important for English language learners?
Mean approx. 15.76	Yes 94%
Median 12 (see FIGURE 1)	No 6%

Q2.1. If yes, please estimate as a percentage the ideal time ratio that an English teacher should spend on pronunciation, compared to vocabulary and grammar in a year.		
Vocabulary	Mean 38.1	Median 40
Grammar	Mean 31.9	Median 30
Pronunciation	Mean 30.6	Median 30

Q3. Which traditional (conventional) method/s do you think are best for teaching the following			
Pronunciation		Prosody	
repetition-based methods	35.3%	conversation-based methods	23.5%
audio-based methods	5.9%	audio-based methods	29.4%
phonics and katakana methods	17.6%	role playing	5.9%
mixed methods	41.2%	mixed methods	41.2%

[5] The data is available from the author upon request.

Q4. Do you think that a teacher of English and his/her students should, as much as possible, all be aligned to a single pronunciation standard?		**Q5.** Do you personally use Elsa Speak?	
Variety	88.2%	Yes	58.8%
Single	11.7%	No	41.2%

Q5.1. If yes, do you use Elsa during, outside class hours, or both?		**Q5.2.** If yes, how often do you use Elsa?	
During	0%	Several times per week	66.6%
Outside	70%	Several times per month	11.1%
Both	30%	Once a month	22.2%

Q5.3. Do you wish you could use Elsa more, less, or the same as now?		**Q5.4.** Do you, as a teacher, need Elsa?	
More	70.6%	Yes	29.4%
Less	0%	No	70.59%
Same as now	29.4%		

Q6. Do you set a time limit for students to use Elsa during class?		**Q6.1.** If yes, how long and how often do your students use Elsa during class?	
Yes	52.9%	5 - 10 min / class	55.5%
No	47.1%	10 - 15 min / class	22.2%
		10 – 15 min, two/three days /week	22.2%

Q7. If your students would use Elsa more during class, would it hinder in any way your teaching practice?		**Q7.1.** Why?
Yes	100%	Answers centered around teaching procedure, lesson plan and allocated time for each activity.
No	0	

Q8. Did you adapt the curriculum or teaching method to accommodate Elsa?		**Q9.** Would you mind if your English pronunciation and that of your students became unsynchronized?	
Yes	64.7%	Yes	11.8%
No	35.3%	No	88.2%

Q10. Do you think automatic repetition is a more efficient method than traditional ones (repeating after teacher, CD etc.)?		**Q10.1.** If yes, do you think traditional repetition methods are still needed for students who learn pronunciation with Elsa?	
Yes	47.1%	Yes	23.5%
No	52.9%	No	76.5%

Q11. Do you think that implicit and indirect correction done by a machine is more efficient than explicit and direct correction done by teachers?		**Q11.1.** If yes, do you think it is still necessary for the teacher to correct the speech of students who study English pronunciation with Elsa?	
Yes	23.5%	Yes	25%
No	76.5%	No	75%

The tallied teachers' age is unknown, but FIGURE 1, *Q1. Distribution of teaching experience (years)* shows that no "experience" group was overrepresented. However, no correlation was found to suggest that teaching experience is a marker for any discernible pattern or consistency in their responses. The data does not indicate a clear association between the teachers' years of experience and the nature of their responses or attitude towards Elsa, during the interviews.

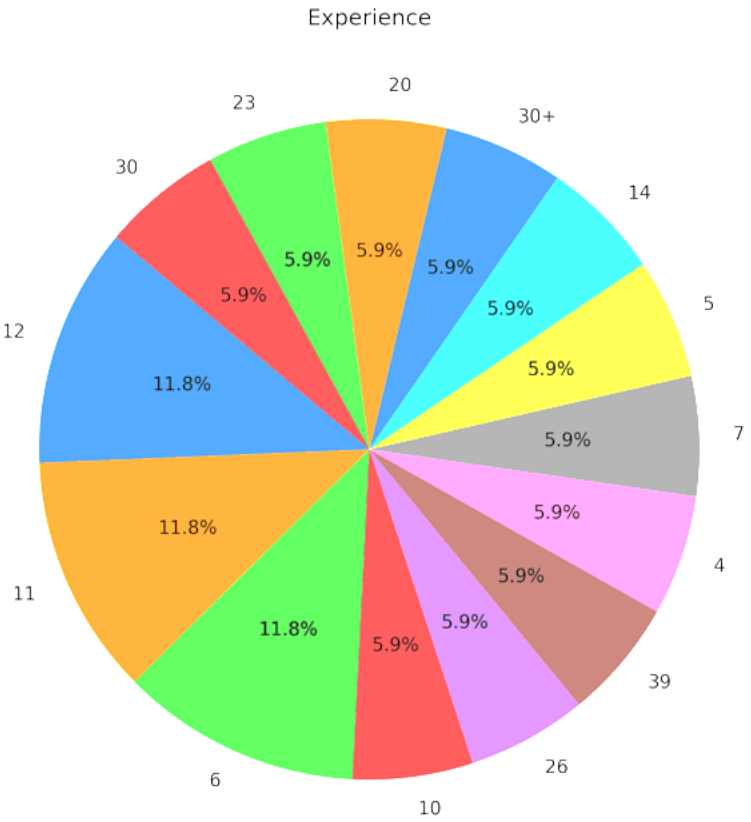

Figure 1: Q1. Distribution of teaching experience (years)

"Hiring" Elsa in Japanese public education, alongside a variety of other English pronunciation tools, and a diverse Japanese and native/native-like English teachers body, point to a real concern, fully confirmed by 94% of teachers who answered "Yes" to Q2 questioning the importance of English pronunciation. Q2.1 however finds most teachers claiming more time for teaching vocabulary and grammar, rather than pronunciation. Perhaps it is because typical schools are not concerned with the systematic evaluation of pronunciation, and that learners seek means of doing it on their own. Elsa, being such a means, may hold the potential to revolutionize formal English education as an *evaluation* tool. One limitation of this study is that, although Q3 sheds some light on teacher selection of pronunciation and prosody teaching methods, their preferred pronunciation and prosody assessment tools are unknown.

Answers to Q4 and Q9 reveal an indirect objection to Elsa's pronunciation standardization approach. 88% of teachers (in both questions) understand their duty being neither aligning students to a single pronunciation nor synchronizing students' pronunciation with teacher's own. Answers to Q5 through Q5.4. show that more than half of respondents do use Elsa for learning themselves, 70% of whom having declared at Q5.4 that they do not need it for teaching. Whether driven by curiosity, or motivated by ambition to improve their own or others' quality of English speech, teachers do not seem easily left behind. Q6 and Q6.1. provide insight into Elsa's practical use in the classroom, which divide respondents quite evenly into two categories: those who limit their students to a few minutes of Elsa use, and those who do not. The latter category is made up almost invariably of teachers in whose class Elsa is not used. Although teachers hardly acknowledge Elsa as a teaching tool, nearly 65% answered that they did make changes to accommodate Elsa (obviously as a learning tool).

There are almost as many teachers who believe that automatic repetition methods are more useful than traditional ones as there are teachers who believe the opposite, the latter accounting in Q10 and Q10.1. for almost 53% . Yet, out of those in the former group, almost 63% believe traditional methods are still needed. Finally, responses for Q11 and Q11.1 reveal that almost 77% of teachers recognize that direct and explicit speech correction done by humans is superior to the indirect and implicit one performed by machines, and only 3 teachers considered there is no need for a teacher to correct students' speech if Elsa is doing it. All respondents answered "Yes" to Q7, that Elsa hinders the educational objectives if used randomly during lessons, because pronunciation coaching can be allotted only a limited time in the lesson plan.

5. CONCLUSIONS

This study contributes to a better understanding of the impact that Elsa Speak AI platform has on Japanese junior high school teachers of English who integrated it into their classrooms. While the literature on the effects of such apps on learners is rich, knowledge voids on the perspectives of teachers, which this research aimed to bridge, can hardly be denied.

Interview results show that a consensus emerged among surveyed teachers, emphasizing that Elsa's unrestricted and/or prolonged use during class is disruptive of the teaching process. The majority of respondents expressed the view that Elsa is neither necessary nor sufficient as a repetition or correction method. The teachers also considered that the notion of aligning students to a single English pronunciation is not an educational goal.

The study concludes that the learning benefits Elsa offers may not equally surface for both students and teachers in educational settings where teaching and learning are interdependent. Despite Elsa's widely recognized potential as a learning tool, interviewed teachers do not seem to have adopted it as a teaching tool, expressing concerns about its disruptive effect and its limitations in serving educational objectives, suggesting that Elsa's integration into schools may have sped past the app's readiness.

6. ACKNOWLEDGEMENTS

This work is funded by the JSPS KAKENHI Grant-in-Aid for Early-Career Scientists, under grant #JP20K13900.

References

[1] Luo X, Qin MS, Fang Z, Qu Z. Artificial Intelligence Coaches for Sales Agents: Caveats and Solutions. J Mark. 2021;85:14-32.

[2] Farrús M. Automatic Speech Recognition in L2 Learning: A Review Based on Prisma Methodology. Languages. 2023;8:242.

[3] https://www.adb.org/sites/default/files/institutional-document/826606/adou2022bn-female-entrepreneurship-gender-equity-examples.pdf

[4] https://elsaspeak.com/en/

[5] Moxon S, Free P. Pronunciation Coach 3D. Comput Assist Lang Learn. 2023;24:205-221.

[6] Coulange S. Computer-Aided Pronunciation Training in 2022: When Pedagogy Struggles to Catch Up. In: Proceedings of the 7th international conference on English Pronun-ciation: Issues and Practices. 2023:11-22.

[7] Frost D, O'Donnell J. Evaluating the Essentials: The Place of Prosody in Oral Production. In: Vol´ın, J., Skarnitzl, R. (eds.) The Pronunciation of English by Speakers of Other Languages. Cambridge Scholars Publishing; 2018:228-259.

[8] Karakaya K, Bozkurt A. Mobile-Assisted Language Learning (Mall) Research Trends and Patterns Through Bibliometric Analysis: Empowering Language Learners Through Ubiquitous Educational Technologies. System. 2022;110:102925.

[9] McGurk H, MacDonald J. Hearing Lips and Seeing Voices. Nature. 1976;264:746-748.

[10] Hisanaga S, Sekiyama K, Igasaki T, Murayama N. Audiovisual Speech Perception in Japanese and English: Interlanguage Differences Examined by Event-Related Potentials. In: Proceedings of the auditory-visual speech processing. 2009:38-42.

[11] Sekiyama K, Burnham D. Impact of Language on Development of Auditory-Visual Speech Perception. Dev Sci. 2008;11:306-320.