Chi, Nhan Phan; von Zansen, Anna; Kautonen, Maria; Grósz, Tamás; Kurimo, Mikko

CaptainA self-study mobile app for practising speaking: task completion assessment and feedback with generative AI

Published in:
Interspeech 2024

Published: 01/09/2024

Document Version
Publisher's PDF, also known as Version of record

Interspeech 2024
1-5 September 2024, Kos, Greece

CaptainA self-study mobile app for practising speaking: task completion assessment and feedback with generative AI

Nhan Phan[*1], *Anna von Zansen*[*2], *Maria Kautonen*[*3], *Tamás Grósz*[1], *Mikko Kurimo*[1]

[1]Aalto University, Finland [2]University of Helsinki, Finland [3]University of Jyväskylä, Finland

firstname.lastname@aalto.fi, anna.vonzansen@helsinki.fi, maria.e.e.kautonen@jyu.fi

Abstract

We introduce the CaptainA mobile app, designed to meet the needs of second language (L2) learners engaged in self-study of Finnish, with potential applicability to other languages. Our app can provide automatic speaking assessment (ASA) of task completion in picture-based tasks, along with grading explanations and corrective feedback. It can also automatically generate pictures for visual tasks, providing users with unlimited practice opportunities. The mobile app is based on our framework that combines visual natural language generation (NLG), automatic speech recognition (ASR), and prompting large language model (LLM) for low-resource language. Our goal is to promote the development of next-generation speech-based computer-assisted language learning (CALL) systems capable of providing automatic scoring with feedback for learners, even when minimal speech data of L2 learners is available. While the mobile app demonstration is designed for Finnish, the app can also be tested in English.

Index Terms: low-resource language, L2 speaking, content feedback, Automatic Speech Assessment, mobile app

1. Introduction

Due to the limited time and resources dedicated to practising L2 speaking in traditional language classrooms, self-study provides L2 learners with a valuable source for learning. Mobile-assisted language learning can be seen as an effective way to enhance language learning due to the informal nature of learning, possibilities for individualised learning, portability and wide availability [1]. L2 learners often benefit from corrective feedback and prefer to receive it from teachers [2], but self-study tools with ASA cannot usually explain the grading or provide corrective feedback based on the user's mistakes (see Gu et al. [3]), thus reducing the effectiveness of self-studying.

For many languages, such as Finnish, the resources available for learning the language are scarce, making developing an effective self-study application beneficial for L2 learners. However, the lack of L2 speech data for training poses a significant challenge in creating such applications.

In this article, we demonstrate how generative AI can be used for designing picture-based tasks [4] and providing automatic feedback on task completion. In speaking assessment, pictures are used e.g. for description, comparing or narrating tasks [4]. Before generative AI, language testers and teachers have drawn the pictures themselves or selected from photo banks available. Selecting suitable pictures takes time and often raises copyright, comparability, and test-security issues.

*Equal contribution

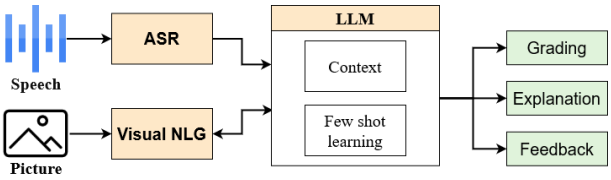

Figure 1: *Overview of our task completion grading framework*

In collaboration between speech technology and language assessment research, we designed and developed a proof-of-concept, building on an existing mobile pronunciation practise app [5] to provide L2 Finnish learners ASA including feedback and explanation on their task completion in picture-based narrative speaking tasks [4] (see Fig. 1).

2. Theoretical background

ASA tools already exist, however, they often provide corrective feedback only on pronunciation [3]. This article demonstrates how the existing CaptainA mobile app [5], which focuses on L2 Finnish pronunciation, could be extended for practising picture-based speech production. Specifically, it focuses on using generative AI to assess task completion, i.e. how well the learner described the given picture.

Conventional ASA models typically rely only on audio data, which is not suitable for picture-based ASA. Based on the proposal by Salaberria et al. that combines image captioning with language models to solve the visual question answering [6], we use a visual NLG model, capable of converting images to natural language based on a given context to enable ASA in picture-based tasks.

One of the significant challenges in providing explanation for automatic scoring and feedback is the lack of training data. LLMs have been trained to evaluate written language tasks in popular languages and also for translation purposes. In theory, those LLMs could address this issue for low-resource languages. They have demonstrated effectiveness in rating, reasoning and providing explanations in complex English texts [7], and can be leveraged for ASA tasks in Finnish.

We implemented our proposed framework (Fig. 1), utilising generative AI, specifically GPT-4, to provide users with transparent scoring and automated feedback on their spontaneous speech practice. The pipeline processes speech data and pictures by integrating the L2 ASR model, Visual NLG, and LLM with in-context learning (Table 1). We evaluated our framework on actual data for one specific task (the task on the right of Fig. 2) and achieved substantial agreement with human experts [8].

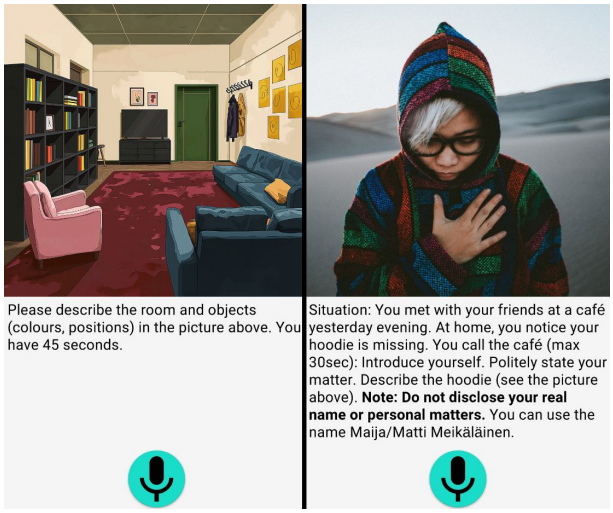

Figure 2: *Interface of visual speaking tasks. The task on the left targets A1-A2 levels, the task on the right A2-B1 levels*

Table 1: *The prompt structure for the LLM.*

Based on the provided transcript of the user's description of a picture, assign a grade as 'Excellent', 'Good', 'Partially' or 'None' for the task completion. **{task question}** **{focus only on L2 speaking content assessment}** Explain your grading in detail and then give the grade. Then **{requirements for corrective feedback}** Grading examples: **{selected examples for each grade}** Grading criteria: **{grading criteria for each grade}** The description of the picture: **{text output of Visual NLG}** User: **{text output of ASR}**

3. Mobile app

The CaptainA mobile app was initially developed for Finnish mispronunciation detection targeting L2 beginners [5]. Recognising the demand for more advanced practice, we expanded the app by adding experimental picture-based tasks [4] targeting A1-B1 (the Common European Framework of Reference for Languages, CEFR) level speakers (see Fig. 2). These tasks involve manually selecting generated pictures and adding in-context learning examples. We also integrated image generation into our pipeline, enabling fully automatic picture generation for the A1-A2 task. While this automatic pipeline offers unlimited practice exercises, it also has lower performance in picture quality and grading accuracy, as well as higher costs.

In this proof-of-concept, automatic scoring (including explanation) focuses on task completion and ignores other rating criteria. However, the automated feedback can correct mistakes in other dimensions, including pronunciation, grammar, etc. (see Table 2). Moreover, users can quickly generate and listen to the corrected version with OpenAI's text-to-speech model. The cost of processing one answer is approximately 5 cents, including text-to-speak generation, but not including automatic picture generation.

4. Conclusion

In this paper, we showed how the existing CaptainA mobile app [5] could be extended to cover ASA also for spontaneous speech in picture-based tasks. Our proof-of-concept addresses many of the challenges outlined in the introduction by combining text-to-image generation, LLM, Visual NLG, and ASR to automatically score task completion in picture-based description tasks, and provides explanations and automated feedback to L2 learners. We developed the application for a language with limited L2 data and hope to attract more similar work on ASA and CALL systems for language learners. Future work includes exploring the use of generative AI to design tasks for more advanced speakers and implement other rating criteria.

Table 2: *Some examples of explanation and feedback from CaptainA. The tasks are on Fig. 2. Corrective feedback is highlighted for non-Finnish readers.*

Answer: Huoneessa on sininen sova ja punainen nojatuoli... **Explanation:** They have mentioned key items such as the blue sofa, red armchair... They also did not mention the colors of the item correctly. For example, the armchair is pink, not red. **Feedback:** Huoneessa on iso sininen sohva, jonka vieressä on pieni vaaleanpunainen nojatuoli...
Answer: ...multa on hukkunut tämmönen raidallinen ja värillinen huppari... **Explanation:** The user has completed the task excellently... The user mentioned that the hoodie is striped and colorful, which matches the description... **Feedback:** Moi, mä olin eilen illalla ystävien kanssa teidän kahvilassa ja huomasin, että multa on hukkunut raidallinen ja värillinen huppari...

5. Acknowledgements

We would like to thank the following projects and funding agencies: Aalto Science-IT, Kielibuusti, NordForsk (project 103893) and Research Council of Finland (grant no 355586, 355587, 355588, 322619, 322625, 322965).

6. References

[1] F. Cakmak, "Mobile learning and mobile assisted language learning in focus," *Language and Technology*, vol. 1, pp. 30–48, 2019.

[2] R. Lyster, K. Saito, and M. Sato, "Oral corrective feedback in second language classrooms," *Language teaching*, vol. 46, no. 1, pp. 1–40, 2013.

[3] L. Gu and L. Davis, "Providing SpeechRater feature performance as feedback on spoken responses," in *Automated Speaking Assessment: Using language technologies to score spontaneous speech*, K. Zechner and K. Evanini, Eds. Routledge, 2020, pp. 159–175.

[4] S. Luoma, *Assessing speaking*. Cambridge University Press, 2004.

[5] N. Phan, T. Grósz, and M. Kurimo, "CaptainA - A mobile app for practising Finnish pronunciation," in *Nordic Conference on Computational Linguistics*. University of Tartu Library, 2023.

[6] A. Salaberria, G. Azkune, O. L. de Lacalle, A. Soroa, and E. Agirre, "Image captioning for effective use of language models in knowledge-based visual question answering," *Expert Systems with Applications*, vol. 212, 2023.

[7] L. Zheng, W.-L. Chiang, Y. Sheng, S. Zhuang, Z. Wu *et al.*, "Judging LLM-as-a-judge with MT-bench and chatbot arena," *Advances in Neural Information Processing Systems*, vol. 36, 2024.

[8] N. Phan, A. von Zansen, M. Kautonen, E. Voskoboinik, T. Grósz, R. Hildén, and M. Kurimo, "Automated content assessment and feedback for Finnish L2 learners in a picture description speaking task," in *INTERSPEECH*, 2024.