



This is a repository copy of *Speak & improve: L2 English speaking practice tool*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/213150/>

Version: Published Version

Proceedings Paper:

Nicholls, D., Knill, K., Gales, M.J.F. et al. (2 more authors) (2023) *Speak & improve: L2 English speaking practice tool*. In: *Proceedings of Interspeech 2023. INTERSPEECH 2023, 20-24 Aug 2024, Dublin, Ireland. International Speech Communication Association (ISCA)* , pp. 3669-3670.

© 2023 ISCA. Reproduced in accordance with the publisher's self-archiving policy.

Reuse

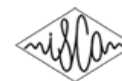
Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>



Speak & Improve: L2 English Speaking Practice Tool

Diane Nicholls^{1,*}, Kate Knill^{2,3*}, Mark J.F. Gales^{2,3}, Anton Ragni², Paul Ricketts¹

¹English Language iTutoring Ltd., Cambridge, UK

²Enhanced Speech Technology Ltd., Cambridge, UK

³ALTA Institute / Cambridge University Engineering Department, Cambridge, UK

diane.nicholls@englishlanguageitutoring.com, kmk1001@cam.ac.uk

Abstract

A problem for building and studying approaches to automatically assess and give feedback to L2 English learners on their speaking ability is a lack of suitable data sets and platforms to run experiments on. This paper describes the Speak & Improve (S&I) speaking practice tool, which has been designed to meet the needs of researchers while also offering learners the opportunity to practise their English speaking and improve their confidence. S&I has tasks that allow the learner to demonstrate and improve their proficiency across the English speaking construct. Most of the tasks are free speaking, that is, the learner is not constrained in what they have to say. Over 400,000 learners worldwide have tried the alpha version of S&I. This paper presents the next version which provides more choice and flexibility to the users and more feedback on performance.

Index Terms: spoken language assessment, computer assisted language learning, speaking practice

1. Introduction

Automated spoken language assessment for free speaking (unscripted responses to prompts) and read aloud speech tasks is gradually being deployed by companies including Pearson, ETS, Cambridge Assessment English, Liulishuo and Duolingo for L2 English speakers. These learners often lack opportunities to practise their speaking skills, especially across the variety of communicative speaking tasks that occur outside the classroom. Whilst for researchers working in this area a major hurdle is the lack of data to train and evaluate assessment and feedback systems on, and platforms to investigate different feedback approaches. This paper presents *Speak & Improve* which is attempting to address some of these issues.

Speak & Improve (S&I)¹ is a research project from the University of Cambridge in association with Cambridge University Press & Assessment and English Language iTutoring Ltd. It is an always available, free, web app. Learners can practise their English speaking and improve their confidence on a wide range of communicative speaking tasks. Users can interact with S&I through many different devices including laptops, tablets and mobile phones, as it is based in the browser. S&I is designed for all proficiency levels, from basic beginner through independent intermediate to proficient learners; on the CEFR [1] scale from below A1 to C1 and above. Launched quietly in 2017, by June 2022, S&I had received 9 million answers submitted by 400,000

users worldwide and was growing at a rate of 70,000 answers per month, reaching a peak of 105,000 answers in March 2023.

The underlying spoken language processing technology in S&I is derived from technology transferred from the ALTA Institute at the University (see [2] for examples and further references). It provides researchers with a tool to gather data and assess approaches to automated spoken language assessment and feedback. A public data set release is being planned.

The first, alpha, version of S&I offered five complete practice tests based on the Linguaskill Speaking exam [3]. This paper introduces the next, beta, version of S&I. Learners can choose to focus on a particular speaking skill and/or to take a complete practice test. Apart from interactive competence, the learner can practice a wide range of communicative speaking skills and get feedback on their level of proficiency.

2. Speak & Improve Tool

After the splash (landing) page, where they first meet the speech robot, Sandi, a user sees the layout in Figure 1.

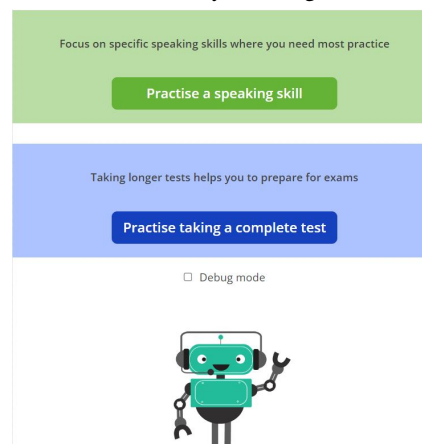


Figure 1: *S&I* selection screen

Practice a speaking skill approach. If the learner wants to focus on a particular aspect of their speaking, they can select from one of seven skills:

1. Answer questions about yourself
2. Read aloud
3. Give your opinion
4. Give a presentation about a graphic
5. Give a presentation about something personal
6. Give advice or make a recommendation
7. Answer questions about a topic

The authors marked with* made equal contributions to this paper. This paper reports on research supported by Cambridge University Press & Assessment (CUP&A), a department of The Chancellor, Masters, and Scholars of the University of Cambridge.

¹<https://speakandimprove.com>

As they work through each task in a skill section, users are given a level for their speech related to the CEFR scale and some motivational feedback from Sandi.

All of these tasks are free speaking tasks except for the Read Aloud task. Figure 2 illustrates a typical question from the "Give your opinion" task. If users register with the tool and log in they have a choice of 5 or 6 practice tasks per category. Otherwise they will have access to 1 task per category.

Give your opinion #1

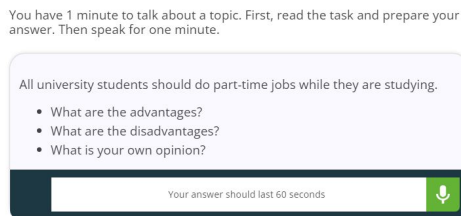


Figure 2: Example of S&I "Give your opinion" speaking task

Practise a complete test approach. A learner who selects to do a complete test has access to 1 or 5 free tests if logged out or in, respectively. They are taken through a flow of 5 full tasks to make a complete test that takes about 15 minutes. At the end, the learner receives a level on the CEFR scale and some motivational feedback. They also receive a level for each of the tasks in the test so that they can identify their weaker and stronger skills.

Assessment and Feedback. An automated spoken language assessment system is run to provide feedback to help the user know how they are doing and guide them on what they are doing well/need to improve. A deep density network (DDN) based [2] auto-marker is used to score the task and/or a complete test. This is similar to that presented in [3] but fully automatic with no human involvement.

Audio is passed through an API to an audio processing docker. The audio quality is first checked. If ok, ASR is run to generate a transcript and then word, syllable and grapheme level time alignments are generated. From these and the audio a range of statistics and information is gathered. These are passed to the auto-marker docker which derives the auto-marker input features from them. The features aim to cover as much of the speaking test construct as is possible. Features [4] relate to fluency (e.g. fraction of disfluencies), pronunciation and intelligibility (e.g. phone distance features, ASR word confidence scores), language resource and discourse management (e.g. grade specific LM perplexity). Appropriateness to the topic is checked through an off-topic response detector [5].

Auto-marking is run at a task and/or overall test level. The regression-based auto-marker returns a score for each task (or test) on a continuous scale of 0-6 which the S&I tool converts into a score based on the CEFR scale (A1, A2, B1, B2, C1, C2)². The score is presented to the user with some motivational feedback provided by Sandi as shown in Figure 3. This will be further expanded as S&I develops. The learner can track their progress through a history display, Figure 4.

Generally the auto-marker gives an accurate prediction of the speaker's achievement on a task. A number of factors, however, can affect the scoring, such as issues with the recording,

²The CEFR level strictly applies to a full speaking test which assesses all aspects of the speaking construct. To avoid confusing learners, the same grades are used for individual tasks as an indicator only.

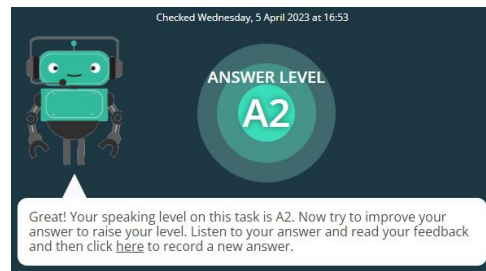


Figure 3: Task score with motivational feedback

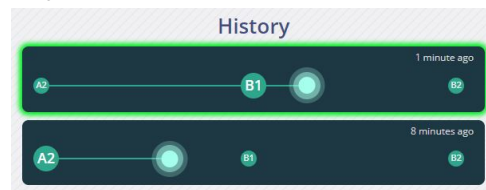


Figure 4: Performance level history indicator

background noise or outlier speakers. To try to prevent the system giving a badly mismatched score in these cases, the auto-marker provides the tool with a measure of its confidence in its own prediction [6]. If the confidence score is below a threshold, the tool can opt to withhold the score from the user. The S&I tool can also detect poor recording conditions and advise the learner to move to a quieter location and/or change to a better internet connection.

3. Conclusions

This paper has presented the Speak & Improve L2 English speaking practice tool which allows learners to practise a range of communicative speaking functions, receiving a score and motivational feedback. Researchers gain data - a public release is planned for 2023/4 - and an experimental platform.

4. Acknowledgements

Thanks to all the developers at ELiT who have contributed to the Speak & Improve tool. This paper reports on research supported by Cambridge University Press & Assessment, a department of The Chancellor, Masters, and Scholars of the University of Cambridge.

5. References

- [1] Council of Europe, *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge, UK: Cambridge University Press, 2001.
- [2] S. Bannò *et al.*, "L2 proficiency assessment using self-supervised speech representations," 2022.
- [3] J. Xu *et al.*, "Linguaskill: Building a validity argument for the speaking test," Linguaskill Research Reports, UCLES, Tech. Rep., 2020, <https://www.cambridgeenglish.org/Images/589637-linguaskill-building-a-validity-argument-for-the-speaking-test.pdf>.
- [4] Y. Wang *et al.*, "Towards automatic assessment of spontaneous spoken English," *Speech Communication*, vol. 104, pp. 47–56, 2018.
- [5] V. Raina and M. Gales, "Off-topic detection systems for spoken language assessment," in *Proc. Language Testing Forum (LTF) 2021*, 2021.
- [6] X. Wu *et al.*, "Ensemble Approaches for Uncertainty in Spoken Language Assessment," in *Proc. Interspeech 2020*, 2020, pp. 3860–3864.