



(<https://www.kieliverkosto.fi/fi>)

The Aasis research project: automatically assessing spoken interaction in L2 Finnish

This article presents aims and methods of the Aasis [Automatic assessment of spoken interaction in a second language] research project, which is funded by the Research Council of Finland (2023–2027). The multidisciplinary project brings together researchers from several fields, namely education, language technology and phonetics, to investigate L2 Finnish learners' interaction in dialogue speaking tasks.

Julkaistu: 5. joulukuuta 2023 | Kirjoittanut: Anna von Zansen

How to assess spoken interaction?

In the context of second language (L2) learning and language assessment, *speaking* refers to an individual language learner's skill, whereas *spoken interaction* involves two or more people talking to each other. In spoken interaction, both speakers act as speakers and listeners constructing the discussion together (Luoma 2004, 20).

As Galaczi and Taylor (2018) conclude, the construct of *interactional competence* is multifaceted and challenging to define. Speakers' personal characters, their interactional skills as well as the context and the purpose of talking to each other are factors shaping the interaction. For a comprehensive overview of empirical research, theoretical approaches and assessment practices related to spoken interaction, see Galaczi & Taylor (2018).

In speaking assessment, the purpose of assessment, i.e., what will the assessment be used for, directs decisions related to construct definition, task design and rating scale development (e.g., Luoma 2004). Usually, one of the first decisions deals with the test format: should spoken interaction be tested one individual at a time, paired or in a group?

Both holistic and analytic scales (e.g., Luoma 2004) can be used to assess L2 speaking including spoken interaction. It is common to focus on dimensions such as pronunciation, fluency, range and accuracy, which can be rated by humans or the machine (Kautonen & von Zansen 2020). The analytic scale “Qualitative features of spoken language” (Council of Europe 2020, appendix 3 p. 183–185) is one example.

However, in paired speaking tests, other dimensions such as topic development, turn-taking and interactive listening strategies become important for the raters (Borger 2019). Since both speakers are expected to interact in the dialogue, it is to be decided whether raters should give individual or shared scores on interaction to the speakers (Borger 2019, May 2011).

Nevertheless, diverse formats can be used side-by-side in order to balance between validity concerns related to variability, authenticity and construct coverage in speaking assessment (Galaczi & Taylor 2018). In the same way, the Aasis project plans to use a combination of different task types (e.g., monologues and dialogues) to ensure sufficient demonstration of L2 learner’s spoken interaction.

Computer-assisted L2 speaking assessment

Technology provides new opportunities for language assessment (Suvorov & Hegelheimer 2013). For example, automated speaking assessment (Zechner & Evanini 2020) has many advantages compared to traditional assessment dependent on human and physical resources. Computer-assisted speaking tests may also refer to online assessments led by the examiner using videoconferencing (e.g., Nakatshura et al. 2021) or online environments enabling test-takers’ interaction (e.g., Ockey et al. 2017). More about the use of dialogic speaking tasks in spoken dialogue systems can be found in Ramanarayanan et al. (2020) and Ockey et al. (2023).

The motivation of the Aasis project derives from the consortium’s previous project, DigiTala (<https://www.helsinki.fi/en/projects/digital-support-learning-and-assessing-second-language-speaking>) (Kautonen

& von Zansen 2020, von Zansen 2023), in which the research team developed an ASR-based (*automatic speech recognition*) online tool (von Zansen et al. 2022) for assessing L2 Swedish and Finnish learners' speech automatically and providing automated feedback to the learners.

The previous project recorded L2 Finnish and Swedish learner's speech and explored which features of speech can automatically be measured. As a result, the automatic scoring system produced a holistic score (estimation of the speaker's proficiency level) and four analytic scores (task completion, pronunciation, fluency, range and accuracy) based on the speech sample recorded by the language learner using Moodle (von Zansen 2023).

However, the targeted speaking tasks were limited to monologue tasks including read-aloud and short production tasks. Next, the Aasis project will expand ASR-based L2 speaking assessment to cover Finnish learners' spoken interaction. Therefore, the main data collected for Aasis will be videoed.

Research gap: non-verbal communication in spoken interaction

Regardless of test format (individual or paired, live or online) and task types, assessment of spoken interaction has traditionally neglected to take notice of non-verbal communication, although it is acknowledged to be essential in the interaction between people.

To illustrate, the scales in the Common European Framework for Reference for Languages (Council of Europe 2020) mention *non-verbal signals* only a few times: mainly in descriptors for the A1 level on mediation scales. Nevertheless, on the above-mentioned scale of Qualitative features of spoken language (Council of Europe 2020, appendix 3 p. 183–185), *non-verbal and intonational cues* are mentioned in the descriptor for the C2 level in the criterion of interaction.

Usually, the rating scales tend to use quite general terms in describing interaction, as the descriptor for B1+ level on the scale of overall spoken interaction (Council of Europe 2020, 72) shows:

“Can communicate with some confidence on familiar routine and non-routine matters related to their interests and professional field. Can exchange, check and confirm information, deal with less routine situations and explain why something is a problem. Can express thoughts on more abstract, cultural topics such as films, books, music, etc.”

Furthermore, as another example, the scale of online conversation and discussion (Council of Europe 2020, 85) defines the B1+ level as follows:

“Can engage in real-time online exchanges with more than one participant, recognising the communicative intentions of each contributor, but may not understand details or implications without further explanation. Can post online accounts of social events, experiences and activities referring to embedded links and media and sharing personal feelings.”

The Aasis (<https://www.helsinki.fi/en/projects/automatic-assessment-spoken-interaction-second-language>) project seeks to fill this gap by investigating both verbal and nonverbal features of spoken interaction in L2 Finnish. The conceptual aim of the project is to improve authenticity of assessment by including the assessment of interaction skills in ASR-based L2 speaking assessment. The methodological aim of the project is to support the reliability of L2 speaking assessment by developing a research-based automatic tool for assessing L2 spoken interaction. Scores produced by the machine could help human raters' work and enable providing automated feedback to the learners.

Starting points for assessing L2 Finnish learners' interaction automatically

In order to reach the goal of the project, assessing spoken interaction automatically, the consortium brings together researchers of language education, phonetics and signal processing. Although the project aim is challenging, the project benefits from results and experiences gained in the previous DigiTala project (Kautonen & von Zansen 2020).

Similarly, the multidisciplinary project starts with designing and pilot testing speaking tasks based on the defined construct, choosing rating scales for human raters as well as drafting questionnaires for

investigating stakeholder beliefs. Human ratings are namely needed for training automatic assessment models that predict the scores using machine learning methods (for automatic scoring models of the DigiTala project see Al-Ghezi et al. forthcoming). However, this time the data include videoed dialogues from academic L2 Finnish learners. Both visual and acoustic signals are analysed in order to investigate features relevant for assessing spoken interaction.

In addition to L2 learners' speech, the research interests of the project concern *non-verbal communication*, which includes body language (e.g., gestures, facial expressions, eye contact) and the use of extra-linguistic speech sounds (e.g., "sh" for silence). Some prosodic qualities or a combination of them, such as voice quality (e.g., gruff or breathy), pitch (e.g., whining) and loudness (e.g., whispering) are also part of non-verbal communication (Council of Europe 2001, 88–90). Moreover, the phonetic research of the project will focus on *interactional phonetics*, i.e., aspects related to the speaking time (e.g., silences, overlapping speech) and turn-taking. Furthermore, the project will explore novel research methods such as algorithm-based facial expression analysis and eye-tracking in the context of L2 education.

The project will reform L2 speaking assessment by investigating features relevant to language learners' spoken interaction and finding ways to measure these automatically. Findings of the project can be used e.g., for creating more accurate assessment methods as well as training teachers and raters. Automated feedback provided by an online tool would increase opportunities for practising speaking by supporting language learning at any time and place with or without access to teacher guidance.

Anna von Zansen is a post-doctoral researcher at the University of Helsinki. Her research interests include computer-assisted language testing, multimodality, L2 speaking and listening, visual cues and nonverbal communication.

The consortium of the Aasis project (Automatic assessment of spoken interaction in a second language, Research Council of Finland 2023–2027) combines expertise in speech and language processing, language education and phonetics at the University of Helsinki (grant number 355586), Aalto University (grant number 355587) and the University of Jyväskylä (grant number 355588). The current project builds on

experiences gained during the consortium's previous project, DigiTala (Academy of Finland 2019–2023, grant numbers 322619, 322625, 322965).

References

- Al-Ghezi, R., Voskoboinik, K., Getman, Y., von Zansen, A., Kallio, H., Kurimo, M., Huhta, A., Hilden, R. (forthcoming). Automatic Speaking Assessment of Spontaneous L2 Finnish and Swedish.
- Borger, L. (2019). Assessing interactional skills in a paired speaking test: Raters' interpretation of the construct. *Apples - Journal of Applied Language Studies*, 13(1), 151–174.
<https://doi.org/10.17011/apples/urn.201903011694>
- Council of Europe (2001) Common European Framework of Reference for Languages: learning, teaching, assessment. Council of Europe, Strasbourg. <https://rm.coe.int/16802fc1bf>
<https://rm.coe.int/16802fc1bf>
- Council of Europe (2020) Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume. Council of Europe Publishing, Strasbourg. www.coe.int/lang-cefr
<http://www.coe.int/lang-cefr>
- Galaczi, E. & Taylor, L. (2018) Interactional Competence: Conceptualisations, Operationalisations, and Outstanding Questions, *Language Assessment Quarterly*, 15:3, 219-236.
<https://doi.org/10.1080/15434303.2018.1453816>
<https://doi.org/10.1080/15434303.2018.1453816>
- Kautonen, M. & von Zansen, A. (2020) DigiTala research project: Automatic speech recognition in assessing L2 speaking. *Kieli, koulutus ja yhteiskunta*, 11(4). <https://www.kieliverkosto.fi/fi/journals/kieli-koulutus-ja-yhteiskunta-kesakuu-2020/digitala-research-project-automatic-speech-recognition-in-assessing-l2-speaking>
<https://www.kieliverkosto.fi/fi/journals/kieli-koulutus-ja-yhteiskunta-kesakuu-2020/digitala-research-project-automatic-speech-recognition-in-assessing-l2-speaking>
- Luoma, S. (2004) *Assessing speaking*. Cambridge University Press.
- May, L. (2011) Interactional Competence in a Paired Speaking Test: Features Salient to Raters, *Language Assessment Quarterly*, 8:2, 127-145.
<https://doi.org/10.1080/15434303.2011.565845>
<https://doi.org/10.1080/15434303.2011.565845>

- Nakatsuhara, F., Inoue, C., Berry V. & Galaczi, E. (2021) Video-conferencing speaking tests: do they measure the same construct as face-to-face tests?, *Assessment in Education: Principles, Policy & Practice*, 28:4, 369-388. <https://doi.org/10.1080/0969594X.2021.1951163> (<https://doi.org/10.1080/0969594X.2021.1951163>)
- Ockey, G., Gu, L. & Keehner, M. (2017) Web-Based Virtual Environments for Facilitating Assessment of L2 Oral Communication Ability, *Language Assessment Quarterly*, 14:4, 346-359. <https://doi.org/10.1080/15434303.2017.1400036> (<https://doi.org/10.1080/15434303.2017.1400036>)
- Ockey, G., Chukharev-Hudilainen, E. & Hirsch, R. (2023) Assessing Interactional Competence: ICE versus a Human Partner, *Language Assessment Quarterly*, <https://doi.org/10.1080/15434303.2023.2237486> (<https://doi.org/10.1080/15434303.2023.2237486>)
- Ramanarayanan, V., Evanini, K., & Tsuprun, E. (2020). Beyond monologues: Automated processing of conversational speech. In K Zechner & K. Evanini (Eds.) *Automated speaking assessment*, pp. 176-191, New York : Routledge.
- Suvorov, R., & Hegelheimer, V. (2013). Computer-assisted language testing. *The companion to language assessment*, 2, 594-613.
- von Zansen, A. (2023). DigiTala – Miten automaattinen palaute voi tukea puhumisen harjoittelua? *Kieli, koulutus ja yhteiskunta*, 14(3). <https://www.kieliverkosto.fi/fi/journals/kieli-koulutus-ja-yhteiskunta-toukokuu-2023/digitala-miten-automaattinen-palaute-voi-tukea-puhumisen-harjoittelua> (<https://www.kieliverkosto.fi/fi/journals/kieli-koulutus-ja-yhteiskunta-toukokuu-2023/digitala-miten-automaattinen-palaute-voi-tukea-puhumisen-harjoittelua>)
- von Zansen, A., Alanen, Al-Ghezi, R., Erkkilä, J., Harjunpää, T., Heijala, M. & Kallio, H. (2022). *DigiTala Moodle plugin*. https://github.com/aalto-speech/moodle-mod_digitala (https://github.com/aalto-speech/moodle-mod_digitala)
- Zechner, K. & Evanini, K. (2020). *Automated speaking assessment: Using language technologies to score spontaneous speech*. New York: Routledge, Taylor & Francis Group.

Artikkeliin viittaaminen

von Zansen, A. (2023). The Aasis research project: automatically assessing spoken interaction in L2 Finnish. *Kieli, koulutus ja yhteiskunta*, 14(7).

Saatavilla: <https://www.kieliverkosto.fi/fi/journals/kieli-koulutus-ja-yhteiskunta-joulukuu-2023/the-aasis-research-project-automatically-assessing-spoken-interaction-in-l2-finnish>

(<https://www.kieliverkosto.fi/fi/journals/kieli-koulutus-ja-yhteiskunta-joulukuu-2023/the-aasis-research-project-automatically-assessing-spoken-interaction-in-l2-finnish>)