

**This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.**

**Author(s):** Kautonen, Maria; von Zansen, Anna

**Title:** DigiTala research project : automatic speech recognition in assessing L2 speaking

**Year:** 2020

**Version:** Published version

**Copyright:** © Kieliverkosto ja kirjoittajat, 2020

**Rights:** CC BY 4.0

**Rights url:** <https://creativecommons.org/licenses/by/4.0/>

**Please cite the original version:**

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>



# DigiTala research project: Automatic speech recognition in assessing L2 speaking

---

**In this article, we discuss the premise of assessing oral language skills with the help of automatic speech recognition. We present the main goals and working methods in DigiTala project (2019–2023) – a research project between University of Helsinki, Aalto University and University of Jyväskylä that aims at developing a digital tool for assessing and practicing oral language skills. Ultimately, this digital tool would enable implementing tasks measuring oral skills into the language tests of the Matriculation Examination.**

Julkaistu: 10. kesäkuuta 2020 | Kirjoittanut: Maria Kautonen ja Anna von Zansen

## The importance of oral language skills and assessment

Speaking and interacting in a second and foreign language (L2) is an essential skill in our modernised society, as globalisation and technological improvements have increased the demand for one's language skills. According to the European Commission's overview of labour market demand for foreign language proficiency (COE 2015, 30–31), employers usually require applicants to have high (C1–C2, see CEFR 2001) or medium level (B1–B2, see CEFR 2001) proficiency in the foreign language useful to their business. Moreover, the report (COE 2015, 30–31) indicates that especially for higher proficiency levels excellent oral skills are essential.

Teaching and assessing oral and interaction skills has also been one key aspect in the Finnish national core curricula for upper secondary schools for the last decades (NCC 2003; 2015, 2019). Conversely, assessing written skills has been emphasized in Finnish language education as well as in language tests of the Matriculation Examination (MEB 2020), which is a national high-stakes examination. The examination is considered high-stakes since it is used to gain entry into higher education. The examination is taken at the end of the Finnish upper secondary school and its purpose is to discover whether upper secondary school students have reached the learning goals of the national core curriculum. (MEB 2020.) The Common European Framework of

Reference for Languages (CEFR 2001) has had a strong impact on Finnish language education. As a result, the curricula (NCC 2003; 2015, 2019) present local applications of the CEFR's (2001) levels.

Assessment tends to have an impact on teaching. Therefore, speaking assessment is often needed to encourage teachers and students to practise oral skills. The Finnish Matriculation Examination (MEB 2020) has a great impact on language teaching especially in upper secondary schools in Finland. However, an oral test has so far been considered to be too challenging to include in language exams in the Matriculation Examination, partly because it is time-consuming and can be expensive to arrange (see e.g. Ministry of Education 2006). As a consequence, new tools are needed for practising and assessing oral skills in language education in Finland.

## Measuring oral language skills with the help of automated assessment

To address the above-mentioned challenges with assessing oral skills in a large-scale and high-stakes test, researchers and educational organisers have turned to the benefits of automated speaking assessment during the last couple of decades (Eskenazi 1999, Zechner & Evanini 2020, 3). Studies show that computer-assisted pronunciation training and automatic speech recognition (ASR<sup>[1]</sup>) can enhance foreign language learning for example in terms of improving pronunciation and providing feedback (Golonka et al. 2014, 70, 81).

Outside of Finland, automated methods are already used to assess pronunciation, fluency, vocabulary, grammar, content and discourse coherence in oral language tests in English (e.g. TOEFL iBT Speaking, see Zechner & Evanini 2020). Research on automated assessment of English as L2 is more extensive than research and development of automated assessment for other languages such as Finnish and Swedish. With current methods, pronunciation can be assessed automatically with the help of features measuring pronunciation of individual sounds (vowels and consonants), durations of individual sounds and syllables, stress and tone, whereas fluency can be measured in terms of e.g. pauses, articulation rate and repetitions (Hsieh, Zechner & Xi 2020). When it comes to grammar and vocabulary, the range and complexity of grammatical structures and lexical diversity can also be measured with automated methods (Yoon, Lu & Zechner 2020). In addition, content and discourse coherence can be measured e.g. by comparing the content of the speech sample with different reference materials and by examining the speaker's use of coherence markers such as conjunctions and discourse connectives (Wang & Evanini 2020).

When it comes to task types, read-aloud speech can at the moment be assessed in a more reliable way than spontaneous speech because the responses are more controlled and what the learners are trying to say can be predicted better (see e.g. Zechner & Evanini 2020). Assessing spontaneous speech samples in a highly reliable way is, however, even more relevant, as it is a more natural speech form in interaction and communication.

In this article, we discuss how the advances in automatic speech recognition and machine learning have a potential to make self-regulated learning as well as teaching and assessing oral language skills more effective. In the field of language learning, ASR enables independent practising of oral skills by comparing the student's pronunciation acoustically with target pronunciation and by providing feedback (Golonka et al. 2014, 73–74). At its best, automated assessment can lead to more standardised assessment, to be used in high-stakes testing such as the Matriculation Examination. Although teachers often question the fairness of automatic scoring, a hybrid scoring model that combines human and machine rating might increase the consistency of assessment while ensuring that also the aspects of performance which are difficult or impossible to measure by automatic systems are included in the assessment (Luo et al. 2016).

## Starting point for developing automated assessment in DigiTala

The purpose of the research project DigiTala (<https://helsinki.fi/en/digitala>) is to develop a digital tool that can be used to assess oral language skills with the help of automatic speech recognition. An additional aim is to use the tool for self-regulated learning purposes, so that the students can practice especially pronunciation with it. In the first stages of the project, the digital assessment tool will be developed for Finnish and Swedish as second languages as Finnish and Swedish are the national languages of Finland. Later, the tool can be further developed to assess other foreign languages.

In order to develop a digital tool to analyse and assess L2 speech, a large corpus of L2 speech will be collected from upper secondary school students studying Finnish or Swedish as a second language, consisting of speech in different types of speaking tasks. Human raters will then rate the performances, so that the tool can be trained to imitate and predict human rating of a speech sample. The digital tool will focus on assessing an overall CEFR-level and assessing pronunciation, fluency, grammar and vocabulary (cf. NCC 2003). At a later stage of the project, the aim is to provide learners with personalised feedback on their oral skills concerning these four dimensions.

Implementing pronunciation assessment and pronunciation feedback in particular is seen valuable in the project as pronunciation plays an important role for overall oral proficiency (Kautonen 2019), comprehensibility of speech (Heinonen 2018) and foreign accent (Kuronen & Zetterholm 2017, Kuronen & Kautonen 2018, Toivola 2011). Poor pronunciation and heavily accented speech can lead to negative listener judgements on not just language skills but also on one's personal traits and other competences in the labour market (Boyd 2003). Pronunciation is thus an important learning goal that learners need to practice. Also, without being able to interpret the speaker's pronunciation, the digital tool will not be able to assess the lexical or grammatical competence of the speaker irrespective of how well the speaker would master these. Pronunciation has, however, a history of being neglected in L2 teaching also in Finland, and language teachers say that they need more knowledge of methods to teach pronunciation (Tergujeff 2013, Huhtamäki & Zetterholm 2017).

At the moment, the project team is working on defining the construct, i.e. specifying which characteristics of oral skills are to be measured. Moreover, the team is outlining the computer-assisted assessment tool that is being developed. The team is also searching for suitable already existing open source platforms (such as Moodle), since the course exam system Abitti (2015) published by the Finnish Matriculation Examination Board cannot be used for this research due to practical reasons. Based on lessons learned in the previous project (<https://blogs.helsinki.fi/digitala-projekti/>) the team will start planning research methods. The current curricula (NCC 2015, 2019) pose an obvious starting point for writing test specifications and designing tasks. On the other hand, after compiling first drafts of assessment criteria to be used by human raters, the scale developed for the previous curriculum (NCC 2003) appears clearer to use for assessment purposes. For assessing pronunciation, a rating tool combining the content of an existing tool (Heinonen & Kautonen 2017) and new descriptors from the Companion volume (COE 2018) will be developed.

Hand in hand with the task design, the team is considering options for providing corrective but encouraging feedback to the students. The project team is also outlining questionnaires and interviews for mapping user experiences. In the future, the tool developed in this project could even be evolved into an intelligent tutoring system<sup>[2]</sup> that provides automatic corrective feedback, which could increase students' motivation and confidence to use the language (see Golonka et al. 2014, 80–81).

In addition to developing the digital assessment tool, the project team will also investigate e.g. which task types and assessment criteria would be feasible in an automated speaking test intended for measuring Finnish upper secondary school students' oral skills. In other words, the project team conducts small proof of concepts in order to find the most practical ways to test and assess oral skills. When it comes to implementing an automated speaking test into the language tests of the Matriculation Examination, necessary precautions need to be taken in order to prevent quality issues (see Bachman & Palmer 1996, 17–42 for test usefulness) in a large-scale and high-stakes test. Close collaboration from the beginning with different stakeholders ensures the usability of the project's findings.

[1] "A technology that allows a computer to identify the words a person speaks into a microphone. ASR is often a component of speech pronunciation software, and as such, identifies particular parameters of the learner's output, such as prosody or specific sounds, and provides feedback on these aspects of performance."(Golonka et al. 2014, 73–74.)

[2] "A program that simulates a tutor by providing direct, customized instruction and/or feedback to a learner. Such a system is generally comprised of four components: an interface (platform), an expert model (domain of knowledge the student is intended to acquire), a student model (current state of student's knowledge), and a tutor model (which provides appropriate feedback and instruction by using the identified gaps between the student and the expert models)." (Golonka et al. 2014, 73.)

*Maria Kautonen is a post-doctoral researcher at the University of Jyväskylä.*

*Anna von Zansen is a post-doctoral researcher at the University of Helsinki.*

*The DigiTala (2019–2023) research project [Digital support for learning and assessing second language speaking] develops a digital tool to assess oral language in high-stakes tests in Finland. The tool can also be used for students to practice foreign or second language pronunciation or producing speech. Pilot versions will be made for Swedish and Finnish language. The project is financed by the Academy of Finland 2019–2023, and combines expertise in speech and language processing, language education and phonetics at the University of Helsinki (grant number 322619), Aalto University (grant number 322625) and the University of Jyväskylä (grant number 322965). The current project builds on lessons learned during a pilot project, see *DigiTala (2015–2017)* (<https://blogs.helsinki.fi/digitala-projekti/>).*

## References

- Abitti (2015). Course exam system. The Matriculation Examination Board. <https://www.abitti.fi/>  
(<https://www.abitti.fi/>)
- Bachman, L. & Palmer, A. (1996). *Language Testing in Practice*. Oxford: Oxford University Press.
- Boyd, S. (2003). Foreign-born teachers in the multilingual classroom in Sweden: The role of attitudes to foreign accent. *International Journal of Bilingual Education and Bilingualism*, 6 (3–4), 283–295.
- CEFR (2001). *Common European Framework of Reference for Languages: learning, teaching, assessment*. Strasbourg: Council of Europe. <https://rm.coe.int/16802fc1bf>  
(<https://rm.coe.int/16802fc1bf>)
- COE (2018). *Common European Framework of Reference for Languages: learning, teaching, assessment. Companion volume with new descriptors*. Strasbourg: Council of Europe. <https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989>  
(<https://rm.coe.int/cefr-companion-volume-with-new-descriptors-2018/1680787989>)
- COE (2015). *Study on foreign language proficiency and employability. Final Report*. Luxembourg: Publications Office of the European Union. <https://op.europa.eu/s/n3GQ>  
(<https://op.europa.eu/s/n3GQ>)
- Eskenazi, M. (1999). Using automatic speech processing for foreign language pronunciation tutoring: Some issues and a prototype. *Language Learning & Technology*, 2(2), 62–76.
- Golonka, E., Bowles, A., Frank, V., Richardson, D. & Freynik, S. (2014). Technologies for foreign language learning: a review of technology types and their effectiveness. *Computer Assisted Language Learning*, 27:1, 70–105. <https://doi.org/10.1080/09588221.2012.700315>  
(<https://doi.org/10.1080/09588221.2012.700315>)

Heinonen, H. & Kautonen, M. (2017). Miten ääntämistä arvioidaan? Käytännön työkalu opettajille. *Kieli, koulutus ja yhteiskunta*, 8(4). <https://www.kieliverkosto.fi/fi/journals/kieli-koulutus-ja-yhteiskunta-syyskuu-2017-1/miten-aantamista-arvioidaan-kaytannon-tyokalu-opettajille> (<https://www.kieliverkosto.fi/fi/journals/kieli-koulutus-ja-yhteiskunta-syyskuu-2017-1/miten-aantamista-arvioidaan-kaytannon-tyokalu-opettajille>)

Heinonen H. (2018). Uttalsfärdigheter och begriplighet i finskspråkiga gymnasisters L2-svenska. In: B. Silén, A. Huhtala, H. Lehti-Eklund, J. Stenberg-Sirén och V. Syrjälä (eds.), *Svenskan i Finland 17. Föredrag vid den sjuttonde sammankomsten för beskrivningen av svenskan i Finland*. Nordica Helsingensia. Helsinki: University of Helsinki. 32–45.

Hsieh, C.-N., Zechner, K. & Xi, X. (2020). Features Measuring Fluency and Pronunciation. In: Zechner, K. & Evanini, K. (eds.), *Automated speaking assessment: Using language technologies to score spontaneous speech*. New York: Routledge. 101–122. <https://doi.org/10.4324/9781315165103-1> (<https://doi.org/10.4324/9781315165103-1>)

Huhtamäki, M. & Zetterholm, E. (2017). Uttalets plats i undervisningen av svenska som andraspråk. *AFinLA-E: Soveltavan kielitieteen tutkimuksia*, 10, 45–60.

Kautonen, M. (2019). *Finskspråkiga inlärares uttal av finlandssvenska i fritt tal på olika färdighetsnivåer*. JYU Dissertations 90. Jyväskylä: University of Jyväskylä.

Kuronen, M. & Kautonen, M. (2018). Foneettisten piirteiden ja vieraan aksentin yhteydestä suomen kielessä. *Lähivertailuja* (28) 2018. 207–241.

Kuronen, M. & Zetterholm, E. (2017). Olika fonetiska drags relativa betydelse för upplevd inföddlighet i svenska. *Nordand* 2, 2017. 134–156.

Luo, D., Gu, W., Luo, R. & Wang, L. (2016). Investigation of the effects of automatic scoring technology on human raters' performances in L2 speech proficiency assessment. *Chinese Spoken Language Processing (ISCSLP) 2016*, 1–5

MEB (2020). The Finnish Matriculation Examination. [Website.] The Matriculation Examination Board. <https://www.ylioppilastutkinto.fi/en/> (<https://www.ylioppilastutkinto.fi/en/>)

Ministry of Education. (2006). *Lukiokoulutuksen suullisen kielitaidon arviointityöryhmän muistio. Opetusministeriön työryhmämuistioita ja selvityksiä 2006:26*. <http://urn.fi/URN:ISBN:952-485-213-6> (<http://urn.fi/URN:ISBN:952-485-213-6>)

NCC (2003). [The Finnish National Core Curriculum for Upper Secondary Schools 2003. Finnish National Agency for Education.] *Lukion opetussuunnitelman perusteet 2003*. Opetushallitus. [https://www.oph.fi/sites/default/files/documents/47345\\_lukion\\_opetussuunnitelman\\_perusteet\\_2003.pdf](https://www.oph.fi/sites/default/files/documents/47345_lukion_opetussuunnitelman_perusteet_2003.pdf) ([https://www.oph.fi/sites/default/files/documents/47345\\_lukion\\_opetussuunnitelman\\_perusteet\\_2003.pdf](https://www.oph.fi/sites/default/files/documents/47345_lukion_opetussuunnitelman_perusteet_2003.pdf))

NCC (2015). [The Finnish National Core Curriculum for Upper Secondary Schools 2015. Finnish National Agency for Education.] Lukion opetussuunnitelman perusteet 2015. Opetushallitus. [https://www.oph.fi/sites/default/files/documents/172124\\_lukion\\_opetussuunnitelman\\_perusteet\\_2015.pdf](https://www.oph.fi/sites/default/files/documents/172124_lukion_opetussuunnitelman_perusteet_2015.pdf)

([https://www.oph.fi/sites/default/files/documents/172124\\_lukion\\_opetussuunnitelman\\_perusteet\\_2015.pdf](https://www.oph.fi/sites/default/files/documents/172124_lukion_opetussuunnitelman_perusteet_2015.pdf))

NCC (2019). [The Finnish National Core Curriculum for Upper Secondary Schools 2019. Finnish National Agency for Education.] Lukion opetussuunnitelman perusteet 2019. Opetushallitus. [https://www.oph.fi/sites/default/files/documents/lukion\\_opetussuunnitelman\\_perusteet\\_2019.pdf](https://www.oph.fi/sites/default/files/documents/lukion_opetussuunnitelman_perusteet_2019.pdf)

([https://www.oph.fi/sites/default/files/documents/lukion\\_opetussuunnitelman\\_perusteet\\_2019.pdf](https://www.oph.fi/sites/default/files/documents/lukion_opetussuunnitelman_perusteet_2019.pdf))

Toivola, M. (2011). Vieraan aksentin arviointi ja mittaaminen Suomessa. Helsinki: Helsingin yliopisto. <http://urn.fi/URN:ISBN:978-952-10-7217-8> (<http://urn.fi/URN:ISBN:978-952-10-7217-8>)

von Zansen, A. (2019). New approaches to assessing listening – Pictures and video in the language tests of the Finnish Matriculation Examination. JYU Dissertations 2019, 136. Jyväskylä: University of Jyväskylä. <http://urn.fi/URN:ISBN:978-951-39-7961-4> (<http://urn.fi/URN:ISBN:978-951-39-7961-4>)

Wang, X. & Evanini, K. (2020). Features Measuring Content and Discourse Coherence. In: Zechner, K. & Evanini, K. (eds.), Automated speaking assessment: Using language technologies to score spontaneous speech. New York: Routledge. 138–156. <https://doi.org/10.4324/9781315165103-1> (<https://doi.org/10.4324/9781315165103-1>)

Yoon, S.-Y., Lu, X. & Zechner, K. (2020). Features Measuring Vocabulary and Grammar. In: Zechner, K. & Evanini, K. (eds.), Automated speaking assessment: Using language technologies to score spontaneous speech. New York: Routledge. 123–137. <https://doi.org/10.4324/9781315165103-1> (<https://doi.org/10.4324/9781315165103-1>)

Zechner, K. & Evanini, K. (2020). Automated speaking assessment: Using language technologies to score spontaneous speech. New York: Routledge, Taylor & Francis Group.

## Artikkeliin viittaaminen

Kautonen, M. & von Zansen, A. (2020). DigiTala research project: Automatic speech recognition in assessing L2 speaking. *Kieli, koulutus ja yhteiskunta*, 11(4).

Saatavilla: <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>)