

Information on the processing of personal data in the research project entitled Automatic assessment of spoken interaction in second language (AASIS)

The research project entitled Aasis (financed by the Research Council of Finland 2023-2027) involves processing of personal data. The purpose of this data protection notice is to provide information on the personal data to be processed, from where they are obtained and how they are used. Detailed information on the rights of data subjects will be provided at the end of this notice.

Your participation in the research project and provision of personal data are voluntary. If you do not wish to participate in the project or you wish to withdraw from it, you can do so without negative consequences. Participation in the study does not affect other language assessments (e.g. your course grade, YKI test results).

1. Data Controller

- University of Helsinki Address: P.O. Box 3 (Fabianinkatu 33), 00014 University of Helsinki, Finland
- University of Jyväskylä, Address: Seminaarinkatu 15, P.O. Box 35, 40014 University of Jyväskylä, Finland
- Aalto University Foundation sr, Address: P.O. Box 11000, FI-00076 Aalto, Finland

2. Contact person and principal investigator

Contact person in matters concerning the research project:

- Name: Raili Hilden
- Institution/Faculty/Department: University of Helsinki / Faculty of Educational Sciences
- Address: P.O. Box 9 (Siltavuorenpenger 3A), 000014 University of Helsinki
- Tel.: +358504482514
- E-mail: raili.hilden@helsinki.fi

3. Contact details of the data protection officer

The Data Protection Officer of the University of Helsinki can be reached at tietosuoja@helsinki.fi, the Data Protection Officer of the University of Jyväskylä at tietosuoja@juu.fi and the Data Protection Officer of Aalto University at tietosuojavastaava@aalto.fi.

4. Description of the research project and the purpose of processing personal data

The Aasis project aims at automatically assessing Finnish as a second language learners' spoken interaction. In addition to Finnish learners' speech, the Aasis project will study non-verbal features such as gaze, gestures and body movements in interaction and language assessment. The aim is to extend the ASR-based (automatic speech recognition) tool developed by the consortium's previous project, DigiTala (Academy of Finland 2019–2023), to cover automatic assessment of spoken interaction and non-verbal

features. Automatic assessment could support teachers' and language testers' work and allow learners to practice speaking on their own.

Participants of the research are adult Finnish learners whose speaking performances are recorded and videoed and human raters (i.e. language teachers or experts) who will assess the speaking performances. Human ratings and codings of interaction are used for training automatic assessment models that predict the scores using machine learning methods. Moreover, surveys and interviews are used to collect participants' background information and views on e.g. automatic assessment and functioning of the speaking tasks or rating scales. Furthermore, the project will use different research methods such as algorithm-based facial expression analyses and explore novel methods such as eye-tracking. Participants will be informed in detail about the methods used before data collection.

The Aasis project employs researchers from the University of Helsinki, Aalto University, and the University of Jyväskylä. The University of Helsinki is responsible for the pedagogical content, analyses of visual cues and for collaboration with learners. Aalto University is responsible for developing automatic speech recognition and automatic assessment. Moreover, Aalto University is responsible for storing research data. The University of Jyväskylä is responsible for analyzing speech samples and training raters.

5. Personal data included in the research data

A) Data being collected from Finnish as a second language learners

- Learners' names, contact information, consents
- Learners' recordings (video of facial expressions and gestures, speech)
- Annotations of learner's interaction
- Learners' test performances (answers, grades)
- Learners' background information (e.g., how many courses they have taken, possibly questions on language background, language learning, oral language skills, gender, first language)
- Learners' self-assessments of their language skills / test performance
- Learners' opinions e.g. on the speaking tasks or automated feedback (questionnaire, interview)
- Learners' performance assessment (automatic assessment)
- Learner's physiological measurements during the speaking performance (e.g. gaze data)
- User behaviour in online environments (e.g. activity logs, mouse clicks)

B) Data being collected from human raters (i.e. language teachers or experts)

- Names and contact information for human raters, consents
- Human raters' background information (e.g. first language, language skills, experience)
- Human raters' assessment of learners' performance (individual dimensions, overall rating)
- Feedback from raters e.g. on rating scales used or the tool that is being developed (questionnaire, interview)
- User behaviour in online environments (e.g. activity logs, mouse clicks)

In addition, we will collect the names and email addresses of Finnish teachers at universities for contacting Finnish learners. However, participation in the study does not affect other language assessments (e.g. Finnish course grade, YKI test results).

Furthermore, other information from the raters may be collected for payment of the fees. However, this information will not be processed as part of the research project.

Moreover, the project may reuse previously collected data, e.g. speech data and human ratings collected

during the DigiTala project (Svenska Folkskolans vänner 2015-2017) and (Academy of Finland 2019-2023).

6. Sources of personal data

The project creates speaking tasks measuring oral language skills in Finnish. With the help of the universities' language teachers, we ask language learners to participate in the speaking test consisting of monologue and dialogue speaking tasks. We record learners' speaking performances using universities' equipment (videocameras, mics, recorders) and premises (classrooms, studios, labs).

In addition, learners respond to a survey that elicits information on their background, their views on the test they have taken, and their self-assessments of their language skills. Some learners are also interviewed for feedback e.g. on the tool and development suggestions.

Data is also collected from human raters who listen to and evaluate learners' speaking performances in an online environment (e.g. Moodle). Raters fill in a background survey and participate in a training organised by the project. Some of them are also interviewed or asked to respond to a survey to gather feedback and suggestions for improvement.

During the 4-year project, some participants may be asked to participate in a study including physiological measurements such as eye-tracking. Gaze data will be collected using commercial research devices for eye-tracking which the participant can easily remove or stop using if they wish. These participants will be informed in detail about the use of the measuring equipment before the data collection starts.

For practical reasons, part of the data collection may occur online using video conferencing (e.g. Zoom) and online environments (Moodle) hosted by the universities. We may also explore online interaction using simulated conversations, where the learner responds to prompts provided by an avatar. In online environments, user behaviour such as mouse clicks and activity logs will be recorded.

Moreover, the project receives previously collected data from the Language Bank of Finland.

7. Sensitive personal data

In this research, special categories of personal data (i.e., sensitive data) will not be processed. The speaking tasks, as well as survey and interview questions, are designed so that they will not disclose any sensitive personal information (such as information related to ethnic origin, political opinions or health). Nor are the questions on the participant's language skills intended to collect information on the participant's ethnic background.

Moreover, we do not collect biometric data for identifying purposes. However, special attention will be paid to collecting and storing of video data (participants facial expressions and gestures combined with their speech) as well as physiological measurements (e.g. gaze data) to ensure privacy protection.

No data considered as special category data under Article 9 of the General Data Protection Regulation will be processed in the study.

The processing of sensitive personal data is based on Article 9(2)(j) of the General Data Protection Regulation (processing is necessary for scientific research purposes), as well as Section 6, Subsection 1, Paragraph 7 of the Finnish Data Protection Act.

8. Lawful basis for processing personal data

Personal data are processed on the following basis (Article 6(1) of the GDPR): performance of a task carried out in the public interest: scientific or historical research purposes or statistical purposes.

9. Recipients of data

Known recipients at the time of writing the Privacy Notice:

- Project researchers (all data)
- Service providers related to the data collection: Zoom video conferencing tool provided by the universities (online connections, interviews, feedback discussions, trainings). Webropol tool for making surveys provided by the Aalto University (surveys for consent and background information). CSC's Funet Filesender provided by the universities (sending large files).
- Teacher in-service training (samples that are used for illustrative purposes with learners' consents)
- Human raters recruited by the project (assessing speaking performance)
- Transcribers/ coders recruited by the project (speaking performances, interviews)
- Once the project is complete, the material will be archived in the Language Bank of Finland from participants that have given their consent for this purpose (names and contact information will not be stored in the Language Bank)
- Project's collaborators may access the data via the partner universities (without transferring personal data)

10. Transfer of data to countries outside the European Economic Area

Data will not be transferred to countries outside the European Economic Area, they are processed only within the EEA.

11. Automated decision-making

The research project involves no automated decision-making that has a significant effect on data subjects. However, the research includes profiling in terms of automatically or partly automatically evaluating the oral language skills of an individual (see section 4, the project develops ways to assess second language learners' spoken interaction automatically).

12. Protection of personal data

The personal data is processed and stored in such a way that only persons who need the data for research purposes can access it. The personal data register including personal information on the participants will be stored separately from the collected speaking performances and ratings (in different systems).

Learners' speaking performances are referred to with research identifiers instead of names. Participants are advised to avoid giving real names, places and sensitive information such as political opinions (see section 7 above).

The data processed in data systems will be protected using the following:

- ☒ Username and password
- ☒ Registration/log of use

- ☒ Access control
- ☒ Encryption (when needed)
- ☒ Two-factor identification
- ☐ Other, please specify:

Physical material (e.g., paper-based data or data in another physical form) is protected in the following ways: In a locked locker that only the project leaders at the University of Helsinki and at the University of Jyväskylä (Raili Hilden and Mikko Kuronen) can access.

Processing direct identifiers:

- ☐ The controller collects the personal data without direct identifiers.
- ☒ Direct identifiers will be removed during the analysis stage and kept separate from the analysed research data.
- ☐ The data will be analysed using direct identifiers, because (give grounds for preserving the direct identifiers):

13. Duration of the processing of personal data in this research project:

The criteria for defining storage of research data containing personal data is based on good scientific practice. In scientific research, the aim is to store the research data so that the research results can be verified and previously collected data can be used for further scientific research on the same subject or for scientific research in other fields.

Personal data collected in the project will be processed for five years after the project has been completed (2032) in order to complete research-related publications.

14. Processing of personal data when the research project ends

- ☒ The research data will be deleted from the universities' storage solutions after five years after the end of the project (2032).
- ☒ The research data with participant's permission will be archived for later, compatible scientific research in accordance with the requirements of the GDPR:
 - ☐ without identifiers ☒ identifiers included (no names or contact details but voice and videos with including facial images)

The storage of the research data is based on Article 5(1)(b) and (e) of the GDPR.

Data subjects will receive a new data protection notice on the new use of the research data, unless the controller can no longer identify the subjects from the data.

In addition, the data subjects will not be informed of the new research if delivering this information to them is impossible or involves a disproportionate effort or renders impossible or seriously impairs the achievement of the research objectives (Article 14(5)(b) of the GDPR).

Where and for how long will the data be archived: the Language Bank of Finland or other similar, trusted and curated data archive, permanently.

The material will be stored in the Language Bank labeled as material with the highest level of protection (RESTRICTED). Restricted materials can only be accessed for personal research upon

application. Prior to new research use, The Language Bank will ensure that the new research purpose is compatible with the original use of the material in accordance with the regulation requirements. The Language Bank provides persistent identifiers for the data.

15. Rights of data subjects and derogations from those rights

The contact person in matters related to research subjects' rights is the contact person stated in section 1 of this notice.

Rights of data subjects

Under the General Data Protection Regulation, data subjects have the following rights:

- Right of access to their own data
- Right to rectification of their data
- Right to the erasure of their data and to be forgotten
- Right to the restriction of processing of their data
- Right to data portability from one controller to another
- Right to object to the processing of their data
- Right not to be subject to automated decision-making

However, data subjects cannot exercise all their rights in all circumstances. The circumstances are affected by, for example, the legal basis for processing personal data.

Further information on the rights of data subjects in various circumstances can be found on the website of the Data Protection Ombudsman: <https://tietosuoja.fi/en/what-rights-do-data-subjects-have-in-different-situations>.

Derogations from rights

The General Data Protection Regulation and the Finnish Data Protection Act enable derogations from certain rights of data subjects if personal data are processed for the purposes of scientific research and the rights are likely to render impossible or seriously impair the achievement of the research purposes.

The need for derogations from the rights of data subjects will always be assessed on a case-by-case basis.

Right to appeal

If you consider that the processing of your personal data has been carried out in breach of data protection laws, you have the right to appeal to the Office of the Data Protection Ombudsman.

Contact details:

Office of the Data Protection Ombudsman
Street address: Ratapihantie 9, 6th floor, 00520 Helsinki
Postal address: PO Box 800, 00521 Helsinki
Phone (switchboard): 029 56 66700
Fax: 029 56 66735
Email: tietosuoja(at)om.fi