

This privacy notice contains the information that must be shared with research participants when you process any personal data in connection with scientific research and the information contained in the record of processing activities. The obligation to inform is laid down in articles 12–14 of the EU's General Data Protection Regulation (EU 2016/679), and the rules governing the record of processing activities are laid down in article 30.

This privacy notice describes the processing of personal data and demonstrates that research complies with data protection legislation. This privacy notice can also be used as a template when you are preparing the information provided to research participants.

1. Name of the research

Automatic Speech Recognition for Assessing Children's Speech

2. Research controller and contact information

- Collaborative research/project with external funding.

University of Eastern Finland, P.O. Box 1627, FI-70211 Kuopio, switchboard +358 (0)294 45 1111, email: kirjaamo@uef.fi

- Research conducted with basic funding or a personal grant

Assistant Professor Satu Saalasti
Department of Humanities, Faculty of Philosophy, University of Eastern Finland
Agora, Yliopistokatu 4, 80100 Joensuu
satu.saalasti@uef.fi, tel. +358 50 320 6178

- Thesis (bachelor's or master's thesis, doctoral dissertation)
- Joint controllers

3. Research parties and the division of responsibilities

The research is conducted by a team including researchers and data scientist. The principal investigator is Satu Saalasti. The research group additionally includes Professor Tomi Kinnunen, Post-doctoral researcher Ruchi Pandey, Laboratory Engineer Petri Pulli and PhD researcher/data scientist Edem Ahadzi.

All members of the research team are committed to:

- secure data handling
- protecting participants' privacy
- complying with GDPR and Finnish legislation

4. Research's supervisor, responsible leader or the group responsible for it

Assistant Professor Satu Saalasti
Department of Humanities, Faculty of Philosophy, University of Eastern Finland
Agora, Yliopistokatu 4, 80100 Joensuu
satu.saalasti@uef.fi, tel. +358 50 320 6178

5. Research implementers

- Assistant Professor Satu Saalasti
- Professor Tomi Kinnunen
- Ruchi Pandey
- Edem Ahadzi

6. Contact information of the data protection officer

Helena Eronen, helena.eronen@uef.fi, +358 29 4458187

7. Contact person for matters related to the processing of personal data

Assistant Professor Satu Saalasti
Department of Humanities, Faculty of Philosophy, University of Eastern Finland
Agora, Yliopistokatu 4, 80100 Joensuu
satu.saalasti@uef.fi, tel. +358 50 320 6178

8. Storing, destroying or archiving research material

Storage time of research material for research purposes (storage time is determined by the purpose for using the material):

Storage time of research material for research purposes:

The research material obtained from the Language Bank of Finland (Kielipankki) will be stored and processed for the duration of the aTOOR project and the associated doctoral research, estimated until November 2027. After this period, all copies of the corpus data held by the research group will be deleted. The original corpus data remains stored and available in the Language Bank of Finland.

How the personal data will be processed after the research has ended:

- Any research materials containing personal data will be destroyed
- Any research materials containing personal data will be archived
 - without identifiers
 - with identifiers

9. For what purpose will personal data be processed?

The purpose for the processing of personal data is developing new analysis methods for scientific research and clinical practice.

Our aim is to develop technology-based methods for remote speech therapy, addressing the shortage of speech therapists in Eastern Finland. We are developing an automatic speech recognition system that produces faithful phonetic transcriptions of Finnish children's speech, including pronunciation errors, to support the clinical assessment of speech sound disorders. The Lahjoita Puhetta corpus (obtained through the Language Bank of Finland) will be used to fine-tune a speech model for Finnish in two stages: first on adult speech to adapt the model to the Finnish language, then on the children's speech subset to adapt the model to the acoustic characteristics of children's speech. The audio recordings will be used solely for model training; no individual speakers will be identified or analysed.

10. On what basis will personal data be processed?

The processing of personal data always requires a legal basis. In this research, the processing basis ^Q is:

- a task carried out in the public interest / exercising public authority belonging to the controller, more specifically:
 - scientific or historical research or the compilation of statistics
 - the archival of research materials and cultural heritage materials
 - performance of a public authority task in the public interest
- the implementation of the legitimate interests of the controller or a third party specify the legitimate interest:
- consent granted by the data subject ^Q
- compliance with the controller's legal obligations related acts:

The research includes the processing of special categories of personal data or personal data relating to criminal convictions and offences ^Q. The special basis required for their processing is:

- archival for the purposes of the public interest, scientific or historical research, statistical purposes
- specific consent granted by the data subject
- a reason related to an important public interest on the basis of legislation
- on the grounds of public interest in the area of public health

11. What personal data is included in the research material?

The research material consists of the Lahjoita Puhetta corpus obtained from the Language Bank of Finland (Kielipankki). The corpus contains audio recordings of spoken Finnish and associated metadata, which may include: speaker age group, gender, regional background, and recording conditions. The audio recordings constitute voice data, which is personal data. No direct identification data (names, personal identity codes, contact information) is included in the data obtained from the Language Bank.

12. What sources will be used to collect personal data?

The following sources will be used to collect personal data:

- Directly from research participant
 - Interview
 - Video recording
 - Digital questionnaire (Webropol, Redcap or similar)
 - Questionnaire sent by mail
 - Other method, please specify:

From other sources than participants, where and what data:

The personal data (audio recordings and speaker metadata) was collected by the Lahjoita Puhetta project and is obtained through the Language Bank of Finland (Kielipankki). No personal data is collected directly from research participants in this part of the research.

The research participant is not obliged to provide the necessary personal data, participation is voluntary

13. Transfer/disclosure of the data outside the research group

No personal data from the Lahjoita Puhetta corpus will be transferred or disclosed outside the research group. The data will be processed only by the named research implementers (Section 5) on CSC's computing infrastructure in Finland.

14. Data transfer/disclosure outside the EU or the EEA

No personal data will be transferred or disclosed outside the EU or EEA. All data processing takes place on CSC's computing infrastructure located in Finland.

15. Automated decision-making and profiling

In this research, the processing of personal data is not used in any decisions concerning the participants nor in profiling. The audio recordings are used solely for training a speech recognition model and no individual-level analysis or evaluation of speakers is performed.

16. Personal data protection principles

A preliminary ethical assessment has been conducted on the research:

- Yes
 - Research ethics committee that provided the assent:
- No

The data that is to be processed in the information systems has been protected using the following:

- user ID password user registration access control (physical location)
- other method, please specify: Data is processed on a CSC supercomputer, which provides multi-factor authentication.

Pseudonymisation and anonymisation:

When conducting scientific research, it is necessary to store the research materials used, so that the results of the research can be verified and that the research materials that have already been collected can be used in further research and new scientific research

initiatives. The research materials will be anonymised or pseudonymised whenever possible. The research results will be published in such a way that no individual persons can be identified from the results. In special cases, such as when an artist is interviewed about their work, there may be reason to identify the person.

In this research:

- The materials will be anonymised (all identification data will be removed completely to ensure that the identification data cannot be reintroduced to the materials and that no new data can be linked to the materials)
- Direct identification data will be removed (pseudonymisation)
 - The identification data can be reintroduced, and new data can be linked to the materials
(key code or similar)
 - The materials are analysed using only indirect identifiers
- The materials contain only indirect or pseudonymous identifiers
- The materials will be analysed using direct identification data, because (specify justification for retaining direct identification data):

Safeguards for any personal data:

- Research plan
- Responsible person for the research, please specify:
- All personal data will be processed and disclosed only for research purposes and all the participants in the research agree to ensure that no data about a specific individual will be disclosed to any third parties
- A protection impact assessment has been conducted on the research (PIA)
- A data protection impact assessment has been conducted on the research (DPIA)

17. The rights of research participants and limitations to these rights

According to the General Data Protection Regulation (GDPR), data subjects have the right:

- to obtain information on the processing of their personal data, unless a specific exception is provided in the law
- to access their data
- to rectify any data concerning them
- to the erasure of their data (not applicable if the processing basis is compliance with a legal obligation or a task carried out in the public interest)
- to restrict the processing of their data
- to object to the processing of their data if the processing basis is public interest or legitimate interests
- to request that the personal data they have provided themselves is transferred directly from one system to another if the processing basis is consent
- to withdraw their consent
- the controller's notification obligation regarding rectification or erasure of personal data or restriction of processing
- to not be subject to automated decision-making (data subjects may allow automated decision-making with their consent)
- to notify the Office of the Data Protection Ombudsman if they believe that the processing of personal data that pertains to them has violated any data protection legislation in force.

Data subjects can exercise their rights by contacting the research's contact person or data protection officer. For more information on the rights of data subjects, contact the research's contact person and/or data protection officer.

If the processing of personal data in the research does not require the identification of the data subject without additional data and the controller is unable to identify the data subject, the right to access, rectify, erase and restrict the use of personal data, as well as any notification obligations and the right to transfer data, do not apply.

A derogation in the rights of research participants is necessary and justified if the research is meant for the purposes of the public interest and the rights of the research participant would be likely to prevent reaching these goals or hinder them greatly and these derogations are necessary to ensure these objectives.

- There are no derogations from the rights of data subjects in this research

This research is likely to derogate from the following rights specified in the EU General Data Protection Regulation:

- Data subjects' right to obtain information on the processing of personal data (the obligation to inform may be derogated from in scientific research according to article 14(5), point (b) of the General Data Protection Regulation, in which case a data protection impact assessment must be carried out and the research data must be made publicly available)
- The data subject's right to access any data concerning them
- The data subject's right to rectify any data concerning them
- The data subject's right to restrict the processing of personal data
- The data subject's right to object to the processing of personal data

Justified reason for the derogation from the rights of data subjects:

The following safeguards have been enacted in this research so that the derogation in the rights of the research participants can be made:

- The research announcement has been made publicly available and a data protection impact assessment has been carried out concerning the research (when derogating from informing the participants)
- The processing of personal data is based on the research plan
- The research has a responsible leader or group
- The personal data is used and disclosed only for the purposes of historical or scientific research or some other compatible purpose, and all parties involved in the research agree to ensure that no data pertaining to any specific individuals is disclosed to any third parties
- The research involves the processing of so-called sensitive materials and an impact assessment on this has been made and submitted to the Office of the Data Protection Ombudsman before the processing is set to begin (when derogating from the right of access, rectification, restriction and objection)