Plan Overview

A Data Management Plan created using DMPTuuli

Title: Predictive Processing Approach to Modelling Prosodic Hierarchy for Speech Synthesis

Creator: Juraj Simko

Principal Investigator: Juraj Simko

Data Manager: Juraj Simko

Project Administrator: Juraj Simko

Affiliation: University of Helsinki

Funder: The Research Council of Finland (former The Academy of Finland)

Template: Academy of Finland data management plan guidelines (2021-2023)

Project abstract:

Speech communication profoundly relies on hierarchically organized melodic, rhythmical, and temporal indices to the information relevant in the given situation and context, that fall under common descriptor of prosody. Speech prosody maintain context-dependent cohesion and coherence of speech communication by encoding a wide array of linguistic, paralinguistic and extralinguistic features in a parallel and interconnected fashion.

Speech synthesis technology has recently achieved almost human-like quality of synthesized output, primarily for prosodically neutral utterances. Despite this progress, the state-of-the-art systems fall short of reproducing prosodic characteristics ubiquitous in human speech. This shortcoming arises primarily from architectural and conceptual decisions failing to recognize the hierarchical nature of prosody as a crucial feature of spoken interaction.

Here we propose a novel speech prosody modelling architecture, and its implementation within a speech synthesis system. The architecture explicitly uses hierarchically encoded prosodic information and is an instantiation of the highly influential Predictive Processing cognitive modelling paradigm within the domain of speech communication. Unifying the treatment of speech perception and production within a single system allows for quantification of high-level parameters capturing aspects of rudimentary situationawareness, including a conversational setting. The use of an ecologically and cognitively grounded model in connection with a linguistically valid descriptive framework is believed to provide more explanatory power than either entirely data-oriented or purely linguistically motivated treatments of prosody in use today.

A parallel objective of the project is to contribute to our theoretical understanding of features and wide-range interdependencies that shape speech prosody and give rise to its context-sensitive realization readily achieved by humans. For this task, the developed deep learning platform will serve as a complex statistical model capturing representations that guide our communicative behaviour in terms of interactions between various hierarchically organized prosodic units. The development and validation of both technological and theoretical platforms will be assisted by the wide expertise in speech technology and prosodic analysis provided by the host research team and our academic partners.

ID: 22273

Start date: 01-09-2023

End date: 31-08-2027

Last modified: 03-07-2023

Grant number / URL: 357262

Predictive Processing Approach to Modelling Prosodic Hierarchy for Speech Synthesis

1. General description of data

1.1 What kinds of data is your research based on? What data will be collected, produced or reused? What file formats will the data be in? Additionally, give a rough estimate of the size of the data produced/collected.

The prime object of research for this project is human speech. Consequently, most of the data used will be speech recordings (usually in wav format).

The main data source used by the project will be the corpus FinSyn that has been collected by the host research group in close collaboration with Fin-CLARIN/CLARIAH research infrastructure. The corpus consists of speech material from three voice talents (2 speakers of Finnish, and one of Finland Swedish), approximately 25 hours of speech from each. The corpus has been processed and the resulting material (approx. 60 GB in total) is stored primarily in multiple hard drives of the host group computers and at the allocated storage place (scratch) at CSC computational infrastructure.

The CLARIAH network is working on necessary technical and legal steps that would lead to a release of the corpus (including appropriate annotations) for public use.

The project will also make use of publicly available speech corpora, primarily LJSpeech (https://keithito.com/LJ-Speech-Dataset/); this database (approx.20 GB) is a part of LibriVOX project, and its usage will be guided by the Public Domain licence.

Processing these databases for the purposes of this project will generate annotations of each dataset (typically plain-text annotations of total approximate size of approx. 100 Mb). These will be downloaded/stored locally (at the researcher's computer) and are covered by the Public Domain licences.

There will be a relatively small amount of data collected within the project (from human subjects) namely subjective evaluation rankings (data consisting of continuous valued rankings). These rankings will be obtained at a letter stage of the project, through evaluations on the basis of existing speech material taken from available speech databases using a suitable crowdsourcing platform. These data are not assumed to contain any sensitive information, such as subject identifiers. The collected data will be stored in plain text (.txt) format, or hdf5/pkl for python binary data objects. The size of the collected data will be small. Specifically, this will be on the scale of few megabytes both in the case of .txt documents and in the case of a proprietary format (e.g., .mat).

There might be also a small amount of other behavioural data collected at the latter stages of the project. At this stage, it is not possible to be more specific regarding their nature or amount. When we know more about the data needed in the future, the DMP will be updated accordingly.

1.2 How will the consistency and quality of data be controlled?

If collected, the quality and consistency behavioural data will be ensured by following a standard protocols. Specifically, before the data collection takes place, and for each subject, it is made sure that the setup and briefing procedure (briefing is designed to provide information quickly and effectively about the experiment) are exactly the same for each participant (following an experimental protocol designed for the experiment). The quality of the data is ensured by keeping notes during the data collection process, through the debriefing process, and through manual check to inspect that all data are in good order. For each experiment, a documentation folder is kept with information regarding the data collection process.

For the existing data, no lossy conversion steps during the project execution are envisaged at this stage.

2. Ethical and legal compliance

2.1 What legal issues are related to your data management? (For example, GDPR and other legislation affecting data processing.)

The University of Helsinki will act as the controller of personal data. The data gathering and storing activity will be carried out whilst respecting the following principles in addition to the legal obligations in Finland and the ethical guidelines of our institutions:

1. Status of personal data: the activities take into account the principle that personal data form part of the personality of the individual, and must not be treated as mere objects of commercial transaction. All personal data will be removed.
2. Confidentiality / privacy: all data of collaborators or users, information on measurements of variables associated with end-users and their behaviour and practices will be kept in secure computer files. During the project the access to personal data is limited to researchers in charge of data acquisition and processing and to those third parties who can demonstrate a legitimate use.
3. Informed consent of the individual will be required for the collection and release of personal data.
4. Principle of legitimate purpose: the principle of a strict relationship between the collection and processing of personal data and handling and the legitimate purpose to which those data are used will guide the collection of data. The ground forming the legitimate purpose is public interest.
5. Security: the security of computer systems is an ethical imperative to ensure the respect for human rights and freedoms of the individual, in particular the confidentiality of data and the reliability of computer systems used for research. During the duration of the project, security of computer systems will be rigorously maintained.
6. The project does not include any activities or produce any results raising security issues and will not use or produce classified information as background or results.

2.2 How will you manage the rights of the data you use, produce and share?

An Undertaking on Transfer of Rights agreement will be signed with the University of Helsinki, granting the UH parallel rights to the data. All team members that will have access to research data will be employees of UH, and will have equal access to the data.

The collected data will be made available for reuse (following the ethical principles outlined in 2.1). Information sheets and consent forms will inform participants of the prospective use of the data and if they agree to grant the right for sharing the collected data. The appropriate consent forms and Privacy Notes will be drafted in collaboration with University of Helsinki Data Protection Officer.

3. Documentation and metadata

3. How will you document your data in order to make the data findable, accessible, interoperable and re-usable for you and others? What kind of metadata standards, README files or other documentation will you use to help others to understand and use your data?

The data collection process will be prepared and detailed in a document before the data collection takes place. This includes details of the material to be collected and forms that make sure that the process can be tracked appropriately. The standard process for this is to create a file with the data documentation and a separate file with the data collected. All data will be stored within a project folder and then in an experiment-specific folder.

The public distribution and reusability of the data will be coordinated with FIN-CLARIN/CLARIAH infrastructure and realized through Kielipankki portal. The project will use metadata standards recommended by Kielipankki, and standard in the field of speech synthesis.

4. Storage and backup during the research project

4.1 Where will your data be stored, and how will the data be backed up?

The immediate research data will be stored in the University of Helsinki centralised IT infrastructure. The data will be stored in the password-protected Home Folder (Z: drive) as well as the newly set up group folder (P: drive). Both are protected by automated hourly backups allowing for the data to be restored in case the primary storage would fail.

A portion of the data, namely a research code, will be shared and backed up using Github hosting services.

4.2 Who will be responsible for controlling access to your data, and how will secured access be controlled?

The primary responsibility for controlling access to the data falls with the project PI. All data will be stored anonymously in the password- protected IT infrastructure managed by University of Helsinki.

5. Opening, publishing and archiving the data after the research project

5.1 What part of the data can be made openly available or published? Where and when will the data, or its metadata, be made available?

The project will make an effort to make all the data obtained publicly available.

The research code will be stored and distributed through Github hosting services under MIT Licence as recommended by the UH.

The speech data (including relevant metadata and annotations) will be distributed in collaboration with Fin_CLARIN through Kielipankki portal.

The human behavioural data will be made available through appropriate channels suggested/provided by journals where the results will be published. Throughout the project we will aim at publications in open-access journals.

5.2 Where will data with long-term value be archived, and for how long?

The speech data will be stored and made available for potential reuse (through FIN-CLARIN infrastructure) as long as they are technologically relevant.

The human behaviour data will be archived for at least a a verification period of 15 years and subsequently destroyed; consent for this will be sought from participants.

6. Data management responsibilities and resources

6.1 Who (for example role, position, and institution) will be responsible for data management?

The data management responsibilities for this project will be allocated to the principal investigator of the project (speech data, annotations, updating the DMP), and to the postdoctoral researcher (TBH; code). The appropriate responsibilities regarding data management will be allocated to members of the project who will be hired during its execution.

6.2 What resources will be required for your data management procedures to ensure that the data can be opened and preserved according to FAIR principles (Findable, Accessible, Interoperable, Re-usable)?

The data preparation should not incur extra overheads beyond what is already estimated in the project. The preservation and sharing of the data will require the allocation of storage space and limited computing resources that should not incur significant extra costs. This is due to the relatively limited storage space that is necessary in order to preserve such files (text-based data). Specifically:
(1) For data documentation a time estimate of approximately one week throughout the project will be allocated.
(2) For data cleaning (anonymizing data) a time estimate of approximately one day per experiment will be allocated.
(3) Data will be stored in the University of Helsinki centralised IT infrastructure. For long-term storage, the FIN-CLARIN infrastructure will be used.