SFB 1287 Limits of Variability in Language
Cognitive, Computational, and Grammatical Aspects

DATA MANAGEMENT PLAN
SFB 1287 / 2025 / PHASE 2

PROJECT B01

Table of contents

General Information

Overview

Project number

B01

Name of Experiment / Acronym / Number

WP2: Behavioral Tasks and Cognitive Test Battery

Involved persons

Andrea Hofmann, Iana Palacheva, Isabell Wartenburger

PI or responsible person (head of the study)

Andrea Hofmann, Isabell Wartenburger

Subject area

Psycholinguistics

Method / Type of data

General participant information: age, gender, exclusion of speech, voice and language disorders, exclusion of neurological disorders, exclusion of hearing impairment, language background, educational level, occupation; musicality questionnaire General tests: audiometry, handedness index Perturbation Task: real-time voice pitch perturbation of auditory feedback; data: audio recordings Auditory acuity: just-noticeable-difference adaptive staircase task (discrimination or identification); data: accuracy data, difference score Repetition task: stimulus imitation; data: audio recordings, rating data Perception task: gating paradigm; data: accuracy data, reaction times Cognitive test battery: test scores, phase locking values; data: audio and tapping recordings, accuracy data, reaction times

Participants (of the study)

Healthy participants German natives Age range: 18-45 years No neurological disorders No speech, voice or hearing impairment

Short description (of the study)

In a series of different experimental tasks, we will collect a range of behavioral measures for the same sample of at least 50 healthy participants. All experimental tasks will make use of the same stimuli which are name coordinates consisting of three different two-syllable names. Each task will shed light on a different aspect of producing and perceiving the same stimuli, thus providing a comprehensive overview of individual performances. Moreover, we are introducing an extensive cognitive test battery – which taken together should enable us to find potential co-variation profiles between perception and production and relate them to underlying general cognitive abilities.

Comments (optional)

Data Management Requirements

Are there requirements regarding the data management from your scholarly / scientific community?

yes

If yes, what are the requirements?

- *DFG Guidelines on the Handling of Research Data*

- *„Handlungsempfehlungen zum Umgang mit Forschungsdaten" University of Potsdam*

- *„Technische und organisatorische Maßnahmen" (TOM) gemäß Art. 32 Abs. 1 DSGVO.*

Financial Support

Who is funding the project?

DFG - Deutsche Forschungsgemeinschaft e.V. (German Research Foundation) - https://www.dfg.de/en/

In which funding line and / or which funding program is the project funded?

Collaborative Research Centre 1287 - Project number 317633480

Dataset Information

Data Origin

Is the dataset being created or re-used?

created

If re-used, who created the dataset and under which address, PID or URL is the data set available?

-

Data Collection

When does data collection start? (approximately / tentatively)

09.01.2023

When does data collection end? (approximately / tentatively)

31.08.2023

Data Handling

Where is the dataset stored during the project?

- *Laptop*

- *Computer Acoustic Laboratory*

- *SFB Server*

- *GitHub*

If data is stored on lab or personal computers, please describe the backup strategy.

Raw data is backed up on the SFB file server. After finalizing data collection and analyses, the analyzed data will also be stored on the SFB file server.

Which file formats are used?

*.csv; *.pdf; *.txt; *.docx;*.py; *.rmd; *.md; *.json; *.wav; *.TextGrid; *.html

Which measures of quality assurance are taken for this dataset?

Use of a fixed study-protocol per task; use of the same data base for all basic experimental information, like time, participant-ID, comments for test session etc., feedback (same *.csv file for whole work-package); pre-fixed pre-processing and analysis scripts; use of scripts to check if data looks consistent and variables are coded correctly; use common file formats; very good documentation; code review; use of repositories as version control.

Data Analysis

When does data analysis start? (approximately / tentatively)

27.02.2023

When does data analysis end? (approximately / tentatively)

30.09.2023

Data Reuse

Which individuals, groups or institutions could be interested in re-using this dataset? What consequences does the reuse potential have for the provision of the data later?

All anonymized data (scores, measures, descriptives), experiment files and code will be made available - thus the whole procedure and analysis is reproducable. The only type of data not publicly available is the participants wav-files. All other experimental files - stimuli, model recordings etc. will also be publicla avaible. Every researcher that is interested in the project can repeat the process, adapt the scripts and use the stimuli.

Legal and Ethics

Personal Data

Does this dataset contain personal data?

no

Are these data anonymised?

Property Rights

Does the project use and/or produce data that is protected by intellectual or industrial property rights?

yes

If yes, please explain which data protected by intellectual or industrial property rights?

- *certain Python, R / R-Studio*

- *Matlab packages*

- *Audapter (manipulation software used, adapted code will be made available)*

- *cognitive test battery tasks (WAIS)*

- *speech-synchronization task (Assaneo et al.)*

Publication

Publishing or Sharing Data

Will this dataset be published or shared?

yes

If yes, the principal investigator of the study ensured that the consent form / subject information sheets support publishing of the data?

yes

If yes, under which terms of use or license will the dataset be published or shared?

Creative Commons Namensnennung (CC-BY)

If yes, when will the data be published?

Recommended procedure: Upload data and obtain digital identifier (e.g., DOI, OSF link) when submitting the first paper; thus, you can cite the data in the paper. If necessary, restrict public access (embargo) until last paper published (max. 2 years).

If no, please explain why not. Please differentiate between legal and contractual reasons and voluntary restrictions.

-

Storage and Long-Term Preservation

Archive

Does this dataset have to be preserved for long-term?

yes

How long does the data need to be stored?

The DFG expects primary data that is the basis of a publication to be stored in the researcher's own institution or an appropriate nationwide infrastructure long-term (for at least 10 years).

What are the reasons this dataset must be preserved for the long-term?

- *Use in a publication / Evidence of good scientific practice*

- *Reuse (if anonymizable data) in subsequent projects or by others*

- *Legal obligations*

- *Documentation because it is socially relevant*

- *Self-commitment*

- *Evidence of good scientific practice*

- *DFG requirements*

Where will the data (including metadata, documentation, and relevant code) be stored or archived after the end of the project?

- *SFB 1287 File-Server*

- *Research Data Server from Project IN-FDM-BB (a.t.m. not available)*

- *GitHub*

- *ResearchBox*