DATA MANAGEMENT PLAN

CRC 1287 | 2021 | Phase 1

Project B01

TABLE OF CONTENTS

GENERAL INFORMATION

Overview

Project number

B01

Name of Experiment / Acronym / Number

WP1 - Coordinate Sequences

Involved persons

Huttenlauch, de Beer, Hansen, Aivazian

PI or responsible person (head of the study)

Isabell Wartenburger, Sandra Hanne

Subject area

- Production data of prosodic boundary cues

- Perception data of prosodic boundary cues

Method / Type of data

- Elicited productions, wav files plus information about speakers

- Perception (2AFC, Gating, Rating)

Participants (of the study)

- Young adults (students, 19 34 years old)

- Elderly adults (61 80 years old)

Short description (of the study)

- Production of coordinate name sequences ("Moni und Lilli und Manu" vs. "(Moni und Lilli) und Manu") to a virtual interlocutor in 5 contexts.

- Contexts consist of four different female interlocut ors: a young adult (C1 adult), a child (C2 child), an elderly adult (C3 elderly), a young non native adult speaker of German (C4 non native), and a noisy environment (the adult of C1 with white background noise, C5 noise).

- Perception check (see Huttenlauch et al., 2021) Gating study (MSC Thesis Marie Hansen) 2AFC Drift Diffusion study (MSC Thesis Valentina Aivazian)

Comments (optional)

- WP1_Produktion: Experiment 1 (young), Experiment 1b (elderly)

- Master Thesis Valentina Aivazian (Exp_DriftDiffusion)

- Master Thesis Marie Hansen (Exp_Gating)

- Experiment 1 (young adult): n=60 productions per participant

- Experiment 1b (elderly adult): n=120 productions: double amount of productions (1 repetition per stimulus)

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 317633492

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)

2018

When does data collection end? (approximately / tentatively)

2021

Data Handling

Where is the dataset stored during the project?

- CRC file server

- Box.UP university cloud

- researcher's computer

- computer in the laboratory

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

All data (raw and finally analyzed data) are stored on the CRC file server. The same holds for the analysis scrip ts and code.

Which file formats are used?

A folder containing:

- Production data (raw and processed): *.wav files, praat TextGrids

- Presentation folder including script, stimuli (videos of interlocutors,2 videos per context), Logfiles

- Instructions for participants

- Description of procedure

- analysis scripts for praat and R

- Participant information

- Empathy Quotient scores

- Scores from working memory test

- Hearing screening

Analysis steps:

- praat scripts can be run in praat, version 6.0.32

- praat scripts need to be run first extracted data will be saved into a table (txt file

- replace all ""?"" by ""NA""

- txt file is read in R

- further data processing and analyses in R script

Audio recordings:

- Recordings took place in a sound attenuated booth in the acoustics laboratory of the University of Potsdam via an Alesis io12 interface.

- Speakers wore a headset HSC 271 (AKG Acoustics by Harman, www.akg.com) with over ear headphones and a condenser microphone and were seated in front of a wide screen monitor with 1920 x 1200 resolution and saw the stimuli in Arial font of size 50. The experiment was run from a Dell laptop standing outside of the recording booth using the software Presentation (Neurobehavioural Systems https://www.neurobs.com/; Version 20.1)

Which measures of quality assurance are taken for this dataset?

- Data are analyzed for outliers and biases in the responses (by the person who analyzed the data).

- Wav files are annotated.

- Production data (young): A perception check encompassed 32 listeners, who had not taken part in the production experiment (sex: 22 female, 10 male; age range: 18 41 years, mean: 24.25 years, SD: 5.8). They were recruited at the University of Potsdam and were reimbursed or received course credits. Each listener judged 267 productions. The majority of all productions was perceived with the intended grouping (689 out of 960). The mean ratio of congruent rates in the subse t which was rated by all participants was 0.84 (SD: 0.29), for the remaining sets of productions the mean ratio was 0.94 (SD: 0.148). 38 productions were excluded since their ratios fell more than 2 SD below the mean (3 productions of the subset and 35 of the remaining items).

- Production data (elderly): For the productions of the older speakers, the perception check was transferred onto Open Sesame (Mathôt et al. 2012) and was run as a web based study on Jatos (Lange et al. 2015) in individual sessions. Dat a of 49 listeners (29 female, 9 male, 11 other/no information; age range: 18 63, mean: 24.63 years, SD: 6.3) were analysed. Another five listeners took part in the study, but had to be excluded due to technical problems. 66 productions (4%) fell below the threshold and were excluded from further analyses.

Data Analysis

When does data analysis start? (approximately / tentatively)

2018

When does data analysis end? (approximately / tentatively)

2021

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?

Non anonymizable data:

- Raw data (audio r ecordings from participants) are non anonymizable.

- Video stimuli and audio stimuli of experimental interlocutors (1 interlocutor per context): 2 videos + 1 audio file per context (5 contexts)

- Reuse of raw data and stimuli by other researchers/scientists is NOT possible because the identity of the participants/recorded person could be revealed (personal data). Raw data and stimuli cannot be shared.

Anonymizable data:

- TextGrids (praat) and extracted measures (duration, f0) could be shared.

- Test scores on working memory assessments (digit span, block span, digit symbol substitution, Mehrfachwahlwortschatztest, picture span (only young adults), MOCA (only elderly adults)), hearing test (single sound audiogram), Empathy Quotient, handedness (Edinburgh Handedn ess Inventory)

LEGAL AND ETHICS

Personal Data

Does this dataset contain personal data?

yes

If yes, are these data anonymised?

yes

Property Rights

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

no

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

-

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?

CC0 (Public Domain)

If yes, when will the data be published?

when the paper is published

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?

- CRC 1287 File-Server

- OSF

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

coordinated by:

Universität Potsdam

https://www.uni-potsdam.de

funded by:

DFG Deutsche Forschungsgemeinschaft

https://www.dfg.de/

in cooperation with:

RUHR UNIVERSITÄT BOCHUM RUB

https://www.ruhr-uni-bochum.de