**ReadMe document for Data on Hui & Godfroid (in press)**

**Last updated: April 26, 2020**

**Content**

1. Citation information
2. Directory of the data packet
3. Flow chart of data processing
4. Codebook for cleaned data sets
5. **Citation information**

|  |
| --- |
| **Please cite the paper if substantive content is discussed:**  Hui, B. & Godfroid, A. (in press). Testing the role of processing speed and automaticity in second language listening. *Applied Psycholinguistics*.  **Please cite the data if you use them for any secondary analysis:**  Hui, B. & Godfroid, A. (2020). Data and codebook for Applied Psycholinguistics article: Testing the role of processing speed and automaticity in second language listening [Data set]. Open Science Framework. <http://doi.org/10.17605/OSF.IO/35VFX> |

1. **Directory of the data packet**

* **DataPacket\_Version1.0**
  + **Combined Analysis (overall, combined analysis)**
    - **Combined\_Script (R script for data analysis for the sentence construction task)**
    - **combined\_clean1 & combined\_clean2 (two cleaned data sets)**
    - **cor\_mat (correlation matrix)**
    - **CombinedAnalysis.RData (R Data file for final models)**
  + **Construction (data and analysis for the sentence construction task)**
    - **CleanedData (folder with 2 cleaned data sets)**
    - **RawData (folder with raw data in .txt format from Superlab)**
    - **con\_final (data set for combined analysis)**
    - **Construction\_Script (R script for data analysis for the sentence construction task)**
    - **con\_StiDur (data set containing stimulus durations)**
  + **Listening (data and analysis for the listening test)**
    - **listening\_final (data set for combined analysis)**
  + **Verification (data and analysis for the sentence verification task)**
    - **CleanedData (folder with 2 cleaned data sets)**
    - **RawData (folder with raw data in .txt format from Superlab)**
    - **ver\_final (data set for combined analysis)**
    - **Ver\_script (R script for data analysis for the sentence verification task)**
    - **ver\_StiDur (data set containing stimulus durations)**
  + **YesNo (data and analysis for the Yes-No RT Test)**
    - **CleanedData (folder with 2 cleaned data sets)**
    - **RawData (folder with raw data in .txt format from Superlab)**
    - **yesno\_final (data set for combined analysis)**
    - **YesNo\_script (R script for data analysis for the Yes-No RT test)**
    - **yesno\_StiDur (data set containing stimulus durations)**

1. **Flow charts of data processing**

*Overall*

Analysis for the Yes-No RT Test

Analysis for the sentence verification task

Analysis for the sentence construction task

Analysis for the listening test

Combined analysis

*Sentence Construction Test*

|  |
| --- |
| **Step 1 - Reading and Merging Raw Data** |

**↓**

|  |
| --- |
| **Exported Con\_clean1.csv** |

**↓**

|  |
| --- |
| **Step 2 - Cleaning Data**  **Read data Con\_clean1.csv**  **Filter relevant rows (test trials)**  **Add group info (monolinguals vs bilinguals)**  **Extract item no.**  **Identify and exclude unsatisfactory items** |

**↓**

|  |
| --- |
| **Exported Con\_clean2.csv** |

**↓**

|  |
| --- |
| **Step 3 - Main Analysis**  **Read data Con\_clean2.csv**  **Compute accuracy**  **Compute mean rection time**  **Compute CV**  **Compute reliability**  **Generate CV-RT Correlation and descriptives** |

**↓**

|  |
| --- |
| **Exported con\_final.csv** |

*Sentence Verification Test*

|  |
| --- |
| **Step 1 - Reading and Merging Raw Data** |

**↓**

|  |
| --- |
| **Exported Ver\_clean1.csv** |

**↓**

|  |
| --- |
| **Step 2 - Cleaning Data**  **Read data Ver\_clean1.csv**  **Filter relevant rows (test trials)**  **Extract item no. from event log**  **Add group info (monolinguals vs bilinguals)** |

**↓**

|  |
| --- |
| **Exported Ver\_clean2.csv** |

**↓**

|  |
| --- |
| **Step 3 - Main Analysis**  **Read data Ver\_clean2.csv**  **Identify and exclude unsatisfactory items**  **Compute accuracy**  **Compute mean rection time**  **Compute CV**  **Compute reliability**  **Generate CV-RT Correlation and descriptives** |

**↓**

|  |
| --- |
| **Exported ver\_final.csv** |

*Yes-No RT Test*

|  |
| --- |
| **Step 1 - Reading and Merging Raw Data** |

**↓**

|  |
| --- |
| **Exported YesNo\_clean1.csv** |

**↓**

|  |
| --- |
| **Step 2 - Cleaning Data**  **Read data YesNo\_clean1.csv**  **Filter relevant rows (test trials)**  **Code word vs nonword trials**  **Extract item no. from event log**  **Add group info (monolinguals vs bilinguals)** |

**↓**

|  |
| --- |
| **Exported YesNo\_clean2.csv** |

**↓**

|  |
| --- |
| **Step 3 - Main Analysis**  **Read data YesNo\_clean2.csv**  **Identify and exclude unsatisfactory items**  **Compute accuracy**  **Compute mean rection time**  **Compute CV**  **Compute reliability**  **Generate CV-RT Correlations and descriptives** |

**↓**

|  |
| --- |
| **Exported yesno\_final.csv** |

*Combined analysis*

|  |
| --- |
| **Step 1 - Loading data sets from individual tasks** |

**↓**

|  |
| --- |
| **Exported combined\_clean1.csv** |

**↓**

|  |
| --- |
| **Step 2 – Data processing**  **inspect correlation matrix**  **transformation of variables** |

**↓**

|  |
| --- |
| **Exported combined\_clean2.csv** |

**↓**

|  |
| --- |
| **Step 3 - Main Analysis**  **Regression (stage 1)**  **Regression (stage 2)**  **Mediation analysis** |

1. **Codebook for cleaned data sets**

**List of cleaned data sets**

*Yes-No RT Test*

* YesNo\_clean1.csv
* YesNo\_clean2.csv

*Sentence Construction Task*

* Con\_clean1.csv
* Con\_clean2.csv

*Sentence Verification Task*

* Ver\_clean1.csv
* Ver\_clean2.csv

*Combined Analysis*

* combined\_clean1.csv
* combined \_clean2.csv

**List of summary data sets for each task and for combined analysis**

* listening\_final.csv
* yesno\_final.csv
* ver\_final.csv
* con\_final.csv

**List of other data sets**

* yesno\_StiDur.csv
* ver\_StiDur.csv
* con\_StiDur.csv

**Code books**

***Listening Test:*** *listening\_final.csv*

|  |  |  |
| --- | --- | --- |
| **Column** | **Variable Name** | **Description** |
| A | sub | Participant ID |
| B - AO | items 1 – 40 | Accuracy for each item (1 = Correct; 0 = Incorrect) |
| AP | listen | Raw overall accuracy (out of 40) |
| AQ | listenacc | Overall accuracy in the listening test in percentage |

***Sentence Construction Task:*** *Con\_clean1.csv*

|  |  |  |
| --- | --- | --- |
| **Column** | **Variable Name** | **Description** |
| A | sub | Participant ID |
| B | trial | Trial logs by Super lab |
| C | event | Event logs by Superlab |
| D | acc | Response Accuracy (C= Correct; E=Error made) |
| E | rt | Reaction Time (RT) in millisecond |

***Sentence Construction Task:*** *Con\_clean2.csv*

|  |  |  |
| --- | --- | --- |
| **Column** | **Variable Name** | **Description** |
| A | sub | Participant ID |
| B | acc | Response Accuracy (C= Correct; E=Error made) |
| C | rt | Reaction Time (RT) in millisecond |
| D | group | Participant Group (mono = Native Speakers; bi = Nonnative speakers) |
| E | item | Item identification |
| F | duration | Stimulus duration |
| G | score | Score for the trial (corresponds to Column B) (1 for correct; 0 for incorrect) |

***Sentence Construction Task:*** *con\_final.csv*

|  |  |  |
| --- | --- | --- |
| **Column** | **Variable Name** | **Description** |
| A | sub | Participant ID |
| B | synacc | Accuracy for the Sentence Construction Task |
| C | syncv | CV for the Sentence Construction Task |
| D | synrt | Mean reaction time in millisecond for the Sentence Construction Task |

***Sentence Construction Task:*** *con\_StiDur.csv*

|  |  |  |
| --- | --- | --- |
| **Column** | **Variable Name** | **Description** |
| A | trial | Trial logs by Super lab |
| B | item | Item identification |
| C | duration | Duration of the stimulus in millisecond |

***Sentence Verification Task:*** *Ver\_clean1.csv*

|  |  |  |
| --- | --- | --- |
| **Column** | **Variable Name** | **Description** |
| A | sub | Participant ID |
| B | event | Event logs by Superlab |
| C | acc | Response Accuracy (C= Correct; E=Error made) |
| D | rt | Reaction Time (RT) in millisecond |

***Sentence Verification Task:*** *Ver\_clean2.csv*

|  |  |  |
| --- | --- | --- |
| **Column** | **Variable Name** | **Description** |
| A | sub | Participant ID |
| B | event | Event logs by Superlab |
| C | acc | Response Accuracy (C= Correct; E=Error made) |
| D | rt | Reaction Time (RT) in millisecond |
| E | plau | Plausibility (sensible; nonsense) |
| F | item | Item identification |
| G | group | Participant Group (mono = Native Speakers; bi = Nonnative speakers) |

***Sentence Verification Task:*** *ver\_final.csv*

|  |  |  |
| --- | --- | --- |
| **Column** | **Variable Name** | **Description** |
| A | sub | Participant ID |
| B | proacc | Accuracy for the Sentence Verification Task |
| C | procv | CV for the Sentence Verification Task |
| D | prort | Mean reaction time in millisecond for the Sentence Verification Task |

***Sentence Verification Task:*** *ver\_StiDur.csv*

|  |  |  |
| --- | --- | --- |
| **Column** | **Variable Name** | **Description** |
| A | item | Item identification |
| B | duration | Duration of the stimulus in millisecond |

***Yes-No RT Test:*** *YesNo\_clean1.csv*

|  |  |  |
| --- | --- | --- |
| **Column** | **Variable Name** | **Description** |
| A | sub | Participant ID |
| B | event | Event logs by Superlab |
| C | acc | Response Accuracy (C= Correct; E=Error made) |
| D | rt | Reaction Time (RT) in millisecond |

***Yes-No RT Test:*** *YesNo\_clean2.csv*

|  |  |  |
| --- | --- | --- |
| **Column** | **Variable Name** | **Description** |
| A | sub | Participant ID |
| B | acc | Response Accuracy (C= Correct; E=Error made) |
| C | rt | Reaction Time (RT) in millisecond |
| D | word | Word Type (word = real word; non = nonword) |
| E | level | Levels in the original paper-based Yes-No Test (3 – 5) |
| F | item | Item identification |
| G | group | Participant Group (mono = Native Speakers; bi = Nonnative speakers) |
| H | score | Score for the trial (corresponds to Column B) (1 for correct; 0 for incorrect) |

***Yes-No RT Test:*** *yesno\_final.csv*

|  |  |  |
| --- | --- | --- |
| **Column** | **Variable Name** | **Description** |
| A | sub | Participant ID |
| B | lexacc | Accuracy for the Yes-No RT test |
| C | lexcv | CV for the Yes-No RT test |
| D | lexrt | Mean reaction time in millisecond for the Yes-No RT test |

***Yes-No RT Test:*** *yesno\_StiDur.csv*

|  |  |  |
| --- | --- | --- |
| **Column** | **Variable Name** | **Description** |
| A | item | Item identification |
| B | duration | Duration of the stimulus in millisecond |

**Combined Analysis:** combined\_clean1 & combined\_clean2 (after transformation)

|  |  |  |
| --- | --- | --- |
| **Column** | **Variable Name** | **Description** |
| A | sub | Participant ID |
| B | listenacc | Overall accuracy in the listening test in percentage |
| C | lexacc | Accuracy for the Yes-No RT test |
| D | lexrt | Mean reaction time in millisecond for the Yes-No RT test |
| E | lexcv | CV for the Yes-No RT test |
| F | synacc | Accuracy for the Sentence Construction Task |
| G | synrt | Mean reaction time in millisecond for the Sentence Construction Task |
| H | syncv | CV for the Sentence Construction Task |
| I | proacc | Accuracy for the Sentence Verification Task |
| J | prort | Mean reaction time in millisecond for the Sentence Verification Task |
| K | procv | CV for the Sentence Verification Task |