



# OPEN SpEx: a German-language dataset of speech and executive function performance

Julia A. Camilleri<sup>1,2,7</sup>✉, Julia Volkening<sup>1,2,3,7</sup>, Stefan Heim<sup>4,5,6</sup>, Lisa N. Mochalski<sup>1,2</sup>, Hannah Neufeld<sup>1,5</sup>, Natalie Schlothauer<sup>1,2</sup>, Gianna Kuhles<sup>1,2</sup>, Simon B. Eickhoff<sup>1,2</sup> & Susanne Weis<sup>1,2</sup>

This work presents data from 148 German native speakers (20–55 years of age), who completed several speaking tasks, ranging from formal tests such as word production tests to more ecologically valid spontaneous tasks that were designed to mimic natural speech. This speech data is supplemented by performance measures on several standardised, computer-based executive functioning (EF) tests covering domains of working-memory, cognitive flexibility, inhibition, and attention. The speech and EF data are further complemented by a rich collection of demographic data that documents education level, family status, and physical and psychological well-being. Additionally, the dataset includes information of the participants' hormone levels (cortisol, progesterone, oestradiol, and testosterone) at the time of testing. This dataset is thus a carefully curated, expansive collection of data that spans over different EF domains and includes both formal speaking tests as well as spontaneous speaking tasks, supplemented by valuable phenotypical information. This will thus provide the unique opportunity to perform a variety of analyses in the context of speech, EF, and inter-individual differences, and to our knowledge is the first of its kind in the German language. We refer to this dataset as SpEx since it combines speech and executive functioning data. Researchers interested in conducting exploratory or hypothesis-driven analyses in the field of individual differences in language and executive functioning, are encouraged to request access to this resource. Applicants will then be provided with an encrypted version of the data which can be downloaded.

Research in the field of executive functioning (EF) and speech has suggested a strong relationship between the two, with studies implying the former to be a basic requirement for the latter<sup>1–5</sup>. Consistent with this relationship, clinical studies have shown an association between executive function impairment and various communication disorders including aphasia and language pragmatic disturbances<sup>6</sup>. Such communication disorders have been shown to result in symptoms across different levels of language including, but not limited to, processes involving lexicon, semantics, syntax, phonology, and prosody<sup>7</sup>.

Although the presence of a general relationship between language and cognitive performance is known, further studies can help to improve the understanding of the nuances of speech and their specific relationship with different aspects of EF. In the clinical context such analyses can be potentially beneficial for the identification of speech biomarkers for specific psychiatric disorders<sup>8–11</sup>.

Understanding the relationship between EF and speech inherently relies on measuring performance of both domains in a standardised manner. Throughout the years, several neuropsychological tests have been designed, with the primary purpose of capturing different executive abilities. Traditionally, such tests have been performed using paper and pencil, making them time-consuming, while requiring verbal administration and interpretation

<sup>1</sup>Institute of Neuroscience and Medicine (INM-7 Brain and Behaviour), Forschungszentrum Jülich, Wilhelm-Johnen-Str., 52428 Jülich, Germany. <sup>2</sup>Institute of Systems Neuroscience, Heinrich-Heine University, Moorenstr. 5, 40225 Düsseldorf, Germany. <sup>3</sup>PeakProfiling GmbH, Eschenallee 36, 14050 Berlin, Germany. <sup>4</sup>Institute of Neuroscience and Medicine (INM-1 Structural and Functional Organisation of the Brain), Forschungszentrum Jülich, Wilhelm-Johnen-Str., 52428 Jülich, Germany. <sup>5</sup>Department of Psychiatry, Psychotherapy and Psychosomatics, Medical Faculty, RWTH Aachen University, Pauwelsstraße 30, 52074 Aachen, Germany. <sup>6</sup>Department of Neurology, Medical Faculty, RWTH Aachen University, Pauwelsstraße 30, 52074 Aachen, Germany. <sup>7</sup>These authors contributed equally: Julia A. Camilleri and Julia Volkening. ✉email: j.camilleri@fz-juelich.de

by trained administrators. Furthermore, standard pen-and-paper tests rely on manual scoring which could introduce errors and may thus lack sensitivity and specificity. Recent technological advancements have brought about the development of computerised versions of these tests which include automated scoring while also increasing the ease of administration. Studies have demonstrated reliability of individual assessments when comparing the computerised tests to the pen-and-paper versions<sup>12–14</sup>. In the case of speech, the most popular tests that are used in both the research and the clinical setting are formal speaking tests such as word generation and picture naming tasks. Depending on the type of speaking task, different aspects of speech can be extracted and different symptoms that can indicate different diseases can be identified. These tasks can also vary in the degree of experimental control that is used. Very structured tasks, such as counting from 1 to 10 or listing weekdays<sup>15</sup>, can provide insights into motor speech functions including respiration and phonation, articulation, resonance or prosody and potentially indicate diseases such as Parkinson's diseases<sup>16</sup> or Ataxia<sup>17</sup>. Additionally, verbal fluency tasks have been used extensively to assess planning ability and cognitive flexibility in diseases, such as aphasia or dementia<sup>18</sup>. Such formal tests tend to be of a controlled experimental nature, allowing for an easy extraction of variables of interest, focusing mostly on number of correct responses or errors, and reaction times. However, within the last few years qualitative analyses and related objective parameters were shown to provide deeper insights into the complex involvement of cognitive processes<sup>19,20</sup>. Such parameters can be extracted from tasks that allow participants more freedom in the speech that they produce due to less experimental control. One such example is the picture description task, which is commonly used to gain insights into syntactic structure as well as pragmatic competencies<sup>21–23</sup>. While the picture description task is framed by the content of the specific picture, interview situations and open questions provide more varied content as well as insights into more complex aspects of speech. Here, lexical selection, syntactic complexity, pragmatic aspects as well as voice modulation can be investigated<sup>24</sup>. However, the analysis of more qualitative aspects of speech is known to require manual transcription of audio files including labelling of sentence structures. Thus, qualitative speech analysis is extremely time-consuming and not feasible in the clinical context<sup>25</sup>. Recent technological advancements in speech recognition and computer-aided speech feature extraction provide a solution for the objective transformation and complex analysis of speech signals, present in audio data using *Natural Language Processing*, which results in the quantification of the data into vectors that represent the information that is related to the speech attribute of interest<sup>26</sup>. Such methods permit the time-efficient extraction of a multitude of variables that go beyond the ones that were traditionally analysed, thus pushing the boundaries of what can be analysed in the context of speech. Moreover, these automated extraction techniques, in combination with multivariate data-driven analytical techniques, pose fewer limits on the data that can now be used for analysis, allowing the analysis of rich, complex, and ecologically valid data such as spontaneous speech.

Over the years, research investigating the relationship between cognition and speech has been dominated by studies that average data over a group of participants<sup>27</sup>, ultimately treating variability as “noise”. As a result, findings yielding from such univariate within-group analyses fail to account for the rich variability that exists across individuals and thus lack generalizability. Recent work has shown that accounting for inter-individual variability, rather than disregarding it, can provide valuable information to the field<sup>28–31</sup>. This has led studies to explore relationships between different speech features and EF while using an individual-difference approach<sup>32–36</sup>. However, despite the noticeable move towards individual-difference studies, their number is still limited when compared to group-level studies. One of the main reasons for this is that analyses that take individual differences into account naturally require a larger number of participants. This, combined with the aforementioned need for manual coding associated with speech production data, has confined previous studies to smaller sets of data which has in turn limited the analyses that could be performed. Another issue contributing to the limitations of studies in the field is that since such studies are usually language-specific they must rely on datasets that contain speech data in the language of interest. To our knowledge, the only publicly available and well-sized dataset containing both speech and cognitive data is in the Dutch language<sup>37</sup>.

Considering all the points mentioned above, the goal of this work was to generate a German language dataset that would allow a rich variety of analyses in the field of cognition and speech with the possibility of taking individual differences into account. To this aim, the dataset capitalises on both the rich variety of attributes that can be found in speech as well as the modern methods that allow the extraction of such attributes. The generated dataset consists of speech data from 148 German native speakers with an age range of 20–55, supplemented by performance measures on several standardised, computer-based EF tests. The speaking tasks performed in this study ranged from controlled tasks such as word generation and interference tasks, to less controlled spontaneous speaking tasks that were designed in a way that mimics a natural conversation. The speech and EF data are complemented by a rich collection of demographic data as well as hormone information. This dataset is thus a large and expansive collection of data that spans a large age-range, that will allow a variety of analyses in the context of speech, EF, and inter-individual differences, and to our knowledge is the first of its kind in the German language. Some studies have linked free verbal reports (stories) to experience in an extension of the commonly used self-report surveys, e.g.,<sup>38</sup>. It could prove highly interesting to combine such finding with the present data to gain further insights into how experience is translated into language.

The data were collected at the Forschungszentrum Jülich in Jülich, Germany, between January and September 2018 in the context of a large-scale project aimed at investigating the relationship between speech and executive functioning. For each of the tests we provide the raw data output and the speech recordings. This data descriptor comprehensively describes the acquisition and curation of the dataset including the individual tests, experimental procedures and the folder structure of the data. The *Data Records* section describes how this data can be accessed. Researchers interested in performing exploratory and/or hypothesis-driven analyses in the field of language and cognitive performance are invited to make use of the collection presented here.

The dataset is owned by the Institute of Neuroscience and Medicine (INM-7, Brain and Behaviour) at the Forschungszentrum Jülich.

Methods

Ethics statement

This study, including the acquisition and sharing of the data, was approved by the ethics committee of the Heinrich-Heine-University Düsseldorf (Study number: 6055R). All procedures in this study were performed in accordance with the declaration of Helsinki, including but not limited to obtaining informed consent forms from each participant before conducting experimental measurements and keeping all private information anonymised. Participants were also informed that they could quit the study at any time if they wished to do so. All participants whose data is included in the dataset provided consent for the sharing of the data.

Participants

This dataset includes 148 healthy participants with an age range of 20–55 [mean age  $37.2 \pm 11.1$ ; 53 males (mean age =  $35.6 \pm 10.7$ ); 95 females (mean age =  $38.61 \pm 11.4$ )]. Eligible participants included native German speakers who had not acquired an additional language before starting school and had no neurological or psychiatric diagnoses. Early bilingual participants were excluded from the study. Information on the ethnic make-up of the sample was not collected. Participants had different levels of education (finished secondary school = 4; professional school/ job training = 45; finished high school with a university-entrance diploma = 43; university degree = 56). Recruitment took place in North Rhine-Westphalia (Germany) via social networks and the Forschungszentrum Jülich mailing list. Testing sessions took place at the Forschungszentrum Jülich and took a duration of 150–180 min depending on the time needed for instructions and the speed with which the participants performed the tests. A remuneration fee of €50 was paid.

Procedures

Data collection was performed by four examiners, all of whom were required to conduct several pilot tests and were instructed by the study leader to ensure a common standard. Each examiner gave standardised instructions before starting each test and help was provided by the examiner whenever the participant had any questions. The testing session included 4 speaking tasks and 14 EF tests. Acquired measures are provided in Fig. 1. Additional details for each of the tests are described below.

Speaking test battery

The speaking test battery included in this study consisted of a number of well-known tools commonly used to test verbal abilities as well as spontaneous speaking tasks that were used to elicit discourse that is as close to natural speech as possible. When selecting the tests, care was taken to ensure that they cover a spectrum of tests that range from formal speaking tests such as word generation and picture naming to less structured tests such as picture description and spontaneous speech. All speaking tasks were presented and automatically recorded using Presentation software (Neurobehavioral Systems, Inc.; Version 20.1, Build 12.04.17) on an HP ProBook 4730s and using a Logitech Stereo USB Headset as a microphone.

General Information		Speech Data		Cognitive Data	
Age  Sex  Education  Family status  Questions regarding physical and psychological well-being	Word generation	Lexical Verbal Fluency Semantic Verbal Fluency		Cognitive flexibility	Trail Making Test Ravens Standard Progressive Matrices Wisconsin Card Sorting Test Tower of London
	Picture naming	Picture-Word Interference Paradigm			Cued Task Switching
	Picture description	Cookie Theft Picture Task			
Hormone Measures		Spontaneous speech	Story Retelling Story Generation	Working memory	N-back Non-verbal Test Non-verbal Learning Test Corsi Block Tapping Task
Cortisol  Estradiol  Progesterone  Testosterone	Spontaneous speech	Behavioral Measures		Inhibition	Response Inhibition Simon Task Stroop Interference Test
			NEO Five Factor Inventory Beck Depression Inventory - II	Attention / Vigilance	Mackworth Clock Test Divided Attention Test Spatial Attention Test

Figure 1. Summary of the measures that were acquired.

*Verbal fluency (VF)*

The VF tasks used in this study were based on the *Regensburger Wortflüssigkeitstest*<sup>39</sup> which is equivalent to the English *Controlled Oral Word Association Test*<sup>40</sup>. The implementation used in this study comprised two types of VF: lexical VF, and semantic VF, each consisting of three separate sub-tasks. The lexical VF task consisted of two simple tasks where participants were required to generate as many German words as possible that start with the letters “M” and “K” respectively. The decision for the selection of the specific letters was based on the difficulty level associated with the search for words starting with the respective letter. While words with the initial letter “M” provide an abundant search space, the letter “K” represents a higher difficulty level due to less available words<sup>39</sup>. An additional, more demanding task involved a switching component where the participants were required to switch between words that start with the letter “G” and words that start with the letter “R”. Each of the three tasks were performed for two minutes. Participants were not allowed to use proper nouns or repeat words more than once. Words with the same root were considered to be the same word and were thus also not allowed. Additionally, a word was only considered as correct if it would be found in a German book or newspaper. The instruction was given in German and was as follows:

‘Bei dieser Aufgabe sollen Sie innerhalb von zwei Minuten möglichst viele verschiedene Wörter nennen, die mit dem Anfangsbuchstaben “M” beginnen. Dabei sollen Sie verschiedene Regeln beachten: Sie sollen nur Wörter nennen, die in einer deutschen Zeitung oder einem deutschen Buch verwendet werden könnten. Dabei sollen Sie keine Wörter mehrfach nennen. Die Wörter dürfen aber auch nicht mit dem gleichen Wortstamm beginnen, also “Müll, Mülleimer, Müllabfuhr, Mülltonne”/“Kerze, Kerzenschein, Kerzenständer, Kerzenlicht” gelten nur als ein Wort. Weiterhin dürfen Sie auch keine Eigennamen nennen, also “Miriam, Max, Madrid, Malta”/“Kerstin, Kurt, Köln, Kreta” gelten nicht. Bitte versuchen Sie, möglichst schnell viele verschiedene Wörter mit dem Anfangsbuchstaben “M/K” zu nennen.’<sup>39</sup>

The semantic VF task consisted of two simple tasks where the participants were required to name animals and jobs respectively. The third task involved a switching component where the participants were required to switch between naming fruit and sports. Each of the three tasks were performed for two minutes and the rules specified in the lexical VF task still applied. The instruction was given in German and was as follows:

‘Bei dieser Aufgabe sollen Sie innerhalb von zwei Minuten möglichst viele verschiedene Wörter aus der Kategorie “Tiere”/“Berufe” nennen. Dabei sollen Sie keine Tiere mehrfach nennen. Bitte versuchen Sie, möglichst schnell viele Tiere/Berufe zu nennen.’<sup>39</sup>

*Picture-word interference paradigm*

Participants were shown 64 different pictures (obtained from<sup>41</sup>), each accompanied by a spoken word which was either semantically related or semantically unrelated to the picture shown. The pictures were shown for 4500 ms and were followed by a fixation cross that was shown for 3000 ms. The auditory stimuli were spoken by a 23-year-old German female. Participants were required to name the picture that was shown as quickly as possible, and their audio was recorded as soon as the picture faded in. The list of picture names and auditory distractors can be found in the data repository as a separate file. For target and feature selection, items were controlled for an unequal onset and distractor items were not used as target items. Moreover, items were controlled for frequency and semantic relatedness using *GermaNet Pathfinder*<sup>42</sup>.

The instruction was given in German and was as follows:

“In dieser Aufgabe benennen Sie bitte wieder die Bilder, die Sie sehen. Während Sie das Bild sehen, werden Sie gleichzeitig Wörter hören. Diese Wörter können manchmal helfen, das Bild zu benennen oder sie erschweren es. Versuchen Sie trotzdem, das Bild so schnell es geht zu benennen.”

*Picture description task*

Participants were shown the *Cookie Theft Picture* obtained from<sup>43</sup>, and asked to describe it in as much detail as possible in 90 s. The instruction was given in German and was as follows:

“Bitte beschreiben Sie in 90 Sekunden dieses Bild so ausführlich wie möglich.”

The answer given by the participants was then recorded for 90 s.

*Spontaneous speaking task*

Participants were first told that they will be asked two questions to which they were required to reply in as much detail as possible for 5 min. This instruction was given in German and was as follows:

“Ich werde Ihnen heute insgesamt zwei Fragen stellen, bei denen ich Sie bitte, etwas ausführlicher für ca. 5 min zu antworten.”

The first question required the participants to either describe a book that they have read recently or to talk about something that they watched on television the night before. In case participants could not respond to this, they were asked to report any events happening within the last weeks. The question was asked in German and was as follows:

“Was haben Sie gestern Abend im Fernsehen geschaut oder welches Buch haben Sie gelesen?”

The second question required the participants to describe a vacation that they would like to take if time and money were no object. The question was asked in German and was as follows:

Test	Battery	Form	Version	Testing time (minutes)
Corsi Block Tapping Test	SCHUHFRIED Wiener Testsystem	S1	28	10
Response Inhibition	SCHUHFRIED Wiener Testsystem	S1	23.04	10
Mackworth Clock Test	PsyToolkit	–	–	5
N-back Non-verbal Test	SCHUHFRIED Wiener Testsystem	S1	22.02	9
Non-verbal Learning Test	SCHUHFRIED Wiener Testsystem	S2	24.01	9
Simon Task	PsyToolkit	–	–	5
Raven's Standard Progressive Matrices	SCHUHFRIED Wiener Testsystem	S5	32.01	18
Stroop Interference Test	SCHUHFRIED Wiener Testsystem	S7	29	15
Cued Task Switching	PsyToolkit	–	–	5
Trail Making Test	SCHUHFRIED Wiener Testsystem	S1	51.2	3
Tower of London	SCHUHFRIED Wiener Testsystem	S11	23.04	11
Divided Attention	SCHUHFRIED Wiener Testsystem	S2	50.2	12
Spatial Attention & Neglect	SCHUHFRIED Wiener Testsystem	S1	23.05	12
Wisconsin Card Sorting Test	PsyToolkit	–	–	5

**Table 1.** A list of the EF tests performed by the participants.

“Wo und wie würden Sie Ihren schönsten Urlaub verbringen? Erzählen Sie uns etwas darüber.“

Both answers given were recorded for 5 min each and the examiners asked for more detail in the case of participants that did not talk for that long.

### EF test battery

The EF test battery consisted of computerised versions of commonly used neuropsychological tests covering different subdomains of EFs either from the SCHUHFRIED Wiener Testsystem or Psytoolkit (<https://www.psytolkit.org/experiment-library/mackworth.html>;<sup>44,45</sup>). The tests included in the battery were chosen to capture a broad range of subdomains of cognitive performance such as cognitive flexibility, planning, working memory, attention, and inhibition. There is overlap in the general areas covered by the tests. However, each test has properties that make it unique compared to the other tests in the battery. Table 1 outlines the specific battery and version that was used for each test while Table 2 provides an overview of the different variables that were measured for each test together with descriptive statistics for each of the variables (mean, standard deviation and range). Values being shared represent raw data for each test.

#### *Corsi block tapping test (CORSI)*

Participants were presented with nine cubes arranged in an irregular order on the screen followed by a pointer that points to three cubes in a specific order. At the end of this sequence a signal sounded prompting the participants to repeat the given sequence. The length of the sequence was increased by one cube each time the participants completed the sequence successfully.

#### *Response inhibition (INHIB)*

The test consisted of two parts. In the first part of the test an arrow was displayed on the screen and participants were asked to respond to the direction in which the arrow was pointing. In the second part of the test the participants were asked to repeat the task as in the previous part but were additionally asked to suppress their motoric response whenever they heard an auditory signal.

#### *Mackworth clock test (MACK)*

Participants were presented with a large green clock hand displayed on a black screen. The hand moved like the second hand of a clock, approximately every second. At infrequent and irregular intervals, the hand made irregular “jumps”. Participants were requested to detect and quickly react to these irregular “jumps” by pressing a button. The irregular “jump” of the clock hand was around 10% of the circle and the duration of the test was 1 min comprising 60 total moves of the clock hand.

#### *N-back non-verbal test (NBN)*

A sequence of 100 abstract figures were presented one by one. The task consisted of indicating whether the figure that was currently displayed was identical to the one shown two places back (2-back paradigm). If it was, the participant was expected to press a button as quickly as possible.

#### *Non-verbal learning test (NVLT)*

Nonsensical, irregular, and geometric figures were presented on the screen. During the course of the test some figures were shown multiple times. For each figure the participants were required to decide whether the current figure has already appeared or whether this figure is being shown for the first time.



Test	Variable	Mean $\pm$ SD	Min to Max
Corsi block spanning test	Block span	5.64 $\pm$ 1.16	3 to 8
	Sum of correct items	10.50 $\pm$ 2.93	2 to 17
	Sum of false items	4.61 $\pm$ 1.37	3 to 9
	Sum of missed items	0.05 $\pm$ 0.21	0 to 1
	Sum of sequence errors	2.59 $\pm$ 1.31	0 to 6
Response inhibition	Reaction time <sup>a</sup>	0.53 $\pm$ 0.09	0.32 to 0.78
	Mean stop signal delay <sup>a</sup>	0.30 $\pm$ 0.07	0.05 to 0.35
	Stop signal reaction time <sup>a</sup>	0.22 $\pm$ 0.08	0.03 to 0 0.58
	Sum of commission errors	13.76 $\pm$ 6.33	1 to 31
	Sum of omission errors	1.33 $\pm$ 2.32	0 to 18
Mackworth clock test	Sum of missed jumps	8.17 $\pm$ 4.85	0 to 23
	Sum of false alarms	3.08 $\pm$ 3.80	0 to 28
N-back non-verbal test	Sum of correct items	8.27 $\pm$ 3.10	0 to 14
	Sum of commission errors	5.73 $\pm$ 3.10	0 to 14
	Sum of errors	8.45 $\pm$ 7.03	0 to 36
	Mean reaction time of correct items <sup>a</sup>	0.77 $\pm$ 0.19	0 to 1.36
	Mean reaction time of errors <sup>a</sup>	0.86 $\pm$ 0.25	0 to 1.64
Non-verbal learning test	Sum of correct responses	31.84 $\pm$ 4.91	11 to 39
	Sum of false responses	11.72 $\pm$ 8.35	0 to 48
	Sum of difference between correct and false responses	20.12 $\pm$ 7.75	– 13 to 35
	Processing time <sup>a</sup>	118.97 $\pm$ 33.14	74 to 303
Simon task	Sum of errors in congruent trials	1.01 $\pm$ 1.48	0 to 8
	Sum of errors in incongruent trials	3.16 $\pm$ 2.90	0 to 19
Ravens standard progressive matrices	Sum of correct items	27.73 $\pm$ 3.64	14 to 32
	Processing time <sup>a</sup>	638.74 $\pm$ 162.12	328 to 905
Stroop interference	Reading interference <sup>a</sup>	0.15 $\pm$ 0.09	– 0.04 to 0.40
	Naming interference <sup>a</sup>	0.13 $\pm$ 0.08	– 0.02 to 0.38
	Interference difference <sup>a</sup>	– 0.06 $\pm$ 0.12	– 0.40 to 0.23
	Number of false reactions (reading baseline)	2.14 $\pm$ 2.58	0 to 13
	Number of false reactions (naming baseline)	2.72 $\pm$ 2.68	0 to 18
	Number of false reactions (reading interference)	3.71 $\pm$ 8.06	0 to 96
	Number of false reactions (naming interference)	3.54 $\pm$ 3.18	0 to 23
	Processing time	405.84 $\pm$ 59.88	294 to 616
Cued task switching	Sum of errors	2.46 $\pm$ 3.13	0 to 19
	Sum of timeouts	0.29 $\pm$ 0.87	0 to 7
	Sum of errors in incongruent trials	2.02 $\pm$ 2.40	0 to 14
	Mean reaction time for incongruent trials <sup>a</sup>	0.6 $\pm$ 0.14	0.37 to 1.34
	Mean reaction time for congruent trials <sup>a</sup>	0.55 $\pm$ 0.13	0.31 to 1.11
	Switch costs (RT incongruent–RT congruent) <sup>a</sup>	0.05 $\pm$ 0.08	– 0.19 to 0.39
Trail making test	Processing time for part A <sup>a</sup>	17.44 $\pm$ 3.48	9.69 to 27.72
	Processing time for part B <sup>a</sup>	25.70 $\pm$ 8.58	12–24 to 57.55
	Difference part B–part A <sup>a</sup>	8.25 $\pm$ 7.14	– 3.32 to 40.57
	Quotient of B/A <sup>a</sup>	1.47 $\pm$ 0.39	0.84 to 3.39
	Sum of errors in part A	0.07 $\pm$ 0.25	0 to 1
	Sum of errors in part B	0.68 $\pm$ 1.09	0 to 9
Tower of London	Planning ability	7.47 $\pm$ 2.31	1 to 12
	Sum of correct responses	10.45 $\pm$ 1.80	2 to 12
	Sum of self-corrections	1.77 $\pm$ 2.09	0 to 11
	Sum of wrong pole choices	0.70 $\pm$ 2.76	0 to 31
	Sum of blocked pole choices	1.13 $\pm$ 2.73	0 to 22
	Sum of impossible position choices	0.81 $\pm$ 2.85	0 to 28
Continued			

Test	Variable	Mean ± SD	Min to Max
Divided attention	Sum of missed items (unimodal visual)	2.12 ± 3.19	0 to 16
	Sum of false alarms (unimodal visual)	3.24 ± 4.51	0 to 42
	Mean reaction time (unimodal visual) <sup>b</sup>	454.30 ± 97.29	251 to 826
	Sum of missed items (cross modal)	2.79 ± 3.13	0 to 17
	Sum of false alarms (cross modal)	3.26 ± 5.99	0 to 49
	Mean reaction time (cross modal) <sup>b</sup>	469.49 ± 111.57	240 to 880
Spatial attention	Mean reaction time of unannounced items <sup>b</sup>	384.53 ± 50.72	289 to 526
	Sum of missed items when correct item is announced	0.99 ± 1.48	0 to 8
	Mean reaction time when correct item is announced <sup>b</sup>	311.15 ± 50.88	207 to 494
	Sum of missed items when incorrect item is announced	0.17 ± 0.43	0 to 2
	Mean reaction time when incorrect item is announced <sup>b</sup>	345.75 ± 52.43	238 to 483
	Mean reaction time (short SOA) <sup>b</sup>	353.32 ± 47.09	261 to 485
	Mean reaction time (Long SOA) <sup>b</sup>	345.18 ± 46.58	256 to 474
	Sum of errors	3.27 ± 3.05	0 to 16
Wisconsin card sorting test	Sum of errors	12.83 ± 5.86	6 to 36
	Sum of perseveration errors	8.02 ± 3.23	4 to 21
	Sum of non-perseveration errors	4.81 ± 3.31	0 to 18
	Sum of timeouts	0.53 ± 1.05	0 to 9

**Table 2.** An overview of the different variables measured for each test. SOA = stimulus onset asynchrony. <sup>a</sup>Time measured in seconds. <sup>b</sup>Time measured in milliseconds.

*Simon task (SIMON)*

Participants were asked to press the “l” key if they read the word “rechts” (German word for *right*) and the “a” key if they read the word “links” (German word for *left*). Each word was displayed either on the right or the left part of the screen meaning that the stimulus could be congruent or incongruent to its position.

*Ravens standard progressive matrices (SPM)*

The participants were shown eight separate items that follow a pattern. The task required the participants to identify one missing item out of 6 choices to complete the pattern. The difficulty in pattern recognition increased during the course of the test.

*Stroop interference test (STROOP)*

Names of colours were displayed on the screen in a colour which was incongruent to the name (e.g., the word “blau” (German for *blue*) printed in red). The test consisted of two conditions. In the naming condition the participants were asked to respond to the colour of the words. In the reading condition participants were asked to respond to the meaning of the word. A baseline measure for the reaction speed and accuracy of the participants was established at the start of the test by presenting colour words without colouring or simple colour bars.

*Cued task switching (SWITCH)*

This task consisted of a shape and a colour task. A cue stimulus informed the participant which task to perform on every trial. The cue for the colour task was the word “COLOR” and the cure for the shape task was the word “SHAPE”. In the colour task participants were asked to respond to the colour of the presented figure while ignoring the shape. In the shape task participants were required to respond to the shape of the presented figure while ignoring the colour. For selecting the respective colour or shape, two letters of the keyboard were determined (the letter “b” was used for the answers *circle* and *yellow* and the letter “n” was used for the answers *rectangle* and *blue*. Depending on the answer, the respective letter was pressed by the participant.

*Trail making test (TMT)*

The task consisted of 2 parts: part A and part B. In part A numbers ranging from 1 to 25 were displayed randomly across the screen. Participants were asked to click on the numbers in ascending order and as quickly as possible. In part B, the numbers displayed on the screen ranged from 1 to 13 and were accompanied by alphabetic letters ranging from A to L, both of which were presented in a random order. Part B required participants to click on numbers and letters alternately and in ascending order.

*Tower of London (TOL)*

Participants were presented with an image that depicts a three-dimensional wooden model with three rods on which three balls of different colors are placed. The left rod holds three balls, the middle rod takes two balls, and the right rod has room for one ball. The participants were asked to move the balls from the starting state to a target position using a minimum number of moves. The target state was always shown in the upper part of the screen and the starting state in the lower part. The minimum number of moves required to achieve this was

shown to the left of the starting state. Various rules were to be observed, one of which was the rule that only one ball can be moved at a time.

#### *Perception and attention functions test: divided attention (WAFG)*

The participants were required to focus on two geometric figures and one auditory stimulus. At certain intervals the stimuli change their intensity (i.e., figure gets lighter and/or auditory stimulus gets louder). The participants were asked to respond when two stimuli became lighter/louder twice in succession.

#### *Perception and attention functions test: spatial attention & neglect (C)*

Four triangles were presented in four spatial positions. The participants were required to react if a triangle changes intensity (i.e., gets darker). In the neglect test an interfering or matching visual cue was also given.

#### *Wisconsin card sorting test (WCST)*

The task used here is not the actual Wisconsin Card Sorting Test, as copyrighted in the US, but rather a computer-based task that is inspired by the original test<sup>46</sup>. Four stimulus cards illustrating different geometrical figures were presented. The figures on the cards differ in number, colour, and form. The task of the participants was to figure out the classification rule to be able to match a newly presented card to one of the four cards. Participants were given feedback for every card that they matched. The classification rule was changed every 10 cards, requiring the participants to shift rules accordingly.

### **Additional data**

In addition to the main set of speaking and EF tasks, phenotypical data were collected through questionnaires including the German version of the Beck Depression Inventory (BDI-II<sup>47</sup>) used to collect information regarding depressive symptoms, and the NEO Five Factor Inventory (NEO-FFI<sup>48</sup>). Furthermore, participants were asked general questions about their background, habits, and their physical and psychological well-being before commencement of the testing session. Saliva samples were collected at the beginning and at the end of the test session, stored in a refrigerator and sent to an external lab for analysis. The two saliva samples of each participant were then pooled at an external lab which carried out quantification analyses for cortisol, progesterone, oestradiol, and testosterone.

### **Data records/usage notes**

The dataset presented in this paper is stored on GDPR-compliant and protected servers of the Forschungszentrum Jülich, housed at the Jülich Super Computing Centre, as agreed upon by the participants. The dataset complies with the four basic principles of FAIR. The dataset is clearly described with metadata, that are accessible on Jülich DATA (<https://data.fz-juelich.de/dataset.xhtml?persistentId=doi:https://doi.org/10.26165/JUELICH-DATA/CHWZDZ>) making it findable, accessible, interoperable and reusable.

Researchers who wish to acquire access to the data are kindly asked to contact the authors at [spexdata@fz-juelich.de](mailto:spexdata@fz-juelich.de). Applicants will be asked to submit an approved ethics application together with a project outline. Additionally, applicants will be asked to ensure that the requested data will be only used for the research project specified and that it will not be passed on to third parties. Once the request is approved applicants will receive temporary access to an encrypted version of the data which they can then download.

The dataset repository contains 76.81 GB of data and includes five main folders, four of which contain the different measures depicted in Fig. 1 (i.e., EF data; speech data; hormone data; questionnaires). A fifth folder presents publications that have already made use of the dataset<sup>32,33</sup>. The folder containing the EF data contains a sub-folder for each of the different tests used (i.e., the 14 tests listed in Tables 1 and 2). Each sub-folder contains a comma-separated-value file with the corresponding raw data as well as a text file consisting of information about the specific measure, details on how it was acquired as well as details of the hardware and software used for the acquisition. On the other hand, the folder containing the speech data contains a sub-folder for each of the participants. Each of these sub-folders contains further sub-folders for each of the 6 speaking tasks, which in turn contain the corresponding raw waveform audio files in the uncompressed format RIFF WAVE (WAV). All the speech utterances were recorded with a bit rate of 2822 kBit/s, a sample size of 32 bit, and a sampling rate of 44.100 kHz.

For each of the 148 participants, 6 min of Lexical Verbal Fluency, 6 min of Semantic Verbal Fluency, 4.13 min of the Picture-Word Interference Paradigm, 1.5 min of the Picture Description Task, 5 min of Story Retelling, and 5 min of Story Generation were recorded. This corresponds to a total of 27.63 min of recorded speech data per subject. In total, the dataset provides 68.16 h of speech recordings.

We expect this dataset to be of interest to researchers conducting exploratory or hypothesis-driven research in the field of individual differences in language and executive functioning. The dataset has already been used to predict verbal fluency scores from EF performance<sup>32</sup>, and to predict EF performance from a comprehensive set of verbal fluency features<sup>33</sup>.

The data were collected at the Forschungszentrum Jülich in Jülich, Germany, between January and September 2018 in the context of a large-scale project aimed at investigating the relationship between speech and executive functioning. For each of the tests we provide the raw data output and the speech recordings. This data descriptor comprehensively describes the acquisition and curation of the dataset including the individual tests, experimental procedures and the folder structure of the data. The *Data Records* section describes how this data can be accessed. Researchers interested in performing exploratory and/or hypothesis-driven analyses in the field of language and cognitive performance are invited to make use of the collection presented here.



# Data availability

The dataset presented here will be made available to interested researchers upon request, as described in this data descriptor. Researchers who wish to acquire access to the data are kindly asked to contact the authors at [spxdata@fz-juelich.de](mailto:spxdata@fz-juelich.de).

Received: 17 July 2023; Accepted: 1 April 2024

Published online: 24 April 2024

# References

1. Baddeley, A. Working memory: Looking back and looking forward. *Nat. Rev. Neurosci.* **4**(10), 829–839. <https://doi.org/10.1038/nrn1201> (2003).
2. Delgado-Álvarez, A. *et al.* Cognitive processes underlying verbal fluency in multiple sclerosis. *Front. Neurol.* **11**, 1951. <https://doi.org/10.3389/FNEUR.2020.629183/BIBTEX> (2021).
3. Hagoort, P. MUC (memory, unification, control) and beyond. *Front. Psychol.* **4**(JUL), 416. <https://doi.org/10.3389/FPSYG.2013.00416/BIBTEX> (2013).
4. Levelt, W. J. M. Accessing words in speech production: Stages, processes and representations. *Cognition* **42**(1–3), 1–22. [https://doi.org/10.1016/0010-0277\(92\)90038-J](https://doi.org/10.1016/0010-0277(92)90038-J) (1992).
5. Tonér, S. & Gerholm, T. N. Links between language and executive functions in Swedish preschool children: A pilot study. *Appl. Psycholinguist.* **42**(1), 207–241. <https://doi.org/10.1017/S0142716420000703> (2021).
6. Ardila, A. The executive functions in language and communication. *Cogn. Acquir. Lang. Disord.* <https://doi.org/10.1016/B978-0-323-07201-4.00016-7> (2012).
7. Hecker, P., Steckhan, N., Eyben, F., Schuller, B. W. & Arnrich, B. Voice Analysis for neurological disorder recognition-A systematic review and perspective on emerging trends. *Front. Digit. Health* <https://doi.org/10.3389/FDGTH.2022.842301> (2022).
8. Creaufmüller, M., Heim, S., Habel, U. & Mühlhaus, J. The influence of semantic associations on sentence production in schizophrenia: An fMRI study. *Eur. Arch. Psychiatry Clin. Neurosci.* **270**(3), 359–372. <https://doi.org/10.1007/S00406-018-0936-9/TABLES/4> (2020).
9. Cummins, N., Sethu, V., Epps, J., Schnieder, S. & Krajewski, J. Analysis of acoustic space variability in speech affected by depression. *Speech Commun.* **75**, 27–49. <https://doi.org/10.1016/J.SPECOM.2015.09.003> (2015).
10. Holshausen, K., Harvey, P. D., Elvevåg, B., Foltz, P. W. & Bowie, C. R. Latent semantic variables are associated with formal thought disorder and adaptive behavior in older inpatients with schizophrenia. *Cortex* **55**(1), 88–96. <https://doi.org/10.1016/J.CORTEX.2013.02.006> (2014).
11. Nicodemus, K. K. *et al.* Category fluency, latent semantic analysis and schizophrenia: A candidate gene approach. *Cortex* **55**(1), 182–191. <https://doi.org/10.1016/J.CORTEX.2013.12.004> (2014).
12. Arce-Ferrer, A. J. & Guzmán, E. M. Studying the equivalence of computer-delivered and paper-based administrations of the Raven standard progressive matrices test. *Educ. Psychol. Meas.* **69**(5), 855–867. <https://doi.org/10.1177/0013164409332219> (2009).
13. Dahmen, J., Cook, D., Fellows, R. & Schmitter-Edgecombe, M. An analysis of a digital variant of the Trail Making Test using machine learning techniques. *Technol. Health Care* **25**(2), 251–264. <https://doi.org/10.3233/THC-161274> (2017).
14. Khaligh-Razavi, S.-M., Sadeghi, M., Khanbagi, M., Kalafatis, C. & Nabavi, S. M. A self-administered, artificial intelligence (AI) platform for cognitive assessment in multiple sclerosis (MS). *BMC Neurol.* **20**(1), 1–13. <https://doi.org/10.1186/S12883-020-01736-X> (2020).
15. Amunts, K. *et al.* Analysis of neural mechanisms underlying verbal fluency in cytoarchitecturally defined stereotaxic space—The roles of Brodmann areas 44 and 45. *Neuroimage* **22**(1), 42–56. <https://doi.org/10.1016/J.NEUROIMAGE.2003.12.031> (2004).
16. Norel, R. *et al.* Speech-based characterization of dopamine replacement therapy in people with Parkinson's disease. *NPJ Parkinson's Dis.* **6**(1), 1–8. <https://doi.org/10.1038/s41531-020-0113-5> (2020).
17. Vogel, A. P. *et al.* Features of speech and swallowing dysfunction in pre-ataxic spinocerebellar ataxia type 2. *Neurology* **95**(2), e194–e205. <https://doi.org/10.1212/WNL.0000000000009776> (2020).
18. Marczyński, C. A. & Kertesz, A. Category and letter fluency in semantic dementia, primary progressive aphasia, and Alzheimer's disease. *Brain Lang.* **97**(3), 258–265. <https://doi.org/10.1016/J.BANDL.2005.11.001> (2006).
19. De Looze, C. *et al.* Cognitive and structural correlates of conversational speech timing in mild cognitive impairment and mild-to-moderate Alzheimer's disease: Relevance for early detection approaches. *Front. Aging Neurosci.* **13**, 207. <https://doi.org/10.3389/FNAGI.2021.637404/BIBTEX> (2021).
20. Vincze, V. *et al.* Linguistic parameters of spontaneous speech for identifying mild cognitive impairment and Alzheimer disease. *Comput. Linguist.* **48**(1), 119–153. [https://doi.org/10.1162/COLI\\_A\\_00428](https://doi.org/10.1162/COLI_A_00428) (2022).
21. Grande, M. *et al.* From a concept to a word in a syntactically complete sentence: An fMRI study on spontaneous language production in an overt picture description task. *NeuroImage* **61**(3), 702–714. <https://doi.org/10.1016/J.NEUROIMAGE.2012.03.087> (2012).
22. Meffert, E., Gallus, M., Grande, M., Schönberger, E. & Heim, S. Neural correlates of spontaneous language production in two patients with right hemispheric language dominance. *Aphasiology* **35**(11), 1482–1504. <https://doi.org/10.1080/02687038.2020.1819955> (2020).
23. Schönberger, E. *et al.* The neural correlates of agrammatism: Evidence from aphasic and healthy speakers performing an overt picture description task. *Frontiers in Psychology* **5**(MAR), 246. <https://doi.org/10.3389/FPSYG.2014.00246/ABSTRACT> (2014).
24. Robin, J. *et al.* Evaluation of speech-based digital biomarkers: Review and recommendations. *Digit. Biomark.* **4**(3), 99–108. <https://doi.org/10.1159/000510820> (2020).
25. Le, D., Licata, K. & Mower Provost, E. Automatic quantitative analysis of spontaneous aphasic speech. *Speech Commun.* **100**, 1–12. <https://doi.org/10.1016/J.SPECOM.2018.04.001> (2018).
26. Madan, A. Speech feature extraction and classification: A comparative review. *Int. J. Comput. Appl.* **90**(9), 975–8887 (2014).
27. Levelt, W. A history of psycholinguistics: The pre-Chomskyan Era. In *A History of Psycholinguistics: The Pre-Chomskyan Era*, 1–656. <https://doi.org/10.1093/ACPROF:OSO/9780199653669.001.0001> (2012).
28. Andrews, S. & Lo, S. Not all skilled readers have cracked the code: Individual differences in masked form priming. *J. Exp. Psychol. Learn. Mem. Cogn.* **38**(1), 152–163. <https://doi.org/10.1037/A0024953> (2012).
29. Dąbrowska, E. Experience, aptitude and individual differences in native language ultimate attainment. *Cognition* **178**, 222–235. <https://doi.org/10.1016/J.COgnITION.2018.05.018> (2018).
30. Engelhardt, P. E., Alfridijanta, O., McMullon, M. E. G. & Corley, M. Speaker-versus listener-oriented disfluency: A re-examination of arguments and assumptions from autism spectrum disorder. *J. Autism Dev. Disord.* **47**(9), 2885–2898. <https://doi.org/10.1007/S10803-017-3215-0/TABLES/4> (2017).
31. Kidd, E., Donnelly, S. & Christiansen, M. H. Individual differences in language acquisition and processing. *Trends Cogn. Sci.* **22**(2), 154–169. <https://doi.org/10.1016/J.TICS.2017.11.006> (2018).
32. Amunts, J., Camilleri, J. A., Eickhoff, S. B., Heim, S. & Weis, S. Executive functions predict verbal fluency scores in healthy participants. *Sci. Rep.* <https://doi.org/10.1038/s41598-020-65525-9> (2020).

33. Amunts, J. *et al.* Comprehensive verbal fluency features predict executive function performance. *Sci. Rep.* <https://doi.org/10.1038/s41598-021-85981-1> (2021).
34. Huettig, F. & Janse, E. Individual differences in working memory and processing speed predict anticipatory spoken language processing in the visual world. *Lang. Cogn. Neurosci.* **31**(1), 80–93. <https://doi.org/10.1080/23273798.2015.1047459> (2015).
35. Jongman, S. R., Roelofs, A. & Meyer, A. S. Sustained attention in language production: An individual differences investigation. *Q. J. Exp. Psychol.* **68**(4), 710–730. <https://doi.org/10.1080/17470218.2014.964736> (2015).
36. Shao, Z., Roelofs, A. & Meyer, A. S. Sources of individual differences in the speed of naming objects and actions: The contribution of executive control. *Q. J. Exp. Psychol.* **65**(10), 1927–1944. <https://doi.org/10.1080/17470218.2012.670252> (2012).
37. Hintz, F., Dijkhuis, M. & van 't Hoff, V., McQueen, J. M., & Meyer, A. S., A behavioural dataset for studying individual differences in language skills. *Sci. Data* **7**(1), 1–18. <https://doi.org/10.1038/s41597-020-00758-x> (2020).
38. Chasteen, A. L., Tagliamonte, S. A., Pabst, K. & Brunet, S. Ageist communication experienced by middle-aged and older Canadians. *Int. J. Environ. Res. Public Health* **19**(4), 2004. <https://doi.org/10.3390/IJERPH19042004/S1> (2022).
39. Aschenbrenner, S., Tucha, O., & Lange, K. W. *Regensburger Wortflüssigkeits-Test: RWT*. Hogrefe, Verlag für Psychologie. [https://www.spielundlern.de/product\\_info.php/products\\_id/45519](https://www.spielundlern.de/product_info.php/products_id/45519) (2001).
40. Ruff, R. M., Light, R. H., Parker, S. B. & Levin, H. S. Benton controlled oral word association test: Reliability and updated norms. *Arch. Clin. Neuropsychol.* **11**(4), 329–338. <https://doi.org/10.1093/ARCLIN/11.4.329> (1996).
41. Snodgrass, J. G. & Vanderwart, M. A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *J. Exp. Psychol. Hum. Learn. Mem.* **6**(2), 174–215. <https://doi.org/10.1037/0278-7393.6.2.174> (1980).
42. Cramer, I., & Finthammer, M. *Tools for Exploring GermaNet in the Context of cl-teaching*. 195–208. (2008). <https://doi.org/10.1515/9783110211818.3.195>
43. Goodglass, H., Kaplan, E., & Barresi, B. *The Assessment of Aphasia and Related Disorders* (Lippincott Williams & Wilkins, 1972).
44. Stoet, G. PsyToolkit: A software package for programming psychological experiments using Linux. *Behav. Res. Methods* **42**(4), 1096–1104. <https://doi.org/10.3758/BRM.42.4.1096> (2010).
45. Stoet, G. PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teach. Psychol.* **44**(1), 24–31. <https://doi.org/10.1177/0098628316677643> (2016).
46. Berg, E. A. A simple objective technique for measuring flexibility in thinking. *J. Gen. Psychol.* **39**(1), 15–22. <https://doi.org/10.1080/00221309.1948.9918159> (1948).
47. Beck, A. T., Steer, R. A., & Brown, G. K. *Beck Depression Inventory (BDI-II)*, 2nd ed., vol. 10 (Pearson, 1996).
48. Costa, P. T., & McCrae, R. R. NEO five-factor inventory (NEO-FFI). *Psychol. Assess. Resour.* **3** (1989).

## Acknowledgements

This study was supported by the Deutsche Forschungsgemeinschaft (DFG, GE 2835/2-1, EI 816/16-1 and EI 816/21-1), the National Institute of Mental Health (R01-MH074457), the Helmholtz Portfolio Theme "Supercomputing and Modeling for the Human Brain", the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement 785907 (HBP SGA2), 945539 (HBP SGA3), The Virtual Brain Cloud (EU H2020, no. 826421), and the National Institute on Aging (R01AG067103).

## Author contributions

Experimental design: J.A.C., J.V., S.B.E., S.H., S.W. Participant recruitment: J.V. Data collection: H.N., J.V., L.N.M., N.S. Quality control of primary data: H.N., J.V., L.N.M., N.S. Data curation: J.A.C., J.V., G.K. Manuscript writing: J.A.C., J.V. with contributions from S.W. and all other authors.

## Funding

Open Access funding enabled and organized by Projekt DEAL.

## Competing interests

Julia Volkening is employed by PeakProfiling GmbH. The other authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to J.A.C.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024