## Data Request form YOUth (version 6.0, February 2020)

**Introduction**

The information you provide here will be used by the YOUth Executive Board, the Data Manager, and the Data Management Committee to evaluate your data request. Details regarding this evaluation procedure can be found in the Data Access Protocol.

All data requests will be published on the YOUth researcher's website in order to provide a searchable overview of past, current, and pending data requests. By default, the publication of submitted and pending data requests includes he names and institutions of the contact person and participating researchers as well as a broad description of the research context.

After approval of a data request, the complete request (including hypotheses and proposed analyses) will be published. If an applicant has reasons to object to the publication of their complete data request, they should notify the Project Manager, who will evaluate the objection with the other members of the Executive Board and the Data Management Committee. If the objection is rejected, the researcher may decide to withdraw their data request.

**Section 1: Researchers**

In this section, please provide information about the researchers involved with this data request.

- Name, affiliation and contact information of the contact person
- Name and details of participating researchers (e.g. intended co-authors)
- Name and details of the contact person within YOUth (if any)

| **Contact person for the proposed study:** | |
|---|---|
| Name: | Sita ter Haar |
| Institution: | Utrecht University |
| Department: | TLC/ Humanities |
| Address: | Trans 10 |
| Email: | s.m.terhaar@uu.nl |
| Phone: | |

| **Participating researcher:** | |
|---|---|
| Name: | Prof. René Kager (temporarily on sick leave) |
| Institution: | Utrecht University |
| Department: | TLC/ Humanities |
| Address: | Trans 10 |
| Email: | Please do not email yet, due to sick leave r.w.j.kager@uu.nl |
| Phone: | |

| **Participating researcher:** | |
|---|---|
| Name: | Anika van der Klis |

| | |
|---|---|
| Institution: | Utrecht University |
| Department: | TLC/ Humanities |
| Address: | Trans 10 |
| Email: | a.vanderklis@uu.nl |
| Phone: | |

| **Participating researcher:** | |
|---|---|
| Name: | dr. Caroline Junge |
| Institution: | Universiteit Utrecht |
| Department: | FSW |
| Address: | Martinus J. Langeveldgebouw, Heidelberglaan 1 |
| Email: | c.m.m.junge@uu.nl |
| Phone: | |

| **Participating researcher:** | |
|---|---|
| Name: | |
| Institution: | |
| Department: | |
| Address: | |
| Email: | |
| Phone: | |

| **Contact person within YOUth (if any)** | |
|---|---|
| Name: | |
| Institution: | |
| Department: | |
| Address: | |
| Email: | |
| Phone: | |

**Section 2: Research context**

In this section, please briefly describe the context for your research plans. This section should logically introduce the next section (hypotheses). As mentioned, please note that this section will be made publicly available on our researcher's website after submission of your request.
Please provide:

- The title of your research plan
- A very brief background for the topic of your research plan
- The rationale for and relevance of your specific research plan
- The specific research question(s) or aim(s) of your research (Please also provide a brief specification)
- A short description of the data you request

References can be added at the end of this section (optional).

| **Title of the study** |
|---|
| **Early predictors for phonological development: babbles and birds** |

**Background of the topic of your research plan, rationale, relevance (max. 500 words)**

Language development in young infants and children is crucial for later linguistic, cognitive and social development. Language skills influence social competence and vice versa. For example, children with language disorders may engage in social interactions less frequently, resulting in lower language input. Early identification of language problems may help prevent such a downward spiral.

We aim to reveal if developmental trajectories in early child vocal production are predictive of later language outcome. Studies on perception already point in this direction (stokes et al), but production data may complement perceptual data and might be an easier 'at home' marker.

The longitudinal data in the youth cohort will provide us with the opportunity to test within-subject developmental trajectories. This individual data can reveal the range of normal individual variation in speech development, as well as impaired or delayed development. Moreover, we can test the possibility to use early vocalizations as an early marker for later speech and (a subset of) languages outcomes. Potentially, we can also test if we can distinguish between speech- language delay with and without later recovery (fig. 1; Lang et al., 2019; Oller et al., 2010).

Early human speech- and language acquisition is thought to develop from universal (experience-independent) to experience-dependent and language specific (Doupe & Kuhl, 1999; Kager et al., 2004; Ter Haar & Levelt, 2018). We expect that the transition will occur later in development in children with lower language outcome (at age '3' and '6'). Children with language delay are expected to recover afterwards, but some children with DLD are expected to remain having more difficulty with non-universal phonology.

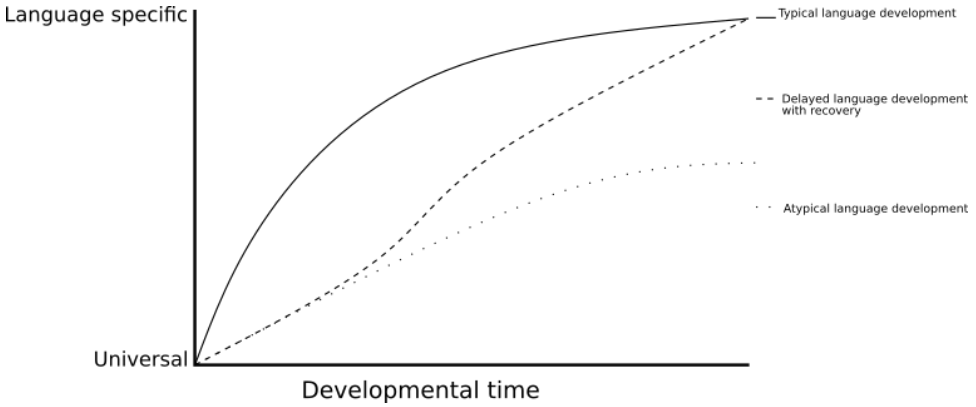


**Fig. 1**

**Schematic representation of expected results**

Over the course of development, universal speech sounds are gradually complemented with more and more language specific sounds. We expect this trajectory to be delayed and not recover in children with DLD (atypical language development)

In addition to human vocal development, we perform similar longitudinal analyses in songbirds. Earlier findings suggest that perceptual 'phonological' development in songbirds predicts later song production learning (ter haar et al., 2014). This perceptual development parallels development in human infants, characterized by a switch from a perceptual preference for universal phonological feature early in development towards language-specific ones at later age (ter haar et al., 2018).

In addition, we investigate song development in sleep deprived songbirds. Disrupted sleep has been identified as a risk factor for Developmental

Language Disorder (DLD, Earle et al). Songbirds can be used as a model for neural plasticity related to sleep and vocal learning.
Together, this research can potentially lead to new methods of early prediction of certain language problems.

**The specific research question(s) or aim(s) of your research**

Aim: Investigate if language scores (phonology, word learning) can be predicted from early vocal productions.
Q1: can the early developmental trajectory of vocalizations (between 5 and 10 months) predict later phonological processing and word learning (3y: peabody, NCDI)?
Q2: can phonological information from NCDI at 10m predict (phonological features in) vocabulary at later age (3y: peabody, NCDI)?
Q3: is there an association between risk of language disorders (based on questionnaire) and predictive value of early phonological development (Q1 and Q2)?

**Summary of the data requested for your project:** Please indicate which data you request to answer your research question.

The exact amount of data used will depend on the amount and quality of useful data, since we expect that not every child vocalizes in the 15 minutes PCI. We start with 15% top, 15% middle and 15% lowest scores on NCDI at 3 years and 10 months, and only of those children that were present at 5m, 10m and 3y. If quality and quantity turns out to be too limited, we may have to ask for an amendment.

**Around 0: 5m:** *all*
parent child interaction (PCI) video's to analyse vocalizations (**Q1,Q3**)

**10m:** *all*
PCI (**Q2**)
NCDI (**Q2,Q3**)

**Around 3:** *initially those children that were included at 5 and/or 10m*
NCDI (**Q1,Q2,Q3**)
Peabody (**Q1,Q2,Q3**)
(maybe later looking-while-listing and PCI if time allows, eg automatic analysis works)

**Parental questionnaires:** *initially those children that were included at 5 and/or 10m*
to select at-risk individuals if necessary and to answer Q3
-Language situation and pragmatics
-medical
-socioeconomic

**Age of testing** (which week, or more detailed): to determine at which developmental stage they are. We expect profound differences in vocal development within age group R3 due to the broad range of ages within group.

| **References** (optional) |
| --- |
| |

## Section 3: Hypotheses

In this section, please provide your research hypotheses. For each hypothesis:

- Be as specific as possible
- Provide the anticipated outcomes for accepting and/or rejecting the hypothesis.

| **Hypotheses** |
| --- |
| *Q1: can the early developmental trajectory of vocalizations (between 5 and 10 months) predict later phonological processing and word learning?*<br>**H1**: We hypothesize that the ratio between universal and language specific phonological features changes between 5 and 10 months (towards more non-universals) and that the strength of this change positively predicts NCDI and peabody scores at 3y<br><br>*Q2: can phonological information from NCDI at 10m predict (phonological features in) vocabulary at later age (3y: peabody, NCDI)*<br>**H2**: We hypothesize that the amount of language-specific features in the passive lexicon at 10 months age is positively related to NCDI (production and perception) and peabody scores at 3y<br><br>*Q3: is there an association between risk of language disorders (based on parental questionair) and predictive value of early phonological development (Q1 and Q2)*<br>**H3**: We hypothesize that indicators for language problems on the questionair are associated with weaker change of phonological features (Q1) and lower receptive vocabulary (Q2) at 10 months, and with lower NCDI and peabody scores at 3y. |

## Section 4: Methods

In this section, you should make clear how the hypotheses are tested. Be as specific as possible. Please describe:

- The study design and study population (Which data do you require from which subjects?)
- The general processing steps (to prepare the data for analysis)
- The analysis steps (How are the data analysed to address the hypotheses? If possible, link each description to a specific hypothesis)
- Any additional aspects that need to be described to clarify the methodological approach (optional)

| **Study design, study population and sample size** (e.g. cross-sectional or longitudinal; entire population or a subset; substantiate your choices) |
| --- |
| Longitudinal individual analyse will be applied because individual developmental trajectories are expected to predict individual outcome.<br><br>Two strategies for analyses will be initiated simultaneously: |

1) manual analyses of vocalizations will be applied to verify and/or replace automatic analyses, depending on the automatic analyses' quality: selection of data based on languages scores at 3y (low/mid/high or top/bottom)
2) automatic analyses of vocalizations: all data for which 3y is available

**General processing steps to prepare the data for analysis**

The first step is selecting data based on NCDI scores at 3y and 10m: 15% top, 15% middle and 15% lowest scores on NCDI at 3 years and 10 months, and only of those children that were present at 5m, 10m and 3y. Once data is selected, the following steps will be performed:

PCI
- audio extraction
- filtering
- segmentation (automatic or manual)
- annotation phonological information: phonological transcription (automatic or manual)
- calculate phonological universal to non-universal ratio (corrected for chance level)

N-CDI
- extract phonological information: match phonological transcription in CELEX corpus to N-CDI items

Peabody task:
- extract and calculate scores (using script Anika van der Klis)

Parental questionnaires:
Extract
- language problems
- language background (multilingual etc)
- Social economic status

**Specific processing and analysis steps to address the hypotheses**

- PCI, N-CDI: calculate phonological universals
- 4 generalized mixed models with

Dependent variables:
- 1) N-CDI at 3y
- 2) peabody scores at 3y
- 3) looking while listening
- 4) (vocal production at 3y)

Independent variables
- (chance corrected) ratios at 5 and 10m, or difference between 5 and 10m
- N-CDI at 10m
- age
- risk at language disorder (parental questionnaires)

Random variable(s):
- individual

◦ (recording location)

**Additional methodological aspects** (optional)

Data will first need to be explored in order to determine if quality of recordings and number of vocalizations etc. are sufficient.
Age groups (especially at 10m) may have to be split up into multiple groups, or variation should be included in the model.

If possible, we would also like to make extra (audio) recordings during the visits at 5 and 10 months of age, and possibly at home in the second year

## Section 5: Data request

In this section, please specify as detailed as possible which data (and from which subjects) you request.

**Data requested**

The exact amount of data used in the potential publication will depend on an initial exploratory phase in which we determine the amount and quality of useful data, as well as effectivity of analyses techniques.

**Around 0:**
**5m:** *top 15%, mid 15% and lower 15% based on NCDI at 3y and 10m*

- parent child interaction (PCI) videos to analyse vocalizations

**10m:** *initially all children that were included at 5m*

- PCI
- Language development questionnaire (=NCDI)

**Around 3:** *initially all children that were included at 5 and 10m*

- Language development questionnaire (=NCDI)
- Peabody

**Questionnaires:** *initially those children that were included at 5 and/or 10m*
to select at-risk individuals if necessary and to answer Q3
see data selection template

**Age of testing** (which week, or more detailed): to determine at which developmental stage they are. We expect profound differences in vocal development within age group R3 due to the broad range of ages within group.


youth