Children's theory of mind in a naturalistic story-listening paradigm: An fNIRS neuroimaging dataset

Authors
Chi-Lin Yu[a], Rachel Eggleston[a], Kehui Zhang[a], Nia Nickerson[a], Xin Sun[a], Rebecca A. Marks[a], Xiaosu Hu[a], Jonathan R. Brennan[b], Henry M. Wellman[a], and Ioulia Kovelman[a]

Affiliations
[a] Department of Psychology, University of Michigan
[b] Department of Linguistics, University of Michigan

Abstract
This article documents a functional Near-Infrared Spectroscopy (fNIRS) neuroimaging dataset deposited in Deep Blue Data. The dataset includes 51 children (age range = 6-12 years) who listened to the first chapter of Alice's Adventures in Wonderland during fNIRS neuroimaging. We also provide the text of the story with several word-by-word predictors motivated by research in Theory of Mind development and language. These annotated, naturalistic datasets can be used to replicate prior work and test new hypotheses about everyday social cognition and natural language comprehension in the developing brain.

Keywords: fNIRS, Children, Theory of Mind, Language, Naturalistic Neuroscience, Story Listening

Specifications Table

Subject	Developmental and Educational Neuroscience
Specific subject area	fNIRS neuroimaging of children's theory of mind in a naturalistic story-listening paradigm
Type of data	Tables, Images, fNIRS hemodynamic data
How data were acquired	TechEn Inc. CW6 fNIRS system with 690 and 830 nm wavelengths, 12 signals, 24 detectors, 46 channels.
Data format	fNIRS data are stored in .nirs files. The text of the story with several word-by-word predictors is provided in an excel sheet. Demographic raw data are stored in another excel sheet.
Parameters for data collection	All participants are children growing up in the US and attending English-only schools
Description of data collection	Participants ($N = 51$) listened to the first chapter of Alice's Adventures in Wonderland during fNIRS neuroimaging
Data source location	University of Michigan, Department of Psychology, Ann Arbor, MI.
Data accessibility	Repository: Deep Blue Data
Related research article	Yu, C. L., Eggleston, R., Zhang, K., Nickerson, N., Sun, X., Marks, R., Hu, X.S., Brennan, J.R., Wellman, H. M., & Kovelman, I. (*Submitted*) Neural Processing of Children's Theory of Mind in a Naturalistic Story-Listening Paradigm.

Value of the Data

- This child dataset allows researchers to explore Theory of Mind and language processing simultaneously in children's emerging neural architectures during natural language comprehension.
- This dataset is the first fNIRS study to use the widely validated story (i.e., Alice's Adventures in Wonderland) in a naturalistic story-listening paradigm. The results can be compared to prior work using electroencephalography (EEG), functional magnetic resonance imaging (fMRI), and magnetoencephalography (MEG).
- The dataset will equip researchers in the fields of developmental, educational, and cognitive neuroscience to address questions about children's neurocognitive profiles of everyday cognitive and social processing more generally.

Overall Data Description

All data are available in the Deep Blue repository under the name "Children's theory of mind in a naturalistic story-listening paradigm: An fNIRS neuroimaging dataset". For a list of the Deep Blue files and contents, see Table 1.

Table 1.

Data/Measure	File Name in Deep Blue	Content
Demographics	Participant_Demographics.csv	Demographic information, including age of testing, gender, and other behavioral variables (see below for more details), etc.
Story listening	AIW.csv	Text of the story with several word-by-word predictors
fNIRS imaging	NIRSfiles.zip	.nirs files by ID for the participants

Demographics

Participants Participants included 51 children (23 females; age range = 5.6-11.97 years; mean age = 8.88; SD = 1.80), all native English speakers with normal hearing, no diagnosed or teacher/parent-suspected intellectual, cognitive, or motor disabilities, and no history of brain injury. The sample was 87% white and 13% multi-ethnic. This study was approved by Institutional Review Boards at the University of Michigan. Parents provided written consent before data collection and were paid for participation.

Language proficiency. The Peabody Picture Vocabulary Test (PPVT) was used to assess children's vocabulary. During administration, children pointed to the best pictorial match for an experimenter-spoken word from among four choices. Standardized PPVT scores indicated all participants had good English proficiency, evidenced by a standard score of 85 or greater. The average PPVT standard score was 111.76 (*SD* = 17.56). We used raw PPVT scores to reflect language proficiency more directly in the neuroimaging analyses. The average PPVT raw score was 166.24 (*SD* = 23.64).

Working Memory. The backward digit span task from the Wechsler Intelligence Scale for Children–Fifth Edition (Wechsler, 2014) was used to assess children's working memory. During administration, an examiner spoke a series of numbers, and the child attempted to repeat the numbers in backward order. As this task is a good proxy for general intelligence (e.g., Alloway & Alloway, 2010), we used it as a control variable in the neuroimaging analyses. The average score was 7.16 (*SD* = 1.77).

Demographic data are presented in an Excel sheet, named "Participant_Demographics.csv". In the .csv file, the "Subject" column indicates subject IDs; the "Gender" column indicates each subject' biological sex (1= female and 0 = male) the "Age" column indicates each subject's age of testing, the "PPVT_Raw" column indicates each subject's raw PPVT score; the "PPVT_SS" column indicates each subject's standardized PPVT score; the "DigitSpan" column indicates each subject's score from the backward digit span task. Note that this project is a subset of a broader project investigating children's

language and literacy development, so some subjects also completed other tasks. Those data are provided in Sun, Zhang, et al. (2022) with identical subject IDs across both repositories.

Neuroimaging Story Listening

The story-listening stimulus was the first chapter of Lewis Carroll's *Alice's Adventures in Wonderland* (read by Kristen McQuillan; available on librivox.org; lasting 12.4 min). This stimulus has been widely validated (Brennan et al., 2012, 2016, 2019; Brennan & Hale, 2019) and is available in Brennan et al. (2016). The text (i.e., the first chapter of Alice's Adventures in Wonderland) of the story with several word-by-word predictors are presented in a sheet named "AIW.csv." Our submitted manuscript, which details those word-by-word predictors, can be made available upon request. Please see below for a brief description for the nine columns in the sheet.

- Word: the text of the story in a word-by-word format.
- OffsetTime: the offset time of each word in the audio story-listening stimulus.
- Surprisal: a surprisal index computed based on GPT language model.
- Mental: the mental word class
- Perception: the perception word class
- Physical-Sensations: the physical-sensations word class
- Emotion: the emotion word class
- Word Rate: an index to indicate the presence of each word
- Word Frequency: a word frequency index estimated based on the HAL written-language corpus

Neuroimaging Data acquisition

Brain data were collected via a TechEN-CW6 system using 36 optodes: 12 emitters of near-infrared light sources and 24 detectors (spaced ~2.7 cm apart). These source-detector pairings resulted in 23 channels per hemisphere. Optodes were mounted in a custom-built cap constructed from 2 mm silicone-rubber material. As shown in Figure 1, our fNIRS optodes cover the frontal, temporal, and temporoparietal regions (Marks et al., 2022; Sun, Marks, et al., 2022).

The optodes were applied using the international 10-10 system positioning (Jurcak et al., 2007). The participant's nasion, inion, Fpz, left/right pre-auricular points, and head circumference were measured; F7, F8, T3, and T4 were used as anchors.

Our signal-to-noise ratio minimum and maximum were set to the standard 80 dB and 120 dB range, respectively. Before the start of the experimental task, a data quality check was completed by examining the participant's cardiac signal across channels. Data were collected at a sampling frequency of 50Hz.

Neuroimaging data are raw data fNIRS files. The onset of the story-listening paradigm is indicated by a stimulus mark (Note that some files don't have a mark; this means that the fNIRS recording and the striy-0listening paradigm starts at the same time). Each fNIRS file is stored in an individual folder named after participant ID. All fNIRS neuroimaging data are *.nirs* files and can be easily read into MATLAB and fNIRS specific software (e.g., Homer).

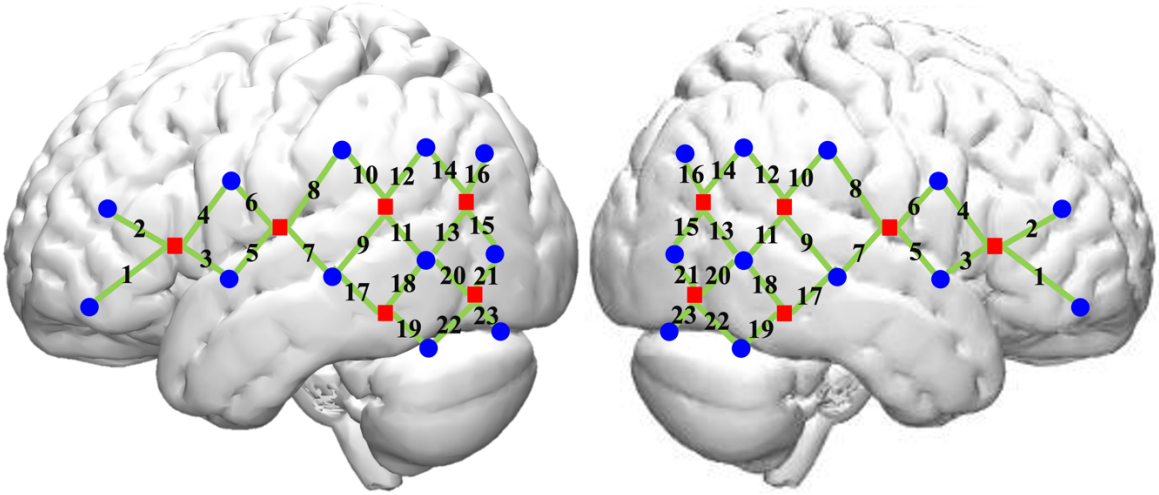

Figure 1. Optode placements and channel configuration. Blue circles represent the sources, and red squares correspond to the detectors. Green lines with numbers indicate the channels (23 channels per hemisphere).

Ethics Statements

Informed consent was obtained from all participating children and their guardians. In addition, all research protocols were approved by the Institutional Review Board at the University of Michigan Ann Arbor and the protocol number is HUM00033727. The dataset has also removed all identifiable information to protect participant privacy.

Acknowledgments

The authors thank members of the Language and Literacy Laboratory at the University of Michigan who helped with participant recruitment, scheduling, and data acquisition. We also thank the National Institutes of Health for funding this work (Kovelman, PI: R01HD092498).

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology*, *106*(1), 20–29.

Brennan, J., & Hale, J. T. (2019). Hierarchical structure guides rapid linguistic predictions during naturalistic listening. *PloS One*, *14*(1), e0207741.

Brennan, J., Lajiness-O'Neill, R., Bowyer, S., Kovelman, I., & Hale, J. T. (2019). Predictive sentence comprehension during story-listening in autism spectrum disorder. *Language, Cognition and Neuroscience*, *34*(4), 428–439.

Brennan, J., Nir, Y., Hasson, U., Malach, R., Heeger, D. J., & Pylkkänen, L. (2012). Syntactic structure building in the anterior temporal lobe during natural story listening. *Brain and Language*, *120*(2), 163–173.

Brennan, J., Stabler, E. P., Van Wagenen, S. E., Luh, W.-M., & Hale, J. T. (2016). Abstract linguistic structure correlates with temporal activity during naturalistic comprehension. *Brain and Language*, *157*, 81–94.

Jurcak, V., Tsuzuki, D., & Dan, I. (2007). 10/20, 10/10, and 10/5 systems revisited: their validity as relative head-surface-based positioning systems. *Neuroimage*, *34*(4), 1600–1611.

Marks, R. A., Eggleston, R. L., Sun, X., Yu, C.-L., Zhang, K., Nickerson, N., Hu, X.-S., & Kovelman, I. (2022). The neurocognitive basis of morphological processing in typical and impaired readers. *Annals of Dyslexia*, *72*(2), 361–383. https://doi.org/10.1007/s11881-021-00239-9

Sun, X., Marks, R. A., Zhang, K., Yu, C.-L., Eggleston, R. L., Nickerson, N., Chou, T.-L., Hu, X.-S., Tardif, T., Satterfield, T., & Kovelman, I. (2022). Brain bases of English morphological processing: A comparison between Chinese-English, Spanish-English bilingual, and English monolingual children. *Developmental Science*, e13251.

Sun, X., Zhang, K., Marks, R., Karas, Z., Eggleston, R., Nickerson, N., Yu, C. L., Wagley, N., Hu, X., Caruso, V., Chou, T. L., Satterfield, T., Tardif, T., & Kovelman, I. (2022). Morphological and phonological processing in English monolingual, Chinese-English bilingual, and Spanish-English bilingual children: An fNIRS neuroimaging dataset. *Data in Brief*, *42*, 108048. https://doi.org/10.1016/j.dib.2022.108048

Wechsler, D. (2014). Wechsler intelligence scale for children--Fifth Edition (WISC-V). *Bloomington, MN: Pearson*.