ARTICLE IN PRESS

Scientific Data

https://doi.org/10.1038/s41597-026-06579-8

Article in Press

Kymata Soto Language Dataset: an electro-magnetoencephalographic dataset for natural speech processing

Received: 16 April 2025

Accepted: 8 January 2026

Cite this article as: Yang, C., Parish, O., Klimovich-Gray, A. *et al.* Kymata Soto Language Dataset: an electro-magnetoencephalographic dataset for natural speech processing. *Sci Data* (2026). https://doi.org/10.1038/s41597-026-06579-8

ChenTianyi Yang, Oliver Parish, Anastasia Klimovich-Gray, Cai Wingfield, William D. Marslen-Wilson, Chao Zhang, Alexandra Woolgar & Andrew Thwaites

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

© The Author(s) 2026. **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/.

ARTICLE IN PRESS

Kymata Soto Language Dataset: an electro-magnetoencephalographic dataset for natural speech processing

ChenTianyi Yang[1], Oliver Parish[2], Anastasia Klimovich-Gray[3], Cai Wingfield[1,4], William D. Marslen-Wilson[5], Chao Zhang[6], Alexandra Woolgar[1,5], Andrew Thwaites[1,2]

[1]MRC Cognition and Brain Sciences Unit, University of Cambridge, UK
[2]Department for Speech Hearing and Phonetic Sciences, UCL, London, UK
[3]School of Psychology, University of Aberdeen, Aberdeen, UK
[4]Institute for Interdisciplinary Data Science and AI, University of Birmingham, Birmingham, UK
[5]Department of Psychology, University of Cambridge, Cambridge, UK
[6]Department of Electronic Engineering, Tsinghua University, Beijing, China

† **Corresponding author.** Andrew Thwaites, SHaPS, UCL, London UK (andrew.thwaites@ucl.ac.uk)

ARTICLE IN PRESS

Abstract

The Kymata Soto Language Dataset comprises raw electroencephalographic (EEG) and magnetoencephalographic (MEG) recordings from 15 native Russian speakers and 20 native English speakers as they listened to approximately seven minutes of conversational speech in their respective native languages. Each participant heard the same conversational speech stimulus multiple times (four repetitions for Russian speakers and eight for English speakers). The dataset includes transcriptions of the recordings, along with timestamp annotations for each phoneme and word. Organized according to the Brain Imaging Data Structure (BIDS), this dataset facilitates in-depth research into brain responses to naturalistic speech. To validate the dataset and our preprocessing pipeline, we employed Python-based analyses, revealing consistent low-level loudness perception trends across both language groups. All EEG and MEG data, audio recordings, transcriptions with timestamp annotations, and validation codes are open source, promoting transparency and reproducibility.

ARTICLE IN PRESS

Background & Summary

Understanding how the human brain processes language has long intrigued researchers. While functional Magnetic Resonance Imaging (fMRI) and functional Near-Infrared Spectroscopy (fNIRS) have been widely used, EEG and MEG have gained popularity due to their high temporal resolution (often reaching 1 kHz) and the relatively low cost of EEG. Over the past 15 years, there has been a shift toward studying language comprehension in more naturalistic contexts. Instead of analysing EEG and MEG responses to isolated words or sentences, researchers increasingly focus on continuous speech, such as audio narrations of classic novels like *The Adventures of Sherlock Holmes* and *Alice's Adventures in Wonderland* (e.g. [1, 2] for EEG, [3, 4, 5, 6, 7] for MEG). This approach enables the examination of more complex language processing mechanisms and provides valuable data for applications such as decoding speech from brain recordings [8, 9].

In this paper, we introduce a dataset of simultaneous EEG and MEG (EMEG) recordings from participants listening to multi-participant conversational speech in different languages. This dataset is particularly suited for investigating intra-language linguistic comparisons and cross-language differences in speech perception.

This dataset offers several features that make it particularly valuable for studying cortical language processing. First, it includes data from two distinct language groups: native English speakers listening to English and native Russian speakers listening to Russian. This design allows for meaningful cross-linguistic comparisons (see Table 1).

Second, the dataset features naturalistic conversational speech rather than scripted narrations. Unlike narrated speech, which tends to be slower and more deliberate, spontaneous conversations exhibit more variable pacing influenced by context and individual differences [10, 11]. As conversation is the dominant form of spoken communication, studying natural speech is crucial for understanding real-world language processing.

Third, it is a simultaneous EEG and MEG Recording. By combining these two modalities, the dataset has enhanced sensitivity to different sources of neural activity and enables more accurate

ARTICLE IN PRESS

source reconstruction [12, 13, 14], providing a more comprehensive understanding of neural responses to auditory stimuli.

During the auditory task, participants fixated on a cross while a dynamic dot-field stimulus appeared in their visual periphery, changing colour over time. This ensured consistent visual engagement across trials.

Name of Dataset	**Name of subset**	**Audio stimuli**	**Native language of participant listening to stimuli**	**Visual stimuli**
Kymata SOTO	English-native-English-conversation	BBC podcast of English speakers discussing the history of ice cream	English	Moving dots on a grey background, randomly horizontal oscillation, and also changing in colour. Participants were instructed to look at fixation cross in the centre of the screen.
	Russian-native-Russian-conversation	BBC podcast of Russian speakers discussing the history coffee	Russian	Moving dots on a grey background, separated into quadrants, randomly oscillating in both the horizontal and vertical dimension, and changing in colour. Participants were instructed to look at fixation cross in the centre of the screen.

ARTICLE IN PRESS

Table 1: Overview of the two data subsets contained in the Kymata Soto Dataset.

Preregistrations for both studies can be found on OSF.io: English-native-English-conversation [15]; Russian-native-Russian-conversation [16].

Methods

ARTICLE IN PRESS

Participants

Fifteen Russian-speaking adults (7 male; 8 female, mean age = 24 years, range = 18–30) and twenty English-speaking adults (11 male; 9 female, mean age = 22 years, range = 18–35) were recruited from the volunteer panel of the Medical Research Council Cognition and Brain Sciences Unit, Cambridge University. Written, informed consent was collected for all participants, and they were paid for their participation. All the participants had normal hearing and no hearing-related disorders. All the participants were native speakers of the language of the dataset. Proficiency in additional languages was not an exclusion criterion for either group of participants. 4 repetitions of the recording were played for the Russian dataset in one recording session, and 8 repetitions (6 repetitions for participant E8) were played for the English dataset in also one recording session. One participant (E3) came for a second session with the same 8 repetitions of the original recording. The studies were approved by the Cambridge Psychology Research Ethics Committee (CPREC; approvals PRE.2010.46 and PRE.2018.101 respectively).

ARTICLE IN PRESS

Stimuli

Audio

For both datasets, the audio stimuli were selected from BBC radio-podcast studio discussions, each with a length 400 seconds (6 min 40 sec). The stimuli were presented at a sampling rate of 44.1 kHz with 16-bit resolution. For the English-native-English-conversation dataset, the stimulus is an edited excerpt from an episode of *You're Dead to Me* titled *The History of Ice Cream*—a three-way discussion (two male, one female) in English concerning the history of ice cream. The original recording is available at https://www.bbc.co.uk/programmes/p0b3kpn9. For the Russian-native-Russian-conversation dataset the stimulus is an edited excerpt from *БибиСева* (BBC Russian Service, 2014), comprising a three-way discussion (two male, one female) in Russian about the history of Colombian coffee. The *БибиСева* archive is no longer publicly hosted, researchers who wish to obtain a copy may contact the BBC directly to inquire about access; however, availability is at the BBC's discretion and cannot be guaranteed. As the stimuli used in this study were edited versions of these recordings, we have provided a script that automatically reproduce the precise excerpts used in the experiment with the dataset.

ARTICLE IN PRESS

To facilitate the use of the dataset in different scenarios, the transcriptions of both recordings were first generated by using Whisper large v2 model [17] using the default parameters and then corrected manually. We further obtained the timestamp information at both word and phoneme level for each recording by feeding the generated transcriptions to the Kaldi-based Montreal Forced Aligner [18] with the English MFA dictionary/acoustic model v3_0_0 and Russian MFA dictionary/acoustic model v3_1_0 respectively for the three datasets, which is also being made available.

Both stimuli were selected to provide engaging and naturalistic examples of conversational speech involving multiple participants of different genders. Its content is largely extemporaneous, ensuring that it would not be familiar to participants in the way a well-known public-domain story might be. The topics of the two stimuli were food-related: the history of ice cream for the English conversation and a discussion of coffee for the Russian conversation.

This choice was made for several reasons. First, the topic is sufficiently engaging to maintain participants' attention throughout the experiment. Second, it is unlikely to elicit strong emotional responses. Third, it is likely to be of interest to a broad range of listeners, including children, should it be required to reuse the stimulus in future studies.

ARTICLE IN PRESS

Visual

For the *Russian-native-Russian-conversation* dataset a pattern of randomly placed colored dots on a grey-masked annular field with a black fixation cross in the center to control eye movements (Fig. 1A). Each dot being 10 x 10 pixels, the central fixation mask being 140 x 140 pixels and the circular edge of the annular field extending to upper and lower edges of the frame, with 100 pixels of linear falloff of the colored dots at the outer edge of the annulus to avoid edge artefacts. Within the annular mask, the horizontal displacement and color of these dots fluctuated pseudo-randomly. The colour moved only in the achromatic and red-green dimensions, and not the yellow-blue dimension. The visual stimulus was 420 seconds long; it started 10 seconds before and ended 10 second after the 400 second auditory stimulus, to avoid the sudden appearance and disappearance of the stimulus during the first and last trial of this auditory stimulus. The stimulus had a refresh rate of 60Hz at a resolution of 1280 x 1024 pixels.

ARTICLE IN PRESS

For the *English-native-English-conversation* dataset, a pattern of randomly placed colored dots on a quadranted grey-masked annular field with a black fixation cross in the center to control eye movements (Fig. 1B). Each dot being 15 x 15 pixels, the central fixation mask being 140 x 140 pixels and the circular edge of the annular field extending to upper and lower edges of the frame, with 100 pixels of linear falloff of the colored dots at the outer edge of the annulus to avoid edge artefacts. Unlike the *Russian-native-Russian-conversation* data set, in which the dots all moved in the same direction and changed color over the entire visual field, in this dataset the visual stimulus was different in each quadrant. Within each quadrant, the horizontal and vertical displacement and color of these dots fluctuated pseudo-randomly. Also, unlike the *Russian-native-Russian-conversation* data, set this dataset's colors moved through the entire color gamut. The visual stimulus started and ended 10 seconds before and after the stimulus to avoid the sudden appearance and disappearance of the stimulus during the first and last trial.

Each frame of the stimulus (in .png format) is available from the dataset repository (see *data availability*).

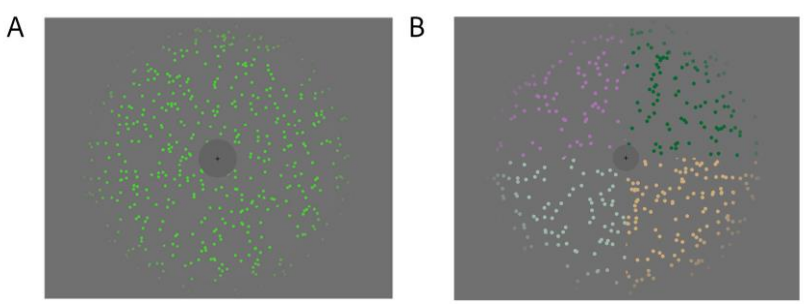


Figure 1. Visual stimuli for the sub-datasets A *Russian-native-Russian-conversation*, **B** *English-native-English-conversation*

The task in all studies was to answer 1–2 multiple-choice questions after each repetition, in the participants native language. These questions served only to keep the subject engaged in the stimulus, and the responses were not recorded.

ARTICLE IN PRESS

A dynamic dot field was chosen to facilitate investigation of early visual responses to color and motion (e.g. [19, 20]) without eliciting object-detection or other visuosemantic processes. The visual stimulus was not meaningfully synchronised with the audio stimuli, beyond the fact that each stimulus was repeated simultaneously, meaning both could elicit simultaneous stimulus-onset effects.

Presentation Equipment details

The sound was presented through ER3 earpieces with 2.5 m tubes (Etymotic Research, Elk Grove Village, Illinois) controlled with Psychophysics Toolbox [21, 22, 23] in Matlab.

The stimuli were presented on a Panasonic PT-D7700 DLP projector (for the Russian stimulus) and a VPixx ProPiXX (for the English stimulus), both with a refresh rate of 60 Hz. The screen size was 0.5 x 0.4m. It was placed 1.3 m from the participant, resulting in a viewing angle of 22º. The screen's white-light luminance was 162 cd/m^2.

The delays in both sound and visual stimuli presentations are corrected for in the configuration files.

Procedure

Within each recording session, concurrent EEG and MEG data were recorded for participants, and sampled at 1,000 Hz while they listened to the podcast through ER3 earpieces. The MEG data were recoded using a 306 channel Elekta-Neuromag VectorView (*English-native-English-conversation*) or MEGIN Truix system (*Russian-native-Russian-conversation*) (MEGIN, Espoo, Finland) containing 102 identical sensor triplets (two orthogonal planar gradiometers and one magnetometer) in a hemispherical array situated in a light magnetically shielded room. The EEG data were recorded from 70 Ag-AgCl electrodes placed in an elastic cap (EASYCAP GmbH, Herrsching-Breitbrunn, Germany) according to the 10/20 system, using a nose electrode as reference. Electrocardiogram (ECG), horizontal Electrooculogram (hEOG), and vertical Electrooculogram (vEOG) were also recorded for potential artifact removal during data pre-processing. The stimuli were binaurally presented at approximately 65 dB SPL for both datasets.

ARTICLE IN PRESS

Before the experiment, a hearing test was performed to confirm the normal function of the equipment and normal hearing of the participants, and then the opening introduction of the podcast was played for 20 s as a practice session to familiarize the participants with the voices of the speakers and the experimental environment.

The repetitions of the recordings were realized through blocks. Two repetitions were grouped in one block with breaks in between. During these breaks, we talked to the participants to ensure that they felt comfortable before we moved on to the next block. The design of the repetitions was to (i) enable data averaging to enhance signal-to-noise ratio; (ii) give researchers the ability to study individual differences reliably with reference to the variability due to noise across each repetition. The second session for participant E3 was also included to validate analyses of interindividual and intraindividual difference.

To ensure that the participants were paying attention to the recording, we designed two two-alternative forced-choice questions about the content of the podcast after each repetition. We also monitored participants through a camera in the MEG room. If they closed their eyes for prolonged intervals or reported feeling sleepy or unfocused we paused the experiment at the breaks and resumed only when they felt more attentive.

Participants had a T1-weighted Magnetic-Resonance-Imaging (MRI) scan taken, either on a Tim Trio (*English-native-English-conversation*) or Primsa (*Russian-native-Russian -conversation*) 3T scanner (Siemens, Erlangen, Germany) at the MRC Cognition and Brain Sciences Unit. The scanner used 1-mm isotropic voxels to obtain MRI images using GRAPPA 3D MPRAGE sequence (TR=2250 ms; TE=2.99 ms; flip-angle=9 degrees; acceleration factor=2) either on the day of their EMEG session or a later depending on the availability of the MRI facility.

Before each EMEG session, the head shape with the position of the EEG electrodes was recorded with a 3-D digitizer (Fastrak Polhemus Inc, Colchester, VA), and co-registered with five head-position indicator (HPI) coils. The head position of participants was checked at the start of each block to ensure that the participant did not move excessively.

Preprocessing

ARTICLE IN PRESS

MEG. The MEG dataset and its annotations are shared raw (i.e. not preprocessed) organized according to the Brain Imaging Data Structure (BIDS: [24, 25]). Note that researchers wanting to use this data will thus need to pass the data through Signal-Space Separation (SSS: [26, 27]) and/or maxwell filtering, to remove environmental artefacts.

MRI. The individual structural MRI data of the English dataset is provided after being defaced using PyDeface [28]. Due to the wording of the Russian dataset participant consent, the MRI structurals of this dataset are not shared. Researchers wishing to do cortical localisation with these data should use an average cortical mesh such as Freesurfer's FSaverage [29].

Stimuli. For the auditory stimuli, we include in the dataset: the original audio ('stimuli/audio/stimulus.wav'), the word-level transcription ('stimuli/audio/mfa_word.txt') with the starting time of each word ('stimuli/audio/mfa_word_stime.txt'), and the phonetic transcription ('stimuli/audio/mfa_phone.txt') with the starting time of each phoneme ('stimuli/audio/mfa_phone_stime.txt'). '<sp>' is used in the transcriptions to denote short pauses, and 'spn' in the phonetic transcriptions are used for words whose phonetic transcriptions are not available in the pronunciation dictionary.

For the visual stimuli, we save each frame of the patterns on the screen in 'stimuli/visual/png_frames_over_stimulus'. Each sub-folder inside contains the stimulus by frame in .png format (60 in total) for each second.

Triggers. The trigger channel is 'STI101'. The trigger value '1' is used for the onset of the practice audio stimulus, the trigger value '2' is used to denote each second of the video stimulus, and the trigger value '3' is used for the onset of each repetition of the 400 second audio conversation stimulus file. Button-press trigger values related to the forced-choice questions between runs are between 4000 and 20000, and are not mapped to specific responses.

Computing environment. The processing of the present data is based on the free and open-source ecosystem of the neuroimaging community. We used: MNE BIDS [25] (https://mne.tools/mne-bids) and Bids-Validator (https://github.com/bids-standard/bids-validator)

Data Records

ARTICLE IN PRESS

The dataset is organized according to Brain Imaging Data Structure (BIDS) 1.9.0 [24] and publicly available on the Open Science Framework data repository (https://osf.io/qdzgh/) [30] under a Creative Common Attribution 4.0 International Licence (CC-BY). An image of the folder structure is provided in Fig. 2. The detailed description of the BIDS file system is available at http://bids.neuroimaging.io/. In summary:

- ‘./dataset_description.json’ describes the dataset
- ‘ct_sparse_vectorview.fif’, ‘sss_cal_vectorview.dat’ in the Russian-native-Russian-conversation dataset and ‘ct_sparse_triux.fif’, ‘sss_cal_triux.dat’ in the English-native-English-conversation dataset are the MEG-machine-specific files that can be used to perform Maxwell filtering [27, 26].
- ‘./stimuli/’ contains the original podcasts and visual stimuli used during the experiments
- Each ‘./sub-XXX’ contains the brain recordings of a unique participant. ‘R’ in front of the numbers means native Russian speakers and ‘E’ means native English speakers
- In each participant folder lies the anatomical and the meg data
- ‘en’ in task means listening to English podcast and ‘ru’ in task means listening to Russian podcast
- The dataset can be read directly using MNE-BIDS

ARTICLE IN PRESS

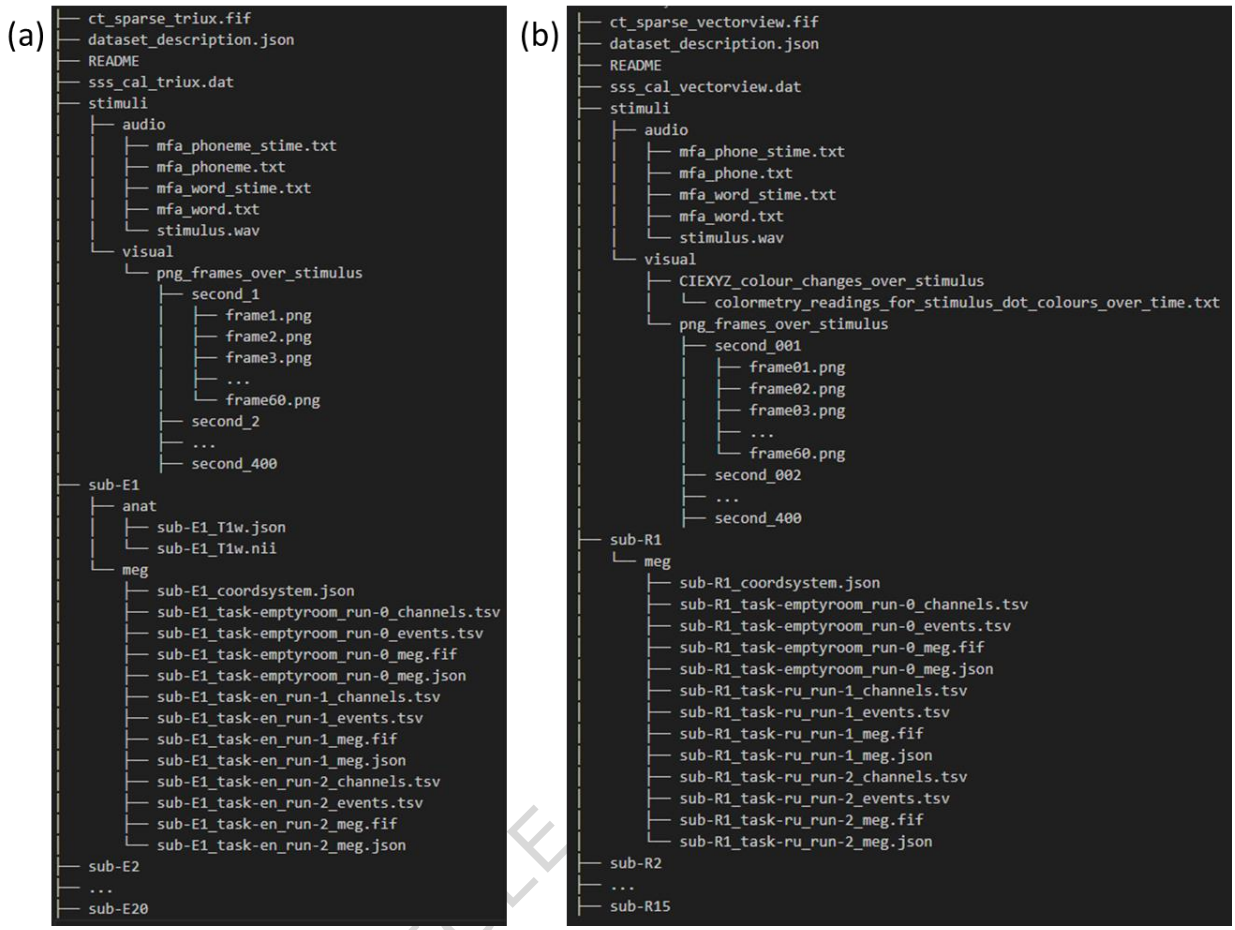


Figure 2: File structure of (a) *Russian-native-Russian-conversation* dataset (b) *English-native-English-conversation* dataset

Technical Validation

We checked that the dataset complies with the standardized brain imaging data structure by using the Bids-Validator (https://github.com/bids-standard/bids-validator).

MEG recordings are subject to large amount of measurement noise, and are thus a challenge to validate empirically. We followed the same analysis procedure as Thwaites et al. [31] and generated the expression plots for three different time-varying loudness (TVL) transformations in the auditory domain – instantaneous loudness, combined instantaneous loudness, and short-term loudness [32]. This preprocessing and validation was performed using the workflows

ARTICLE IN PRESS

provided by the *kymata-core* software toolbox [33]; TVL transforms of the audio stimuli were computed using a Matlab implementation [34] of the Moore et al. model [32]. These describe the timings and locations at which each of these loudness-related mathematical transformations take place in the cortex. The peaks of the different loudness transforms can be seen to occur at very similar latencies for both the *Russian-native-Russian-conversation* and *English-native-English-conversation* dataset (Fig. 3)

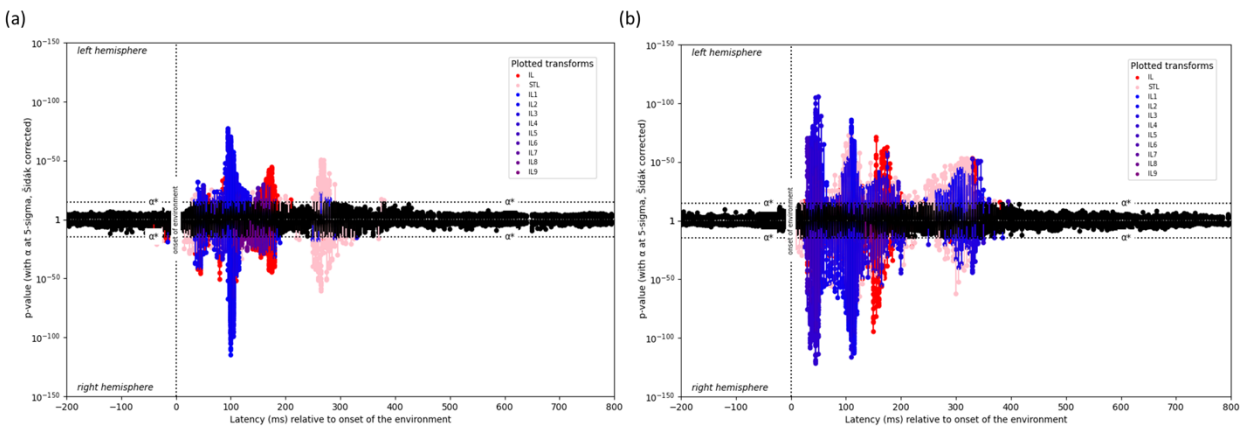


Figure 3: Confirmation that the TVL expression plot is comparable between both (a) *Russian-native-Russian-conversation* dataset (b) *English-native-English-conversation* dataset.

The *Russian-native-Russian-conversation* dataset has also been used in numerous existing studies [31, 35, 19, 20].

Usage Notes

The following Python code will read the first run of the first subject of the English-native-English-conversation dataset.

```
from mne_bids import BIDSPath, read_raw_bids

bids_path = BIDSPath(subject='E1', task='en', run='1', datatype='meg',

                     extension = 'fif', root='my/data/path', suffix = 'meg')
```

ARTICLE IN PRESS

```
raw = read_raw_bids(bids_path)
```

Code Availability

Only the open-source MNE BIDS python package [25] (https://mne.tools/mne-bids) is required to open the files.

Data Availability

The dataset is available on OSF at http://doi.org/10.17605/OSF.IO/QDZGH

Acknowledgements

For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission. This work was supported by UKRI MRC intramural funding (SUAG/093/G116768) to AW, an ERC Advanced Grant (Neurolex) to WDM, a grant from Neuroverse FZCO to AT and CZ, and a National Natural Science Foundation of China (NSFC) grant to CZ (62476151).

ARTICLE IN PRESS

Author contributions

CY: Writing – review & editing; Data curation; Methodology; Software; Validation; OP: Software; Methodology; Data curation; AK-G: Writing – review & editing; Data curation; CW: Writing – review & editing; Software; WM-W: Funding acquisition; CZ: Writing – review & editing; Funding acquisition; AW: Writing – review & editing; Supervision; AT: Conceptualization; Data curation; Methodology; Writing – original draft; Writing – review & editing; Supervision

Competing interests

ARTICLE IN PRESS

The authors declare no competing interests.

Additional Information

Correspondence and requests for materials should be addressed to AT.

References

[1] J. R. Brennan and J. T. Hale, “Hierarchical structure guides rapid linguistic predictions during naturalistic listening.,” *Plos One,* vol. 14, no. 1, p. e0207741, 2019j.

[2] M. P. Broderick, A. J. Anderson, G. M. Di Liberto, M. J. Crosse and E. C. Lalor, “Electrophysiological correlates of semantic dissimilarity reflect the comprehension of natural, narrative speech,” *Current Biology,* vol. 28, no. 5, p. 803–809, 2018.

[3] K. Armeni and J. Schoffelen, “A 10-hour within-participant magnetoencephalography narrative dataset to test models of naturalistic language comprehension. Version 1.,” Radboud University, Radboud, 2022.

[4] L. Gwilliams, G. Flick, A. Marantz, L. Pylkkänen, P. D. and J. King, “Introducing MEG-MaSC a high- quality magneto-encephalography dataset for evaluating natural speech processing,” *Scientific Data,* vol. 10, no. 862, 2023.

[5] D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. Behrens, E. Yacoub, K. Ugurbil and W.-M. H. Consortium, “The WU-Minn human connectome project: an overview,” *Neuroimage,* vol. 80, p. 62–79, 2013.

[6] J. M. Schoffelen, R. Oostenveld, N. H. Lam, J. Uddén, A. Hultén and P. Hagoort, “A 204-subject multimodal neuroimaging dataset to study language processing,” *Scientific Data,* vol. 6, no. 1, 2019.

[7] H. Cai, Z. Yuan, Y. Gao and e. al., “A multi-modal open dataset for mental-disorder analysis.,” *Sci Data,* vol. 9, no. 178, pp. https://doi.org/10.1038/s41597-022-01211-x, 2022.

[8] A. Défossez, C. Caucheteux, J. Rapin, O. Kabeli and J. R. King, “Decoding speech perception from non-invasive brain recordings,” *Nature Machine Intelligence,* vol. 5, no. 10, p. 1097–1107, 2023.

[9] Y. Yang, Y. Duan, Q. Zhang, R. Xu and H. Xiong, “Decode neural signal as speech,” arXiv preprint arXiv:2403.01748, 2024.

[10] J. Yoo, H. Oh, S. Jeong and J. I.K., “omparison of Speech Rate and Long-Term Average Speech Spectrum between Korean Clear Speech and Conversational Speech,” *J Audiol Otol,* vol. 23, no. 4, p. 187–192, 2019.

[11] J. Yuan, M. Liberman and C. Cieri, “Towards an Integrated Understanding of Speaking Rate in Conversation,” in *Interspecch,* 2006.

[12] J. G. Samuelsson, S. Khan, P. Sundaram, N. Peled and M. S. Hämäläinen, “Cortical Signal Suppression (CSS) for Detection of Subcortical Activity Using MEG and EEG,” *Brain Topogr,,* vol. 32, no. 2, p. 215–228, 2019.

ARTICLE IN PRESS

[13] R. N. Henson, E. Mouchlianitis and K. J. Friston, “MEG and EEG data fusion: simultaneous localisation of face-evoked responses,” *Neuroimage,* vol. 47, no. 2, p. 518–589, 2009.

[14] A. Molins, S. M. Stufflebeam, E. N. Brown and M. S. Hämäläinen, “Quantification of the benefit from integrating MEG and EEG data in minimum l2-norm estimation,” *Neuroimage,* vol. 42, no. 3, p. 1069–1077, 2008.

[15] A. Thwaites, C. Zhang and A. Woolgar, “Kymata Entrainment-to-English-Language-Narratives Study,” 29 April 2023. [Online]. Available: https://doi.org/10.17605/OSF.IO/45DCG.

[16] A. Thwaites, “Kymata Entrainment-to-Russian-Language-Narratives Study,” 27 July 2024. [Online]. Available: https://doi.org/10.17605/OSF.IO/TC825.

[17] A. Radford, J. W. Kim, T. Xu, G. Brockman, C. McLeavey and I. Sutskever, “Robust speech recognition via large-scale weak supervision,” in *International conference on machine learning,* 2023.

[18] M. McAuliffe, M. Socolof, S. Mihuc, M. Wagner and M. Sonderegger, “Montreal forced aligner: Trainable text-speech alignment using kaldi,” in *Interspeech,* 2017.

[19] A. Thwaites, C. Wingfield, E. Wieser, A. Soltan, W. Marslen-Wilson and I. Nimmo-Smith, “Tracking cortical entrainment to the CIECAM02 and CIELAB color appearance models in the human cortex,” *Vision Research,* vol. 145, 2018.

[20] C. Wingfield, A. Soltan, I. Nimmo-Smith, W. D. Marslen-Wilson and A. Thwaites, “Tracking cortical entrainment to stages of optic-flow processing,” *Vision Research,* vol. 226, 2025.

[21] D. Brainard, “The Psychophysics Toolbox,” *Spat. Vis.,* vol. 10, p. 433e436, 1997.

[22] D. Pelli, “The VideoToolbox software for visual psychophysics: transforming numbers into movies,” *Spat. Vis.,* vol. 10, p. 437e442, 1997.

[23] M. Kleiner, D. Brainard and D. Pelli, “What's new in Psychtoolbox-3?,” *Perception,* vol. 36, 2007.

[24] K. Gorgolewski, et al., “The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments.,” *Scientific Data,* vol. 3, 2016.

[25] S. Appelhoff, et al., “MNE-BIDS: Organizing electrophysiological data into the BIDS format and facilitating their analysis,” *Journal of Open Source Software,* vol. 4, no. 44, 2019.

[26] S. Taulu and J. Simola, “Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements,” *Physics in Medicine and Biology,* vol. 51, no. 7, p. 1759–1768, 2006j.

[27] S. Taulu and M. Kajola, “Presentation of electromagnetic multichannel data: the signal space separation method,” *Journal of Applied Physics,* vol. 97, no. 12, p. 124905, 2005.

[28] O. F. Gulban, D. Nielson, R. Poldrack, J. Lee, K. J. Gorgolewski and S. Vanessasaurus Ghosh, “poldracklab/pydeface: v2.0.0,” Zenodo.

[29] B. Fischl, “FreeSurfer,” *Neuroimage,* vol. 62, no. 2, p. 774–781, 2012.

[30] C. Yang, O. Parish, A. Klimovich-Gray, C. Wingfield, W. D. Marslen-Wilson, C. Zhang, A. Woolgar and A. Thwaites, “Kymata SOTO Language Dataset (English and Russian Conversations), OSF.io,” 2025. Available: https://doi.org/ 10.17605/OSF.IO/QDZGH /.

ARTICLE IN PRESS

[31] A. Thwaites, J. Schlittenlacher, I. Nimmo-Smith, M.-W. W.D. and B. Moore, “Tonotopic representation of loudness in the human cortex,” *Hearing Research,* vol. 344, p. 244–254, 2017.

[32] B. C. Moore, B. R. Glasberg, A. Varathanathan and J. Schlittenlacher, “A loudness model for time-varying sounds incorporating binaural inhibition,” *Trends in hearing,* vol. 20, p. 2331216516682698, 2016.

[33] Kymata Research Group, “kymata-core,” 4 August 2024. [Online]. Available: https://github.com/kymata-atlas/kymata-core. [Accessed 26 September 2025].

[34] Kymata Research Group, “third-party-functions,” 11 January 2017. [Online]. Available: https://github.com/kymata-atlas/third-party-functions. [Accessed 26 September 2025].

[35] A. Thwaites, B. Glasberg, I. Nimmo-Smith and B. Marslen-Wilson W.D.: Moore, “Representation of Instantaneous and Short-Term Loudness in the Human Cortex,” *Frontiers in Neuroscience,* vol. 10, no. 183, 2016.

[36] G. Jenner, A. Gray and R. Osman, Performers, *The History of Ice Cream.* [Sound Recording]. BBC Radio 4. 2023.

ARTICLE IN PRESS